JOINT PRECODING AND ANTENNA SELECTION IN MASSIVE MIMO SYSTEMS

Rafael da Silva Chaves

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Wallace Alves Martins

Rio de Janeiro
Março de 2018

JOINT PRECODING AND ANTENNA SELECTION IN MASSIVE MIMO SYSTEMS

Rafael da Silva Chaves

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Wallace Alves Martins, D.Sc.

Prof. Marcello Luiz Rodrigues de Campos, Ph.D.

Prof. Raimundo Sampaio Neto, Ph.D.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2018

Chaves, Rafael da Silva

Joint Precoding and Antenna Selection in Massive MIMO Systems/Rafael da Silva Chaves. – Rio de Janeiro: UFRJ/COPPE, 2018.

XIII, 90 p.: il.; 29, 7cm.

Orientador: Wallace Alves Martins

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2018.

Referências Bibliográficas: p. 78 – 90.

1. Massive MIMO. 2. Antenna Selection. 3. Precoding. 4. Sparsity-aware Precoding. 5. Joint Precoding and Antenna Selection. I. Martins, Wallace Alves. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

À dona Rosa e ao seu Ideny.

Agradecimentos

Agradeço à minha mãe Rosa e ao meu pai Ideny, por todo carinho, apoio e incentivo que me deram ao longo dos 7 anos da minha vida acadêmica. Sem vocês eu nunca conseguiria chegar tão longe.

Agradeço em especial ao meu irmão e melhor amigo Gabriel, por sempre estar ao meu lado em todos os momentos e por ter muita paciência comigo. Seu papel foi crucial nesta jornada.

Agradeço ao meu orientador Wallace Martins, pelas oportunidades e por todo conhecimento que conseguiu me passar. Obrigado pela confiança que depositou em mim e por toda a ajuda que você me deu na confecção deste trabalho. Desde a graduação você vem me ajudando a evoluir como engenheiro, pesquisador e pessoa, sou muito grato por tudo.

Agradeço aos professores Paulo Diniz, Marcello Campos e Markus Lima, pessoas que contribuem diretamente na minha formação.

Agradeço a todos os meus amigos que me ajudaram direta ou indiretamente na realização deste trabalho. Em especial, agradeço àqueles que acompanharam a minha jornada de perto, ouvindo as minhas reclamações: Vinicius, Matheus, Felipe, Roberto, Igor, Lucas, Wesley, Marcelo Spelta, Rômulo, Renata, Rebeca, Marcelo Castro, Luana e Thaís.

Agradeço aos professores Marcello Campos e Raimundo Sampaio, por aceitarem o convite para compor a banca avaliadora deste trabalho.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo apoio financeiro fornecido durante a confecção desta dissertação.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PRECODIFICAÇÃO E SELEÇÃO DE ANTENAS EM SISTEMAS MIMO MASSIVO

Rafael da Silva Chaves

Março/2018

Orientador: Wallace Alves Martins

Programa: Engenharia Elétrica

Esta dissertação apresenta uma visão geral sobre MIMO (do termo em inglês, *multiple-input multiple-output*) massivo e propõe novos algoritmos que permitem a pré-codificação de sinais e a seleção de antenas de forma simultânea. MIMO massivo é uma nova tecnologia candidata para compor a quinta geração (5G) dos sistemas celulares. Essa tecnologia utiliza uma quantidade muito grande de antenas na estação-base e, sob condições de propagação favorável ou assintoticamente favorável, pode alcançar taxas de transmissão elevadas, ainda que utilizando um simples processamento linear. Entretanto, os sistemas MIMO massivo apresentam algumas desvantagens, como por exemplo, o alto custo de implementação das estações-bases. Uma maneira de lidar com esse problema é utilizar algoritmos de seleção de antenas na estação-base. Com esses algoritmos é possível reduzir o número de antenas ativas e consequentemente reduzir o custo nas estações-bases. Essa dissertação também apresenta uma classe pouco estudada de pré-codificadores não-lineares que buscam sinais pré-codificados esparsos para realizar a seleção de antenas conjuntamente com a pré-codificação. Além disso, este trabalho propõem dois novos pré-codificadores pertencentes a essa classe, para os quais o número de antenas selecionadas é controlado por um parâmetro de projeto. Resultados de simulações mostram que os pré-codificadores propostos conseguem uma BER (do termo em inglês, *bit-error rate*) menor que os algoritmos clássicos usados para selecionar antenas. Além disso, resultados de simulações mostram que os pré-codificadores propostos apresentam uma relação linear com o parâmetro de projeto que controla a quantidade de antenas selecionadas; tal relação independe do número de antenas na estação-base e do número de terminais servidos por essa estação.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

JOINT PRECODING AND ANTENNA SELECTION IN MASSIVE MIMO SYSTEMS

Rafael da Silva Chaves

March/2018

Advisor: Wallace Alves Martins

Department: Electrical Engineering

This thesis presents an overview of massive multiple-input multiple-output (MIMO) systems and proposes new algorithms to jointly precode and select the antennas. Massive MIMO is a new technology, which is candidate for comprising the fifth-generation (5G) of mobile cellular systems. This technology employs a huge amount of antennas at the base station and can reach high data rates under favorable, or asymptotically favorable, propagation conditions, while using simple linear processing. However, massive MIMO systems have some drawbacks, such as the high cost related to the base stations. A way to deal with this issue is to employ antenna selection algorithms at the base stations. These algorithms reduce the number of active antennas, decreasing the deployment and maintenance costs related to the base stations. Moreover, this thesis also describes a class of nonlinear precoders that are rarely addressed in the literature; these techniques are able to generate precoded sparse signals in order to achieve joint precoding and antenna selection. This thesis proposes two precoders belonging to this class, where the number of selected antennas is controlled by a design parameter. Simulation results show that the proposed precoders reach a lower bit-error rate than the classical antenna selection algorithms. Furthermore, simulation results show that the proposed precoders present a linear relation between the aforementioned design parameter that controls the signals' sparsity and the number of selected antennas. Such relation is invariant to the number of base station's antennas and the number of terminals served by this base station.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A Brief History of Wireless Communication

The invention of the first wireless communication system is usually credited to the Italian electrical engineer Guglielmo Marconi. Marconi is the inventor of the wireless telegraphy, which is a system that transmits telegraph messages without wire connections, as those employed by electric telegraphy [1]. Actually, this was not a new idea since numerous inventors had been exploring wireless telegraphy, but none had proven technically and commercially successful. The mathematical theory of electromagnetic waves formulated by James C. Maxwell in 1873 [2] and the experimental confirmation of the existence of these waves by Heinrich Hertz in 1888 made possible Marconi's invention in 1984.

In December of 1894, Marconi made an indoor experiment that consisted in ringing a bell on the other side of a room by pushing a telegraphic button on a bench within this room. In the summer of 1895, Marconi continued his experiment outdoors. In his outdoor experiments he was not able to transmit signals over the distance of 0.8 km, which was predicted by Oliver Lodge as the maximum transmission distance reached by radio waves. However, in the same summer, Marconi found out that much larger distances could be achieved as long as the antennas were made taller and the transmitter and receiver were properly grounded. With these improvements, the resulting system was capable of surpassing the distance of 3.2 km and of transmitting over hills [3].

In May of 1897, Marconi did the world's first ever wireless transmission over the open sea. The experiment consisted in sending a message over the Bristol Channel, UK, from Flat Holm island to Lavernock point in Penarth, reaching a range of 6.0 km. In 1901, Marconi made history by using radio waves for transatlantic transmissions. His communication system sent a message from Poldhu, Cornwall, to Signal Hill in St. John's, Newfoundland (now part of Canada). The distance

between the two points was about $3,500$ km [4].

The Canadian inventor Reginald Fessenden conceived the amplitude modulation (AM) for music and voice broadcasting in 1906. In addition, Fessenden also invented the heterodyne receiver, which is able to rectify and receive AM signals. In 1913, the American electrical engineer Edwin H. Armstrong conceived the superheterodyne receiver, which was used in the first broadcast radio transmission in 1920 at Pittsburgh, USA. In 1921, The Detroit Police Department, USA, was the first one to use land mobile wireless communication. In 1929, the Russian inventor and engineer Vladimir Zworykin performed the first experiment of TV transmission. In 1933, Armstrong invented frequency modulation (FM).

The first public mobile telephone service was introduced in June of 1946 at St. Louis, USA. It was a half-duplex system that used 120 kHz of FM bandwidth [5]. This system had a feature known as press to transmit, which means that only one user could talk at a time using the push-to-talk button. In 1958, the launch of the Signal Communication by Orbital Relay Equipment (SCORE) satellite led to a new era of satellite communications. In the 1960s, automatic channel trunking was introduced, enabling the creation of full-duplex. The most important breakthrough for modern mobile communications happened in the 1970s, when AT&T Bell Laboratories introduced the concept of cellular mobile systems [6].

In the last four decades there was a huge explosion in the rising of radio systems. Wireless communication systems migrated from the first-generation (1G) narrowband analog systems in the 1980s, to the second-generation (2G) narrowband digital systems in the 1990s, followed by the third-generation (3G) wideband multimedia systems in the 2000s, up to the ongoing fourth-generation (4G) systems that provide mobile ultra-broadband (rate of Gbps) access. Nowadays, research and development regarding the fifth-generation (5G) systems are being pursued worldwide.

1.2 Cellular Communications

The 1G mobile cellular systems were analog speech communication systems. They employed frequency division multiple access (FDMA) coupled with frequency division duplexing (FDD) schemes, analog FM for speech modulation, and frequency shift keying (FSK) modulation for control signaling, besides providing full-duplex analog voice services. The main system used in 1G standard was the Advanced Mobile Phone Services (AMPS) [7] that was developed by Bell Labs in 1970. This system was mainly deployed at the frequency bands from 450 MHz to 1 GHz, with a bandwidth of 30 kHz per channel, reaching a data rate of 10 kbps.

The 2G mobile cellular systems were deployed in the early 1990s. They provided digital wireline-quality voice service and short message service (SMS). These sys-

tems were featured by digital implementation, unlike the 1G systems. New access methods, such as time division multiple access (TDMA), and code division multiple access (CDMA), were introduced. The main 2G mobile cellular standards were the Global System for Mobile Communication (GSM) [8], and Interim 95 CDMA (IS-95 CDMA) [9]. The 2G systems were mainly deployed at the frequency bands from 900 MHz to 1.9 GHz.

In 1990, the GSM was introduced by European Telecommunications Standards Institute (ETSI). The GSM system was based on FDMA/TDMA/FDD and Gaussian minimum shift keying (GMSK) modulation. The spectrum was divided into channels with bandwidth of 200 kHz. Each channel was time-divided for eight users and reached a data rate of 270.833 kbps.

The TIA/EIA IS-95 standard was introduced in 1993, and the IS-95 revision was released in 1995. IS-95 operated jointly with the analog AMPS, and was the basis for the 2G CDMA distribution. The first IS-95A system was launched at Hong Kong in 1996. This system employed CDMA/FDD with orthogonal quaternary phase shift keying (OQPSK) modulation. It used the same bands as IS-54, but they were 1.2288-MHz wide. IS-95 allowed variable data rate of 1.2 kbps, 2.4 kbps, 4.8 kbps, and 9.6 kbps. IS-95 was significantly more complex than other 2G technologies. IS-95 employed techniques, such as power control, frequency and delay diversity, variable-rate coding, and soft handoff.

The 3G mobile cellular systems are featured by wideband communications. As general requirements, it demanded a data rate of 2 Mbps for stationary mobiles, 384 kbps for a user with pedestrian speed, and 144 kbps in a mobile vehicle. It was the global system supporting global roaming before the 4G. The 3G network uses packet switching, and is typically deployed at the 2 GHz frequency band.

In June of 1998, International Telecommunication Union Radiocommunication Sector (ITU-R) received 11 competing proposals for terrestrial mobile systems, and approved five. Two mainstream 3G standards are Wideband CDMA (WCDMA) and CDMA 2000, which was administrated by Third-generation Partnership Project (3GPP) and Third-generation Partnership Project 2 (3GPP2), respectively. In October of 2007, ITU-R elected to include WiMAX (802.12e) in the IMT-2000 suite of wireless standards.

In 2005, 3GPP approved the further study of six physical layer proposal: multicarrier WCMDA, multicarrier TD-SCDMA, and four orthogonal frequency division multiple access (OFDMA)-based proposals. In 2008, Long-term Evolution (LTE) was publicized in 3GPP Release 8. LTE uses a number of bandwidths scalable from 1.25 MHz to 20 MHz, and both FDD and time division duplexing (TDD) schemes can be used. Both orthogonal frequency division multiplexing (OFDM) and multiple-input, multiple-output (MIMO) technologies are employed to enhance

the data rate to 172.8 Mbps for the downlink and 86.4 Mbps for the uplink.

In 2008, ITU-R specified a set of requirements for 4G standards, which was called the IMT-Advanced specification. These requirements set the peak of data rate to 100 Mpbs for high mobility and 1 Gpbs for low mobility equipments. Since the first-release version LTE support much less than 1 Gbps of peak data rate, it was not capable of attending the IMT-Advanced compliant, but are often branded 4G by service providers.

The spread spectrum technology used in 3G systems is abandoned in all 4G candidate systems and replaced by OFDMA transmission and single carrier with frequency domain equalization (SC-FD) schemes, making it possible to transfer very high data rates despite extensive multipaths. The peak of data rate is further improved with MIMO [5, 9]. Table 1.1 list the main characteristics of each mobile generation.

Table 1.1: Main characteristics of each mobile generation

Generation	**1G**	**2G**	**3G**	**4G**
Application	Analog voice	Digital voice	Digital voice	Wireless Internet
		SMS	Multimedia	
Data Rate	10 kbps	270 kbps	2 Mbps	1 Gbps
Frequency	450 MHz – 1 GHz	900 MHz – 1.9 GHz	1.6 GHz – 2 GHz	2 GHz – 8 GHz
Bandwidth	30 kHz	1.2 MHz	100 MHz	100 MHz
Multiplexing	FDMA	TDMA	CDMA	OFMD
		CDMA	OFDM	OFDMA
MIMO	No	No	Yes	Yes

1.3 What Is the Next Step for Cellular Communications?

The next step for cellular communication is the development of a new standard capable of supplying the increasing demand for communication at high data rates. This new standard is commercially called 5G, which is currently being developed worldwide. The 5G standard will have to account for a variety of services and emerging new applications. Possible scenarios currently envisioned for 5G networks are: very large data rate wireless connectivity, Internet of things (IoT), tactile Internet, and wireless regional area networks, which are now detailed:

- Very large data rate wireless connectivity: Users will be able to download large amounts of data in a short period of time. A typical application is in the

high-definition video streaming services, like Netflix, YouTube, and Twitch. Another application is in games with augmented reality, like Pokemon GO [10–12].

- Internet of things: IoT will be able to remotely control and connect a lot of devices (things), like TVs, washing machines, air conditioners, and lights in a smart house as well as cars, buses, traffic signals, and smartphones in a smart city. These connected things will have limited processing capabilities, forcing them to sporadically transmit small amounts of data. IoT demands for modulations that are robust to time synchronization errors and effective for short-range communications [13, 14].

- Tactile Internet: It refers to real-time cyber-physical tactile control experiments. The tactile Internet will enable humans and machines to interact with the environment, in real time, while on the move. This system requires reliable communication services with small latency. The target latency is in the order of 1 ms, requiring a physical layer (PHY) latency around $200 \sim 300$ μs [15, 16].

- Wireless regional area networks: It is expected that 5G will also play a crucial role by bringing Internet access to sparsely populated areas. In this scenario, the 5G systems will have very low mobility and latency will not be a key requirement [17].

The 5G standard must be flexible in order to supply all those demands and it is unlikely that 5G requirements can be achieved with a mere evolution of the *status quo*. The keys aspects of 5G networks related to PHY are: waveform design, millimeter wave (mmWave), and massive MIMO.

The discussions related to waveform design are seeking for a substitute for OFDM/OFDMA modulations [13, 18]. The most popular proposals are the filter bank multicarrier (FBMC) modulation [18–26], faster-than-Nyquist (FTN)/ time-frequency-packed (TFP) signaling [26, 27], filtered OFDM (f-OFDM) [18, 22, 28, 29], generalized frequency division multiplexing (GFDM) [26, 30], bi-orthogonal frequency division multiplexing (BFDM) [26, 31, 32], universal filtered multicarrier (UFMC) [26, 33]. All of those proposals circumvent some of the OFDM deficiencies. Moreover, all of these waveforms are OFDM-inspired, which is a huge advantage given that the base station's structure of 4G networks may be reused for 5G systems.

The motivation behind the mmWaves is working in an unused portion of the spectrum. While spectrum has become scarce up to microwave frequencies (1.6 up to 30 GHz), it is still available in the mmWave frequencies (30 up to 300 GHz). MmWaves already have a standard (IEEE 802.11ad) and works for applications

such as small-cell backhaul [34]. However, this a subject that is not fully understood. MmWave technologies can be combined with MIMO to enhance the achievable rates [35]. In addition, due to the high operating frequency, digital processing may be hindered in some cases, demanding for analog processing [36]. MmWave technologies have to deal with two major issues: it does not have sufficiently large coverage due to the propagation nature of mmWaves [35, 37], and it does not have support for mobility in non-line-of-sight (NLoS) environment [35, 38].

Massive MIMO is a technology that employs a very large number of antennas at the base station and serves a considerable number of terminals by using the same time-frequency resource [39]. Traditional MIMO systems usually employ up to a maximum of 12 antennas for transmissions, such as in 4G systems, while current massive MIMO proposals consider using hundreds of antennas in the base station. This quantitative change brings a qualitative change, since it opens up new possibilities for massive MIMO transmissions. Massive MIMO systems are able to focus the radiated energy toward the intended directions while minimizing intra- and inter-cell interferences [34]. Massive MIMO systems are able to achieve high data rates by using simple digital linear processing, under favorable or asymptotically favorable propagation [10, 39, 40].

The use of massive MIMO systems does not prevent the use of new waveforms or mmWaves. Contrariwise, these three technologies may be used together. Massive MIMO and FBMC modulation are used together in [20, 24, 41]. Some propagation characteristics of massive MIMO systems simplify the channel equalization for FBMC modulation. The use of mmWaves in massive MIMO considerably reduces the size of the antennas, increasing the number of antennas per m^2 [35].

Although massive MIMO technology is very promising, it also faces some challenges. Massive MIMO systems have to deal with pilot contaminations, which is induced by the limited number of orthogonal pilots generated by the base stations [39, 42]. Massive MIMO systems rely on TDD schemes due to the guarantee of channel reciprocity. However, the uncertainty in the analog components of the radio frequency chains (RFCs) may unbalance the channel reciprocity, requiring a calibration [43, 44]. Massive MIMO is rather different from everything appearing in previous mobile communication standards, demanding for major changes in the design of base stations [34, 45].

Another important issue related to the deployment of massive MIMO systems is the cost of the base station. The increase in the number of antennas at the base station provides an increasing in the number of RFCs as well, resulting in prohibitively high power consumption and base station's cost [46]. The RFCs are basically compound by power amplifiers, analog-to-digital converters (ADCs), and digital-to-analog converters (DACs), phase shifters, and mixers. An attempt to

solve the issues related to the base stations is reducing the peak-to-average-power ratio (PAPR) of the transmitted signals. Massive MIMO signals usually have high PAPR, demanding for high-quality power amplifiers, which are commonly the most expensive components of RFCs. The decrease of PAPR enables the use of low-quality power amplifiers, which reduces the cost related to the base station [47–55].

Another way to deal with the issues related to the base station's cost is using 1-bit quantizers. In general, the base station uses high-precision (e.g., 10 bits or more) ADCs/DACs [56]. The 1-bit quantization reduces the power consumption on RFCs and reduce the complexity of other analogical components, such as power amplifiers [53, 56–59]. The 1-bit quantization is a solution that both increases the energy efficiency and decreases the bases station's cost.

One more alternative to reduce the base station's cost is reducing the number of RFCs by selecting antennas. With a lower number of active antennas, the number of active RFCs is reduced, increasing the energy efficiency and decreasing the base station's cost [55, 60–65]. Antenna selection for massive MIMO is a topic that deserves special attention. The main algorithms used in massive MIMO was originally developed for point-to-point MIMO systems [60, 61] and might not meet the actual massive MIMO requirements. The goal of this thesis is to tackle antenna selection for massive MIMO systems.

1.4 Main Contributions

The main contributions of this work are:

- Providing an overview of the massive MIMO technology;
- Presenting the precoding stage from a beamforming viewpoint;
- Studying a subject not fully tackled in the literature, which is the joint precoding and antenna selection;
- Proposing two new nonlinear precoding algorithms that perform joint precoding and antenna selection;
- Analyzing the precoding algorithms over sparse multipath channels.

1.5 Organization

The text is organized as follows. Chapter 2 aims to provide a brief overview of massive MIMO technology. The chapter highlights some propagation characteristics innate to massive MIMO systems. These characteristics are related to favorable or

asymptotically propagation. Moreover, this chapter presents the signal model for massive MIMO systems.

Chapter 3 summarizes the main precoders and detectors employed in massive MIMO systems. The chapter also presents the importance of the linear precoding and detection algorithms under favorable or asymptotically favorable propagation. Under this type of propagation, the linear algorithms can reach high data rates. Moreover, the chapter presents the precoding stage from a beamforming viewpoint, which makes easier to bring new ideas to the precoder design.

Chapter 4 presents classical algorithms to select antennas in massive MIMO. These algorithms consist in antenna selection via random choice and channel capacity maximization.

Chapter 5 describes a new class of precoders that are used to jointly precode and select the antennas. This new class of precoders aim to produce sparse precoded signals, being called sparsity-aware precoders. In addition, this chapter proposes two new precoding algorithms.

Chapter 6 presents some simulation results. These simulations aim to evaluate the performance of the proposed algorithms. The results are promising: they show that the bit-error rate of the proposed algorithms are close to the benchmarks that do not use antenna selection as long as some mild conditions hold. Furthermore, the results show an unexpected behavior related to the antenna selection: there exists a linear relation between the number of selected antennas and a design parameter of the proposed algorithms.

Chapter 7 draws some conclusions regarding this work and presents some possible future research directions.

1.6 Notation

Throughout the thesis, vectors and matrices are represented in bold face with lower case and uppercase letters, respectively. The symbols $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{N}$ denote the set of complex, real, and natural numbers, respectively. The symbols $\mathbf{0}_{M\times K}$, $\mathbf{1}_M$, and $\mathbf{I}_M$ denote an $M \times K$ matrix with zeros, an all-one vector with length M, and an $M \times M$ identity matrix, respectively. Given $\mathcal{M} = \{1,\ 2,\ \cdots,\ M\}$, the cardinality of this set is $\text{card}\,(\mathcal{M}) = M$.

Given the matrix $\mathbf{A} \in \mathbb{C}^{M\times K}$, the notations $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{*}$, $\mathbf{A}^{\mathrm{H}}$, and $\mathbf{A}^{-1}$ stand for transpose, conjugate, Hermitian transpose, and inverse operations on $\mathbf{A}$, respec-

tively. Matrix $\mathbf{A}$ can be represented as follows:

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MK} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K \end{bmatrix},\end{aligned}$$

where $\mathbf{a}_k \in \mathbb{C}^{M\times 1}$ is the kth column of $\mathbf{A}$.

The scalar $X \in \mathbb{C}$ stands for a random variable, the vector $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$ stands for a random vector, the scalar $x \in \mathbb{C}$ stands for a realization of X, and the vector $\mathbf{x} \in \mathbb{C}^{M\times 1}$ stands for a realization of $\boldsymbol{x}$. The notation $\mathsf{E}[\boldsymbol{x}]$ stands for the expected value of $\boldsymbol{x}$. The notation $\mathrm{Diag}(\mathbf{x})$ stands for the diagonal matrix composed by the elements of $\mathbf{x}$, i.e.,

$$\begin{aligned}\mathbf{X} &= \mathrm{Diag}(\mathbf{x}), \\ &= \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_M \end{bmatrix}.\end{aligned}$$

The support of a vector $\mathbf{x}$ is defined as the index set of its nonzero entries, i.e.,

$$\mathrm{supp}(\mathbf{x}) = \{m \in \mathcal{M} : x_m \neq 0\}.$$

The notation $\|\mathbf{x}\|_p$ for $p \geq 1$ stands for the l_p-norm of $\mathbf{x}$, which is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{m\in\mathcal{M}} |x_m|^p\right)^{1/p}.$$

For $p = 0$, the l_0-norm[1] of $\mathbf{x}$ is defined as the number of nonzero entries of $\mathbf{x}$, i.e.,

$$\|\mathbf{x}\|_0 = \mathrm{card}(\mathrm{supp}(\mathbf{x})).$$

The vector $\mathbf{x}$ is called K-sparse if at most K of its entries are nonzero, i.e., if

$$\|\mathbf{x}\|_0 \leq K.$$

[1]The l_0-norm is not a norm in a mathematical sense, but this nomenclature will be kept to maintain the coherence with the literature.

Chapter 2

Massive MIMO: A Brief Overview

2.1 Introduction

Massive multiple-input, multiple-output, also called large-scale antenna wireless communication system, was first proposed by Marzetta in [39]. As mentioned in Chapter 1, massive MIMO systems arise as a disruptive technology, with very promising results in terms of sum-rate capacity and spectral efficiency [12, 34, 40, 45, 66]. The main concept of massive MIMO is equipping the base station with a large number of antennas and serving multiple terminals using the same time-frequency resource [67]. This chapter presents the basic concepts of this new technology, highlighting the main differences among massive MIMO and other standard MIMO technologies.

2.2 Preliminary Definitions

2.2.1 Communication Links

A communication link is a connection among two or more devices. This connection may be an actual physical channel or a logical channel that uses one or more actual physical channels. In wireless communications, the links can be cast as forward or reverse links.

Forward Link

The forward link is the communication link from a fixed location to a mobile terminal, for instance, the link from a base station to a smartphone. This communication link is also called downlink. In a multi-user scenario, the fixed location has different communication links with different mobile terminals and, in this case, the downlink channel is often called a broadcast channel [68]. In the broadcast channel, each ter-

minal usually receives different data, but there is a special case when the same data are transmitted to all terminals, which is referred to as a multicast channel [67].

Reverse Link

The reverse link is the communication link from a mobile terminal to a fixed location. This communication link is also called uplink. In a multi-user scenario, there are several mobile terminals communicating with the fixed location and, in this case, the uplink channel is often called a multiple-access channel [68].

2.2.2 Duplexing Schemes

Channel access methods are used in cellular networks for dividing forward and reverse communication channels over the same physical communication medium. They are known as duplexing methods, and the main duplexing schemes employed in wireless communications are time-division duplexing and frequency-division duplexing.

Time-division Duplex

Time-division duplexing is the application of time-division multiplexing to separate the forward and reverse data. In TDD operation, the base station learns the uplink channel from uplink pilots sent by terminals. Moreover, because the channel is reciprocal,[1] once the base station has learned the uplink channel, it automatically has a legitimate estimate of the downlink channel, avoiding the transmission of downlink pilots. There is no standard defined for wireless massive MIMO systems yet, but the first option is a TDD operation mode [39, 67]. Hence, all the MIMO systems addressed in this work will be considered operating in TDD scheme.

Frequency-division Duplex

Frequency-division duplexing means that base station and terminals operate at different carrier frequencies, and use frequency-division multiplexing to separate the forward and reverse data. In FDD operation, the terminals learn the downlink channel from pilots sent by the base station, and communicate the estimated channel state information (CSI) back to the base station over a control channel. This feedback can be very costly, except in special cases, such as in line-of-sight (LoS) propagation, when the CSI can be efficiently quantized [67]. To learn the uplink channel, the base station listens to pilots sent by the terminals. There are a few works using FDD operation mode, but this duplexing scheme is not as popular as TDD [69–73].

[1] The impulse response between any two antennas is the same in both directions, for the same time-instant and frequency range of communication.

2.3 Basic Concepts of MIMO Technology

Multiple-input, multiple-output technology can be divided into three categories namely: point-to-point MIMO, multiuser MIMO (MU-MIMO), and massive MIMO. Point-to-point MIMO and MU-MIMO were very popular in previous communication standards, whereas massive MIMO is a strong candidate to be part of the 5G standard.

2.3.1 Point-to-point MIMO

Point-to-point MIMO emerged in the late 90s [74–81] and represents the simplest form of MIMO system, where the base station equipped with an antenna array serves a terminal also equipped with an antenna array. In point-to-point MIMO, different terminals are orthogonally multiplexed. Figure 2.1 depicts a simplified point-to-point MIMO system with an M-antenna base station and a K-antenna terminal.

A common figure of merit for MIMO systems is the link achievable rate, which is also called channel capacity or sum-rate capacity [67, 82]. In the presence of additive white Gaussian noise (AWGN) at the receiver, the following formulas respectively define the link spectral efficiency measured in b/s/Hz at uplink and downlink:

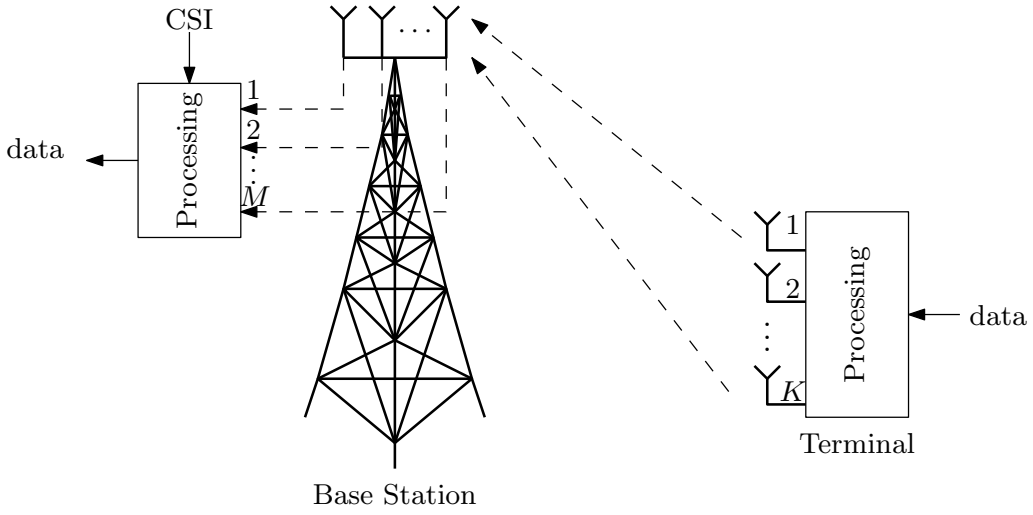
$$C_{\text{ul}} = \log_2 \det\left(\mathbf{I}_M + \frac{\rho_{\text{ul}}}{K}\mathbf{H}\mathbf{H}^{\text{H}}\right), \tag{2.1}$$

$$C_{\text{dl}} = \log_2 \det\left(\mathbf{I}_M + \frac{\rho_{\text{dl}}}{M}\mathbf{H}\mathbf{H}^{\text{H}}\right), \tag{2.2}$$

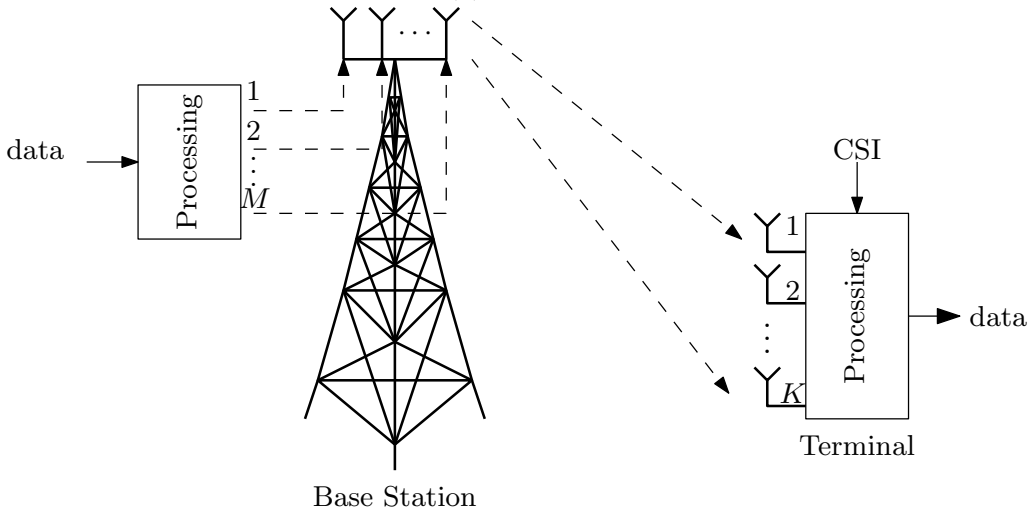
where $\mathbf{H} \in \mathbb{C}^{M\times K}$ is the multiple-access channel matrix, $h_{mk} \in \mathbb{C}$ is the gain between the mth transmitting antenna and kth receiving antenna, $\rho_{\text{ul}} \in \mathbb{R}_+$ and $\rho_{\text{dl}} \in \mathbb{R}_+$ are the reverse link signal-to-noise ratio (SNR) per terminal and the forward link SNR, respectively. The normalizations by M and K mean that, for constant values of ρ_{ul} and ρ_{dl}, the total radiated power is independent of the number of antennas. The channel capacity values in (2.1) and (2.2) require the receiver to know $\mathbf{H}$ but do not require the transmitter to know $\mathbf{H}$ [67, 82]. With complete CSI knowledge at both ends of the link, it is possible to highly improve the related performance [82]. An important fact to be mentioned here is that (2.1) and (2.2) are ideal theoretical bounds, which are calculated assuming ideal channel coding schemes at base station and terminal. Thus, they are rarely achieved in practical situations [67].

In rich scattering propagation environments[2] with sufficiently high SNR values, C_{ul} and C_{dl} scale linearly with $\min(M, K)$ and logarithmically with SNR [82]. Hence, in theory, the link spectral efficiency can be increased by simultaneously using large

[2]Rich scattering means that the receiving antennas receive signal from all directions.



(a) Uplink.



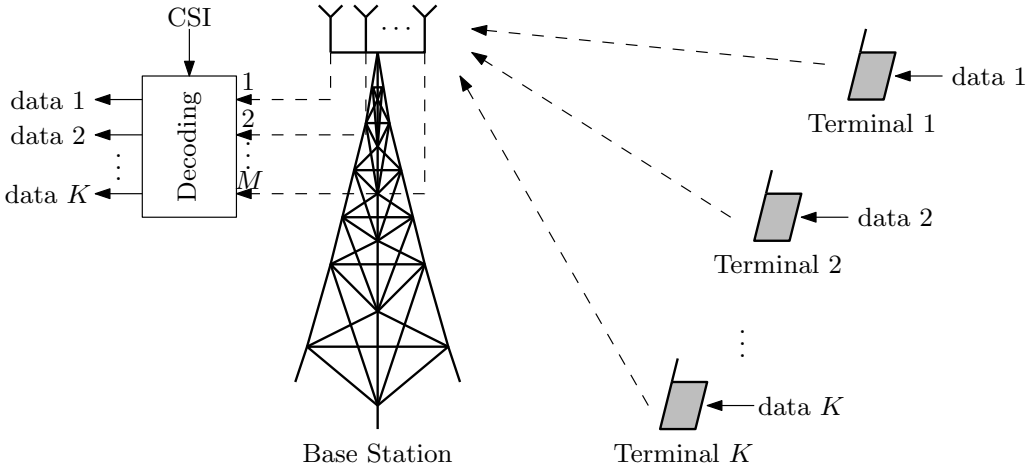
(b) Downlink.

Figure 2.1: Point-to-point MIMO system with an M-antenna base station and a K-antenna terminal.

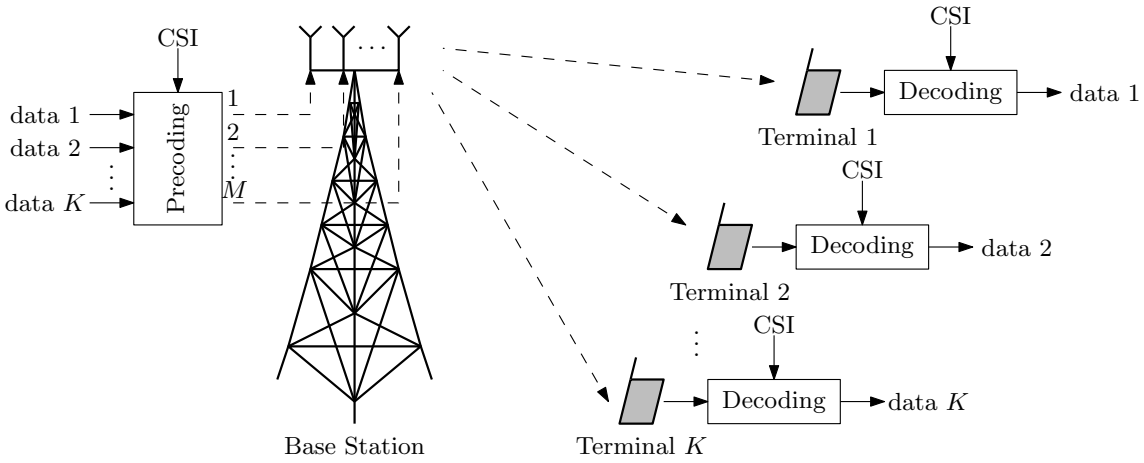
arrays at the transmitter and the receiver. In practice, however, three factors seriously limit the usefulness of point-to-point MIMO, even with large arrays at both ends of the link. First, the terminal equipment requires independent RF chains per antenna as well as the use of advanced digital signal processing to separate data streams, preventing the use of large-scale antenna arrays. Second, the propagation environment must support $\min(M, K)$ independent streams. This is often not the case in practice when compact arrays are used. Third, near the cell edge, where most terminals are usually located, and for which SNRs are typically low due to high path losses, the spectral efficiency scales slowly with $\min(M, K)$ [82].

2.3.2 Multiuser MIMO

MU-MIMO systems enable a single base station to serve a multiplicity of terminals using the same time-frequency resources. In fact, the MU-MIMO scenario can be obtained from the point-to-point MIMO setup by splitting the K-antenna terminal model into multiple autonomous terminals. In general, the terminals in MU-MIMO are single-antenna devices, which are less complex than the K-antenna terminals in point-to-point MIMO. Moreover, the single-antenna terminals are typically separated by many wavelengths, and the terminals cannot collaborate among themselves, either in uplink or downlink. In MU-MIMO, different terminals are spatially multiplexed. Figure 2.2 describes a simplified multiuser MIMO system with an M-antenna base station and K single-antenna terminals.



(a) Uplink.



(b) Downlink.

Figure 2.2: Multiuser MIMO system with an M-antenna base station and K single-antenna terminals.

Assuming a TDD operation, the multiple-access and broadcast channel capacity are given by

$$C_{\text{ul}} = \log_2 \det\left(\mathbf{I}_M + \rho_{\text{ul}}\mathbf{H}\mathbf{H}^{\text{H}}\right), \tag{2.3}$$

and

$$C_{\text{dl}} = \begin{cases} \underset{\mathbf{p}\in\mathbb{R}_+^{K\times 1}}{\text{maximize}} \ \log_2 \det\left(\mathbf{I}_M + \rho_{\text{dl}}\mathbf{H}^{*}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\right) \\ \text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1 \end{cases}, \tag{2.4}$$

where $\mathbf{H} \in \mathbb{C}^{M\times K}$ is the multiple-access channel matrix, $\mathbf{p} \in \mathbb{R}_+^{K\times 1}$ is the power allocation among the users, $\rho_{\text{ul}} \in \mathbb{R}_+$ and $\rho_{\text{dl}} \in \mathbb{R}_+$ are the reverse link SNR per terminal and the forward link SNR, respectively, $\mathbf{1}_K$ stands for an all-one vector with length K, and $\text{Diag}\left(\mathbf{p}\right)$ stands for a diagonal matrix with the elements of $\mathbf{p}$.

The computation of downlink capacity according to (2.4) requires the solution of a convex optimization problem, which appears in many communication applications. Indeed, it is a power allocation problem and can be solved with iterative watter filling algorithms [82, 83]. The derivation of (2.3) and (2.4) assumes CSI knowledge for both uplink and downlink. In the uplink, the base station alone must know the channels, and each terminal has to be aware of its permissible transmission rate separately in order for the capacity in (2.3) to be achieved. In the downlink, both the base station and the terminals must have CSI knowledge in order for the capacity (2.4) to be achieved, as explained in [84]. Obtaining CSI knowledge at both ends of the link might be impracticable, making it hard for the system to achieve the theoretical capacity in practical situations. Additionally, the data rates in (2.3) and (2.4) are calculated assuming an expensive channel coding scheme, which is infeasible in practical situations.

One of the main differences between point-to-point MIMO and MU-MIMO is the cooperative detection of point-to-point MIMO. In MU-MIMO systems there is no cooperation among terminals, preventing sophisticated detection algorithms on forward link. The inability of the terminals to cooperate in the MU-MIMO system does not compromise the multiple-access channel sum-rate capacity as can be straightforwardly verified via the comparison of (2.1) and (2.3).[3] Note also that the broadcast channel capacity in (2.4) may exceed the downlink capacity in (2.2) for point-to-point MIMO, because (2.4) assumes the base station knows $\mathbf{H}$, whereas (2.2) does not. Nonetheless, the reader must keep in mind that CSI knowledge at both ends is necessary to achieve the bounds in (2.3) and (2.4) [67].

MU-MIMO systems have two fundamental advantages over point-to-point MIMO systems. First, it is much less sensitive to assumptions about the propagation en-

[3] Although these expressions are exactly the same, point-to-point MIMO and MU-MIMO are in fact different; for instance, the derivation of (2.3) does not assume cooperation among the terminals.

vironment. Second, MU-MIMO requires only single-antenna terminals. Notwithstanding these virtues, two factors seriously limit the practicality of MU-MIMO in its originally conceived form. First, achieving the capacities in (2.3) and (2.4) requires complicated digital signal processing by both the base station and the terminals. Second, in the downlink, both the base station and the terminals must know $\mathbf{H}$ to achieve the theoretical data rate in (2.4), thus requiring substantial resources to be set aside for transmission of pilots in both directions. It is worth pointing out that practical MU-MIMO systems usually do not possess such information, working below their capacity limits.

2.3.3 Massive MIMO

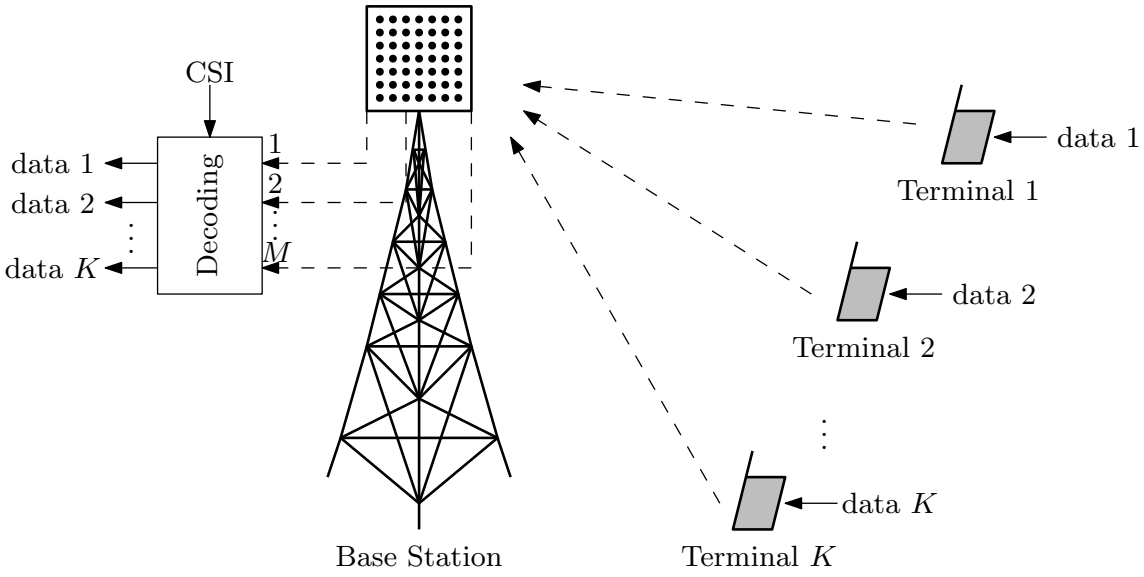
Massive MIMO was originally conceived by Marzetta [39]. Massive MIMO systems are enhanced versions of MU-MIMO systems that aim to overcome the main issues of multiuser MIMO. There are three fundamental distinctions between massive MIMO and conventional MU-MIMO. First, only the base station learns $\mathbf{H}$, so the single-antenna terminals may be cheaper than in MU-MIMO systems. Second, M is typically much larger (typically ranging from 50 to 1000) than K, increasing the sum-rate capacity, while reducing the radiated power by each individual antenna and, simultaneously, increasing the number of terminals that can be served. Third, simple linear digital signal processing is near optimal and it is used in both the uplink and the downlink [39, 40, 42].

Figure 2.3 depicts a simplified single-cell massive MIMO network with an M-antenna base station and K single-antenna terminals. Either in the reverse link or in forward link transmissions, all terminals occupy the full time-frequency resources concurrently. In the reverse link, the base station has to recover the individual signals transmitted by the terminals. In the forward link, the base station has to ensure that each terminal receivers only the signal intended for it. The base station's multiplexing and de-multiplexing signal processing is made possible by utilizing a large number of antennas and by its CSI knowledge.

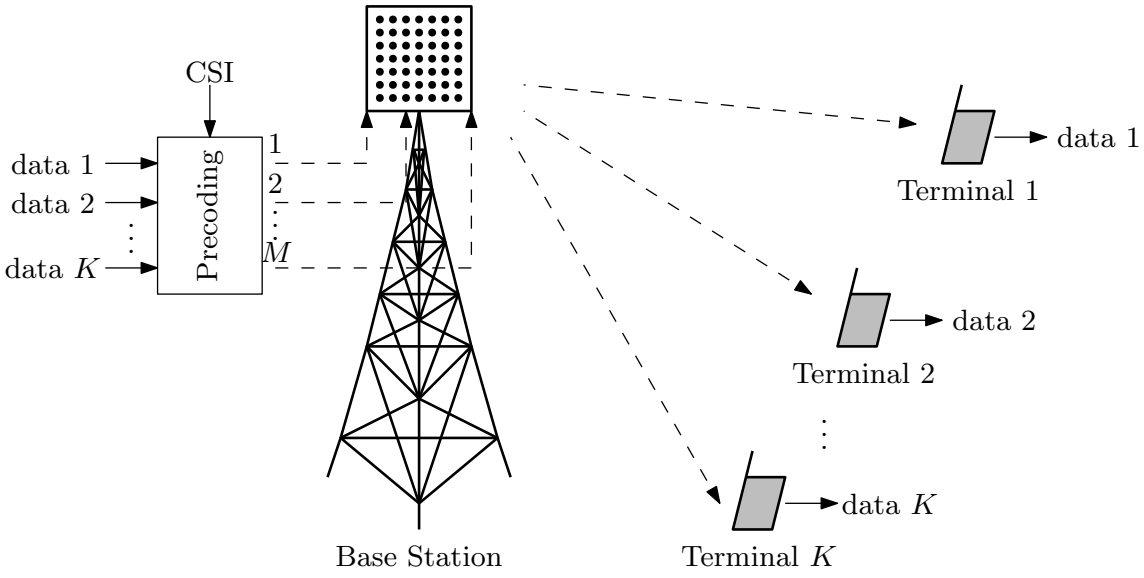
2.3.4 Pilot Signals and Channel Estimation

Point-to-point MIMO, MU-MIMO, and massive MIMO require different degrees of CSI knowledge at the base station and at the terminals. This CSI may be obtained either by estimation from received pilot signals, or by feedback from the receiver to the transmitter, or by combining both strategies.

Learning the channel by sending pilots consumes resources that could otherwise be used to transmit data. To facilitate channel estimation at the receiver, during each segment of the time-frequency plane over the coherence interval, a unique pilot



(a) Uplink.



(b) Downlink.

Figure 2.3: Massive MIMO system with an M-antenna base station and K single-antenna terminals.

waveform needs to be assigned to each transmitting antenna, and all pilots need to be mutually orthogonal. For example, in FDD scheme, if M antennas transmit orthogonal pilots in the forward link, then at least M samples per coherence interval have to be spent on pilots to estimate the equivalent channel.

The number of pilots necessary in each duplexing method is different for each type of MIMO technology. Table 2.1 summarizes the amount of resources consumed by pilot transmission and CSI feedback for point-to-point MIMO, MU-MIMO, and massive MIMO. In Table 2.1, it is possible to see why TDD operation is preferable for massive MIMO, since the number of pilot resources is independent of the number of base station antennas [67]. Moreover, feedback from terminals is entirely avoided.

Consequently, massive MIMO operating in TDD has immeasurable scalability with respect to the number of base station antennas, the cornerstone of massive MIMO concept.

Table 2.1: Resources consumed by pilot transmission

	FDD		**TDD**	
	Uplink	**Downlink**	**Uplink**	**Downlink**
Point-to-Point MIMO (no CSI knowledge)	K pilots	M pilots	K pilots	M pilots
Multiuser MIMO	K pilots + M CSI coefficients	M pilots	K pilots	M pilots
Massive MIMO	K pilots + M CSI coefficients	M pilots	K pilots	not used

Notwithstanding those advantages, massive MIMO has limitations when operating in TDD mode. When the terminals have some mobility, the coherence interval is reduced and there is only time for the creation of a limited set of orthogonal pilots. In a multi-cell scenario, different base stations share some of those pilots, contaminating the channel estimates with information from other cells. This phenomenon is called pilot contamination and it is harmful in multi-cell networks using massive MIMO. Dealing with pilot contamination is a major concern in massive MIMO-system design [42, 85, 86].

2.4 System Model

Consider a generic single-cell massive MIMO wireless communication system, in which K different single-antenna mobile terminals communicate with an M-antenna base station in the uplink, as depicted in Figure 2.4. The signal model for a time-invariant multiple-access channel is given by

$$\mathbf{y} = \sqrt{\rho_{\text{ul}}}\,\mathbf{H}\mathbf{s} + \mathbf{v}, \tag{2.5}$$

where $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is the received signal at the base station, $\mathbf{s} \in \mathbb{C}^{K\times 1}$ is a realization of a random vector $\boldsymbol{s} = [\, S_1 \;\; S_2 \;\; \cdots \;\; S_K \,]^{\mathrm{T}}$ that models the signals transmitted by the terminals, $\rho_{\text{ul}} \in \mathbb{R}_+$ is the SNR for reverse link measured at the base station, $\mathbf{v} \in \mathbb{C}^{M\times 1}$ is a realization of the AWGN random vector $\boldsymbol{v}$, which is assumed to be circularly symmetric complex Gaussian distributed, i.e. $\boldsymbol{v} \sim \mathcal{CN}(\mathbf{0}_{M\times 1}, \mathbf{I}_M)$, and $\mathbf{H} \in \mathbb{C}^{M\times K}$ is the multiple-access channel matrix between the base station antenna array and the set of terminals' antennas. In addition, each terminal is constrained

to have unitary power, i.e.,

$$\mathsf{E}\left[|S_k|^2\right] = 1, \quad \forall\, k \in \mathcal{K}, \tag{2.6}$$

where $\mathcal{K} = \{1,\ 2,\ \cdots,\ K\}$ is the set of the terminals' indexes and $\mathsf{E}[\cdot]$ denotes the expected value.

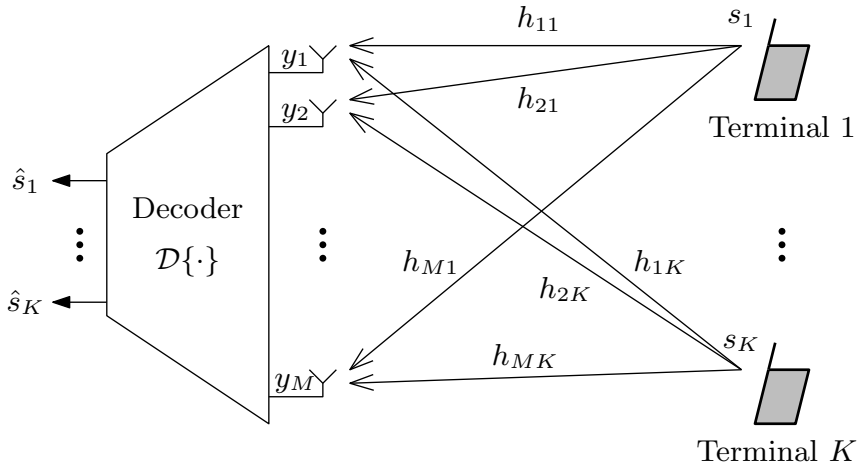


Figure 2.4: Signal model for massive MIMO in uplink.

In order to recover the message sent by the mobile terminals, the base station performs a decoding operation $\mathcal{D}\{\cdot\}$ on the received signal vector $\mathbf{y}$. Then, the reconstructed message $\hat{\mathbf{s}}$ is given by

$$\hat{\mathbf{s}} = \mathcal{D}\left\{\mathbf{y}\right\}. \tag{2.7}$$

There are a lot of possible decoding algorithms for massive MIMO. As will be further described in Section 3.3.1, linear digital signal processing is near optimal for massive MIMO systems in terms of achievable rate. Thus, it is possible to recover $\mathbf{s}$ with low-cost algorithms, while keeping a reasonable performance. Further details regarding linear and nonlinear decoders for massive MIMO systems will be summarized in Section 3.3.

In the downlink, an M-antenna base station communicates with K different single-antenna mobile terminals, as shown in Figure 2.5. Due to the channel reciprocity, the downlink channel matrix is the transpose of the uplink channel matrix. Therefore, the signal model for a time-invariant broadcast channel is given by

$$\mathbf{y} = \sqrt{\rho_{\text{dl}}}\,\mathbf{H}^{\text{T}}\mathbf{x} + \mathbf{v}, \tag{2.8}$$

where $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is the received signal at terminals, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is a realization of the random vector $\boldsymbol{x}$ that models the signal transmitted by the base station, $\rho_{\text{dl}} \in \mathbb{R}_+$ is the SNR for forward link measured at terminals, and $\mathbf{v} \in \mathbb{C}^{K\times 1}$ is a realization

of an AWGN random vector $\boldsymbol{v} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \mathbf{I}_K)$. Furthermore, the total transmit power is independent of the number of antennas, i.e,

$$\mathsf{E}\left[\|\boldsymbol{x}\|_2^2\right] = 1. \tag{2.9}$$

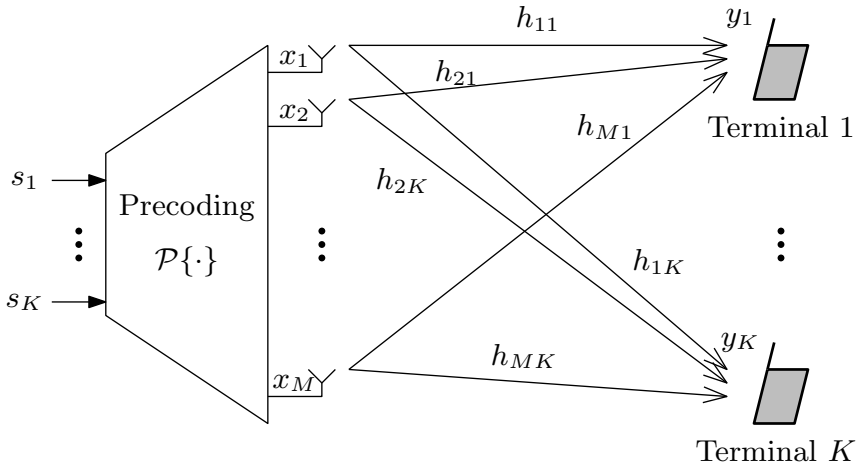


Figure 2.5: Signal model for massive MIMO in downlink.

As shown in Figure 2.5, the terminals do not perform any processing to recover the original message sent by the base station. The base station has to perform a precoding operation, denoted as $\mathcal{P}\{\cdot\}$, on the message $\mathbf{s}$, so that $\mathbf{y} \approx \mathbf{s}$. Hence, the signal $\mathbf{x}$ transmitted by the base station is

$$\mathbf{x} = \mathcal{P}\{\mathbf{s}\}. \tag{2.10}$$

Like decoding, there are many precoding algorithms for massive MIMO and the linear precoding methods are suboptimal solutions as will be described in Section 3.2.1. Further details regarding the most popular precoders for massive MIMO systems will be presented in Section 3.2.

2.5 Propagation in Massive MIMO

Before introducing the main detection and precoding techniques in Chapter 3, it is necessary to describe some propagation characteristics inherent to massive MIMO transmissions. These characteristics are related to the so-called *favorable propagation* that may happen in massive MIMO channels.

2.5.1 Favorable Propagation for Deterministic Channels

Intuitively, to maximize performance from information-theoretic or bit-error rate (BER) perspectives, the uplink channel vectors should be as different as possible,

according to some appropriate metric. This appropriate metric is the so-called *favorable propagation* offered by the channel [67, 87, 88], defined as

$$\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} = 0, \quad k, k' \in \mathcal{K}, \text{ with } k \neq k', \tag{2.11}$$

where $\mathbf{h}_k$ denotes the kth column of the uplink channel matrix $\mathbf{H}$. The result in (2.11) means that the uplink channel vectors of different users are orthogonal.

In practice, the orthogonality requirement in (2.11) usually does not hold, but it can be asymptotically satisfied. In this case, it is said that the environment offers *asymptotically favorable propagation* as long as

$$\frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} \longrightarrow 0, \quad k, k' \in \mathcal{K}, \text{ with } k \neq k', \text{ and } K \ll M \longrightarrow \infty. \tag{2.12}$$

Letting $M \longrightarrow \infty$ has no physical meaning, but taking the limits is useful in order to understand the behavior of the propagation when the number of antennas grows unlimited.

2.5.2 Capacity Upper Bound Under Favorable Propagation

The conditions for favorable propagation described in (2.11) and (2.12) are the preferable scenarios from a channel-capacity perspective. Indeed, the uplink capacity in (2.3) can be written as

$$\begin{aligned} C_{\mathrm{ul}} &= \log_2 \det\left(\mathbf{I}_M + \rho_{\mathrm{ul}}\mathbf{H}\mathbf{H}^{\mathrm{H}}\right) \\ &\overset{(a)}{=} \log_2 \det\left(\mathbf{I}_K + \rho_{\mathrm{ul}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\right) \\ &\overset{(b)}{\leq} \log_2 \left(\prod_{k\in\mathcal{K}} \left\|\mathbf{e}_k + \rho_{\mathrm{ul}}\mathbf{H}^{\mathrm{H}}\mathbf{h}_k\right\|_2\right), \end{aligned} \tag{2.13}$$

where $\mathbf{e}_k \in \mathbb{R}_+^{K\times 1}$ is the kth column of the identity matrix $\mathbf{I}_K$. The Sylvester's determinant theorem [89] is used in (a). This theorem states that, if $\mathbf{A} \in \mathbb{C}^{M\times K}$ and $\mathbf{B} \in \mathbb{C}^{K\times M}$, then

$$\det(\mathbf{I}_M + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_K + \mathbf{B}\mathbf{A}). \tag{2.14}$$

In (b), the Hadamard's inequality [90] is used. This inequality asserts that, if $\mathbf{A} = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_K\,] \in \mathbb{C}^{K\times K}$, then

$$|\det(\mathbf{A})| \leq \prod_{k\in\mathcal{K}} \|\mathbf{a}_k\|_2 . \tag{2.15}$$

According to the Hadamard's inequality, the capacity upper bound in (2.13) is

achieved if only if $\mathbf{H}^{\mathrm{H}}\mathbf{H}$ is diagonal, which happens when the environment induces favorable or asymptotically favorable propagation. Then, under this condition, (2.13) becomes

$$C_{\mathrm{ul}} = \sum_{k \in \mathcal{K}} \log_2 \left(1 + \rho_{\mathrm{ul}} \left\|\mathbf{h}_k\right\|_2^2\right). \tag{2.16}$$

The bound in (2.16) confirms the importance of the favorable propagation condition for massive MIMO systems. Chapter 3 will show that simple digital linear processing is optimum under this condition.

The concept of favorable or asymptotically favorable propagation can also be analyzed for the downlink capacity, but this requires more work, since the corresponding data rate expression in (2.4) involves solving an optimization problem.

2.5.3 Measures of Favorable Propagation

Some channels will not induce favorable or asymptotically favorable propagation. An important question is how far from favorable propagation a given channel model parametrized by matrix $\mathbf{H}$ is. There is a common measure for quantify this, namely: the *distance from favorable propagation* [87, 88].

Distance from Favorable Propagation

The first measure is the "distance" from favorable propagation. This measure uses the ratio between the sum-rate capacity in (2.3) and the upper bound in (2.16), i.e.,

$$\zeta_{\mathrm{C}} = \frac{\log_2 \det \left(\mathbf{I}_M + \rho_{\mathrm{ul}} \mathbf{H}\mathbf{H}^{\mathrm{H}}\right)}{\sum\limits_{k \in \mathcal{K}} \log_2 \left(1 + \rho_{\mathrm{ul}} \left\|\mathbf{h}_k\right\|_2^2\right)}. \tag{2.17}$$

Another measure is the SNR increase that would be needed for the channel capacity offered by $\mathbf{H}$ to reach the upper bound in (2.16), i.e., one must find $\zeta_\rho \in \mathbb{R}_+$ that satisfies the following equation:

$$\sum_{k \in \mathcal{K}} \log_2 \left(1 + \rho_{\mathrm{ul}} \left\|\mathbf{h}_k\right\|_2^2\right) = \log_2 \det \left(\mathbf{I}_M + \zeta_\rho \rho_{\mathrm{ul}} \mathbf{H}\mathbf{H}^{\mathrm{H}}\right). \tag{2.18}$$

Actually, these two measures are not distances strictly speaking, but they are referred to as distances in the literature.

2.5.4 Favorable Propagation for Random Channels

The concept of favorable propagation was presented for a deterministic multiple-access channel $\mathbf{H}$, but in practice, $\mathbf{H}$ will be a realization of a random matrix $\boldsymbol{H}$ due to the stochastic nature inherent to fading. Hence, it is of paramount importance to

examine if favorable propagation takes place on average. There are some alternatives to perform this analysis, for instance, by studying the distribution of the singular values of $\boldsymbol{H}$, or the probability that $\sigma_{\max}(\boldsymbol{H})/\sigma_{\min}(\boldsymbol{H})$ falls below a given threshold. Moreover, the aforementioned distances $\zeta_{\mathrm{C}}(\boldsymbol{H})$ and $\zeta_{\rho}(\boldsymbol{H})$ may also be used as well as their probability to fall below a given threshold. Furthermore, another way to evaluate the favorable propagation is analyzing the behavior of $\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{h}_{k'}$ on average. The favorable propagation will be analyzed for two particular scenarios: independent Rayleigh fading (rich scattering) channel and spatial multipath channel.

Independent Rayleigh Channel

In this scenario the system operates in a dense, rich scattering environment with signal being received from all directions, as illustrated in Figure 2.6. In a rich scattering environment, the multiple-access channel gain between the kth single-antenna terminal and the mth base station antenna is denoted as $h_{mk} \in \mathbb{C}$. This gain can be split into two terms: a complex-valued small-scale fading (or fast fading) coefficient times a large-scale fading coefficient that embodies both range-dependent pathloss (or geometric fading) and shadow fading, i.e.,

$$h_{mk} = g_{mk}\sqrt{\beta_k}, \quad \forall\ (m,k) \in \mathcal{M} \times \mathcal{K}, \tag{2.19}$$

where $\mathcal{M} = \{1,\ 2,\ \cdots,\ M\}$ is the set of the base station antennas' indexes, $g_{mk} \in \mathbb{C}$ is the small-scale fading coefficient, and $\beta_k \in \mathbb{R}_+$ is the large-scale fading coefficient. Both g_{mk} and β_k are realizations of random variables G_{mk} and B_k. The small-scale fading coefficients are assumed to be different for different users and for each different antennas at the base station, whereas the large-scale fading coefficients are the same for different antennas at the base station, but are user-dependent.

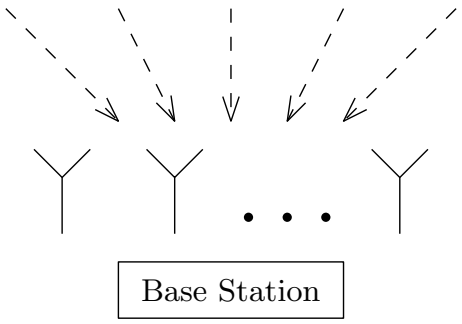


Figure 2.6: Base station located in a propagation environment with rich scattering.

Small-scale fading models range-dependent phase shifts as well as constructive and destructive interferences among different propagation paths. These phenomena happen over intervals of a wavelength or less [40]. The small-scale fading coefficients

are usually assumed to be i.i.d.[4] and drawn from a circularly symmetric complex Gaussian distribution,[5] i.e., $G_{mk} \sim \mathcal{CN}(0,1)$. Rayleigh fading comes as a byproduct of simple physical models. For instance, in rich scattering, the small-scale fading coefficient represents the combined effect of many independent propagation paths; hence, by the superposition principle and the central limit theorem, they will be approximately circularly symmetric complex Gaussian random variables [67].

The large-scale fading coefficient usually is assumed to be constant due to the slow variation of the geometric and shadow fading over the space [39]. Anyway, a possible model for the large-scale fading coefficient is

$$\beta_k = \frac{z_k}{r_k^{\delta}}, \tag{2.20}$$

where $r_k \in \mathbb{R}_+$ is the distance between the kth terminal and the base station, $\delta \in \mathbb{R}_+$ is the decay exponent, and z_k is the realization of a random variable Z_k that models the shadow fading and is log-normally distributed, i.e., $\ln(Z_k) \sim \mathcal{N}(0, \sigma^2_{Z_k})$.

A realization of the multiple-access channel matrix between the base station antenna array and the set of antenna terminals is denoted as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_K \end{bmatrix}, \tag{2.21}$$

where $\mathbf{h}_k = [h_{1k}\ h_{2k}\ \cdots\ h_{Mk}]^{\mathrm{T}}$ is the uplink channel vector of the kth user. Taking into account the model in (2.19), the uplink channel matrix can also be represented in terms of the small-scale fading matrix and the large-scale matrix as

$$\mathbf{H} = \mathbf{G}\mathrm{Diag}\left(\boldsymbol{\beta}\right)^{1/2}, \tag{2.22}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_K \end{bmatrix}, \tag{2.23}$$

with $\mathbf{g}_k = [g_{1k}\ g_{2k}\ \cdots\ g_{Mk}]^{\mathrm{T}}$ being the small-scale fading vector of the kth user, and with $\boldsymbol{\beta} = [\beta_1\ \beta_2\ \cdots\ \beta_K]^{\mathrm{T}}$ denoting the large-scale fading vector. Asymptotically favorable propagation does not hold for independent Rayleigh channel, but it holds in probability when $M \longrightarrow \infty$. Indeed, for independent Rayleigh channel, it is

[4] Independent and identically distributed.

[5] The literature refers to this as Rayleigh fading, despite the small-scale fading coefficients are not drawn from a Rayleigh distribution, but their absolute values. Nevertheless, from now on small-scale fading used in this work will be referred to as Rayleigh fading to keep the coherence with the literature.

possible to write

$$\begin{aligned}\frac{1}{M}\left\|\mathbf{h}_k\right\|_2^2 &= \frac{1}{M}\beta_k \mathbf{g}_k^{\mathrm{H}}\mathbf{g}_k \\ &= \beta_k\left(\frac{1}{M}\sum_{m\in\mathcal{M}}|g_{mk}|^2\right) \\ &\xrightarrow{p} \beta_k, \quad \text{for } M\longrightarrow\infty \text{ and } k\in\mathcal{K}, \end{aligned} \tag{2.24}$$

and

$$\begin{aligned}\frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} &= \frac{1}{M}\sqrt{\beta_k\beta_{k'}}\,\mathbf{g}_k^{\mathrm{H}}\mathbf{g}_{k'} \\ &= \sqrt{\beta_k\beta_{k'}}\left(\frac{1}{M}\sum_{m\in\mathcal{M}} g_{mk}^{*}g_{mk'}\right) \\ &\xrightarrow{p} 0, \quad \text{for } M\longrightarrow\infty, \text{ and } k,k'\in\mathcal{K}, \text{ with } k\neq k', \end{aligned} \tag{2.25}$$

eventually yielding

$$\frac{1}{M}\mathbf{H}^{\mathrm{H}}\mathbf{H} \xrightarrow{p} \mathrm{Diag}\left(\boldsymbol{\beta}\right), \tag{2.26}$$

where the convergence in probability comes from the weak law of the large numbers. Therefore, rich scattering environments induce asymptotically favorable propagation.

Spatial Multipath Channel

The channel model mentioned before considers that the received signals arrive from all directions independently, which means that the environment has rich scattering and no spatial correlation [91]. However, in reality, the received signals may only arrive from some sparse incident angles, which means that the environment has poor scattering and the spatial correlation comes along with the channel sparsity, as illustrated in Figure 2.7.

With the sparsity property of wireless channels, the uplink channel vector $\mathbf{h}_k$ in the spatial domain can be modeled as the superposition of the channel vectors in the angular domain [75]:

$$\mathbf{h}_k = \sum_{n\in\mathcal{N}} g_{kn}\sqrt{\beta_k}\,\mathbf{a}\left(\theta_{kn}\right) \tag{2.27}$$

$$\begin{aligned} &= \sqrt{\beta_k}\left[\mathbf{a}\left(\theta_{k1}\right) \quad \mathbf{a}\left(\theta_{k2}\right) \quad \cdots \quad \mathbf{a}\left(\theta_{kN}\right)\right]\mathbf{g}_k \\ &= \sqrt{\beta_k}\mathbf{A}_k\mathbf{g}_k, \end{aligned} \tag{2.28}$$

where $\mathcal{N} = \{1,\ 2,\ \cdots,\ N\}$ is the set of the multipath indexes, $\theta_{kn} \in [0,\ \pi]$ is a realization of the random variable Θ_{kn} that models the angle of arrival of the

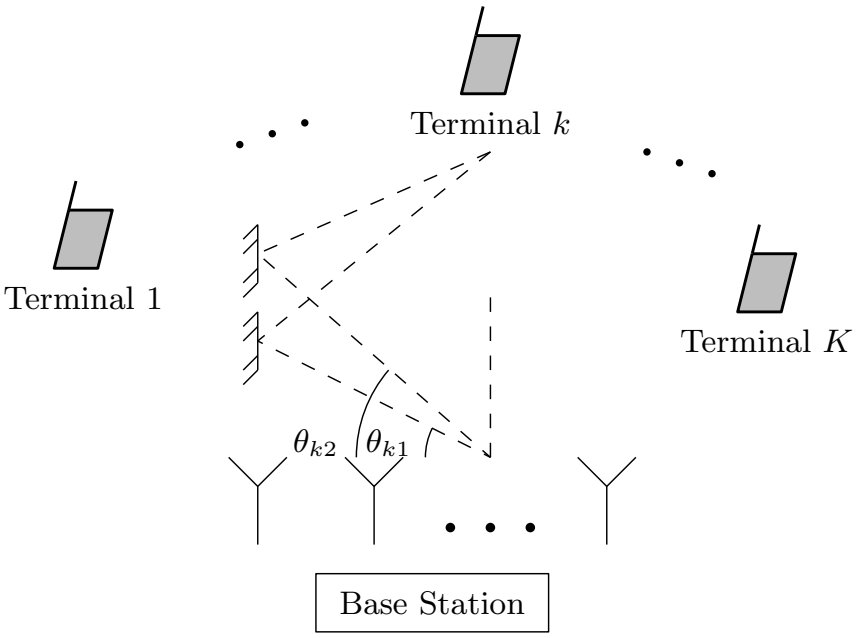


Figure 2.7: Base station located in a propagation environment with multipath.

nth multipath connecting the kth terminal and the base station, whereas $g_{kn} \in \mathbb{C}$ and $\beta_k \in \mathbb{R}_+$ are realizations of the random variables G_{kn} and B_k that model the corresponding small-scale and large-scale fading coefficients, respectively. The random variables G_{kn} and B_k have the same distribution as the small-scale and large-scale fading coefficients in the independent Rayleigh channel case. Vector $\mathbf{a}(\theta) \in \mathbb{C}^{M\times 1}$ is the so-called *array steering vector* [92], which depends on the array geometry. For a uniform linear array (ULA), $\mathbf{a}(\theta)$ is written as [92]

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & \mathrm{e}^{-\mathrm{j}\pi\cos(\theta)} & \cdots & \mathrm{e}^{-\mathrm{j}\pi(M-1)\cos(\theta)} \end{bmatrix}^{\mathrm{T}}. \tag{2.29}$$

By writing $\mathbf{A}_k = [\, \underline{\mathbf{a}}_{k1}\ \underline{\mathbf{a}}_{k2}\ \cdots\ \underline{\mathbf{a}}_{kM}\,]^{\mathrm{T}}$, one has

$$\begin{aligned} \frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} &= \frac{\sqrt{\beta_k\beta_{k'}}}{M}\, \mathbf{g}_k^{\mathrm{H}}\mathbf{A}_k^{\mathrm{H}}\mathbf{A}_{k'}\mathbf{g}_{k'} \\ &= \frac{\sqrt{\beta_k\beta_{k'}}}{M} \sum_{m\in\mathcal{M}} \mathbf{g}_k^{\mathrm{H}}\underline{\mathbf{a}}_{km}^{*}\underline{\mathbf{a}}_{k'm}^{\mathrm{T}}\mathbf{g}_{k'} \\ &= \sqrt{\beta_k\beta_{k'}}\, \mathbf{g}_k^{\mathrm{H}} \underbrace{\left(\frac{1}{M} \sum_{m\in\mathcal{M}} \underline{\mathbf{a}}_{km}^{*}\underline{\mathbf{a}}_{k'm}^{\mathrm{T}} \right)}_{=\mathbf{B}_{k,k'}} \mathbf{g}_{k'} \\ &= \sqrt{\beta_k\beta_{k'}}\, \mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,k'}\mathbf{g}_{k'}. \end{aligned} \tag{2.30}$$

If $k = k'$, then

$$\frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_k = \beta_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{g}_k > 0, \tag{2.31}$$

since $\mathbf{B}_{k,k'}$ is a Hermitian positive-definite matrix and $\mathbf{g}_k \neq 0$. On the other hand, if $k \neq k'$, it is not possible in general to guarantee that $(1/M)\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} \longrightarrow 0$. Nevertheless, one can still write in this case that

$$\frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} = \sqrt{\beta_k\beta_{k'}}\sum_{n\in\mathcal{N}}\sum_{n'\in\mathcal{N}} g_{kn}^{*}\mathbf{B}_{k,k'}(n,n')g_{k'n'}. \tag{2.32}$$

Thus, when the number of multipaths N is sufficiently large, one can state that the following approximation holds in probability:

$$\frac{1}{M}\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_{k'} \approx \sqrt{\beta_k\beta_{k'}}N^2\underbrace{\mathsf{E}\left[G_{kn}^{*}G_{k'n'}\right]}_{=0}\mathsf{E}\left[\boldsymbol{B}_{k,k'}(n,n')\right] = 0, \tag{2.33}$$

in which it is assumed that the random variables Θ_{kn} and G_{kn} are independent allowing one to replace $\mathsf{E}\left[G_{kn}^{*}G_{k'n'}\boldsymbol{B}_{k,k'}(n,n')\right]$ with $\mathsf{E}\left[G_{kn}^{*}G_{k'n'}\right]\mathsf{E}\left[\boldsymbol{B}_{k,k'}(n,n')\right]$.

In summary, spatial multipath channels usually do not induce asymptotically (with respect to the number of antennas) favorable propagation, but when the number of multipaths grows to infinity, one has

$$\frac{1}{M}\mathbf{H}^{\mathrm{H}}\mathbf{H} \xrightarrow{p} \mathrm{Diag}\left(\beta_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{g}_k\right). \tag{2.34}$$

2.6 Conclusion

Massive MIMO is a very promising technology. This chapter presented a summary of the main concepts regarding massive MIMO, pointing out its potential in terms of spectral efficiency and channel capacity. A mathematical description of the uplink and the downlink transmissions was presented. Moreover, some key results concerning the propagation in massive MIMO systems were presented, including the study of favorable and asymptotically favorable propagations. The condition of favorable or asymptotically favorable propagation will play a central role to show the optimality of the linear processing in massive MIMO in the next chapter, which will also address the main precoders and detectors used in massive MIMO.

Chapter 3

Precoding and Detection

3.1 Introduction

This chapter presents a variety of precoding and detection algorithms for massive MIMO systems, namely: matched filter (MF), zero-forcing (ZF), regularized zero-forcing (RZF), minimum mean square error (MMSE), dirty paper coding (DPC), iterative linear filer schemes, random step methods, and tree-based algorithms.

3.2 Precoding

Precoding is a technique which exploits transmission diversity by properly weighing the data stream. This technique will reduce the corrupting effects of the communication channel. For massive MIMO systems, both nonlinear and linear precoding schemes can be used. The function of precoding is almost the same of equalization, but precoding is performed at the transmitter, instead of at the receiver. In massive MIMO, precoding techniques usually aim to maximize the signal-to-interference-plus-noise ratio (SINR). Nonlinear precoding methods, such as dirty paper coding (DPC) [93], vector perturbation [94], and lattice-aided methods [95], have a better performance albeit with higher implementation complexity. In fact, nonlinear precoding techniques are of paramount importance when M is not much larger than K, which is not the case in massive MIMO [40]. Thus, it is more common to use low-complexity linear precoding methods in massive MIMO systems.

3.2.1 Linear Precoding

For linear precoding, the precoding operator $\mathcal{P}\{\cdot\}$ in Figure 2.5 is a matrix $\mathbf{W} \in \mathbb{C}^{M \times K}$. Depending on the application, this matrix can have different purposes, such as right inverting the broadcast channel matrix $\mathbf{H}^{\mathrm{T}}$ or maximizing the SINR related

to the signals received by the terminals. The most common linear precoding methods are the MF, ZF, RZF, and MMSE.

Matched Filter

Matched filter precoding is the simplest linear precoding, where the MF precoding matrix is given by

$$\mathbf{W}_{\text{MF}} = \mathbf{H}^{*}\text{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\text{T}}\right)^{-1/2}\text{Diag}\left(\mathbf{p}\right)^{1/2}. \tag{3.1}$$

This precoder amplifies the signal of interest as much as possible, disregarding interference. If only one terminal were transmitting, this processing would be optimal. Under favorable or asymptotically favorable propagation, MF is also optimal in terms of sum-rate capacity. This result is the cornerstone of massive MIMO theory and is demonstrated bellow.

Under asymptotically favorable propagation, the overall forward link sum-rate capacity for all users becomes

$$\begin{aligned} C_{\text{dl}} &= \begin{cases} \underset{\mathbf{p}\in\mathbb{R}_+^{K\times 1}}{\text{maximize}} \;\; \log_2\det\left(\mathbf{I}_M + \rho_{\text{dl}}\mathbf{H}^{*}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\right) \\ \text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1 \end{cases} \\ &\overset{(a)}{=} \begin{cases} \underset{\mathbf{p}\in\mathbb{R}_+^{K\times 1}}{\text{maximize}} \;\; \log_2\det\left(\mathbf{I}_K + \rho_{\text{dl}}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\mathbf{H}^{*}\right) \\ \text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1 \end{cases} \\ &= \begin{cases} \underset{\mathbf{p}\in\mathbb{R}_+^{K\times 1}}{\text{maximize}} \;\; \log_2\det\left(\mathbf{I}_K + \rho_{\text{dl}}\text{Diag}\left(\mathbf{p}\right)\text{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\text{T}}\right)\right) \\ \text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1 \end{cases}, \end{aligned} \tag{3.2}$$

where (a) uses the Sylvester's determinant theorem. If MF precoding is applied in (2.8), the transmitted precoded signal is given by

$$\mathbf{x}_{\text{MF}} = \mathbf{H}^{*}\text{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\text{T}}\right)^{-1/2}\text{Diag}\left(\mathbf{p}\right)^{1/2}\mathbf{s}. \tag{3.3}$$

Then the received signal vector by the terminals is

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho_{\text{dl}}}\,\mathbf{H}^{\text{T}}\mathbf{H}^{*}\text{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\text{T}}\right)^{-1/2}\text{Diag}\left(\mathbf{p}\right)^{1/2}\mathbf{s} + \mathbf{v} \\ &= \sqrt{\rho_{\text{dl}}}\,\text{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\text{T}}\right)^{1/2}\text{Diag}\left(\mathbf{p}\right)^{1/2}\mathbf{s} + \mathbf{v}. \end{aligned} \tag{3.4}$$

The MF precoding technique separates the signal in the direction of different terminals, avoiding the inter-user interference. Furthermore, the channel capacity

achieved by (3.4) can be maximized by properly choosing $\mathbf{p}$ as in (3.2), which demonstrates that MF precoding is optimum under favorable or asymptotically favorable propagation [39, 96].

The optimality of matched filtering processing in downlink is entirely related to the assumption of favorable or asymptotically favorable propagation. This assumption is based on two conditions: $M \gg K$ and $M \longrightarrow \infty$. The first one is very reasonable and it is not so difficult to hold in practical situations. However, the second one is just a mathematical tool to simplify the theoretical analysis, which does not have any meaning in practice. It is totally fair to ask what is the optimum number of antennas from a cost-effectiveness point of view. Unfortunately, there is no definitive answer and it depends for sure on the details of the propagation, the complexity of the processing, and the cost of antenna elements. In general, works in this area use the number of antennas between 50 and 1000 [40, 42, 56, 57, 97].

Zero-forcing

Zero-forcing precoding is more computationally expensive than MF precoding for it performs a $K \times K$ matrix inversion. The precoded signal by ZF is solution of the following convex optimization problem

$$\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \left\|\mathbf{H}^{\mathrm{T}}\mathbf{x} - \mathbf{s}\right\|_2^2. \tag{P-3.1}$$

The problem (P-3.1) is known as least-squares problem and has infinitely many solutions due to the fact that $\mathbf{H}^{\mathrm{T}}$ is a full-row rank matrix with much more columns than rows. A common choice among these infinity solutions is the minimum l_2-norm solution, which yields the ZF-precoded signal

$$\mathbf{x}_{\mathrm{ZF}} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}\right)^{-1}\mathbf{s}, \tag{3.5}$$

where the ZF precoding matrix is

$$\begin{aligned}\mathbf{W}_{\mathrm{ZF}} &= \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}\right)^{-1} \\ &= \left(\mathbf{H}^{\mathrm{T}}\right)^{\dagger},\end{aligned} \tag{3.6}$$

which is the Moore-Penrose pseudoinverse matrix. Note that ZF precoding inverts perfectly the channel no matter the number of base station antennas, M, unlike the MF precoding. This fact is a significant advantage for ZF precoding because it can guarantee reasonable channel capacity and bit-error rate.

Regularized Zero-forcing

Regularized zero-forcing precoding is very similar to ZF, except the diagonal loading factor added prior to the inversion of the matrix $\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}$. The RZF precoding is also the solution of a convex optimization problem, but now there is a constraint on the power of the precoded signal, i.e., now there is an l_2-norm regularization in the problem. The l_2-norm regularization is known as Ridge/Tikunov regression [98]. The formulation of RZF precoding can be written as

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \left\|\mathbf{H}^{\mathrm{T}}\mathbf{x}-\mathbf{s}\right\|_2^2 \\ &\text{subject to} \ \ \|\mathbf{x}\|_2^2=\xi \end{aligned}, \tag{P-3.2}$$

where $\xi \in \mathbb{R}_{+}$ is the power of $\mathbf{x}$. The solution of (P-3.2) is given by

$$\mathbf{x}_{\mathrm{RZF}} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}+\xi\mathbf{I}_K\right)^{-1}\mathbf{s}, \tag{3.7}$$

and the RZF precoding matrix is given by

$$\mathbf{W}_{\mathrm{RZF}} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}+\xi\mathbf{I}_K\right)^{-1}. \tag{3.8}$$

The RZF precoding performance is bounded by MF and ZF precoding performances. When $\xi \longrightarrow 0$, RZF precoding approaches to ZF precoding, and when $\xi \longrightarrow \infty$ RZF precoding approaches to MF precoding. Thus, RZF precoding can be a flexible alternative to MF and ZF precoders.

Minimum Mean Square Error

Minimum mean square error precoding is formulated as the following convex optimization problem

$$\begin{aligned} &\underset{\mathbf{W}\in\mathbb{C}^{M\times K}}{\text{minimize}} \ \mathsf{E}\left[\left\|\left(\mathbf{H}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{v}\right)-\boldsymbol{s}\right\|_2^2\right] \\ &\text{subject to} \ \mathbf{W}\boldsymbol{s}=\boldsymbol{x} \end{aligned}. \tag{P-3.3}$$

The precoded signal for a given realization $\mathbf{s}$ is then written as

$$\mathbf{x}_{\mathrm{MMSE}} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}+\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I}_K\right)^{-1}\mathbf{s}, \tag{3.9}$$

and the MMSE precoding matrix is given by

$$\mathbf{W}_{\mathrm{MMSE}} = \mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}^{*}+\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I}_K\right)^{-1}. \tag{3.10}$$

The MMSE precoding is a specific case of RZF precoding with $\xi = 1/\rho_{\text{dl}}$. This regularization factor in MMSE precoding takes into consideration the effect of the environment noise, unlike other precoding algorithms described in this section. In terms of bit-error rate, MMSE precoding outperforms MF and ZF precoding in low SNR regimes due the regularization factor based on the SNR. In high SNR regime, the performances of the three precoding schemes tend to be the same.

3.2.2 Nonlinear Precoding

Dirty Paper Coding

In the massive MIMO context, DPC is a nonlinear algorithm used as a precoding scheme to improve the sum-rate capacity, yielding the maximum achievable rate [40, 42]. If the transmitter has perfect knowledge of interference for a given terminal, then the sum-rate capacity is the same as that in the case of no interference to the terminal, or is equivalent to the case where the receiver has perfect knowledge of the interference so that it can subtract it [93]. Based on this idea, the interference can be presubtracted at the transmitter without increasing the transmit power [99]. DPC precoding has practical implementations that are rarely used due to their computational complexity [40].

3.2.3 Precoding as Beamforming

Precoding can be viewed as a generalization of beamforming to support multi-stream transmissions in wireless communication MIMO systems. In conventional linear-in-parameters beamforming, the same signal is emitted from each of the base station antennas with appropriate weighting such that the signal power is maximized at the terminal input, as illustrated in Figure 3.1.

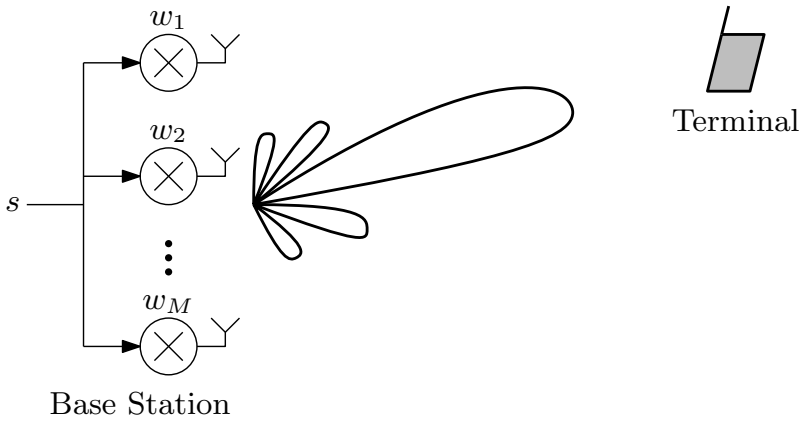


Figure 3.1: Example of a simplified communication system using beamforming.

For a base station having a ULA with M antennas, the beampattern for a given direction $\theta \in [0,\ \pi]$ of a conventional linear-in-parameters beamforming having weighting vector $\mathbf{w} \in \mathbb{C}^{M\times 1}$ can be calculated as

$$b(\theta) = \mathbf{w}^{\mathrm{T}}\mathbf{a}(\theta), \tag{3.11}$$

where $\mathbf{a}(\theta)$ is the underlying steering vector. A plethora of algorithms to project the weighting vector $\mathbf{w}$ can be found in the literature, ranging from algorithms based on simple linear estimation to adaptive algorithms [92, 100–102].

In MU-MIMO systems, there are a lot of terminals and conventional beamforming cannot simultaneously send different signals to each terminal and maximize the signal level at all of the terminals. In order to maximize the throughput in multiple receiving antenna systems, multi-stream transmission is generally required. In multi-stream transmission, the base station has to send different signals in different directions for each terminal. In order to do that it is necessary a precoding operation at the base station. Figure 3.2 shows an MU-MIMO system with an M-antenna base station and K single-antenna terminals using precoding. This figure illustrates the precoding acting in LoS scenario. In the presence of local scattering, the precoding steers the beams in the direction of the multipaths [67].

The precoding in Figure 3.2 can be viewed as M individual conventional linear-in-parameters beamformers, and each one aims to steer the signal in the direction of a terminal. Let x_m be the signal transmitted by the mth antenna of the base station, given as

$$\begin{aligned} x_m &= s_1 w_{m1} + s_2 w_{m2} + \cdots + s_K w_{mK} \\ &= \begin{bmatrix} w_{m1} & w_{m2} & \cdots & w_{mK} \end{bmatrix} \mathbf{s}, \end{aligned} \tag{3.12}$$

where $w_{mk} \in \mathbb{C}$ is the precoding weight for the mth antenna and the kth user signal. The effective transmitted vector is given by

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1K} \\ w_{21} & w_{22} & \cdots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{MK} \end{bmatrix} \mathbf{s} \\ &= \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_K \end{bmatrix} \mathbf{s} \\ &= \mathbf{W}\mathbf{s}. \end{aligned} \tag{3.13}$$

Each $\mathbf{w}_k$ is a weighting vector that steers the desired signal to the respective terminal. The beampattern for a given direction θ and for a precoding matrix $\mathbf{W}$ can be

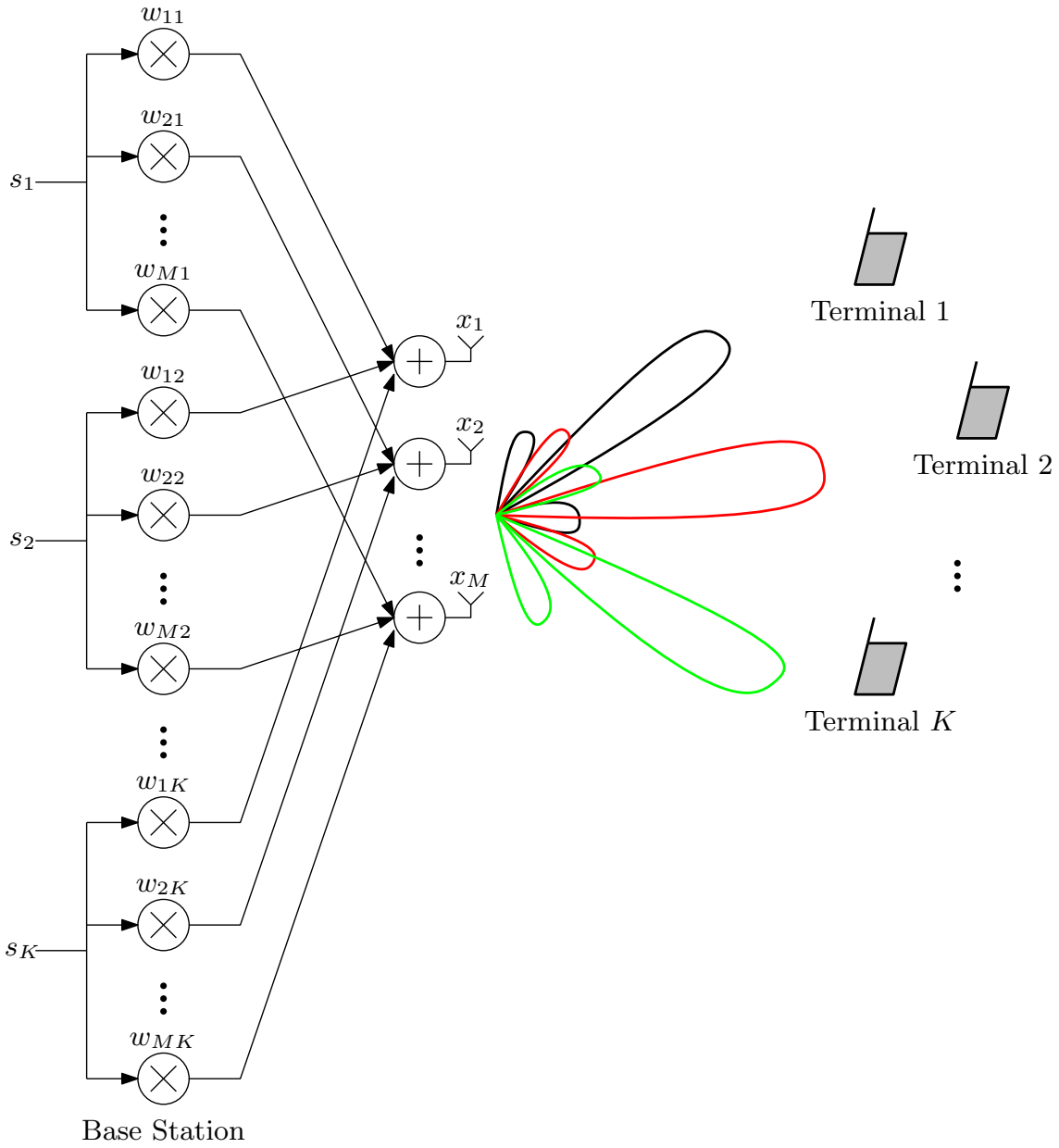


Figure 3.2: Example of a simplified MU-MIMO system using precoding.

calculated as

$$b(\theta) = \mathbf{x}^{\mathrm{T}}\mathbf{a}(\theta) \tag{3.14}$$

$$= \mathbf{s}^{\mathrm{T}} \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}}\mathbf{a}(\theta) \\ \mathbf{w}_2^{\mathrm{T}}\mathbf{a}(\theta) \\ \cdots \\ \mathbf{w}_K^{\mathrm{T}}\mathbf{a}(\theta) \end{bmatrix}$$

$$= \sum_{k\in\mathcal{K}} s_k b_k(\theta), \tag{3.15}$$

where $b_k(\theta) = \mathbf{w}_k^{\mathrm{T}}\mathbf{a}(\theta)$. It is possible to see that $b(\theta)$ is the combination of all $b_k(\theta)$,

which means that the transmitted signal is steered in all terminals' directions. Note that the beampattern could be calculated directly from the weighting vectors $\mathbf{w}_k$, but this could not be possible with nonlinear precoding techniques, which must be calculated from (3.14).

This approach helps analyze the precoding algorithms by another viewpoint, which is the spatial multiplexing provided by the precoders. In Chapter 6, some simulations are performed to evaluate the precoders about the quality of their beampatterns.

3.2.4 Practical Considerations

The aforementioned precoding methods focus on the theoretical aspects of massive MIMO systems, but there are some practical issues that can limit the use of massive MIMO. One of the main issues is the cost of the base station due to many antennas' RF chains. The RF chains contain all analog components in the front-end of a base station, such as power amplifiers, phase shifters, and ADCs/DACs. The RF chains can be reduced using antenna selection algorithms, which is the main focus of this work and will be approached in Chapters 4 and 5. Another alternative is employing low-cost power amplifiers in RF chains. The precoded massive MIMO signals, in general, have a high PAPR, and the use of power-efficient power amplifiers can decrease the signal distortion. For this reason, some precoding algorithms based on per-antenna constant envelope constraint to reduce the PAPR have been developed [47–55]. Those new precoding techniques with low PAPR enable the use of low-cost power amplifiers. There are also some works studying the use of 1-bit ADCs and DACs. Indeed, with 1-bit quantization the PAPR is naturally reduced, working as an alternative to reduce base station's cost [53, 56–59].

3.3 Detection

Similarly to precoding algorithms for massive MIMO systems, simple linear detection algorithms are near optimal under favorable or asymptotically favorable propagation conditions. The most common linear detection algorithms are the same used for precoding, viz.: MF, ZF, and MMSE. The performance of massive MIMO systems using linear detection algorithms has been studied in [85]. In addition to linear detection methods, nonlinear detection algorithms, such as iterative filtering schemes, random step methods, and three based algorithms can be used for detection in massive MIMO systems. Nonlinear algorithms achieve lower levels of bit-error rate than linear detection algorithms, but they have much higher computational complexity.

3.3.1 Linear Detection

For linear detection, the detection operator $\mathcal{D}\{\cdot\}$ in Figure 2.4 is a matrix $\mathbf{D} \in \mathbb{C}^{K\times M}$, which can be used, for instance, to invert the multiple-access channel matrix $\mathbf{H}$ or to maximize the SINR. The most common linear detection algorithms are the MF, ZF, RZF, and MMSE.

Matched Filter

Like in Section 3.2.1, MF detector is the simplest linear detector, where the MF detection matrix is written as

$$\mathbf{D}_{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}. \tag{3.16}$$

Like the MF precoder, under favorable or asymptotically favorable propagation mentioned in Section 2.5, the MF detector is also optimum in terms of achievable rate. This result is demonstrated below.

Applying the MF detection algorithm in (2.5), one has

$$\begin{aligned}\mathbf{H}^{\mathrm{H}}\mathbf{y} &= \sqrt{\rho_{\mathrm{ul}}}\,\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{s} + \mathbf{H}^{\mathrm{H}}\mathbf{v} \\ &= \sqrt{\rho_{\mathrm{ul}}}\,\mathrm{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\mathrm{T}}\right)\mathbf{s} + \mathbf{H}^{\mathrm{H}}\mathbf{v},\end{aligned} \tag{3.17}$$

where $\mathbf{H}^{\mathrm{H}}$ is the MF detection matrix. Note that due to the asymptotically orthogonality of the channel vectors, the MF detector does not color the noise. Moreover, since $\mathrm{Diag}\left(\left[\,\|\mathbf{h}_1\|_2^2 \;\cdots\; \|\mathbf{h}_K\|_2^2\,\right]^{\mathrm{T}}\right)$ is a diagonal matrix, MF separates the signal from different users into different streams, avoiding the inter-user interference. Furthermore, the sum-rate capacity achieved by the matched filtering is the same as in (2.16), which implies that matched filtering is optimum when the number of antennas at the base station, M, grows to infinity.

3.3.2 Nonlinear Detection

Iterative Linear Filtering Schemes

This method works by solving the detection of the received vector $\mathbf{y}$ via iterative linear filtering, using the previous estimate of $\mathbf{y}$ to enhance the current estimate of $\mathbf{y}$. The propagated information can be either hard or soft. The methods typically employ matrix inversions repeatedly during the iterations, which can be very costly when M is large. An alternative to reduce the high computational cost is using the matrix inversion lemma [40]. A popular soft information based detection algorithm is the conditional MMSE with soft interference cancellation (MMSE-SIC) scheme [103]. Another algorithm similar to MMSE-SIC is the block-iterative generalized decision feedback equalizer (BI-GDFE) algorithm [104]. BI-GDFE algorithm can approach

the single user MF within only a few iterations even if the number of antennas is large [42].

Random Step Methods

Random step methods are matrix-inversion free. A basic random step algorithm starts with the initial vector, usually the MMSE solution, and evaluates the mean squared error (MSE) for vectors in its neighborhood. The neighboring vector with smallest MSE is chosen, and the process restarts, continuing like this until achieving the maximum number of iterations. Common random step algorithms are the likelihood ascent search (LAS) algorithm [105] and tabu search (TS) algorithm [106]. The TS algorithm is superior to LAS algorithm because it can avoid local minimum points.

Tree-Based Algorithms

The most prominent algorithm in this class is the sphere decoder (SD) [107]. The SD is, in fact, a maximum likelihood decoder that only considers points inside a sphere with certain radius. If the sphere is too small for finding any appropriate points, it has to be increased. Many tree-based low-complexity algorithms try to reduce the search by only expanding the fraction of the tree-nodes that appear to be the most "promising" ones. Such a method is the stack decoder [108], where the nodes of the tree are expanded in the order of least Euclidean distance to the received signal. Another algorithm in this class is the fixed complexity sphere decoder [109], which is a low complexity, suboptimal version of the SD.

3.4 Conclusion

This chapter presented a big picture of precoding and detecting algorithms, emphasizing their key characteristics. Next chapter addresses a practical issue concerning to massive MIMO wireless systems: the high cost of base stations. To deal with this problem, next chapter introduces a very common technique used in point-to-point MIMO systems: antenna selection.

Chapter 4

Classic Antenna Selection Algorithms

4.1 Introduction

An important problem in massive MIMO systems is related to the base station's cost. The increase in the number of antennas at the base station makes feasible a plethora of theoretical gains, but also imposes many practical challenges [42]. These challenges are mostly related to the RF chains [64, 65, 110–116] and hardware power consumption [47–52, 54, 55]. An RFC contains all analog components before the transmitting antennas, such as power amplifiers, mixers, phase shifters, and ADCs/DACs. There are some alternatives to reduce the base station's cost, like PAPR-awareness precoding and 1-bit quantizer precoding, as briefly discussed in Chapter 3. Moreover, another solution is the hybrid precoders that perform a digital and an analogical processing to reduce the number of RF chains [112–116]. Another technique is selecting specific antennas to transmit the data [60, 61]. The antenna selection technique decreases the number of active RFCs by selecting antennas, which also decreases the number of active analog components. This reduction in the number of active RFCs both increases the energy efficiency and decreases the base station's cost. This antenna selection algorithms was firstly developed to point-to-point MIMO, and the idea has been used in massive MIMO systems as well [62–65]. This chapter presents the main antenna selection algorithms, namely: antenna selection via channel capacity maximization and random selection.

4.2 Reduced-dimension Model and Precoding

4.2.1 Reduced-dimension Model

Consider the single-cell massive MIMO system operating in downlink as depicted in Figure 4.1. This figure illustrates a massive MIMO system with a new processing stage at base station. Now the number of active antennas is reduced and an antenna selection algorithm is run to select the best set of active transmitting antennas. In this case, $L \in \mathbb{N}$ antennas are selected to be active, and the received signal by the single-antenna terminals can be written as

$$\mathbf{y} = \sqrt{\rho_{\mathrm{dl}}}\, \mathbf{H}_L^{\mathrm{T}} \mathbf{z} + \mathbf{v}, \tag{4.1}$$

where $\mathbf{z} \in \mathbb{C}^{L\times 1}$ is a realization of the random vector $\boldsymbol{z}$ that models the reduced-dimension precoded signal, $\rho_{\mathrm{dl}} \in \mathbb{R}_+$ is the SNR for forward link measured at terminal, $\mathbf{v} \in \mathbb{C}^{K\times 1}$ is a realization of an AWGN random vector $\boldsymbol{v} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \mathbf{I}_K)$, and $\mathbf{H}_L^{\mathrm{T}}$ is the reduced-dimension broadcast channel. The subscript L in $\mathbf{H}_L^{\mathrm{T}}$ denotes that the L columns comprising $\mathbf{H}_L^{\mathrm{T}}$ were selected from the complete broadcast channel matrix $\mathbf{H}^{\mathrm{T}}$, corresponding to the L active antennas.[1] As in the signal model in (2.8), the total transmitted power is independent of the number of antennas, i.e.,

$$\mathsf{E}\left[\|\boldsymbol{z}\|_2^2\right] = 1. \tag{4.2}$$

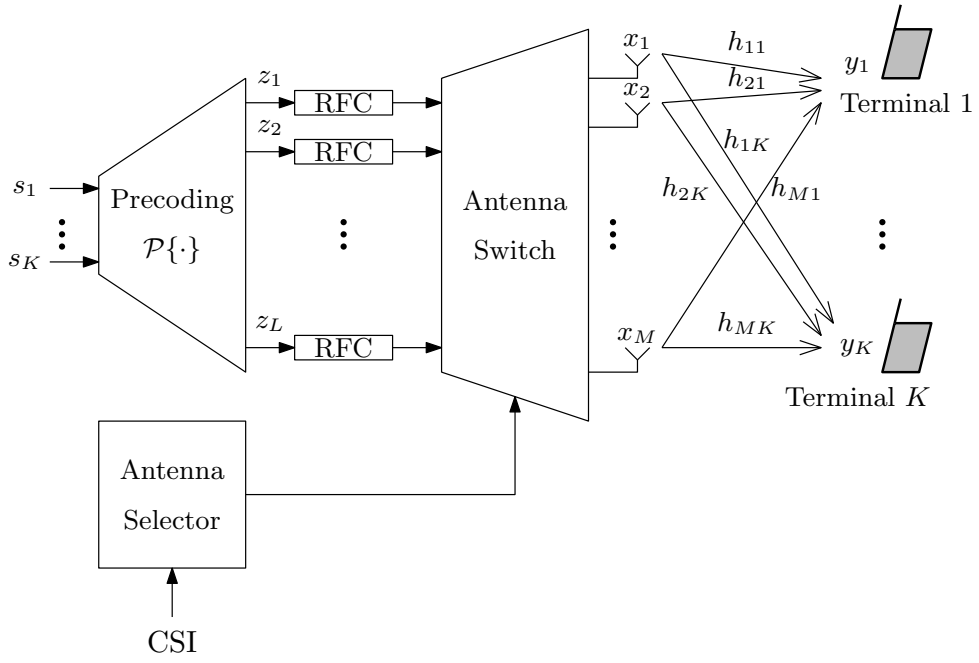


Figure 4.1: Massive MIMO system with antenna selector.

[1] The notation does not specify which columns were kept from the original matrix.

The signal model in (4.1) can also be represented in terms of the complete forward link channel matrix; in this case, the received signal by the terminals can be rewritten as

$$\mathbf{y} = \sqrt{\rho_{\mathrm{dl}}}\,\mathbf{H}^{\mathrm{T}}\mathbf{S}\mathbf{z} + \mathbf{v}, \tag{4.3}$$

where $\mathbf{S} \in \mathbb{R}_{+}^{M\times L}$ is the antenna-selector matrix, which is represented by the antenna-switch block in Figure 4.1. Another representation of (4.1) is in terms of the complete precoded signal, as follows:

$$\mathbf{y} = \sqrt{\rho_{\mathrm{dl}}}\,\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{x} + \mathbf{v}, \tag{4.4}$$

where $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the complete precoded signal, and $\boldsymbol{\delta} \in \mathbb{R}_{+}^{M\times 1}$ is the antenna-selector vector that has the following structure:

$$\delta_m = \begin{cases} 1, \text{ if the } m\text{th antenna is selected} \\ 0, \text{ otherwise} \end{cases}, \ \forall\ m \in \mathcal{M}, \tag{4.5}$$

and satisfying

$$\mathbf{1}_M^{\mathrm{T}}\boldsymbol{\delta} = L. \tag{4.6}$$

4.2.2 The Antenna Selector Matrix

The antenna selector matrix $\mathbf{S}$ introduced in (4.3) is of paramount importance to the algorithms that will be presented in this chapter. The antenna selector matrix is a permuted version of the identity matrix with a reduced number of columns, which are selected from the $\mathrm{supp}\left(\boldsymbol{\delta}\right)$. Moreover, due to the construction of $\mathbf{S}$, the antenna selector matrix and the antenna selector vector have the following property

$$\mathrm{Diag}\left(\boldsymbol{\delta}\right) = \mathbf{S}\mathbf{S}^{\mathrm{T}}. \tag{4.7}$$

The construction of the antenna selector matrix is easier to explain with a toy example, as the following one.

Example 4.1. Consider a MIMO system with $M = 5$ antennas at base station and $L = 3$ active antennas. Consider an antenna selector vector given by

$$\boldsymbol{\delta} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \end{bmatrix}^{\mathrm{T}}. \tag{4.8}$$

First, it is necessary to get the support of $\boldsymbol{\delta}$, which is given by

$$\mathrm{supp}\left(\boldsymbol{\delta}\right) = \{2,\ 4,\ 5\}. \tag{4.9}$$

Then, matrix $\mathbf{S}$ is given by

$$\mathbf{S}= \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_4 & \mathbf{e}_5 \end{bmatrix} \tag{4.10}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.11}$$

thus yielding

$$\mathbf{S}\mathbf{S}^{\mathrm{T}}= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 0 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$

$$= \mathrm{Diag}\left(\boldsymbol{\delta}\right). \tag{4.12}$$

In future analyses, the construction operator $\mathcal{B}\{\cdot,\cdot\}$, which takes the columns of the matrix in the first argument according to the support provided in the second argument, will be used for the sake of simplicity. This operator is responsible for building matrix $\mathbf{S}$ or any matrix with a given structure. Thus, in the previous example, $\mathbf{S}$ can be written as follows:

$$\mathbf{S} = \mathcal{B}\{\mathbf{I}_M, \mathrm{supp}\left(\boldsymbol{\delta}\right)\}. \tag{4.13}$$

4.2.3 Precoding in Reduced-dimension Model

In a massive MIMO system with antenna selection, the precoding algorithm has to use the reduced downlink channel matrix $\mathbf{H}_L$. Although all algorithms of Chapter 3 could be used here for this reduced model, the antenna selection algorithms presented in this chapter will use only MF and ZF precoding algorithms.

Matched Filter

In this case, the MF precoding matrix is given by

$$\mathbf{W}_{\mathrm{MF}} = \left(\mathbf{H}_L^{\mathrm{T}}\right)^{\mathrm{H}}, \tag{4.14}$$

so that the reduced-dimension precoded signal is

$$\begin{aligned} \mathbf{z} &= \mathbf{W}_{\mathrm{MF}}\mathbf{s} && (4.15)\\ &= \left(\mathbf{H}^{\mathrm{T}}\mathbf{S}\right)^{\mathrm{H}}\mathbf{s} \\ &= \mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\mathbf{s}, && (4.16) \end{aligned}$$

whereas the effective transmitted signal is

$$\begin{aligned} \mathbf{x} &= \mathbf{S}\mathbf{z} && (4.17)\\ &= \mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\mathbf{s} \\ &= \mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\mathbf{s}. && (4.18) \end{aligned}$$

Thus, the MF precoding can be computed by either (4.15) followed by (4.17) or directly by (4.18).

Zero-forcing

For ZF precoding, the precoding matrix is

$$\mathbf{W}_{\mathrm{ZF}} = \mathbf{H}_L^{*}\left(\mathbf{H}_L^{\mathrm{T}}\mathbf{H}_L^{*}\right)^{-1}, \tag{4.19}$$

so that the reduced-dimension precoded signal is

$$\begin{aligned} \mathbf{z} &= \mathbf{W}_{\mathrm{ZF}}\mathbf{s} && (4.20)\\ &= \mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\right)^{-1}\mathbf{s} \\ &= \mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right)^{-1}\mathbf{s}, && (4.21) \end{aligned}$$

whereas the effective transmitted signal is

$$\begin{aligned} \mathbf{x} &= \mathbf{S}\mathbf{z} && (4.22)\\ &= \mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right)^{-1}\mathbf{s} \\ &= \mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\left(\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right)^{-1}\mathbf{s}. && (4.23) \end{aligned}$$

4.3 Antenna Selection

The previous section shows the precoders for a given $\mathbf{S}$ and $\boldsymbol{\delta}$, but it does not explain what is the better method to find them. The core question in this chapter is how to find the best (in a sense to be further defined) antenna selector vector $\boldsymbol{\delta}$. The most common algorithms are the random selection and the channel capacity maximization selection.

4.3.1 Random Selection

A possible solution is to perform random antenna selection, which is a naive solution that randomly selects L out of M antennas. However, this strategy may not guarantee the best antenna selector vector. The common method to find the best antenna selector vector $\boldsymbol{\delta}$ is through channel capacity maximization [60, 61] that will be presented in Section 4.3.2. Algorithms 1 and 2 summarize the MF precoding along with random antenna selection (RANDOM-MF), as well as the ZF precoding along with random antenna selection (RANDOM-ZF).

Algorithm 1 : RANDOM-MF

$\text{supp}(\boldsymbol{\delta}) \leftarrow \texttt{randperm}(M, L)$
$\text{supp}(\boldsymbol{\delta}) \leftarrow \texttt{sort}(\text{supp}(\boldsymbol{\delta}))$
$\mathbf{S} \leftarrow \mathcal{B}\{\mathbf{I}_M, \text{supp}(\boldsymbol{\delta})\}$
$\mathbf{H}_L^{\mathrm{T}} \leftarrow \mathbf{H}^{\mathrm{T}}\mathbf{S}$
$\mathbf{W}_{\mathrm{MF}} \leftarrow \mathbf{H}_L^{*}$
$\mathbf{z} \leftarrow \mathbf{W}_{\mathrm{MF}}\mathbf{s}$
$\mathbf{x}_{\text{RANDOM-MF}} \leftarrow \mathbf{S}\mathbf{z}$

Algorithm 2 : RANDOM-ZF

$\text{supp}(\boldsymbol{\delta}) \leftarrow \texttt{randperm}(M, L)$
$\text{supp}(\boldsymbol{\delta}) \leftarrow \texttt{sort}(\text{supp}(\boldsymbol{\delta}))$
$\mathbf{S} \leftarrow \mathcal{B}\{\mathbf{I}_M, \text{supp}(\boldsymbol{\delta})\}$
$\mathbf{H}_L^{\mathrm{T}} \leftarrow \mathbf{H}^{\mathrm{T}}\mathbf{S}$
$\mathbf{W}_{\mathrm{ZF}} \leftarrow \left(\mathbf{H}_L^{\mathrm{T}}\right)^{\dagger}$
$\mathbf{z} \leftarrow \mathbf{W}_{\mathrm{ZF}}\mathbf{s}$
$\mathbf{x}_{\text{RANDOM-ZF}} \leftarrow \mathbf{S}\mathbf{z}$

4.3.2 Channel Capacity Maximization Selection

For the signal model defined in (4.3), the ideal channel capacity is the maximum of the following convex optimization problem, as mentioned in Section 2.3.2:

$$\begin{aligned} &\underset{\mathbf{p}\in\mathbb{R}_{+}^{K\times 1}}{\text{maximize}} \ \log_2 \det\left(\mathbf{I}_M + \rho_{\text{dl}}\mathbf{S}^{\text{T}}\mathbf{H}^{*}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\mathbf{S}\right) \\ &\text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1, \end{aligned} \tag{P-4.1}$$

where $\mathbf{p} \in \mathbb{R}_{+}^{K\times 1}$ is the power allocation vector. Using Sylvester's determinant theorem in (P-4.1) yields

$$\begin{aligned} &\underset{\mathbf{p}\in\mathbb{R}^{K\times 1}}{\text{maximize}} \ \log_2 \det\left(\mathbf{I}_K + \rho_{\text{dl}}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\mathbf{S}\mathbf{S}^{\text{T}}\mathbf{H}^{*}\right) \\ &\text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1. \end{aligned} \tag{P-4.2}$$

Therefore, (P-4.2) can be finally rewritten as

$$\begin{aligned} &\underset{\mathbf{p}\in\mathbb{R}^{K\times 1}}{\text{maximize}} \ \log_2 \det\left(\mathbf{I}_K + \rho_{\text{dl}}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\text{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right) \\ &\text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1. \end{aligned} \tag{P-4.3}$$

Problem (P-4.3) appears often in communications and can be solved using an iterative water filling algorithm, as mentioned in Chapter 2. However, water filling algorithms just solve (P-4.3) in relation to $\mathbf{p}$, assuming a predefined $\boldsymbol{\delta}$. In order to find the optimal antenna selector, (P-4.3) also needs to take into account $\boldsymbol{\delta}$ as an optimization variable. Thus, the optimal antenna selector matrix $\text{Diag}\left(\boldsymbol{\delta}_{\text{o}}\right)$ is solution of the following optimization problem:

$$\begin{aligned} &\underset{\substack{\mathbf{p}\in\mathbb{R}^{K\times 1}\\ \boldsymbol{\delta}\in\mathbb{R}^{M\times 1}}}{\text{maximize}} \ \log_2 \det\left(\mathbf{I}_K + \rho_{\text{dl}}\text{Diag}\left(\mathbf{p}\right)\mathbf{H}^{\text{T}}\text{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right) \\ &\text{subject to } \mathbf{1}_K^{\text{T}}\mathbf{p} = 1, \\ &\qquad\qquad \mathbf{1}_M^{\text{T}}\boldsymbol{\delta} = L, \\ &\qquad\qquad \delta_m \in \{0,\ 1\} \ \forall\ m \in \mathcal{M}. \end{aligned} \tag{P-4.4}$$

Convex Relaxation

Solving (P-4.4) over the power allocation vector $\mathbf{p}$ among the terminals, and the antenna selection vector $\boldsymbol{\delta}$ is a hard task. The problem is no longer convex and there are no computationally efficient algorithms to solve it. Therefore, a common approach to solve (P-4.4) is performing an optimization in two steps:

(i) Firstly, the power distribution among all users is assumed to be equal, i.e., $p_k = 1/K,\ \forall\ k \in \mathcal{K} = \{1,\ 2,\ \cdots,\ K\}$, which allows one to solve the optimization

problem with respect to $\boldsymbol{\delta}$ to find the best antenna selector vector $\boldsymbol{\delta}_{\mathrm{o}}$;

(ii) Then, with the best set of L antennas selected, the optimization over $\mathbf{p}$ is performed to find the maximum sum-rate capacity. For instance using an iterative water filling algorithm.

Although, the strategy above does not guarantee the global solution, because the decoupling of the variables $\mathbf{p}$ and $\boldsymbol{\delta}$, it at least guarantees a lower bound for the channel capacity [65].

Performing step (i) and setting $p_k = 1/K,\ \forall\ k \in \mathcal{K}$, the new optimization problem is given by

$$\begin{aligned} \underset{\boldsymbol{\delta}\in\mathbb{R}^{M\times 1}}{\text{maximize}}\ & \log_2 \det\left(\mathbf{I}_K + \frac{\rho_{\mathrm{dl}}}{K}\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right) \\ \text{subject to}\ & \mathbf{1}_M^{\mathrm{T}}\boldsymbol{\delta} = L, \\ & \delta_m \in \{0,\ 1\}\ \forall\ m \in \mathcal{M}. \end{aligned} \tag{P-4.5}$$

Despite the fact that $\log_2 \det\left(\mathbf{I}_K + (\rho_{\mathrm{dl}}/K)\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right)$ is concave, (P-4.5) is non-convex, for the binary restriction on $\boldsymbol{\delta}$ actually makes the optimization problem NP-hard. NP-hard problems could in principle be solved by exhaustive search, but due to the combinatorial nature of the problem, it is unfeasible to solve it for large values of M, which is the case in practical massive MIMO.

In order to make the problem computationally feasible, a convex relaxation is performed in the binary restriction on $\boldsymbol{\delta}$ [117], yielding

$$\begin{aligned} \underset{\boldsymbol{\delta}\in\mathbb{R}^{M\times 1}}{\text{maximize}}\ & \log_2 \det\left(\mathbf{I}_K + \frac{\rho_{\mathrm{dl}}}{K}\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right) \\ \text{subject to}\ & 0 \leq \delta_m \leq 1\ \forall\ m \in \mathcal{M}, \\ & \mathbf{1}_M^{\mathrm{T}}\boldsymbol{\delta} = L, \end{aligned} \tag{P-4.6}$$

or alternatively

$$\begin{aligned} \underset{\boldsymbol{\delta}\in\mathbb{R}^{M\times 1}}{\text{maximize}}\ & \log_2 \det\left(\mathbf{I}_K + \frac{\rho_{\mathrm{dl}}}{K}\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\left(\boldsymbol{\delta}\right)\mathbf{H}^{*}\right) \\ \text{subject to}\ & 0 \preceq \boldsymbol{\delta} \preceq 1, \\ & \mathbf{1}_M^{\mathrm{T}}\boldsymbol{\delta} = L. \end{aligned} \tag{P-4.7}$$

Now, δ_m can be any real value between 0 and 1, and the best antenna selector $\boldsymbol{\delta}_{\mathrm{o}}$ is obtained by keeping the L largest values of $\boldsymbol{\delta}$, setting them to one and the remaining $M-L$ to zero. After that, the optimal $\mathbf{p}$ can be found with a water filling algorithm.

This convex relaxation yields a near-optimal solution, except for the case when a small number of antennas are selected, i.e., $L \ll M$ [65, 117]. In antenna selection for massive MIMO system, the number of active antennas is reduced via a trade-off

with the achievable data rate. Therefore, in practical situations, L should be relatively large ($L \gg K$) and the convex relaxation will show near-optimal results [65]. Algorithms 3 and 4 summarize the MF precoding along with antenna selection using channel capacity maximization (CM-MF) and the ZF precoding along with antenna selection using channel capacity maximization (CM-ZF).

Algorithm 3 : CM-MF

$$\boldsymbol{\delta} \leftarrow \underset{\substack{0 \preceq \boldsymbol{\delta} \preceq 1 \\ \mathbf{1}_M^{\mathrm{T}} \boldsymbol{\delta} = L}}{\operatorname{argmax}} \log_2 \det \left(\mathbf{I}_K + \frac{\rho_{\mathrm{dl}}}{K} \mathbf{H}^{\mathrm{T}} \operatorname{Diag}(\boldsymbol{\delta}) \mathbf{H}^{*} \right)$$

$[\boldsymbol{\delta}_{\mathrm{o}}, \texttt{indexes}] \leftarrow \texttt{sort}(\boldsymbol{\delta}, \texttt{'descend'})$

$\boldsymbol{\delta}_{\mathrm{o}}(\texttt{indexes}(1:L)) \leftarrow 1$

$\boldsymbol{\delta}_{\mathrm{o}}(\texttt{indexes}(L+1:\texttt{end})) \leftarrow 0$

$\mathbf{S} \leftarrow \mathcal{B}\{\mathbf{I}_M, \operatorname{supp}(\boldsymbol{\delta}_{\mathrm{o}})\}$

$\mathbf{H}_L^{\mathrm{T}} \leftarrow \mathbf{H}^{\mathrm{T}} \mathbf{S}$

$\mathbf{W}_{\mathrm{MF}} \leftarrow \mathbf{H}_L^{*}$

$\mathbf{z} \leftarrow \mathbf{W}_{\mathrm{MF}} \mathbf{s}$

$\mathbf{x}_{\text{CM-MF}} \leftarrow \mathbf{S}\mathbf{z}$

Algorithm 4 : CM-ZF

$$\boldsymbol{\delta} \leftarrow \underset{\substack{0 \preceq \boldsymbol{\delta} \preceq 1 \\ \mathbf{1}_M^{\mathrm{T}} \boldsymbol{\delta} = L}}{\operatorname{argmax}} \log_2 \det \left(\mathbf{I}_K + \frac{\rho_{\mathrm{dl}}}{K} \mathbf{H}^{\mathrm{T}} \operatorname{Diag}(\boldsymbol{\delta}) \mathbf{H}^{*} \right)$$

$[\boldsymbol{\delta}_{\mathrm{o}}, \texttt{indexes}] \leftarrow \texttt{sort}(\boldsymbol{\delta}, \texttt{'descend'})$

$\boldsymbol{\delta}_{\mathrm{o}}(\texttt{indexes}(1:L)) \leftarrow 1$

$\boldsymbol{\delta}_{\mathrm{o}}(\texttt{indexes}(L+1:\texttt{end})) \leftarrow 0$

$\mathbf{S} \leftarrow \mathcal{B}\{\mathbf{I}_M, \operatorname{supp}(\boldsymbol{\delta}_{\mathrm{o}})\}$

$\mathbf{H}_L^{\mathrm{T}} \leftarrow \mathbf{H}^{\mathrm{T}} \mathbf{S}$

$\mathbf{W}_{\mathrm{ZF}} \leftarrow \left(\mathbf{H}_L^{\mathrm{T}}\right)^{\dagger}$

$\mathbf{z} \leftarrow \mathbf{W}_{\mathrm{ZF}} \mathbf{s}$

$\mathbf{x}_{\text{CM-ZF}} \leftarrow \mathbf{S}\mathbf{z}$

4.4 Conclusion

This chapter described the most common antenna selection algorithms. Antenna selection based on channel capacity maximization finds the best selector through a convex optimization problem, but it presents a few issues. As discussed in Chapter 3, practical systems commonly use simple linear precoding techniques to produce the precoded signal $\mathbf{x}$. However, the optimum antenna selector $\boldsymbol{\delta}_{\mathrm{o}}$ was found using the ideal downlink sum-rate capacity. Therefore, this selector may not be the optimum for the linear precoders as MF and ZF. Additionally, the antenna selection is performed using the channel capacity as metric, but this strategy may not be the

best choice for other practical metrics, such as bit-error rate. Furthermore, this antenna selection algorithm does not avoid the necessity of precoding at base station, increasing the data processing performed by the base station. Next chapter proposes two new algorithms to perform joint antenna selection and signal precoding, focusing on the minimization of the estimation error.

Chapter 5

Joint Precoding and Antenna Selection

5.1 Introduction

The previous chapter presented some algorithms to select antennas in massive MIMO systems. The main classic solution finds the best set of active antennas by maximizing the sum-rate capacity achieved by ideal coding schemes. However, this algorithm does not have any integration with the precoder, which, in general, is far from the ideal. This lack of integration between the precoding and the antenna selection algorithm forces the base station to run two algorithms instead of one, which may increase the complexity in the base station, besides the inherent suboptimality of the uncoupled approach.

Considering those aspects, this chapter proposes new algorithms that interpret the antenna selection problem as a sparse recovery problem. In this approach, the selected antennas are the indexes of the nonzero entries of the precoded vector. Hence, the proposed approach performs joint precoding and antenna selection, which may reduce the complexity in base stations.

5.2 Sparse Estimation Problem

Consider the parameter estimation problem for the model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{5.1}$$

where $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the parameter vector to be estimated, $\mathbf{A} \in \mathbb{C}^{K\times M}$ is the measurement matrix with full row-rank and with $K < M$, $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is the measured data vector, and $\mathbf{v} \in \mathbb{C}^{K\times 1}$ the measurement noise. A possible solution to this problem is finding an $\mathbf{x}$ that minimizes the residue corresponding to the signal model in

(5.1), i.e., finding a solution to the least-squares problem

$$\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 . \tag{P-5.1}$$

The least-squares problem is a convex optimization problem, with infinite many solutions since matrix $\mathbf{A}$ does not have full column-rank; the most commonly used value for $\mathbf{x}$ is the one which has minimum l_2-norm. Therefore, the least-squares solution with minimum l_2-norm is given by

$$\mathbf{x}_{\text{LS}} = \mathbf{A}^{\dagger}\mathbf{y}, \tag{5.2}$$

as already seen in Chapter 3 for the ZF precoding.

Equation (5.1) appears frequently in sparse recovery problems and for the vast majority of them the parameter vector $\mathbf{x}$ is supposed to be sparse. However, the least-squares solution does not feature any sparsity in general. Finding sparse solutions for (5.1) encompasses an entire area of research known as sparse recovery/estimation problems.

The goal in sparse recovery problems is to find the sparsest vector $\mathbf{x}$, satisfying some constraints, such as $\mathbf{A}\mathbf{x}=\mathbf{y}$. The popular algorithms used in sparse recovery problems are divided into three classes: greedy methods, thresholding-based methods, and optimization methods. Greedy methods are based on the matching pursuit algorithm [118–120], which computes the best matching projections of multidimensional data onto the span of an over-complete dictionary. The most common greedy algorithms are the orthogonal matching pursuit (OMP) [121–123], compressive sampling matching pursuit (CoSaMP) [124], and subspace pursuit [125]. The most common thresholding-based algorithms are the basic thresholding [126], iterative hard thresholding (IHT) [127], and hard thresholding pursuit (HTP) [128]. For optimization methods, the main algorithms are the basis pursuit (BP) [129], quadratically constrained basis pursuit (QCBP) [126], basis pursuit denoising (BPDN) [130], least absolute shrinkage and selection operator (LASSO) [131], and Dantzig selector [132]. Table 5.1 summarizes the common algorithms used in sparse recovery problems.

Greedy and thresholding-based methods are out of the scope of this work, since they do not have the same flexibility presented by optimization methods when one wants to ensure a minimum sparsity level in the solution. Section 5.3 will clarify the importance of limiting the sparsity and will show how optimization can handle this constraint. More details on greedy and thresholding-based methods can be found in [126] and references therein.

The optimization methods for sparse recovery are derived from the following

Table 5.1: Main algorithms in sparse recovery problems

Greedy Methods	**Thresholding Methods**	**Optimization Methods**
OMP [121–123]	Basic Thresholding [126]	BP [129]
CoSaMP [124]	IHT [127]	QCBP [126]
Subspace Pursuit [125]	HTP [128]	BPDN [130]
		LASSO [131]
		Dantzig Selector [132]

problem

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{x}\|_0 \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}. \end{aligned} \tag{P-5.2}$$

Problem (P-5.2) has the same difficulty presented by (P-4.5): it is NP-hard and, therefore, it is infeasible to solve it in practical cases for large dimensions. The natural approach for solving (P-5.2) is to use some convex relaxation. The most common relaxation for (P-5.2) is to replace the l_0-norm with the l_1-norm, yielding

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}. \end{aligned} \tag{P-5.3}$$

This principle is called l_1-minimization or basis pursuit [129]. In general, l_1-minimizers are sparse for real-valued entries of the optimization variable. However, basis pursuit may not result in sparse solutions for complex-valued entries [126]. The problems (P-5.2) and (P-5.3) can be equivalent in some cases. For instance, (P-5.2) and (P-5.3) are said to be strongly equivalent if (P-5.2) has a unique solution which coincides with the unique solution of (P-5.3). And a sufficient and necessary condition for this to hold is that the range space property holds. Details regarding this topic can be found in [126] and references therein.

Another alternative problem to deal with sparse recovery is a variation of basis pursuit that extends l_1-minimization taking into account the effect of measurement error. The new problem is given by

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ &\text{subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \eta, \end{aligned} \tag{P-5.4}$$

where $\eta \in \mathbb{R}_+$ is the l_2-norm regularization factor. This variation is natural because the measured vector $\mathbf{y}$ is not exactly equal to $\mathbf{Ax}$ in general, as seen in (5.1). Problem (P-5.4) is commonly known as quadratically constrained basis pursuit or as noise-aware l_1-minimization [126]. The solution of (P-5.4) is related to the output of the basis pursuit denoising [130], which consists in solving

$$\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{Ax}-\mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1, \tag{P-5.5}$$

where $\lambda \in \mathbb{R}_+$ is the l_1-norm regularization factor. The solution of (P-5.4) is also linked to the solution of the LASSO estimator [131], which consists in solving

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{Ax}-\mathbf{y}\|_2^2 \\ &\text{subject to} \ \|\mathbf{x}\|_1 \leq \tau \end{aligned}, \tag{P-5.6}$$

where $\tau \in \mathbb{R}_+$ is the l_1-norm regularization factor.

QCBP, BPDN, and LASSO can have equivalent solutions, but they have different motivations and interpretations. QCBP wants to find the sparsest $\mathbf{x}$ that still satisfies a constraint on the noise energy. BPDN minimizes the noise power, regularizing the objective function with the l_1-norm of the solution $\mathbf{x}$. Finally, LASSO minimizes the noise power subject to a constraint on the l_1-norm of the solution $\mathbf{x}$.

Another type of l_1-minimization problem is the Dantzig selector [132], which is given by

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ &\text{subject to} \ \left\|\mathbf{A}^{\mathrm{H}}(\mathbf{Ax}-\mathbf{y})\right\|_\infty \leq \gamma, \end{aligned} \tag{P-5.7}$$

where $\gamma \in \mathbb{R}_+$ is the l_∞-norm regularization factor. The intuition behind the constraint in (P-5.7) is that the residual $\mathbf{Ax}-\mathbf{y}$ should have small correlation with the columns $\mathbf{a}_m$ of the matrix $\mathbf{A}$.

5.3 Sparsity-aware Precoding Algorithms

Consider the signal model in (4.3), which is written again here for the reader's convenience:

$$\mathbf{y} = \sqrt{\rho_{\mathrm{dl}}}\,\mathbf{H}^{\mathrm{T}}\mathrm{Diag}\,(\boldsymbol{\delta})\,\mathbf{x} + \mathbf{v},$$

where $\mathbf{H}^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ is the broadcast channel matrix, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the precoded signal, $\boldsymbol{\delta} \in \mathbb{R}_+^{M\times 1}$ is the antenna selector, $\rho_{\mathrm{dl}} \in \mathbb{R}_+$ is the SNR for forward link measured at terminal, and $\mathbf{v} \in \mathbb{C}^{K\times 1}$ is the noise vector. For massive MIMO downlink, (4.3) is also an undetermined system like (5.1), since $M \gg K$. Additionally,

in antenna selection $\mathrm{Diag}(\boldsymbol{\delta})\,\mathbf{x}$ is an L-sparse vector, where the L nonzero entries of $\mathrm{Diag}(\boldsymbol{\delta})\,\mathbf{x}$ correspond to the selected antennas. Therefore, the effective transmitted signal $\mathrm{Diag}(\boldsymbol{\delta})\,\mathbf{x}$ can be found with sparse recovery algorithms. Besides that, there is no necessity to separate precoding and antenna selection anymore, eventually meaning that the problem comes down to find an L-sparse precoded signal. Table 5.2 shows the relation among sparse recovery and massive MIMO downlink variables.

Table 5.2: Relation among sparse recovery and massive MIMO variables

Sparse Recovery	**Massive MIMO Downlink**
Measurement Matrix ($\mathbf{A}$)	Broadcast Channel Matrix ($\mathbf{H}^{\mathrm{T}}$)
Measured Vector ($\mathbf{y}$)	Received Signal ($\mathbf{y}$)
Parameter Vector ($\mathbf{x}$)	Precoded Signal ($\mathbf{x}$)

Unlike most cases in sparse recovery, antenna selection does not seek the sparsest solution. There is a trade-off between the number of antennas and the quality of service of the communication system. The sparsest solution for antenna selection may severely reduce the sum-rate capacity, and the gain in sum-rate capacity is one of the main reasons to increase the number of antennas in MIMO systems. Antenna-selection schemes aim to decrease the number of antennas keeping acceptable values for channel capacity or bit-error rate. Thus, the formulation in (P-5.2) is slightly different from that desired for antenna selection. Antenna selection aims to find a specific L-sparse vector for a desired value of L. A better formulation for antenna selection problem would be

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \;\; \left\|\mathbf{H}^{\mathrm{T}}\mathbf{x}-\mathbf{s}\right\|_2^2 \\ &\text{subject to} \;\; \|\mathbf{x}\|_0 = L. \end{aligned} \tag{P-5.8}$$

Replacing the l_0-norm with the l_1-norm and the equality constraint with an inequality constraint yield

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \;\; \left\|\mathbf{H}^{\mathrm{T}}\mathbf{x}-\mathbf{s}\right\|_2^2 \\ &\text{subject to} \;\; \|\mathbf{x}\|_1 \leq \tau, \end{aligned} \tag{P-5.9}$$

which is the LASSO estimator presented in Section 5.2. Note that τ is not exactly the number of selected antennas L, but it is possible to verify that there is an approximately linear relation between L and τ, as will be presented in Chapter 6.

LASSO estimator is a very powerful tool and it seems to fit very well to solve the antenna selection problem. Thus, LASSO is the basic building-block of the proposals in this chapter, and the first sparsity-aware precoding algorithm is based on LASSO. This idea of using l_1-norm regularization to promote sparsity in antenna arrays is also employed in [102] within the context of hexagonal arrays to satellite signals. In massive MIMO, the l_1-norm regularization for solving antenna selection is introduced in [55]. The authors used a BPDN-like formulation combined with the replica method from statistical mechanics considering Replica Symmetry to select antennas. Moreover, the proposed solution in [55] is evaluated in terms of a figure of merit called asymptotic distortion, which is unusual to evaluate communication systems performance.

5.4 LASSO Precoding

The first proposed algorithm will be called *LASSO precoding*, which can be viewed as a generalization of the ZF precoding algorithm in (P-3.1). The LASSO precoding adds new constraints to (P-3.1), raising awareness regarding the sparsity in the solution. The LASSO precoding technique is lightly different from the LASSO presented in (P-5.9), i.e.:

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{C}^{M\times 1}}{\text{minimize}} \quad & \left\|\mathbf{H}^{\mathrm{T}}\mathbf{x}-\mathbf{s}\right\|_2^2 \\ \text{subject to} \quad & \|\mathbf{x}\|_1 \leq \tau, \\ & \|\mathbf{x}\|_2 \leq \eta, \end{aligned} \tag{P-5.10}$$

where $\tau \in \mathbb{R}_+$ is the l_1-norm regularization factor, $\eta \in \mathbb{R}_+$ is the l_2-norm regularization factor, and both of them are tuning parameters.

The main idea of this optimization is to shrink the solution based on the value of the l_1-norm regularization factor τ: the smaller the value of τ, the more shrunken the solution will be, thus inducing more sparsity. Although the l_2-norm constraint works like a regularization for the problem, it has a practical meaning related to the energy control of the precoded signal $\mathbf{x}$.

In classical antenna selection algorithms it is possible to choose the desired number of active antennas L. It is therefore desirable that the same could be performed in LASSO precoder. The LASSO precoding promotes a high sparsity degree, but it faces a big issue: there is no straightforward relation between the value of τ and the l_0-norm of the estimate.

The question here is how to choose τ and η in order for the LASSO precoding to achieve the desired number of active antennas. Unfortunately, there is no close expression for τ as a function of the desired number of active antennas L. We

propose to use a strategy based on the l_1- and l_2-norms of the ZF precoded signal: the value of both τ and η can be respectively chosen as fractions of the l_1- and l_2-norms of the ZF solution, for instance:

$$\tau = \alpha \left\| \mathbf{x}_{\mathrm{ZF}} \right\|_1 , \tag{5.3}$$

$$\eta = \left\| \mathbf{x}_{\mathrm{ZF}} \right\|_2 , \tag{5.4}$$

where $\alpha \in [0,\ 1.0]$ is a parameter to adjust the sparsity of the solution, called sparsity factor. Note that when $\alpha = 1$, LASSO precoding produces the same result of ZF precoding. Roughly speaking, the constraints in (5.3) and (5.4) mean that the signal produced by LASSO precoding should have the same instantaneous energy of the ZF-precoded signal and might have a percentage of selected coefficients proportional to α. The percentage of used coefficients of LASSO solution is defined as

$$p = \frac{L}{M}, \tag{5.5}$$

where $L \in \mathbb{N}$ is the sparsity of the LASSO precoding solution (number of selected antennas) and $M \in \mathbb{N}$ is the total number of coefficients (number of antennas). Although related , the parameter α is not exactly equal to p, but simulation results of Chapter 6 indicate that there exists an approximately linear mapping between α and p.

The LASSO precoding has a drawback that might limit its use. It is data-dependent, which means that its solution always varies with the symbol vector $\mathbf{s}$, i.e., it is a nonlinear precoder. In comparison with ZF precoding that solves an optimization per coherence time, and antenna selection algorithms that solve an optimization per coherence time or new SNR value, LASSO precoding solves an optimization per sampling time. However, the proposed LASSO precoding might still be worth using depending on the optimization tools available. Moreover, this data-dependent nature of LASSO precoding must be further studied: if the support of the solution does not vary too much for each sampling time, some partial update approach could be used to reduce the complexity of the LASSO precoding, like in [133]. The LASSO precoding pseudo-code is described in Algorithm 5.

Algorithm 5 : LASSO

$\alpha \leftarrow [0,\ 1.0]$
$\mathbf{x}_{\mathrm{ZF}} \leftarrow \operatorname{argmin} \left\| \mathbf{H}^{\mathrm{T}}\mathbf{x} - \mathbf{s} \right\|_2^2$
$\tau \leftarrow \alpha \left\| \mathbf{x}_{\mathrm{ZF}} \right\|_1$
$\eta \leftarrow \left\| \mathbf{x}_{\mathrm{ZF}} \right\|_2$
$\mathbf{x}_{\mathrm{LASSO}} \leftarrow \operatorname{argmin} \left\| \mathbf{H}^{\mathrm{T}}\mathbf{x} - \mathbf{s} \right\|_2^2 \quad \text{s.t. } \left\| \mathbf{x} \right\|_1 \leq \tau,\ \left\| \mathbf{x} \right\|_2 \leq \eta$

The ZF precoding is one of the best linear precoding algorithms due to its ability

to right invert the downlink channel. This ability of ZF precoding helps it to produce low bit-error rate, as corroborated by the simulations that will be presented in Chapter 6. In order to satisfy the constraints, LASSO precoding loses this ability to right invert the channel, thus the LASSO solution may differ significantly to the ZF solution. If the "best" support vector is previously known, this could be used to improve the performance of LASSO precoding, or at least provide an upper bound for the LASSO precoding solution. This new solution is called LASSO-supported ZF (LASSO-SZF) and its algorithm is described in Algorithm 6.

Algorithm 6 : LASSO-SZF

$\alpha \leftarrow [0,\ 1.0]$
$\mathbf{x}_{\text{ZF}} \leftarrow \text{argmin}\ \left\|\mathbf{H}^{\text{T}}\mathbf{x} - \mathbf{s}\right\|_2^2$
$\tau \leftarrow \alpha \left\|\mathbf{x}_{\text{ZF}}\right\|_1$
$\eta \leftarrow \left\|\mathbf{x}_{\text{ZF}}\right\|_2$
$\mathbf{x}_{\text{LASSO}} \leftarrow \text{argmin}\ \left\|\mathbf{H}^{\text{T}}\mathbf{x} - \mathbf{s}\right\|_2^2 \quad \text{s.t.}\ \left\|\mathbf{x}\right\|_1 \leq \tau,\ \left\|\mathbf{x}\right\|_2 \leq \eta$
$\mathbf{S} \leftarrow \mathcal{B}\{\mathbf{I}_M, \text{supp}\left(\mathbf{x}_{\text{LASSO}}\right)\}$
$\mathbf{H}_L^{\text{T}} \leftarrow \mathbf{H}^{\text{T}}\mathbf{S}$
$\mathbf{W}_{\text{ZF}} \leftarrow \left(\mathbf{H}_L^{\text{T}}\right)^{\dagger}$
$\mathbf{z} \leftarrow \mathbf{W}_{\text{ZF}}\mathbf{s}$
$\mathbf{x}_{\text{LASSO-SZF}} \leftarrow \mathbf{S}\mathbf{z}$

The LASSO-SZF will yield better results in terms of bit-error rate than LASSO precoding, providing an upper bound for LASSO precoding performance. This occurs due to the ability of LASSO-SZF inverting the equivalent channel matrix $\mathbf{H}_L^{\text{T}}$. Indeed, simulations results in Chapter 6 indicate that LASSO-SZF precoding can outperform LASSO precoding in terms of bit-error rate. As LASSO-SZF precoding relies on LASSO precoding, it also has the same aforementioned drawbacks of the LASSO precoding. Furthermore, the LASSO-SZF precoding cannot be included in the class of joint precoding and antenna selection algorithms. This precoding algorithm just combines the antenna selection performed by LASSO with the ZF precoding, but it does not jointly perform them.

5.5 Conclusion

This section presented a different methodology to deal with antenna selection problem, called joint precoding and antenna selection. This idea uses the LASSO estimator, which is very popular in statistics and sparse recovery areas. From the LASSO estimator two new nonlinear sparsity-aware precoders are derived. Next chapter presents simulation results to evaluate the sparsity-aware precoder performances and compare with the classical linear precoders and antenna selection algorithms.

Chapter 6

Simulation Results

6.1 Introduction

This chapter describes the simulation results for the precoding algorithms proposed in Section 5.4. These algorithms are evaluated in terms of both beampattern design and BER performance. The results are compared with those obtained by using the classic linear precoders and the classic antenna selection algorithms. Moreover, simulation results indicate the existence of an approximately linear relation between the sparsity factor (see Section 5.4) and the number of selected antennas.

6.2 Methodology

This chapter evaluates several algorithms presented throughout the thesis, which are listed in Table 6.1, where the two proposals are highlighted in bold. These algorithms are evaluated in terms of beampattern design and BER performance.

The beampattern is calculated by using (3.14) and (3.15) for nonlinear and linear precoding algorithms, respectively. The beampattern design of an M-antenna base station equipped with a ULA, transmitting to K single-antenna terminals is qualitatively evaluated by visually inspecting the steering direction and the spread energy over the space. Moreover, an objective analysis is made by investigating the figure of merit called out of direction emission (ODE), which calculates the energy emissions out of the desired directions with an uncertainty of $\pm 1^\circ$ in the desired directions.

The performance is also evaluated in terms of the average BER per user as a function of the forward link SNR measured at terminals using Monte-Carlo simulation and assuming complete CSI knowledge by the base station. The average BER

Table 6.1: Summary of the algorithms used in the simulations

Acronymous	**Description**	**Section**
MF	Matched filter	3.2.1
ZF	Zero-forcing	3.2.1
LASSO	Least absolute shrinkage and selection operator	5.4
LASSO-SZF	LASSO-supported ZF	5.4
CM-MF	MF precoding along with antenna selection using channel capacity maximization	4.3.2
CM-ZF	ZF precoding along with antenna selection using channel capacity maximization	4.3.2
RANDOM-MF	MF precoding along with random antenna selection	4.3.1
RANDOM-ZF	ZF precoding along with random antenna selection	4.3.1

per user is calculated as a mean of the individual BER_k of each user, i.e.,

$$\overline{\mathrm{BER}} = \frac{1}{K} \sum_{k \in \mathcal{K}} \mathrm{BER}_k, \tag{6.1}$$

and the SNR is defined as

$$\rho_{\mathrm{dl}} = 10 \log_{10} \left(\frac{P_{\mathrm{s}}}{P_{\mathrm{n}}} \right), \tag{6.2}$$

where $P_{\mathrm{s}} \in \mathbb{R}_+$ is the signal power measured at the base station, and $P_{\mathrm{n}} \in \mathbb{R}_+$ is the noise power measured at terminals. Moreover, the percentage of selected antennas is analyzed (see Section 5.4).

6.2.1 Scenario 1: Beampattern Design

For the nonlinear precoding algorithms, a random data block of a 4 quadrature amplitude modulation (QAM) constellation is transmitted. This scenario considers $M \in \{50, 100, 200\}$ antennas, $K = 2$ terminals, and sparsity factor appearing in (P-5.6) as $\alpha \in \{0.50,\ 0.80,\ 0.90,\ 0.95\}$. The signals for the first and second terminals have departure angles of $\theta_1 = 45°$ and $\theta_2 = 135°$, respectively. The SNR for forward link measured at terminals is set to $\rho_{\mathrm{dl}} = 10$ dB for CM-ZF and CM-MF algorithms. The results of ZF and MF precoders are used as benchmarks (corresponding to $\alpha = 1.00$, which means that all antennas are used). It is assumed that the base station has knowledge of both departure angles. Table 6.2 summarizes all key parameters

used in this simulation.

Table 6.2: Simulation parameters of scenario 1

Parameters	**Value**
Full-array geometry	ULA
Constellation	4-QAM
Number of blocks	1
Number of antennas	$M \in \{50,\ 10,\ 200\}$
Number of terminals	$K \in 2$
Sparsity factor	$\alpha \in \{0.50,\ 0.80,\ 0.90,\ 0.95\}$
SNR	$\rho_{\text{dl}} = 10$ dB
Departure angles	$(\theta_1,\ \theta_2) = (45°,\ 135°)$

6.2.2 Scenario 2: Bit-error Rate Performance

A total of 100 random data blocks of a 4-QAM constellation are transmitted. This scenario considers $M \in \{50,\ 100,\ 200\}$ antennas, $K \in \{3,\ 5,\ 10\}$ terminals, and sparsity factor as $\alpha \in \{0.50,\ 0.80,\ 0.90,\ 0.95\}$. The simulations use the SNR values within the set $\{-10,\ -7,\ -4,\ -1,\ 2,\ 5,\ 8\}$ dB. The channel model used in the simulations is the multipath MIMO channel mentioned in the end of Section 2.5. The simulations use $N = 2$ multipaths, $\Theta_{kn} \sim \mathcal{U}([0,\ \pi])$, $G_{mk} \sim \mathcal{CN}(0, 1)$, with a constant power delay profile equal to 1, and $\beta_{kn} = 1$. The Monte-Carlo simulation runs 500 different channels. The results of ZF and MF precoders are used as benchmarks. It is assumed that the base station has complete CSI knowledge. Table 6.3 lists all key parameters used in this simulation.

Table 6.3: Simulation parameters of scenario 2

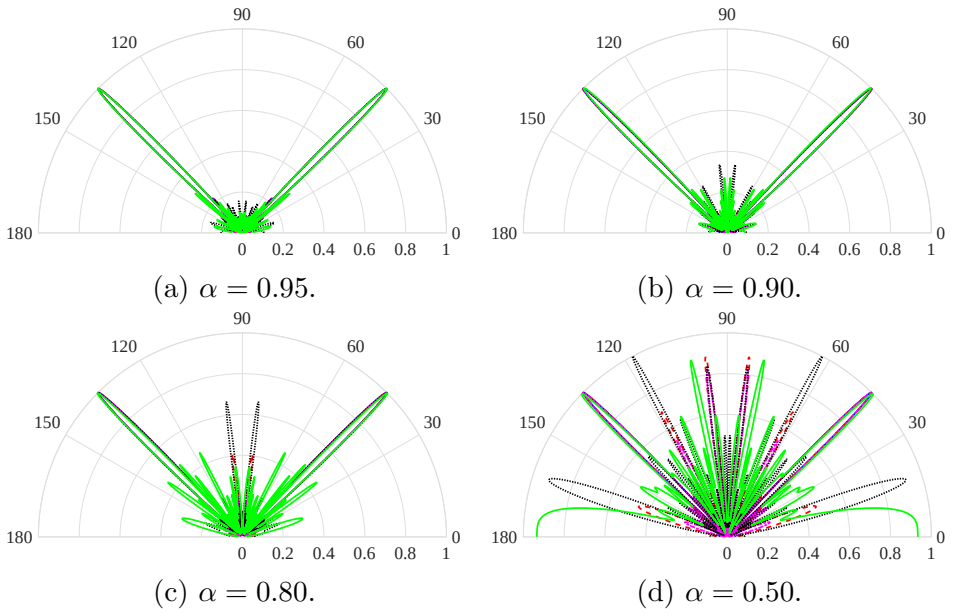
Parameters	**Value**
Full-array geometry	ULA
Constellation	4-QAM
Number of blocks	100
Monte-Carlo runs	500
Number of antennas	$M \in \{50,\ 10,\ 200\}$
Number of terminals	$K \in \{3,\ 5,\ 10\}$
Sparsity factor	$\alpha \in \{0.50,\ 0.80,\ 0.90,\ 0.95\}$
SNR	$\rho_{\text{dl}} = \{-10,\ -7,\ -4,\ -1,\ 2,\ 5,\ 8\}$ dB
Channel model	Spatial multipath channel
Number of multipaths	$N = 2$
Departure angles	$\Theta_{kn} \sim \mathcal{U}([0,\ \pi])$
Small-scale fading coefficient	$G_{mk} \sim \mathcal{CN}(0, 1)$
Power delay profile	1 (Constant)
Large-scale fading coefficient	$\beta_{kn} = 1$

6.3 Beampattern of Sparsity-aware Precoding Algorithms

Figure 6.1 shows the beampattern for the ZF-based precoders considering $M = 50$ and different sparsity levels, whereas Figure 6.2 depicts the beampattern for the MF-based precoders considering $M = 50$ and the same values of α. Table 6.4 shows the number of selected antennas L for each value of α used in the simulation for $M = 50$. The case with $\alpha = 1.00$ represents the ZF and MF precoders, which have the same beampattern due to the asymptotically favorable propagation (see Section 2.5).

Table 6.4: Sparsity factor *versus* number of active antennas for $M = 50$

Sparsity factor	**Selected antennas**	**Percentage of active antennas**
1.00	50	100%
0.95	45	90%
0.90	41	82%
0.80	32	64%
0.50	13	26%



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

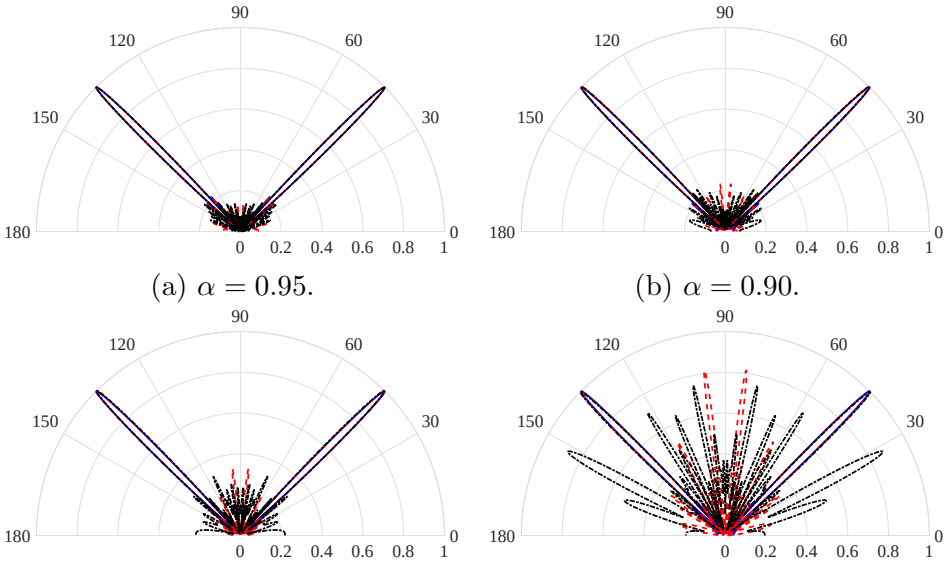
(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.1: Beampatterns of the ZF-based precoders for $M = 50$, and different values of α. Legend: – ZF, – – LASSO, –. LASSO-SZF, : CM-ZF, and – RANDOM-ZF.

Figures 6.1a and 6.2a illustrate the beampatterns for ZF- and MF-based precoders for $\alpha = 0.95$, respectively. In this case, the precoders use 90% of the antennas and resolve the terminals' directions. Visual inspection of these figures tells us that LASSO, LASSO-SZF, and CM-ZF precoders have approximately the same beampattern of the ZF precoder, whereas the CM-MF precoder has the same beampattern

of the MF precoder. These beampatterns are close, but not the same, as can be noticed in Figure 6.3. Additionally, the beampattern of random methods have side lobes spread over the space, with approximately 10% of the energy corresponding to the largest peak. The energy out of the desired directions may be irrelevant, but this amount of energy could be sufficient to induce some inter-user interference.

Figures 6.1b and 6.2b show the beampattens for ZF- and MF-based precoders for $\alpha = 0.90$, respectively. With 82% of the antennas, the precoders still resolve the terminals' directions. However, some precoders are steering to some undesired directions. For instance, the LASSO precoder has low-energy peaks (approximately 10% of the energy corresponding to the largest peak) in the directions of 83.3° and 97°. Moreover, the beampatterns of LASSO-SZF, CM-ZF, and CM-MF precoders are still very similar to the beampatterns of ZF and MF precoders. Furthermore, the beampattern of RANDOM-ZF follows the same behavior shown in Figure 6.1a, with more pronounced secondary lobes, whereas the beampattern of RANDOM-MF has approximately 20% of the energy corresponding to the largest peak focused in some undesired directions, as such 30°, 60°, 83.3°, 97°, 120°, and 150°.



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.2: Beampattern of the MF-based precoders for $M = 50$, and different values of α. Legend: – MF, – – CM-MF, and –. RANDOM-MF.

Figures 6.1c and 6.2c depict the beampatterns of ZF- and MF-based precoders for $\alpha = 0.80$, respectively. The LASSO beampattern has peaks with energy around 40% of the energy corresponding to the largest peak focused in 83.3° and 97°. Other beampatterns that point toward undesired directions are the CM-ZF's and CM-MF's, besides 83.3° and 97° directions, they steer toward 60° and 120°. In addition, the beampattern of the LASSO-SZF precoder is still very similar to the ZF's, even using only 64% of the antennas. Moreover, the beampattern of the random methods

have much more pronounced side lobes. The energy spread of the CM-ZF is more uniform than the energy spread of CM-MF, which is more focused in some specific directions.

Figures 6.1d and 6.2d illustrate the beampatterns of the ZF- and MF-based precoders for $\alpha = 0.50$, respectively. In this case, the precoders use 26% of the antennas. The random methods yield poor beampatterns, spreading a high amount of energy over many directions. Morevoer, the beampatterns of LASSO and CM-ZF are very similar to each other, steering toward the same undesired directions with approximately the same energy. Furthermore, the LASSO-SZF and CM-MF precoders have similar beampatterns, which spend more energy in the correct directions than the LASSO's and CM-ZF's. Despite the reduced number of active antennas, their energy spreading in undesired directions is smaller than that produced by other precoders. With the decrease of α, the array spatial selectivity is reduced and the energy is spread over the space, possibly point toward in some specific undesired directions. This result is harmful to massive MIMO systems, for it may cause inter-user interference in the terminals.

Figure 6.3 shows the ODE as function of the sparsity level for $M = 50$. This figure shows that the random methods have high ODE, as previously discussed. The other precoders have similar ODEs, with the CM-ZF being slightly worse with respect to this figure of merit for some sparsity levels. This fact could not be noticed in the qualitative analysis of the beampattern. Moreover, the level of ODE for LASSO and CM-MF are very close. Furthermore, with the exception of the random-based precoders, the other precoders achieve ODE levels close to the ZF and MF solutions, for sparsity factors close to one, thereby indicating the potential of those algorithms.

Figure 6.4 depicts the beampattern for the ZF-based precoders considering $M = 100$ and different sparsity levels, whereas Figure 6.5 illustrates the beampattern for the MF-based precoders for $M = 100$ and the same values of α. Table 6.5 shows the number of selected antennas L for each value of α used in the simulation for $M = 100$.

Table 6.5: Sparsity factor *versus* number of active antennas for $M = 100$

Sparsity factor	**Selected antennas**	**Percentage of active antennas**
1.00	100	100%
0.95	90	90%
0.90	80	80%
0.80	64	64%
0.50	25	25%

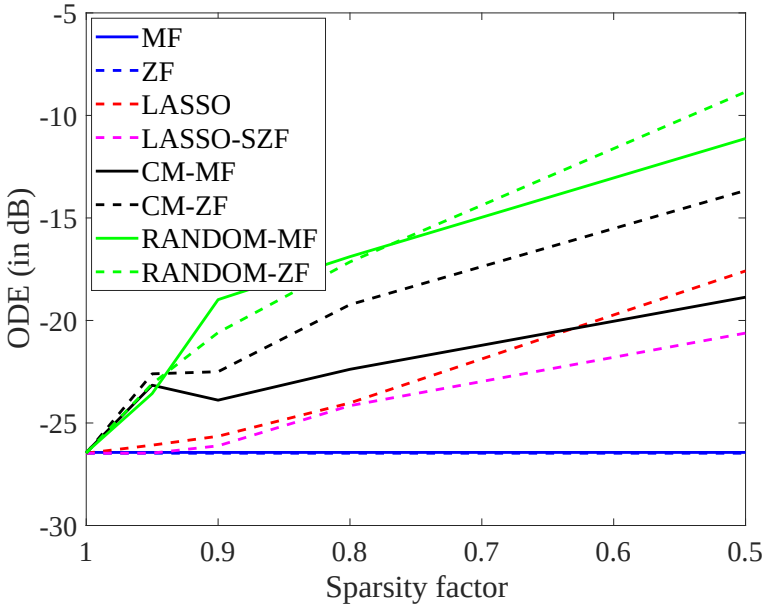
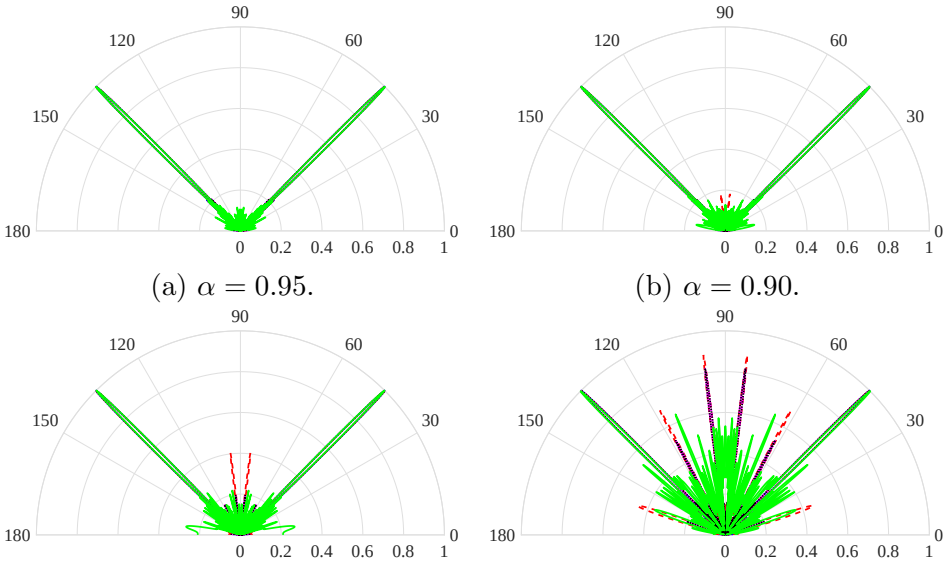


Figure 6.3: Out of direction emissions for $M = 50$.

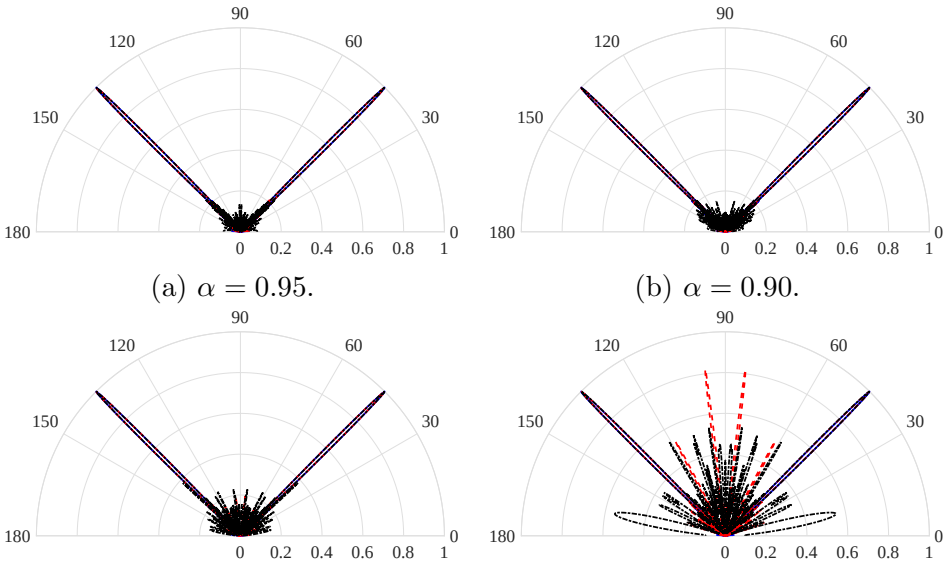


(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.4: Beampattern of the ZF-based precoders for $M = 100$, and different values of α. Legend: – ZF, – – LASSO, –. LASSO-SZF, : CM-ZF, and – RANDOM-ZF.

Figures 6.4 and 6.5 present slight differences compared with Figures 6.1 and 6.2. These differences are related to the increase in number of antennas. As M grows, the array spatial selectivity increases, performing a better steering in the terminal directions, as can be seen in the beampatterns. However, the increase in the array spatial selectivity also promotes the steering toward some undesired directions. For this case with $M = 100$, even the random methods get a much more focused energy spreading with very tight beams.



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.5: Beampattern of the MF-based precoders for $M = 100$, and different values of α. Legend: – MF, – – CM-MF, and –. RANDOM-MF.

Figure 6.6 shows the ODE as function of the sparsity level for $M = 100$. In this case, LASSO-SZF and CM-ZF precoders have a similar ODE for each sparsity value. Besides that, the LASSO, LASSO-SZF, CM-ZF and CM-MF ODEs are near the ZF and MF bound for sparsity levels close to 1. Furthermore, the random methods have similar ODEs that are outperformed by the other precoders.

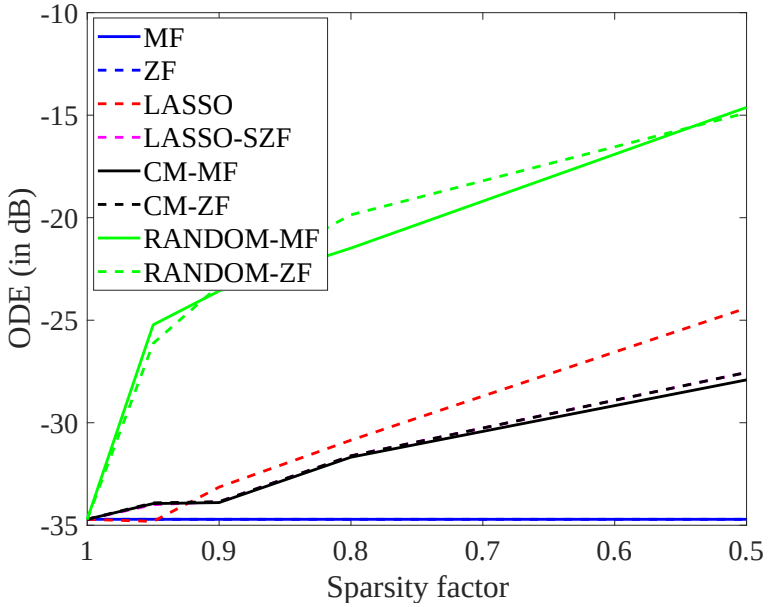


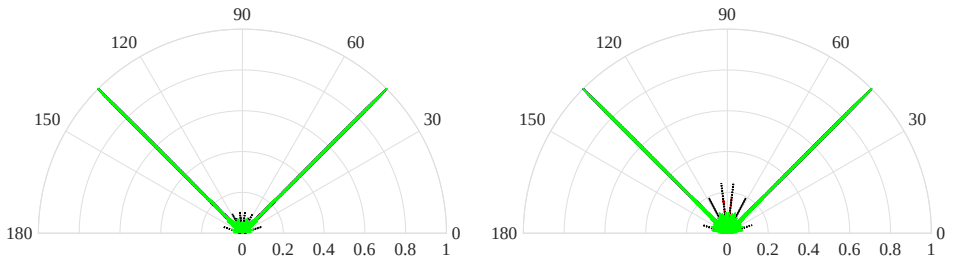
Figure 6.6: Out of direction emissions $M = 100$.

Figure 6.7 depicts the beampattern of the ZF-based precoders considering $M = 200$ different sparsity levels, whereas Figure 6.8 illustrates the beampattern of the

MF-based precoders for $M = 200$ and the same values of α. These results follow the same pattern presented by Figures 6.1, 6.2, 6.4, and 6.5. Table 6.6 shows the number of selected antennas L for each value of α used in the simulation for $M = 200$. Figure 6.9 shows the ODE as function of the sparsity level for $M = 200$. These results depicted by these figures also follow the same pattern as in Figure 6.9.

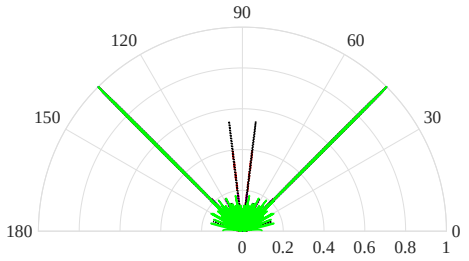
Table 6.6: Sparsity factor *versus* number of active antennas for $M = 200$

Sparsity factor	**Selected antennas**	**Percentage of active antennas**
1.00	200	100%
0.95	180	90%
0.90	161	81%
0.80	126	63%
0.50	49	25%

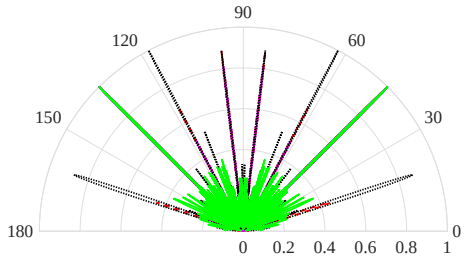


(a) $\alpha = 0.95$.

(b) $\alpha = 0.90$.



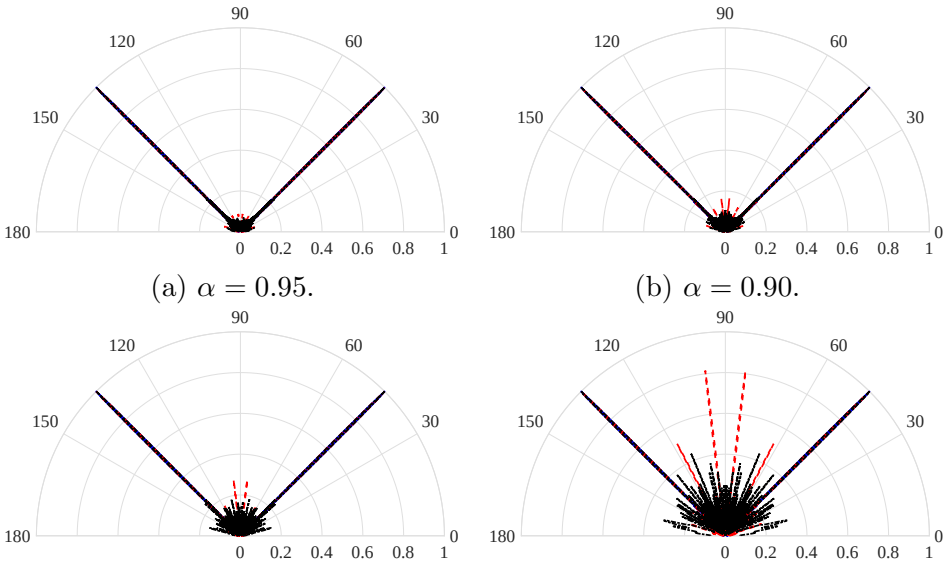
(c) $\alpha = 0.80$.



(d) $\alpha = 0.50$.

Figure 6.7: Beampattern of the ZF-based precoders for $M = 200$, and different values of α. Legend: – ZF, – – LASSO, –. LASSO-SZF, : CM-ZF, and – RANDOM-ZF.

It is worth highlighting some facts about the results in Tables 6.4–6.6. The percentage of active antennas is a good figure of merit because it normalizes the number of selected antennas by the total number of antennas. These results of percentage of active antennas are very promising, showing that, for this specific scenario, there exists an approximately linear relation between the sparsity factor and the percentage of active antennas. Besides that, this relation seems to be independent of the total number of antennas. Although, this linear relation holds in this simple scenario, one cannot guarantee its validity for more complex scenarios. In



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.8: Beampattern of the MF-based precoders for $M = 200$, and different values of α. Legend: – MF, – – CM-MF, and –. RANDOM-MF.

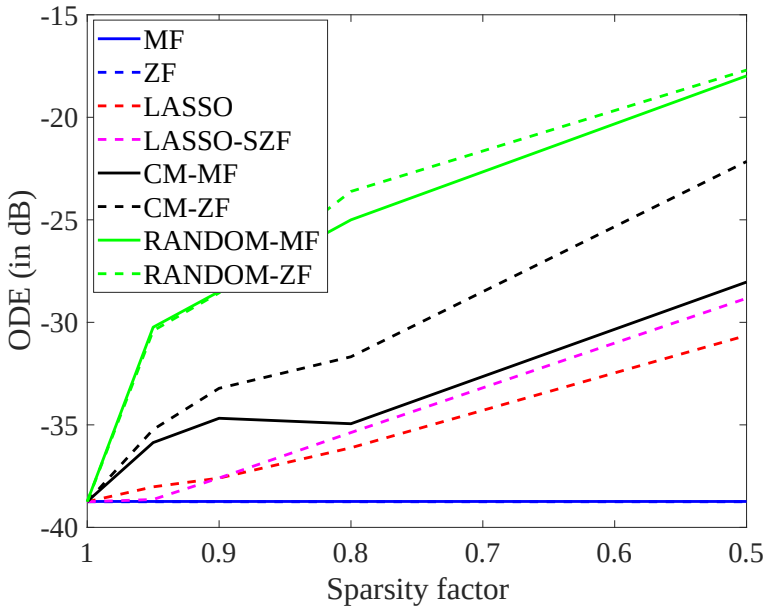


Figure 6.9: Out of direction emissions $M = 200$.

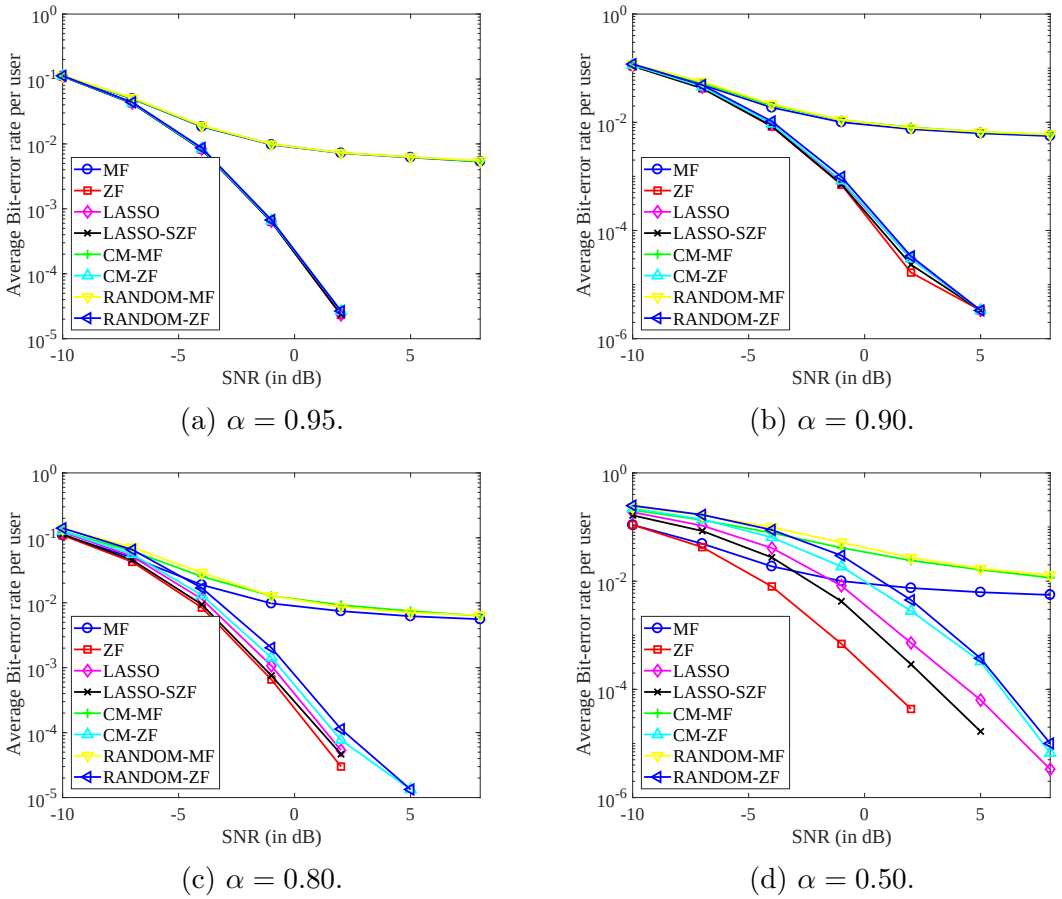
order to verify this linear relation, next section performs a Monte-Carlo simulation to evaluate it.

6.4 Bit-error Rate Performance of Sparsity-aware Precoding Algorithms

Figure 6.10 shows the average BER per user considering $M = 50$ and $K = 3$ for different values of α. Table 6.7 shows the number of selected antennas L, and the percentage of selected antenna p for each value of α for $M = 50$ and $K = 3$.

Table 6.7: Sparsity factor *versus* number of active antennas for $M = 50$ and $K = 3$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	50	100%
0.95	49	98%
0.90	46	92%
0.80	39	78%
0.50	17	34%



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.10: Average BER per user for $M = 50$, $K = 3$ and different values of α.

Figure 6.10a depicts the average BER per user for $\alpha = 0.95$. In this case, the BER of sparsity-aware precoders and antenna selection algorithms is expected to

be close to the BER of linear precoders ($\alpha = 1.00$). Indeed, this happens due to a natural redundancy in massive MIMO systems. The massive MIMO may be able to compensate for the lost of a few antennas. In this case $L = 49$ antennas are selected, so the system uses 98% of the antennas. Additionally, there is a remarkable difference between the performance of the ZF- and MF-based precoders. As mentioned in Section 2.5, the MF precoder is optimum in conditions of favorable propagation or asymptotically favorable. These conditions are not satisfied here due to the small number of multipaths used to generate the channel. In a few cases, the asymptotically favorable propagation might be satisfied for some links, but in the average, the bad conditioned links bias the average BER per user.

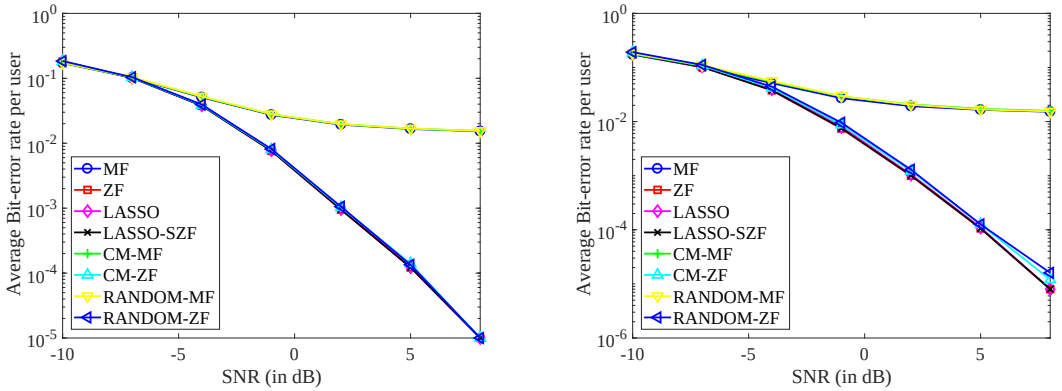
Figure 6.10b illustrates the average BER per user for $\alpha = 0.90$. In this case $L = 46$ antennas are used, representing 92% of the antennas. With 92% of active antennas, the BER of ZF-based methods are still close. Figure 6.10c shows the average BER per user for $\alpha = 0.80$. With this value of α, $L = 39$ antennas are selected, corresponding to 78% of active antennas. This result shows that with $L = 39$ antennas, LASSO and LASSO-SZF precoders achieve BER levels very similar to the ZF's and they have a mild advantage in comparison with the others ZF-based precoders.

Figure 6.10d depicts the average BER per user for $\alpha = 0.50$. For this value of α, the number of selected antennas is $L = 17$, resulting in 34% of active antennas. In this case, the LASSO and LASSO-SZF precoding algorithms outperform the other ZF-based precoding algorithms. Moreover, it LASSO-SZF precoder is a lower bound for LASSO precoder, as discussed in Section 5.4.

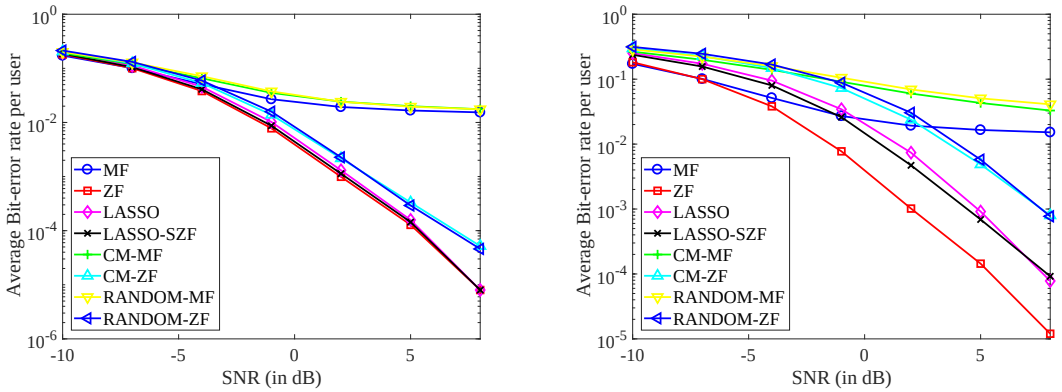
Figure 6.11 illustrates the average BER per user considering $M = 50$ and $K = 5$ for different values of α. Table 6.8 displays the number of selected antennas L and the percentage of selected antenna p for each value of α for $M = 50$ and $K = 5$. The results in Figure 6.11 follow the same behaviors observed in Figure 6.10.

Table 6.8: Sparsity factor *versus* number of active antennas for $M = 50$ and $K = 5$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	50	100%
0.95	49	98%
0.90	46	92%
0.80	39	78%
0.50	17	34%



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.



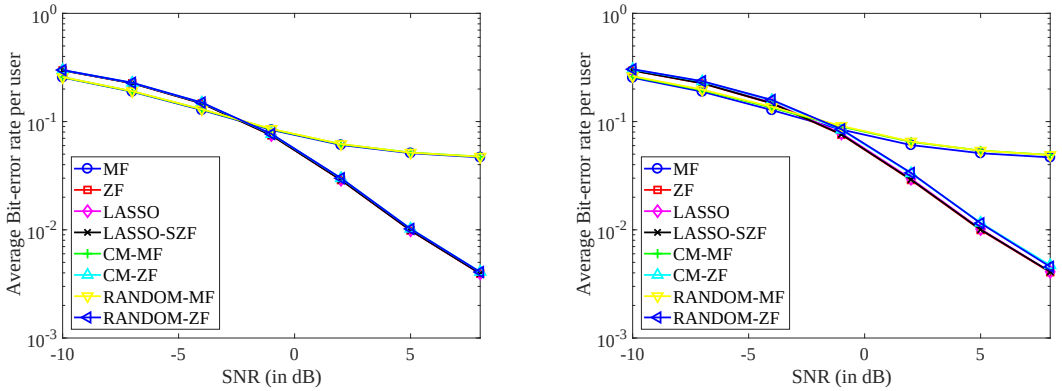
(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.11: Average BER per user for $M = 50$, $K = 5$ and different values of α.

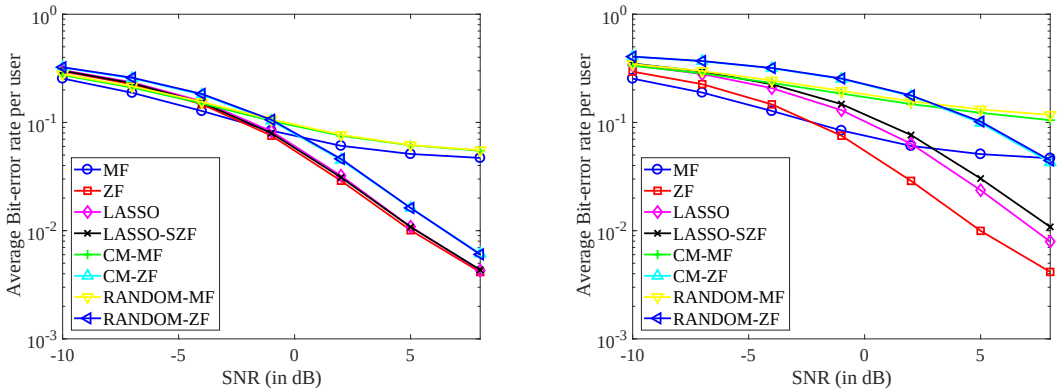
Figure 6.12 shows the average BER per user considering $M = 50$ and $K = 10$. Table 6.9 displays the number of selected antennas L and the percentage of selected antenna p for each value of α used for $M = 50$ and $K = 10$. Once again, Figure 6.12 follows the same patterns observed in Figures 6.10 and 6.11, with minor differences. For $\alpha = 0.50$, LASSO precoder achieves a lower BER than LASSO-SZF precoder. Moreover, CM-ZF and RANDOM-ZF precoders achieves the same BER performance.

Table 6.9: Relation between the sparsity factor and the number of selected antennas for $M = 50$ and $K = 10$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	50	100%
0.95	49	98%
0.90	46	92%
0.80	39	78%
0.50	17	34%



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.



(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.12: Average BER per user for $M = 50$, $K = 10$ and different values of α.

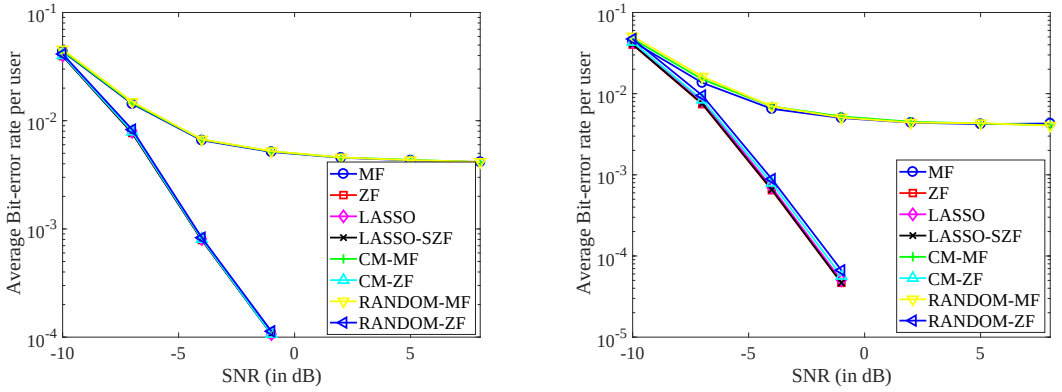
The percentages of active antennas presented in Tables 6.7–6.9 are very related to that ones reported in Section 6.3. Additionally, they seem to be invariant to the number of terminals, having exactly the same number of selected antennas for each value of α. This fact makes possible to build a linear model between α and the number of selected antennas, which is invariant to the number of terminals.

Figures 6.13–6.15 show the average BER per user for $M = 100$ and $K \in \{3,\ 5,\ 10\}$. Tables 6.10–6.12 display the number of selected antennas L and the percentage of selected antenna p for each value of α for $M = 100$ and $K \in \{3,\ 5,\ 10\}$. These results have the same pattern presented by the results for $M = 50$ and $K \in \{3,\ 5,\ 10\}$, having only modest variation in the results. The number of selected antennas is close for each simulation with $M = 100$ antennas. Moreover, they are the same to the simulations with $M = 50$ antennas. This fact is very interesting and indicates that the linear model may be expanded to encompass the average percentage of selected antennas, which is more general than the number of selected antennas. Furthermore, this result seems to be invariant to the number of antennas M.

Figures 6.16–6.18 depict the average BER per user considering $M = 200$ and $K \in \{3,\ 5,\ 10\}$. Tables 6.13–6.15 display the average number of selected antennas

Table 6.10: Relation between the sparsity factor and the number of active antennas for $M = 100$ and $K = 3$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	100	100%
0.95	98	98%
0.90	92	92%
0.80	77	77%
0.50	33	33%



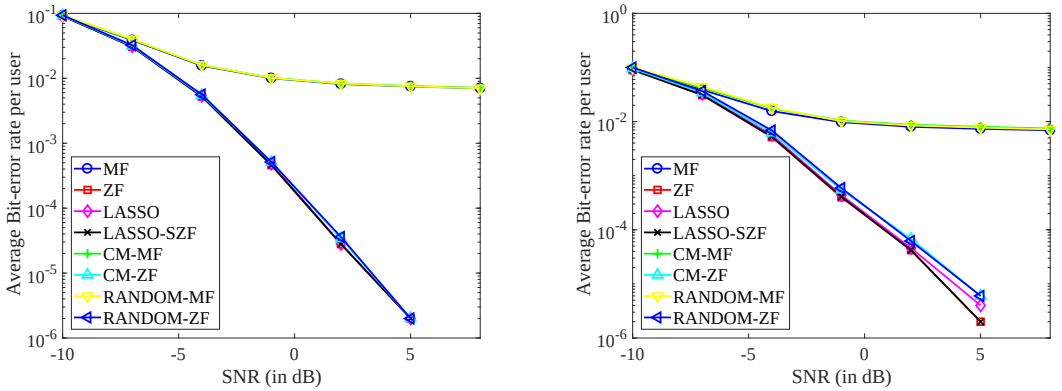
(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

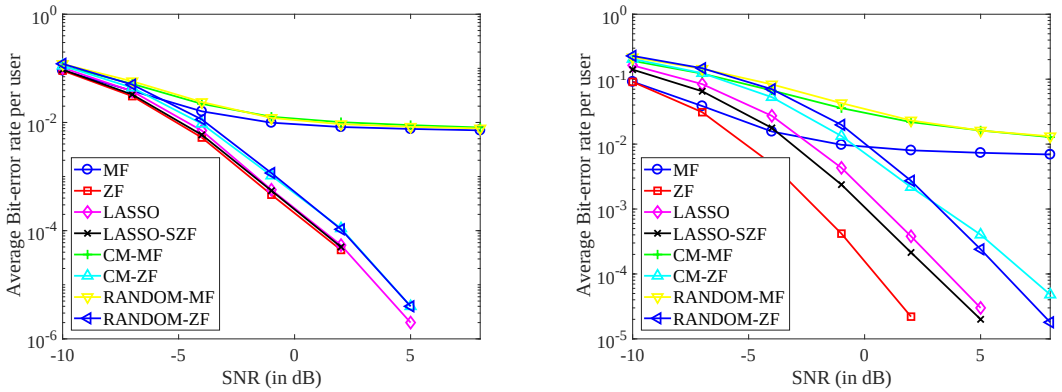
Figure 6.13: Average BER per user for $M = 100$, $K = 3$ and different values of α.

Table 6.11: Relation between the sparsity factor and the number of active antennas for $M = 100$ and $K = 5$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	100	100%
0.95	98	98%
0.90	92	92%
0.80	77	77%
0.50	34	34%



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.



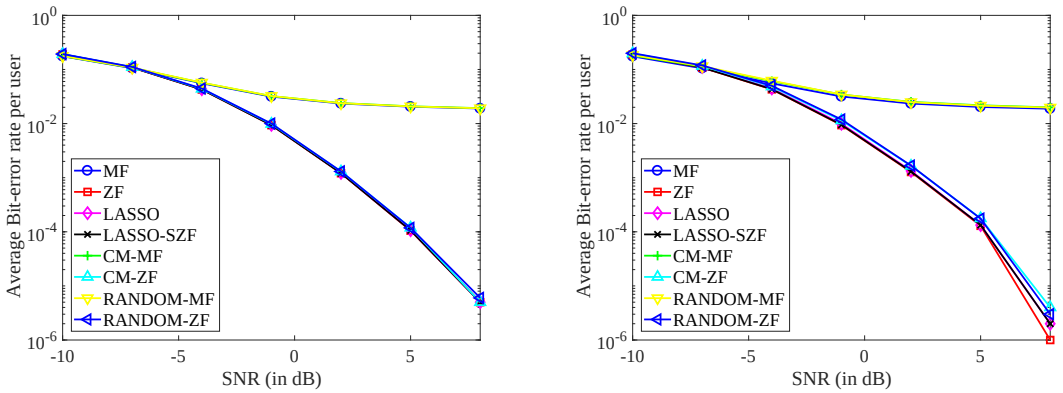
(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

Figure 6.14: Average BER per user for $M = 100$, $K = 5$ and different values of α.

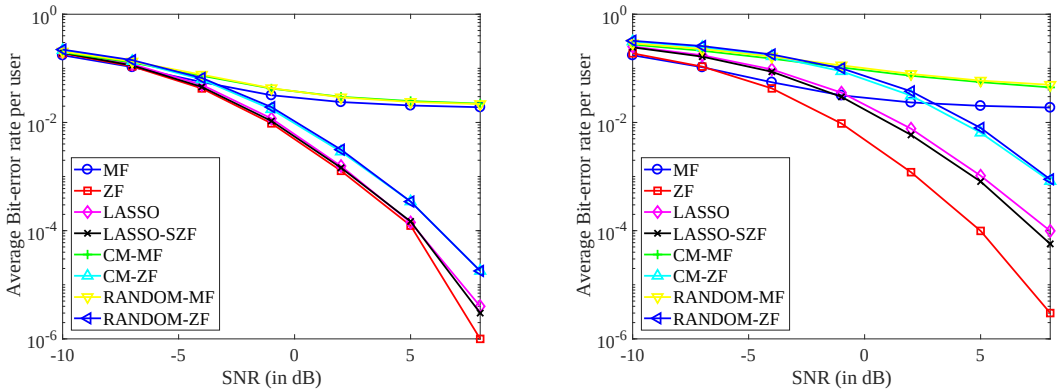
Table 6.12: Relation between the sparsity factor and the number of active antennas for $M = 100$ and $K = 10$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	100	100%
0.95	98	98%
0.90	92	92%
0.80	77	77%
0.50	34	34%

L and the percentage of selected antenna p for each value of α for $M = 200$ and $K \in \{3,\ 5,\ 10\}$. These results have the same pattern presented by the results for $M = 50$, $M = 100$, and $K \in \{3,\ 5,\ 10\}$, having only modest variations in the results. The number of selected antennas is almost the same for each simulation with $M = 200$ antennas. Moreover, they are also close to the simulations with $M = 50$ and $M = 100$ antennas. This result confirms the linear model adopted to α and the percentage of selected antennas.



(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.



(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

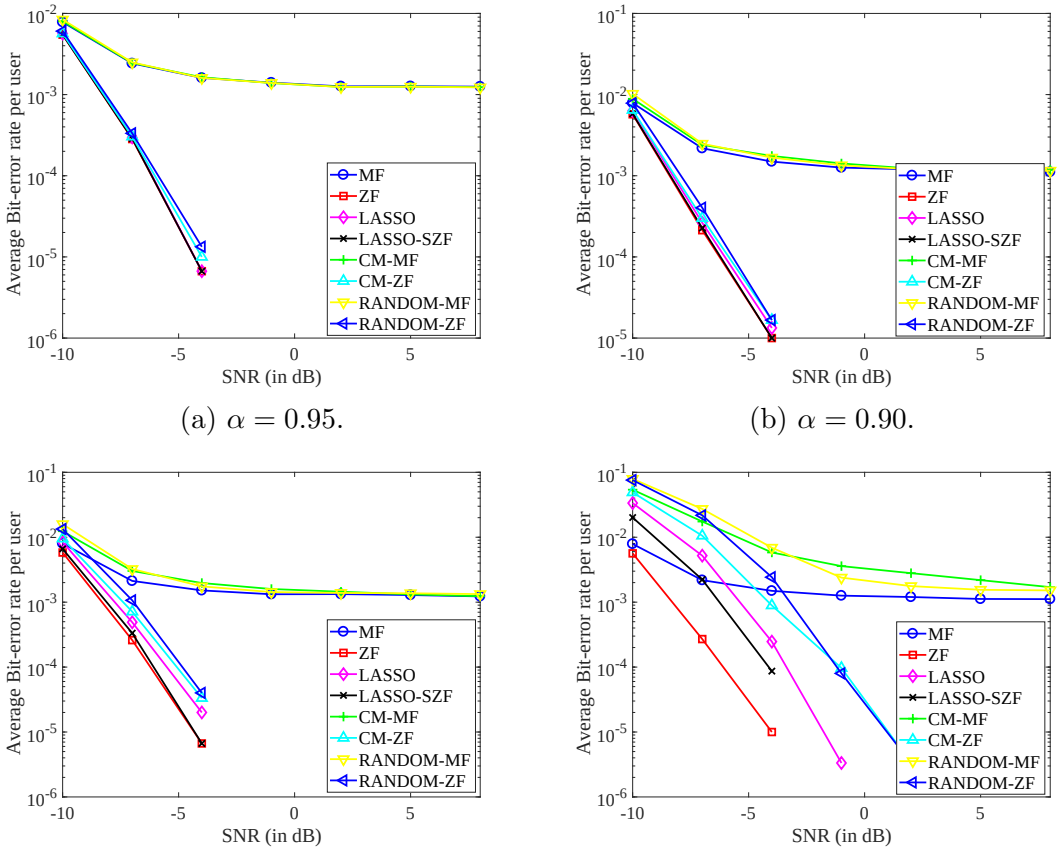
Figure 6.15: Average BER per user for $M = 100$, $K = 10$ and different values of α.

Table 6.13: Relation between the sparsity factor and the number of active antennas for $M = 200$ and $K = 3$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	200	100%
0.95	194	97.5%
0.90	182	91%
0.80	153	76.5%
0.50	64	32%

Table 6.14: Relation between the sparsity factor and the number of active antennas for $M = 200$ and $K = 5$

Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	200	100%
0.95	195	97.5%
0.90	183	91.5%
0.80	154	77%
0.50	66	33%



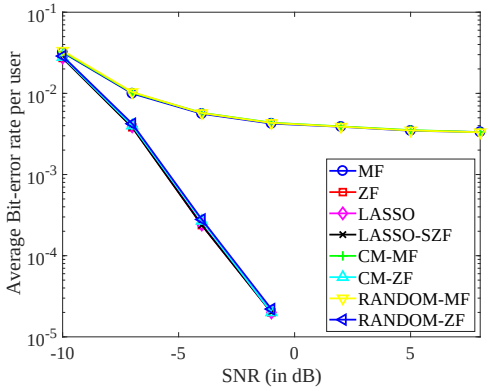
(a) $\alpha = 0.95$. (b) $\alpha = 0.90$.

(c) $\alpha = 0.80$. (d) $\alpha = 0.50$.

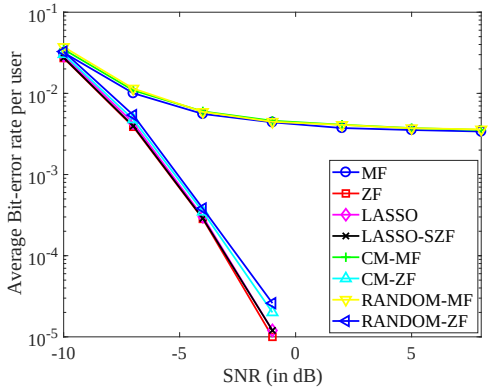
Figure 6.16: Average BER per user for $M = 200$, $K = 3$ and different values of α.

Table 6.15: Relation between the sparsity factor and the number of active antennas for $M = 200$ and $K = 10$

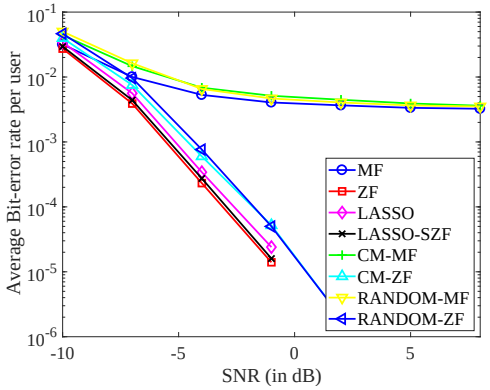
Sparsity factor	**Selected antennas**	**Percentage of selected antennas**
1.00	200	100%
0.95	195	97.5%
0.90	183	91.5%
0.80	154	76.5%
0.50	67	33.5%



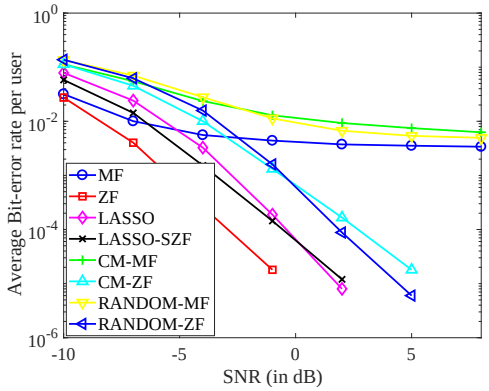
(a) $\alpha = 0.95$.



(b) $\alpha = 0.90$.

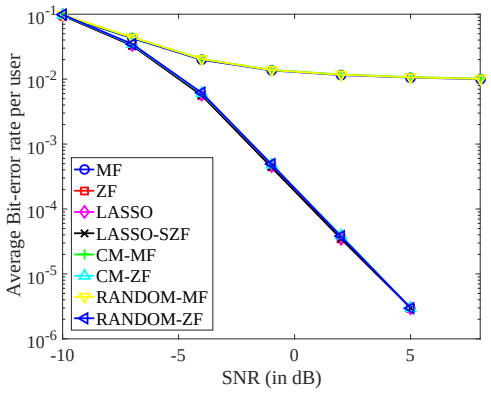


(c) $\alpha = 0.80$.

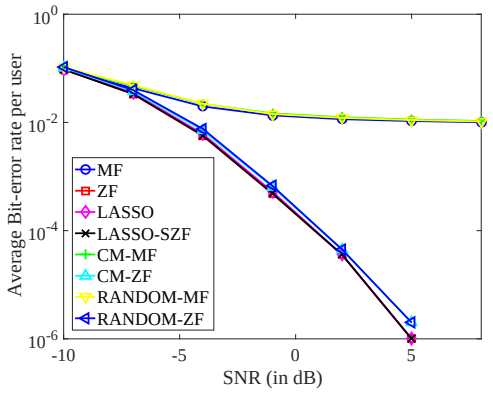


(d) $\alpha = 0.50$.

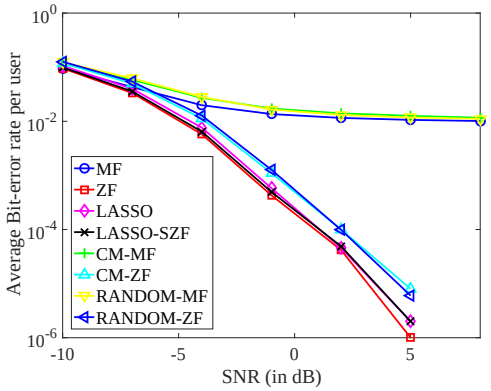
Figure 6.17: Average BER per user for $M = 200$, $K = 5$ and different values of α.



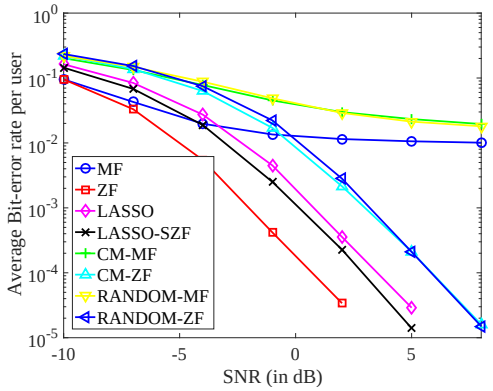
(a) $\alpha = 0.95$.



(b) $\alpha = 0.90$.



(c) $\alpha = 0.80$.



(d) $\alpha = 0.50$.

Figure 6.18: Average BER per user for $M = 200$, $K = 10$ and different values of α.

Chapter 7

Conclusion and Future Works

7.1 Concluding Remarks

This thesis presented a basic overview of massive MIMO technology, showing its potential to achieve high sum-rate capacity. A mathematical description of uplink and downlink was also given, including details regarding the main precoders and detectors. In addition, the thesis discussed the most common antenna selection algorithms for massive MIMO. The main algorithm is based on channel capacity maximization and finds the best selector through a convex optimization problem.

The thesis studied a subject not fully tackled in the literature, which is the joint precoding and antenna selection. Precoders belonging to this class are known as sparsity-aware precoders because they produce sparse precoded signals to select antennas. Furthermore, two new sparsity-aware precoders were proposed, namely LASSO and LASSO-SZF precoding. These precoders are nonlinear and depend on a parameter to adjust the number of selected antennas.

Simulation results indicated that the proposed algorithms achieve bit-error rates close to those classic precoders, such as zero-forcing. The proposed algorithms presented an unexpected behavior related to the parameter to adjust the number of selected antennas. They have an approximately linear mapping with the percentage of selected antennas, which is desired but unexpected due to the nonlinear nature of the sparsity-aware precoders.

7.2 Future Research Directions

Possible directions to future works include:

- The sparsity-aware precoders proposed in this work are data-dependent. However, these precoders might vary slowly with the input data, calling for a detailed study in order to verify how the set of selected antenna varies with the

input data. If the slowly variation is confirmed, sparsity-aware precoders with partial update could be developed, reducing the complexity in these precoders;

- The simulations in this thesis assumed complete CSI knowledge. The performance of the sparsity-aware precoders must be analyzed for partial CSI knowledge;
- Simulations using spatial multipath channels with a large number of multipaths must be conducted;
- Simulations in rich scattering fading must be performed. It is important to verify if the observed behavior is kept for rich scattering channels;
- New constraints might be included in the sparsity-aware precoders, for instance, PAPR-aware constraints or 1-bit quantization;
- Pilot contamination is a topic of major concern for practical massive MIMO systems. It produces inter-cell interference that is harmful for massive MIMO systems. Simulation must be done to evaluate the performance of the sparsity-aware precoders in this type of environment.

Bibliography

[1] HONG, S., BRAND, J., CHOI, J., *et al.*, "Applications of Self-Interference Cancellation in 5G and Beyond", *IEEE Communications Magazine*, v. 52, n. 2, pp. 114–121, February 2014.

[2] MAXWELL, J. C., *A Treatise on Eletricity and Magnetism*. Oxford, Clarendon Press, 1873.

[3] MARCONI, G., "Wireless Telegraphic Communication", *Nobel Lecture*, pp. 196–221, December 1909.

[4] BELROSE, J. S., "Marconi and the History of Radio", *IEEE Antennas and Propagation Magazine*, v. 46, n. 2, pp. 130–131, June 2004.

[5] DU, K.-L., SWAMY, M. N. S., *Wireless Communication Systems: From RF Subsystems to 4G Enabling Technologies*. 1 ed. Cambridge, Cambridge University Press, 2010.

[6] MACDONALD, V. H., "The Cellular Concept", *The Bell System Technical Journal*, v. 58, n. 1, pp. 1–27, January 1978.

[7] HANSON, B. L., BRONELL, C. E., "Human Factors Evaluation of Calling Procedures for the Advanced Mobile Phone System (AMPS)", *IEEE Transactions on Vehicular Technology*, v. 28, n. 2, pp. 126–131, May 1979.

[8] SAUTER, M., "Global System for Mobile Communications (GSM)". In: *From GSM to LTE-Advanced*, chapter 6, Chichester, John Wiley & Sons, Ltd, pp. 1–71, jun 2014.

[9] ROBERTS, M. L., TEMPLE, M. A., MILLS, R. F., *et al.*, "Evolution of the Air Interface of Cellular Communications Systems Toward 4G Realization", *IEEE Communications Surveys and Tutorials*, v. 8, n. 1, pp. 2–23, First Quarter 2006.

[10] BANGERTER, B., TALWAR, S., AREFI, R., *et al.*, "Networks and Devices for the 5G Era", *IEEE Communications Magazine*, v. 52, n. 2, pp. 90–96, February 2014.

[11] BHUSHAN, N., Junyi Li, MALLADI, D., *et al.*, "Network Densification: The Dominant Theme for Wireless Evolution into 5G", *IEEE Communications Magazine*, v. 52, n. 2, pp. 82–89, February 2014.

[12] JIANG, C., ZHANG, H., REN, Y., *et al.*, "Machine Learning Paradigms for Next-Generation Wireless Networks", *IEEE Wireless Communications*, v. 24, n. 2, pp. 98–105, April 2017.

[13] FETTWEIS, G. P., ALAMOUTI, S., "5G: Personal Mobile Internet Beyond What Cellular Did to Telephony", *IEEE Communications Magazine*, v. 52, n. 2, pp. 140–145, February 2014.

[14] DING, Y., JIN, Y., REN, L., *et al.*, "An Intelligent Self-Organization Scheme for the Internet of Things", *IEEE Computational Intelligence Magazine*, v. 8, n. 3, pp. 41–53, August 2013.

[15] FETTWEIS, G. P., "The Tactile Internet: Applications and Challenges", *IEEE Vehicular Technology Magazine*, v. 9, n. 1, pp. 64–70, March 2014.

[16] SCHAICH, F., WILD, T., CHEN, Y., "Waveform Contenders for 5G - Suitability for Short Packet and Low Latency Transmissions". In: *IEEE 79th Vehicular Technology Conference*, pp. 1–5, Seoul, May 2014.

[17] TADAYON, N., AISSA, S., "Modeling and Analysis of Cognitive Radio Based IEEE 802.22 Wireless Regional Area Networks", *IEEE Transactions on Wireless Communications*, v. 12, n. 9, pp. 4363–4375, September 2013.

[18] BANELLI, P., BUZZI, S., COLAVOLPE, G., *et al.*, "Modulation Formats and Waveforms for 5G Networks: Who Will Be the Heir of OFDM?: An Overview of Alternative Modulation Schemes for Improved Spectral Efficiency", *IEEE Signal Processing Magazine*, v. 31, n. 6, pp. 80–93, November 2014.

[19] FARHANG-BOROUJENY, B., "OFDM Versus Filter Bank Multicarrier", *IEEE Signal Processing Magazine*, v. 28, n. 3, pp. 92–112, May 2011.

[20] FARHANG, A., MARCHETTI, N., FIGUEIREDO, F., *et al.*, "Massive MIMO and Waveform Design for 5th Generation Wireless Communication Systems". In: *1st International Conference on 5G for Ubiquitous Connectivity*, pp. 70–75, Levi, November 2014.

[21] FARHANG-BOROUJENY, B., "Filter Bank Multicarrier Modulation: A Waveform Candidate for 5G and Beyond", *Advances in Electrical Engi-*

neering, Advances in Electrical Engineering, v. 2014, pp. 1–25, December 2014.

[22] LI, J., BALA, E., YANG, R., "Resource Block Filtered-OFDM for Future Spectrally Agile and Power Efficient Systems", *Physical Communication*, v. 11, pp. 36–55, June 2013.

[23] FARHANG, A., AMINJAVAHERI, A., REYHANI, A. R., *et al.*, "Time Reversal with Post-Equalization for OFDM without CP in Massive MIMO". In: *International Symposium on Wireless Communication Systems*, pp. 352–358, Poznan, September 2016.

[24] FARHANG, A., MARCHETTI, N., FARHANG-BOROUJENY, B., "Filter Bank Multicarrier for Massive MIMO". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 4, Chichester, Wiley, pp. 67–89, 2016.

[25] DE CANDIDO, O., CHEEMA, S. A., BALTAR, L. G., *et al.*, "Downlink Precoder and Equalizer Designs for Multi-User MIMO FBMC/OQAM". In: *20th International ITG Workshop on Smart Antennas*, pp. 454–461, Munich, March 2016.

[26] BUZZI, S., UGOLINI, A., ZAPPONE, A., *et al.*, "An Introduction to Modulations and Waveforms for 5G Networks". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 1, Chichester, Wiley, pp. 1–23, 2016.

[27] ANDERSON, J. B., "Faster-than-Nyquist Signaling for 5G Communication". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 2, Chichester, Wiley, pp. 24–46, 2016.

[28] ABDOLI, J., JIA, M., MA, J., "Filtered OFDM: A New Waveform for Future Wireless Systems". In: *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 66–70, Stockholm, June 2015.

[29] CHENG, X., HE, Y., GE, B., *et al.*, "A Filtered OFDM Using FIR Filter Based on Window Function Method". In: *IEEE 83rd Vehicular Technology Conference*, pp. 1–5, Najing, May 2016.

[30] MICHAILOW, N., MATTHÉ, M., GASPAR, I. S., *et al.*, "Generalized Frequency Division Multiplexing for 5th Generation Cellular Networks", *IEEE Transactions on Communications*, v. 62, n. 9, pp. 3045–3061, September 2014.

[31] SCHAFHUBER, D., MATZ, G., HLAWATSCH, F., "Pulse-Shaping OFDM/BFDM Systems for Time-Varying Channels: ISI/ICI Analysis, Optimal Pulse Design, and Efficient Implementation". In: *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1012–1016, Lisbon, September 2002.

[32] KASPARICK, M., WUNDER, G., JUNG, P., *et al.*, "Bi-orthogonal Waveforms for 5G Random Access with Short Message Support". In: *European Wireless Conference*, pp. 1–6, Berlin, May 2014.

[33] VAKILIAN, V., WILD, T., SCHAICH, F., *et al.*, "Universal-Filtered Multi-Carrier Technique for Wireless Systems Beyond LTE". In: *IEEE Globecom Workshops*, pp. 223–228, Atlanta, December 2013.

[34] BOCCARDI, F., HEATH, R. W., LOZANO, A., *et al.*, "Five Disruptive Technology Directions for 5G", *IEEE Communications Magazine*, v. 52, n. 2, pp. 74–80, February 2014.

[35] ROH, W., SEOL, J.-Y., PARK, J., *et al.*, "Millimeter-Wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results", *IEEE Communications Magazine*, v. 52, n. 2, pp. 106–113, February 2014.

[36] ALKHATEEB, A., EL AYACH, O., LEUS, G., *et al.*, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 831–846, October 2014.

[37] SAYEED, A. M., BRADY, J. H., "Millimeter-Wave MIMO Transceivers: Theory, Design and Implementation". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 10, Chichester, Wiley, pp. 231–253, 2016.

[38] LI, Q. C., JUNG, H., ZONG, P., *et al.*, "5G Millimeter-wave Communication Channel and Technology Overview". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 15, Chichester, Wiley, pp. 354–371, 2016.

[39] MARZETTA, T. L., "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas", *IEEE Transactions on Wireless Communications*, v. 9, n. 11, pp. 3590–3600, November 2010.

[40] RUSEK, F., PERSSON, D., Buon Kiong Lau, *et al.*, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays", *IEEE Signal Processing Magazine*, v. 30, n. 1, pp. 40–60, January 2013.

[41] FARHANG, A., MARCHETTI, N., DOYLE, L. E., *et al.*, "Filter Bank Multicarrier for Massive MIMO". In: *IEEE 80th Vehicular Technology Conference*, pp. 1–7, Vancouver, September 2014.

[42] LU, L., LI, G. Y., SWINDLEHURST, A. L., *et al.*, "An Overview of Massive MIMO: Benefits and Challenges", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 742–758, October 2014.

[43] WEI, H., WANG, D., ZHU, H., *et al.*, "Mutual Coupling Calibration for Multiuser Massive MIMO Systems", *IEEE Transactions on Wireless Communications*, v. 15, n. 1, pp. 606–619, January 2016.

[44] EDFORS, O., LIU, L., TUFVESSON, F., *et al.*, "Massive MIMO for 5G: Theory, Implementation and Prototyping". In: Luo, F.-L., Zhang, J. C. (eds.), *Signal Processing for 5G: Algorithms and Implementations*, 1 ed., chapter 9, Chichester, Wiley, pp. 192–230, 2016.

[45] ANDREWS, J. G., BUZZI, S., CHOI, W., *et al.*, "What Will 5G Be?", *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, pp. 1065–1082, June 2014.

[46] BJÖRNSON, E., HOYDIS, J., KOUNTOURIS, M., *et al.*, "Massive MIMO Systems With Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits", *IEEE Transactions on Information Theory*, v. 60, n. 11, pp. 7112–7139, November 2014.

[47] STUDER, C., LARSSON, E. G., "PAR-Aware Large-Scale Multi-User MIMO-OFDM Downlink", *IEEE Journal on Selected Areas in Communications*, v. 31, n. 2, pp. 303–313, February 2013.

[48] MOHAMMED, S. K., LARSSON, E. G., "Constant-Envelope Multi-User Precoding for Frequency-Selective Massive MIMO Systems", *IEEE Wireless Communications Letters*, v. 2, n. 5, pp. 547–550, October 2013.

[49] MOHAMMED, S. K., LARSSON, E. G., "Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems", *IEEE Transactions on Communications*, v. 61, n. 3, pp. 1059–1071, March 2013.

[50] ZHANG, J., HUANG, Y., WANG, J., *et al.*, "Per-Antenna Constant Envelope Precoding and Antenna Subset Selection: A Geometric Approach", *IEEE*

Transactions on Signal Processing, v. 64, n. 23, pp. 6089–6104, December 2016.

[51] WANG, Z., BABU, P., PALOMAR, D. P., "Design of PAR-Constrained Sequences for MIMO Channel Estimation via Majorization–Minimization", *IEEE Transactions on Signal Processing*, v. 64, n. 23, pp. 6132–6144, December 2016.

[52] BAO, H., FANG, J., CHEN, Z., *et al.*, "An Efficient Bayesian PAPR Reduction Method for OFDM-Based Massive MIMO Systems", *IEEE Transactions on Wireless Communications*, v. 15, n. 6, pp. 4183–4195, June 2016.

[53] GOKCEOGLU, A., BJORNSON, E., LARSSON, E. G., *et al.*, "Waveform Design for Massive MISO Downlink with Energy-Efficient Receivers Adopting 1-bit ADCs". In: *IEEE International Conference on Communications*, pp. 1–7, Kuala Lumpur, May 2016.

[54] CHEN, J.-C., "Low-PAPR Precoding Design for Massive Multiuser MIMO Systems via Riemannian Manifold Optimization", *IEEE Communications Letters*, v. 21, n. 4, pp. 945–948, April 2017.

[55] BEREYHI, A., SEDAGHAT, M. A., ASAAD, S., *et al.*, "Nonlinear Precoders for Massive MIMO Systems with General Constraints". In: *21th International ITG Workshop on Smart Antennas*, pp. 23–30, Berlin, March 2017.

[56] CASTANEDA, O., GOLDSTEIN, T., STUDER, C., "POKEMON: A Non-Linear Beamforming Algorithm for 1-bit Massive MIMO". In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3464–3468, New Orleans, March 2017.

[57] JEDDA, H., NOSSEK, J. A., MEZGHANI, A., "Minimum BER Precoding in 1-Bit Massive MIMO Systems". In: *Sensor Array and Multichannel Signal Processing Workshop*, pp. 1–5, Rio de Janeiro, July 2016.

[58] STÖCKLE, C., MUNIR, J., MEZGHANI, A., *et al.*, "Channel Estimation in Massive MIMO Systems Using 1-Bit Quantization". In: *17th International Workshop on Signal Processing Advances in Wireless Communications*, n. 1, pp. 1–6, Edinburg, July 2016.

[59] GOKCEOGLU, A., BJÖRNSON, E., LARSSON, E. G., *et al.*, "Spatio-Temporal Waveform Design for Multi-user Massive MIMO Downlink with 1-bit Receivers", *IEEE Journal of Selected Topics in Signal Processing*, v. 11, n. 2, pp. 347–362, March 2017.

[60] HEATH, R., PAULRAJ, A., "Antenna Selection for Spatial Multiplexing Systems Based on Minimum Error Rate". In: *IEEE International Conference on Communications*, pp. 2276–2280, Helsink, June 2001.

[61] HEATH, R. W., SANDHU, S., PAULRAJ, A., "Antenna Selection for Spatial Multiplexing Systems with Linear Receivers", *IEEE Communications Letters*, v. 5, n. 4, pp. 142–144, April 2001.

[62] GAO, X., EDFORS, O., LIU, J., *et al.*, "Antenna Selection in Measured Massive MIMO Channels Using Convex Optimization". In: *IEEE Globecom Workshops*, pp. 129–134, Atlanta, December 2013.

[63] GAO, X., EDFORS, O., TUFVESSON, F., *et al.*, "Multi-Switch for Antenna Selection in Massive MIMO". In: *IEEE Global Communications Conference*, pp. 1–6, San Diego, December 2014.

[64] GKIZELI, M., KARYSTINOS, G. N., "Maximum-SNR Antenna Selection Among a Large Number of Transmit Antennas", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 891–901, October 2014.

[65] GAO, X., EDFORS, O., TUFVESSON, F., *et al.*, "Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally?", *IEEE Transactions on Communications*, v. 63, n. 11, pp. 3917–3928, November 2015.

[66] LARSSON, E. G., EDFORS, O., TUFVESSON, F., *et al.*, "Massive MIMO for Next Generation Wireless Systems", *IEEE Communications Magazine*, v. 52, n. 2, pp. 186–195, February 2014.

[67] MARZETTA, T. L., LARSSON, E. G., YANG, H., *et al.*, *Fundamentals of Massive MIMO*. 1 ed. Cambridge, Cambridge University Press, 2016.

[68] COVER, T. M., THOMAS, J. A., *Elements of Information Theory*. 2 ed. New Jersey, John Wiley & Sons, Inc., 2006.

[69] KUO, P.-H., KUNG, H. T., TING, P.-A., "Compressive Sensing Based Channel Feedback Protocols for Spatially-correlated Massive Antenna Arrays". In: *IEEE Wireless Communications and Networking Conference*, pp. 492–497, Shangai, April 2012.

[70] RAO, X., LAU, V. K. N., "Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems", *IEEE Transactions on Signal Processing*, v. 62, n. 12, pp. 3261–3271, June 2014.

[71] CHOI, J., LOVE, D. J., BIDIGARE, P., "Downlink Training Techniques for FDD Massive MIMO Systems: Open-Loop and Closed-Loop Training With Memory", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 802–814, October 2014.

[72] NAM, J., ADHIKARY, A., AHN, J.-Y., *et al.*, "Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 876–890, October 2014.

[73] CHOI, J., CHANCE, Z., LOVE, D. J., *et al.*, "Noncoherent Trellis Coded Quantization: A Practical Limited Feedback Technique for Massive MIMO Systems", *IEEE Transactions on Communications*, v. 61, n. 12, pp. 5016–5029, December 2013.

[74] FOSCHINI, G. J., GANS, M. J., "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas", *Wireless Personal Communications*, v. 6, n. 3, pp. 311–335, March 1998.

[75] RALEIGH, G. G., CIOFFI, J. M., "Spatio-Temporal Coding for Wireless Communication", *IEEE Transactions on Communications*, v. 46, n. 3, pp. 357–366, March 1998.

[76] TAROKH, V., SESHADRI, N., CALDERBANK, A. R., "Space-time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Transactions on Information Theory*, v. 44, n. 2, pp. 744–765, March 1998.

[77] ALAMOUTI, S. M., "A Simple Transmit Diversity Technique for Wireless Communications", *IEEE Journal on Selected Areas in Communications*, v. 16, n. 8, pp. 1451–1458, October 1998.

[78] MARZETTA, T. L., HOCHWALD, B. M., "Capacity of a Mobile Multiple-antenna Communication Link in Rayleigh Flat Fading", *IEEE Transactions on Information Theory*, v. 45, n. 1, pp. 139–157, January 1999.

[79] GUEY, J.-C., FITZ, M. P., BELL, M. R., *et al.*, "Signal Design for Transmitter Diversity Wireless Communication Systems over Rayleigh Fading Channels", *IEEE Transactions on Communications*, v. 47, n. 4, pp. 527–537, April 1999.

[80] TAROKH, V., JAFARKHANI, H., CALDERBANK, A. R., "Space-time Block Codes from Orthogonal Designs", *IEEE Transactions on Information Theory*, v. 45, n. 5, pp. 1456–1467, July 1999.

[81] TELATAR, E., "Capacity of Multi-antenna Gaussian Channels", *European Transactions on Telecommunications*, v. 10, n. 6, pp. 585–595, November 1999.

[82] TSE, D., VISWANATH, P., "MIMO II: capacity and multiplexing architectures". In: *Fundamentals of Wireless Communication*, chapter 8, Cambridge, Cambridge University Press, pp. 332–382, 2005.

[83] DINIZ, P. S. R., MARTINS, W. A., LIMA, M. V. S., *Block Transceivers: OFDM and Beyond*, v. 5. Colorado Springs, CO, Morgan&Claypool, 2012.

[84] TSE, D., VISWANATH, P., "MIMO IV: multiuser communication". In: *Fundamentals of Wireless Communication*, chapter 10, Cambridge, Cambridge University Press, pp. 425–495, 2005.

[85] HOYDIS, J., TEN BRINK, S., DEBBAH, M., "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?", *IEEE Journal on Selected Areas in Communications*, v. 31, n. 2, pp. 160–171, February 2013.

[86] MULLER, R. R., COTTATELLUCCI, L., VEHKAPERA, M., "Blind Pilot Decontamination", *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, pp. 773–786, October 2014.

[87] NGO, H. Q., LARSSON, E. G., MARZETTA, T. L., "Aspects of Favorable Propagation in Massive MIMO". In: *22nd European Signal Processing Conference*, pp. 1–5, Lisbon, September 2014.

[88] WU, X., BEAULIEU, N. C., LIU, D., "On Favorable Propagation in Massive MIMO Systems and Different Antenna Configurations", *IEEE Access*, v. 3, pp. 5578 – 5593, April 2017.

[89] SYLVESTER, J. J., "On the Relation Between the Minor Determinants of Linearly Equivalent Quadratic Functions", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 1, n. 4, pp. 295–305, 1851.

[90] GARLING, D. J. H., *Inequalities: A Journey into Linear Analysis*. Cambridge, Cambridge University Press, 2007.

[91] GAO, Y., JIANG, W., KAISER, T., "Bidirectional Branch and Bound Based Antenna Selection in Massive MIMO Systems". In: *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, pp. 563–568, Hong Kong, August 2015.

[92] VAN TREES, H. L., *Optimum Array Processing*. 1 ed. New York, John Wiley & Sons, Inc., 2002.

[93] COSTA, M., "Writing on Dirty Paper", *IEEE Transactions on Information Theory*, v. 29, n. 3, pp. 439–441, May 1983.

[94] HOCHWALD, B., PEEL, C., SWINDLEHURST, A., "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication—Part II: Perturbation", *IEEE Transactions on Communications*, v. 53, n. 3, pp. 537–544, March 2005.

[95] WINDPASSINGER, C., FISCHER, R., HUBER, J., "Lattice-Reduction-Aided Broadcast Precoding", *IEEE Transactions on Communications*, v. 52, n. 12, pp. 2057–2060, December 2004.

[96] MÜLLER, R. R., ALFANO, G., ZAIDEL, B. M., *et al.*, "Applications of Large Random Matrices in Communications Engineering", pp. 1–84, October 2013.

[97] PENG, W., ZHENG, L., CHEN, D., *et al.*, "Distributed Precoding for BER Minimization with PAPR Constraint in Uplink Massive MIMO Systems", *IEEE Access*, v. 6, pp. 6668–6676, 2017.

[98] TIKHONOV, A. N., "Solution of Incorrectly Formulated Problems And Regularization Method", *Doklady Akademii Nauk SSSR, 151*, v. 151, n. 3, pp. 501–504, 1963.

[99] EREZ, U., SHAMAI, S., ZAMIR, R., "Capacity and Lattice Strategies for Canceling Known Interference", *IEEE Transactions on Information Theory*, v. 51, n. 11, pp. 3820–3833, November 2005.

[100] DE CAMPOS, M. L. R., WERNER, S., APOLINÁRIO, J. A., "Constrained Adaptive Filters". In: *Adaptive Antenna Arrays Trends and Applications*, 1 ed., chapter 3, Berlin, Springer-Verlag, pp. 46–64, 2004.

[101] DINIZ, P. S. R., *Adaptive Filtering: Algorithms and Practical Implementation*. 4 ed. Boston, Springer, 2013.

[102] DE ANDRADE, J. F., DE CAMPOS, M. L. R., APOLINÁRIO, J. A., "$\mathbf{L}_1$-Constrained Normalized LMS Algorithms for Adaptive Beamforming", *IEEE Transactions on Signal Processing*, v. 63, n. 24, pp. 6524–6539, December 2015.

[103] LAMPE, A., HUBER, J., "On Improved Multiuser Detection with Iterated Soft Decision Interference Cancellation". In: *IEEE Communications Theory Mini-Conference*, pp. 172–176, Vancouver, June 1999.

[104] LIANG, Y.-C., SUN, S., HO, C. K., "Block-Iterative Generalized Decision Feedback Equalizers for Large MIMO Systems: Algorithm Design and Asymptotic Performance Analysis", *IEEE Transactions on Signal Processing*, v. 54, n. 6, pp. 2035–2048, June 2006.

[105] VARDHAN, K. V., MOHAMMED, S. K., CHOCKALINGAM, A., *et al.*, "A Low-Complexity Detector for Large MIMO Systems and Multicarrier CDMA Systems", *IEEE Journal on Selected Areas in Communications*, v. 26, n. 3, pp. 473–485, April 2008.

[106] ZHAO, H., LONG, H., WANG, W., "Tabu Search Detection for MIMO Systems". In: *18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, Athens, September 2007.

[107] FINCKE, U., POHST, M., "Improved Methods for Calculating Vectors of Short Length in a Lattice, Including a Complexity Analysis", *Mathematics of Computation*, v. 44, n. 170, pp. 463, April 1985.

[108] SALAH, A., OTHMAN, G. R. B., OUERTANI, R., *et al.*, "New Soft Stack Decoder for MIMO Channel". In: *42nd Asilomar Conference on Signals, Systems and Computers*, pp. 1754–1758, Pacific Grove, October 2008.

[109] BARBERO, L., THOMPSON, J., "Fixing the Complexity of the Sphere Decoder for MIMO Detection", *IEEE Transactions on Wireless Communications*, v. 7, n. 6, pp. 2131–2142, June 2008.

[110] DONG, K., PRASAD, N., WANG, X., *et al.*, "Adaptive Antenna Selection and Tx/Rx Beamforming for Large-Scale MIMO Systems in 60 GHz Channels", *EURASIP Journal on Wireless Communications and Networking*, v. 1, n. 59, pp. 1–14, August 2011.

[111] FANG, B., QIAN, Z., SHAO, W., *et al.*, "RAISE: A New Fast Transmit Antenna Selection Algorithm for Massive MIMO Systems", *Wireless Personal Communications*, v. 80, n. 3, pp. 1147–1157, February 2015.

[112] NI, W., DONG, X., "Hybrid Block Diagonalization for Massive Multiuser MIMO Systems", *IEEE Transactions on Communications*, v. 64, n. 1, pp. 201–211, January 2016.

[113] BOGALE, T. E., LE, L. B., HAGHIGHAT, A., *et al.*, "On the Number of RF Chains and Phase Shifters, and Scheduling Design With Hybrid Analog–Digital Beamforming", *IEEE Transactions on Wireless Communications*, v. 15, n. 5, pp. 3311–3326, May 2016.

[114] ALKHATEEB, A., NAM, Y.-H., ZHANG, J., *et al.*, "Massive MIMO Combining with Switches", *IEEE Wireless Communications Letters*, v. 5, n. 3, pp. 232–235, June 2016.

[115] SAH, A. K., CHATURVEDI, A. K., "Quasi-Orthogonal Combining for Reducing RF Chains in Massive MIMO Systems", *IEEE Wireless Communications Letters*, v. 6, n. 1, pp. 126 – 129, February 2017.

[116] GARCIA-RODRIGUEZ, A., MASOUROS, C., RULIKOWSKI, P., "Reduced Switching Connectivity for Large Scale Antenna Selection", *IEEE Transactions on Communications*, v. 65, n. 5, pp. 2250–2263, May 2017.

[117] JOSHI, S., BOYD, S., "Sensor Selection via Convex Optimization", *IEEE Transactions on Signal Processing*, v. 57, n. 2, pp. 451–462, February 2009.

[118] FRIEDMAN, J. H., STUETZLE, W., "Projection Pursuit Regression", *Journal of the American Statistical Association*, v. 76, n. 376, pp. 817–823, December 1981.

[119] MALLAT, S. G., ZHANG, Z., "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Transactions on Signal Processing*, v. 41, n. 12, pp. 3397–3415, December 1993.

[120] QIAN, S., CHEN, D., "Signal Representation Using Adaptive Normalized Gaussian Functions", *Signal Processing*, v. 36, n. 1, pp. 1–11, March 1994.

[121] CHEN, S., BILINGS, S. A., LUO, W., "Orthogonal Least Squares Methods and Their Application to Non-linear System Identification", *International Journal of Control*, v. 50, n. 5, pp. 1873–1896, November 1989.

[122] PATI, Y. C., REZAIIFAR, R., KRISHNAPRASAD, P. S., "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition". In: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–44, Pacific Grove, November 1993.

[123] DAVIS, G. M., MALLAT, S. G., ZHANG, Z., "Adaptive Time-Frequency Decompositions", *Optical Engineering*, v. 33, n. 7, pp. 2183–2191, July 1994.

[124] NEEDELL, D., TROPP, J. A., "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples", *Applied and Computational Harmonic Analysis*, v. 26, n. 3, pp. 301–321, May 2009.

[125] DAI, W., MILENKOVIC, O., "Subspace Pursuit for Compressive Sensing Signal Reconstruction", *IEEE Transactions on Information Theory*, v. 55, n. 5, pp. 2230–2249, May 2009.

[126] FOUCART, S., RAUHUT, H., *A Mathematical Introduction to Compressive Sensing*, Applied and Numerical Harmonic Analysis. New York, Springer, 2013.

[127] BLUMENSATH, T., DAVIES, M. E., "Iterative Hard Thresholding for Compressed Sensing", *Applied and Computational Harmonic Analysis*, v. 27, n. 3, pp. 265–274, November 2009.

[128] FOUCART, S., "Hard Thresholding Pursuit: An Algorithm for Compressive Sensing", *SIAM Journal on Numerical Analysis*, v. 49, n. 6, pp. 2543–2563, January 2011.

[129] CHEN, S. S., DONOHO, D. L., SAUNDERS, M. A., "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, v. 20, n. 1, pp. 33–61, January 1998.

[130] DONOHO, D. L., "Compressed Sensing", *IEEE Transactions on Information Theory*, v. 52, n. 4, pp. 1289–1306, April 2006.

[131] TIBSHIRANI, R., "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B*, v. 58, n. 1, pp. 267–288, January 1996.

[132] CANDES, E., TAO, T., "The Dantzig Selector: Statistical Estimation When p Is Much Larger than n", *The Annals of Statistics*, v. 35, n. 6, pp. 2313–2351, December 2007.

[133] GOLLAMUDI, S., Yih-Fang Huang, "Updator-Shared Adaptive Parallel Equalization (U-SHAPE) Using Set-Membership Identification". In: *IEEE International Symposium on Circuits and Systems*, pp. 13–16, Atlanta, May 1996.