Fachgebiet Methoden der Signalverarbeitung
Technische Universität München

Linear Precoding and Analysis of Performance Criteria in MIMO Interference Channels

Samer Bazzi

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. -Ing. Wolfgang Kellerer
Prüfer der Dissertation:	1. Prof. Dr. -Ing. Wolfgang Utschick
	2. Prof. David Gesbert, Ph. D.

Die Dissertation wurde am 28.09.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 04.03.2016 angenommen.

Contents

Acknowledgments

First of all, I would like to express my gratitude to DOCOMO Euro-Labs for giving me the chance to complete my thesis and perform research on various interesting and challenging subjects. Special thanks go to my DOCOMO supervisor Prof. Guido Dietl for his valuable help, advice, and comments on different aspects of my thesis. He gave me his trust, was very supportive through my thesis and was always available for discussions and idea exchange. Special thanks go to Prof. Wolfgang Utschick as well for agreeing to be my thesis supervisor and treating me like an internal member of his group. He provided critical valuable comments, carefully judged my research output, and pointed out to new directions whenever necessary. Both supervisors' excitement to explore new wireless communications topics was a constant motivation for me and their knowledge in different theoretical and practical aspects shaped my work.

I thank Hauke, Serkan, and later Emmanuel for being such easygoing and friendly office mates, with which I could have numerous discussions on work and non-work related topics. I thank my remaining colleagues in the wireless team Jamal, Petra, Patrick, Toshi, Iwamura-san, Marwa, and Gerhard for being friendly and for providing a relaxed working atmosphere in the office in general.

During my six years in Munich, I have met new friends or reunited with old friends. These include but are not limited to David, Abdallah, Layal, Mohammed, Ronnie, Jelena, Noemi, Mari, and Nabil. I greatly cherish them and thank them for the great times we had together and for making life in Munich a pleasant one so far.

Finally, my deepest gratitude goes to my parents and siblings for their unconditional support, care, and love in addition to their trust in me. I am lucky to have them and I dedicate this work to them.

Abstract

A lot of interest has been directed towards wireless multiple-input multiple-output (MIMO) interference channels in the past years. This interest was triggered by interference alignment (IA), a technique which creates noise-limited channels out of interference-limited channels through transmitter cooperation. The existing literature on MIMO interference channels now covers a variety of linear precoding techniques, whose advantages and disadvantages are well understood by now.

On the other hand, much less results exist on the analytical performance characterization of those techniques in terms of achievable rates. Such a characterization is important for different reasons. For instance, closed-form rate expressions reveal how different system parameters affect the resulting performance, an aspect that is not revealed by simulations. Furthermore, closed-form expressions are useful for benchmarking purposes, and save simulations time and cost. Additionally, they can be used in a variety of optimization problems.

Random matrix theory tools alleviate the task of obtaining closed-form expressions substantially. Even though such tools result in expressions that are only exact asymptotically, these expressions usually provide accurate estimates for finite system parameters as well. In this thesis, we use random matrix theory tools to derive closed-form rate expressions of IA techniques, and closed-form rate lower bounds of eigenmode precoding in MIMO interference channels. Additionally, rate lower bounds of the maximum ratio transmission technique are derived. Due to the special nature of the latter, this can be performed without the need for random matrix theory. The performance characterization of the latter two precoding types is especially important in scenarios where transmitters are equipped with large antenna arrays (massive MIMO), in which case these non- cooperative and relatively simple techniques exhibit a good performance.

The ultimate measure to characterize performance is the spectral efficiency, which takes into account any signaling overheads not related to data transmission. The most important overheads to consider are the channel state information acquisition training overheads, taking place over the air link. We calculate the training overheads of the different considered precoding techniques in time-division-duplex mode, which, in addition to the derived rate expressions, allow characterizing the spectral efficiencies of these different techniques.

The spectral efficiency analysis allows investigating how many transmit antennas do the considered non-cooperative techniques require to emulate the per-

formance of a noise-limited system, i.e., to perform similarly to a system where transmitters have a fixed number of antennas and employ IA. The reason for such an investigation is that the considered non-cooperative techniques might be more feasible than IA in practical systems for various reasons. Whereas the literature on massive MIMO envisions transmitters with hundreds or unlimited number of antennas where the simplest precoding techniques result in noise-limited systems, this thesis gives a more realistic and relatively modest upper bound on the number of transmit antennas that simple maximum ratio transmission and eigenmode precoding techniques require to emulate the performance of noise-limited interference channels through the use of IA. Furthermore, the value of this upper bound drops significantly in fast-varying scenarios due to the unavoidable higher training overheads of IA which considerably reduce its theoretical promised gains and resulting spectral efficiency. Such overheads are largely neglected in the literature which mainly considers channels with constant coefficients.

1. Introduction

1.1. Thesis Context

While early cellular networks such as the global system for mobile communications (GSM) mainly focused on ensuring reliable low-rate telephony services, current cellular networks are driven by the need to increase user data rates. Besides the fact that the increasing number of smart-phones connected to the internet will necessitate major improvements to current network architectures, future 5G wireless networks will comprise a multitude of different entities and is expected to support different new services such as the internet of things, device-to-device communication, and car-to-car communication. According to initial forecasts, around 50 billion entities will be connected to the network by 2020 [1], which is around 100 times the current number. This huge number of connected entities will require new physical layer, random access, as well as upper layer architectures.

5G systems envision a 1000-fold increase in data rates compared to current long-term evolution (LTE) systems' rates [1, 2]. This figure includes users with high mobility as well as cell-edge users. This 1000-fold increase can only be achieved through a combination of different features, including network densification, bandwidth expansion, new waveform designs, and advances in multiple-input multiple-output (MIMO) systems. Each of these features has its share of advantages and challenges from a theoretical and a practical point of view. This thesis focuses on MIMO systems.

MIMO systems have received a lot of attention since the pioneering works of Foschini in [3] and Telatar in [4], where the possibility for linear growths in sum-rate and capacity were shown for the point-to-point (P2P) MIMO model, which consists of a transmitter and a receiver with multiple antennas each. Later on, similar conclusions were shown to hold for the MIMO broadcast channel (BC), which consists of a transmitter and a number of decentralized non-cooperating receivers, despite some fundamental differences between these two channel models [5–7]. Despite the natural big gap between theoretical investigations and practical implementations, some simple MIMO techniques are now part of the LTE standards in the downlink and uplink, including single-user as well multi-user scenarios [8]. Additional schemes relying on spatial diversity such as space-frequency block coding schemes are a part of the standards as well, mainly used to improve reliability and combat multi-path fading.

Achieving a 1000-fold increase in data rates is an especially challenging task for cell-edge users, as the rates experienced by such users are still orders of magnitudes lower than the cell average, even with initial coordinated multi-point (CoMP) transmission schemes [9]. This is mainly due to large path losses, large experienced interference power from neighboring base stations, and the lack of accurate channel state information (CSI) at the base stations to jointly design the beamformers/precoders and mitigate interference on a multi-cell level. While network densification and reduced cell sizes can reduce the path loss of such users, advanced cooperative precoding designs are still necessary for efficient inter-cell interference mitigation in current LTE systems. With accurate multi-cell CSI exchanged between the transmitters, interference mitigation techniques promise huge improvements in cell-edge rates [10–13]. A stronger form of cooperation exists when transmitters exchange CSI as well as data symbols, leading to what is called network MIMO [14–16]. With accurate multi-cell CSI, network MIMO converts interference power from interferers into additional useful signal power coming from the latter. In that case, the different transmitters have the function of a distributed transmit array present in multiple cells and the conventional concepts of “serving” and “interfering” transmitters or base stations do not apply anymore. However, this thesis focuses on the former type of cooperation.

An alternative way to fight interference is through the deployment of large transmit antennas arrays serving a much smaller number of users, called massive MIMO systems, as advocated in [17, 18]. The insight behind this approach is that under sufficient antenna spacing and rich fading conditions, interference coming from different transmitters equipped with large antenna arrays will simply "fade out" or add destructively at the receiver terminals, becoming negligible with respect to the signal of interest. In such scenarios, cooperative precoding techniques would no longer be necessary and each transmitter could revert back to single-cell (non-cooperative) processing. Unfortunately, the above mentioned references in addition to others in the literature assume that transmitters are equipped with hundreds or unlimited number of antennas, which might be unfeasible for different reasons including high deployment costs or lack of space in dense urban areas.

1.2. Thesis Overview and Contributions

In this thesis, we aim to find out how much transmit antennas do single-cell precoding techniques necessitate to result in a similar performance to the one yielded by the cooperative interference mitigation precoding techniques. The reason for such an investigation is that the need for accurate multi-cell CSI at the transmitters and receivers, the mainly iterative nature and unpredictable

convergence behavior, and computational complexity are factors which, so far, hinder the implementation of the latter techniques in a real system. On the other hand, we consider two simple single-cell precoding techniques in this paper which require modest CSI knowledge, are simple to compute and are therefore more feasible in practical systems from a signal processing or algorithmic point of view. The main challenge in this case, however, lies in ensuring feasible and low cost hardware implementations of large non-conventional transmit arrays.

The task at hand requires a spectral efficiency analysis of the different precoding techniques in question. That, in turn, includes:

1. an analysis of the achievable rates during data transmission, and
2. an analysis of the CSI acquisition training overheads, which take place over the air link and occupy a portion of time that cannot be used for data transmission.

In what follows, we briefly summarize the contents and contributions of each chapter.

Chapter 2:

In this chapter, we discuss the advantages of MIMO systems and briefly review linear precoding designs in P2P and BC MIMO scenarios. We motivate the use of massive MIMO with a simple example, and present two commonly used duplex transmission and reception modes.

Chapter 3:

In this chapter, we first introduce the thesis system model, the K-user MIMO interference channel (IC). This model, which consists of K interfering transmitter-receiver pairs, efficiently represents communication near cell-edges in a cellular system. We distinguish three main communication modes according to the system configuration in terms of the number of transmit antennas, receive antennas, desired streams, and users. Focusing on the most challenging mode, we review different linear precoding methods such as interference alignment (IA) [10, 11]. Except for special cases, the methods in this mode are iterative and require an alternating optimization procedure. We then discuss the possible transmission strategies in the remaining two modes. At the end of that chapter, we differentiate between *cooperative* and *non-cooperative* methods, listing the advantages, disadvantages, and requirements of each method. Additionally, we discuss how do the discussed cooperative methods relate to current LTE CoMP methods.

Chapter 4:

In this chapter, we focus on cooperative IA methods and derive closed-form expressions of their achievable rates in MIMO ICs in the large system limit. Random matrix theory (RMT) tools such as the Marčenko-Pastur distribution [19–21] and the Shannon transform [21] are integral parts of this chapter, and are first reviewed for completeness. The law of large numbers (LLN), described in Appendix A.4, constitutes an important building block of this chapter as well. Even though the derived expressions are only exact asymptotically, simulation results show that these expressions provide accurate estimates for small and finite system parameters as well. The derived expressions hold for any variant of IA, given the designed precoders and receive filters have orthogonal columns. Additionally, Chapter 4 reveals some interesting observations about the asymptotic water-filling level which, contrary to the finite case, is independent of the instantaneous eigenvalues and is completely determined by the asymptotic eigenvalue distribution.

Chapter 5:

In this chapter, we focus on non-cooperative eigenmode precoding (cf. Appendix B) and maximum ratio transmission (MRT) [22] methods. We derive ergodic (average) lower bounds of the achievable rates of these methods in MIMO ICs using a separate stream decoding assumption and Jensen's inequality. The lower bounds for eigenmode precoding are derived using RMT tools as well, while the special nature of MRT combined with relaxed CSI requirements at the receiver allows deriving lower bounds without the need for RMT tools. The bounds are observed by simulations to get tighter as the number of interferers increases. This has a straightforward interpretation that we discuss, given the separate stream decoding assumption used to obtain the bounds.

Chapter 6:

In this chapter, we perform a training overhead analysis of IA, eigenmode precoding, and MRT in a time-division-duplex system. Along with the derived expressions of Chapters 4 and 5, this allows characterizing the spectral efficiency of these methods. We investigate the required number of transmit antennas for the non-cooperative methods of Chapter 5 to perform similarly to IA. We conclude that massive configurations are not necessary to emulate to the performance of IA, and we find a relatively modest upper bound on that required number. We show how the training overheads of IA can significantly reduce its spectral efficiency in scenarios with short coherence times, and lead to a considerable drop of the value of the found upper bound.

Chapter 7:

In this last chapter, we summarize the main findings of this thesis and we discuss open problems and possible future lines of work.

1.3. Notations

In this thesis, variables are written in italic font. In addition, lower case bold italic letters denote column vectors while upper case bold italic letters denote matrices.

Below, a list of notations frequently used throughout the thesis is presented.

$[\boldsymbol{A}]_{mn}$	(m,n)th entry of $\boldsymbol{A}$
$\lambda_l(\boldsymbol{A})$	lth largest eigenvalue of $\boldsymbol{A}$
$\sigma_l(\boldsymbol{A})$	lth largest singular value of $\boldsymbol{A}$
$\mathrm{tr}(\boldsymbol{A})$	trace of $\boldsymbol{A}$
$\lvert\boldsymbol{A}\rvert$	determinant of $\boldsymbol{A}$
$\boldsymbol{A}^{\perp}$	nullspace of $\boldsymbol{A}$
$\lvert\lvert\boldsymbol{A}\rvert\rvert_{\mathrm{F}}$	Frobenius norm of $\boldsymbol{A}$
$\boldsymbol{A}^{1:l}$	first l columns of $\boldsymbol{A}$
$\boldsymbol{A}^{l}$	lth column of $\boldsymbol{A}$
$\mathbf{0}_N$	zero vector of size N
$\mathbf{1}_N$	ones vector of size N
$\mathbf{0}_{N\times M}$	zero matrix of size $N \times M$
$\mathbf{I}_N$	identity square matrix of size $N \times N$
$\mathrm{diag}(a_1, a_2, \ldots, a_K)$	diagonal matrix with elements $a_1, a_2, \ldots, a_K$
$\mathrm{Bdiag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_K)$	block diagonal matrix with matrix elements $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_K$
$\boldsymbol{e}_t$	column vector of suitable size containing 1 on its tth entry and 0 elsewhere

$(\bullet)^{\mathrm{T}}$	transpose of a vector or matrix
$(\bullet)^{\mathrm{H}}$	conjugate transpose (Hermitian) of a vector or matrix
$\|\|\boldsymbol{x}\|\|_2$	Euclidean norm of $\boldsymbol{x}$
$\mathrm{E}[\bullet]$	expectation
$f(x)\Big\|_a^b$	$f(b) - f(a)$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of strictly positive real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{C}^N$	the N-dimensional complex vector space
$(\bullet)^*$	conjugate
$\log(\bullet)$	base 2 logarithm
$\ln(\bullet)$	natural logarithm
$\|\bullet\|$	absolute value
$\mathrm{var}(\bullet)$	variance
$\Re(\bullet)$	real part
$\Im(\bullet)$	imaginary part
$\lceil\bullet\rceil$	ceiling
$\mathcal{N}(\mu, c)$	real Gaussian (normal) distribution with mean μ and variance c
$\mathcal{N}_{\mathbb{C}}(\mu, c)$	circularly symmetric complex Gaussian distribution with mean μ and variance c
$\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{C})$	circularly symmetric multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$
$(M, N, d)^K$	the K-user MIMO IC with M antennas at each transmitter, N antennas at each receiver, and d communicated streams between each transmitter-receiver pair

2. MIMO Systems: An Overview

This chapter motivates the use of MIMO wireless communications systems. After introducing a simple P2P MIMO model, we discuss the advantages of MIMO communications and briefly review precoding strategies in P2P as well as BC scenarios. We then explain why such strategies fail to deliver enough data rates in interference-limited scenarios. We additionally motivate the use of massive MIMO systems, and introduce two commonly used duplex modes in practical systems.

2.1. MIMO Communications in Noise Limited Networks

2.1.1. Channel Capacity and Fundamental Limits of Communications

In 1948, C. E. Shannon developed the main information theoretic tools required to guarantee reliable communications over a wireless P2P link, where the transmitter and receiver have a single antenna each [23]. By means of the noisy-channel coding theorem, Shannon showed that given a link *capacity* C and a desired transmission rate $R < C$, there exists channel codes that guarantee reliable transmission over the P2P link, i.e., codes that make the bit error probability at the receiver arbitrarily small. The capacity C is therefore the maximum data rate that can be reliably achieved and usually serves as an upper bound for the achievable rates of practical systems. Achieving capacity requires coding over arbitrarily large blocks of data or symbols, which is usually not the case in practice.

C is hard to characterize in the general case. Nonetheless, when the link is impaired by Gaussian noise, we have an additive white Gaussian noise (AWGN) link and choosing symbols from a Gaussian distribution maximizes the transmission rate and gives the corresponding capacity C_{AWGN}. The latter is given by the Shannon-Hartley theorem [24, 25]:

$$C_{\text{AWGN}} = \log\left(1 + \frac{p}{\sigma^2}\right) \text{ bits/sec/Hz} \tag{2.1}$$

where p is the received power and σ^2 is the Gaussian noise power. The ratio p/σ^2 is called the signal-to-noise-ratio (SNR). This simple equation reveals the fundamental limits of communications in this simple single-input-single-output (SISO) scenario. These are the received power, the additive noise power, and

the available bandwidth. Additive noise is a natural phenomenon that can be overcome through advancements in hardware or chip design for instance, but is uncontrollable from an information theoretic point of view. Bandwidth is usually an expensive resource that cannot be acquired easily, and in many practical systems the available bandwidth is fixed. This leaves the received power as a parameter that can be optimized. In the case of an omnidirectional antenna at the transmitter, the transmitted power will radiate in all directions, and only a small fraction will actually reach the receiver. Increasing the received power by a certain factor requires thus increasing the transmit power by the same factor. Keeping in mind that the transmit power cannot be arbitrarily increased due to high power costs as well as practical and environmental reasons, this shows a direct drawback of using single omnidirectional antennas.

2.1.2. Advantages of MIMO Systems

With this short introduction that highlighted the shortcomings of SISO systems, we now introduce and explain the advantages of MIMO wireless systems. MIMO systems offer different benefits, not all of which can be cultivated simultaneously as will be elaborated.

First, we focus on a P2P MIMO link. Such a link consists of a transmitter with M antennas and a receiver with N antennas, as shown in Fig. 2.1. In the frequency domain, the transfer function between the transmitter (input) and receiver (output) is modeled by the complex channel matrix $\boldsymbol{H} \in \mathbb{C}^{N \times M}$, whose (k,l)th entry h_{kl} contains the fading—attenuation or channel response due to transmission over the air link [25]—between transmit antenna l and receive antenna k. The system equation for the P2P link reads:

$$\hat{\boldsymbol{s}} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{2.2}$$

where $\hat{\boldsymbol{s}} \in \mathbb{C}^N$ is the vector of received signals, $\boldsymbol{n} \in \mathbb{C}^N$ is the additive noise present at the receiver, and $\boldsymbol{x} \in \mathbb{C}^M$ is the vector of signals fed to the transmit antennas and sent over the air link. Note that this predominantly used *linear* channel model (cf. Appendix A.1) is only an approximation of a real system as it neglects D/A and A/D conversions, amplifications, etc., which are not linear and are not considered to be part of the channel matrix $\boldsymbol{H}$. The vector $\boldsymbol{x}$ can be designed in a multitude of ways, according to the system configuration and requirements. We start with beamforming.

Beamforming / Array Gain

A way to increase the received power without increasing the transmit power is through the use of beamforming. Through a different phase weighting at each transmit antenna, a beam with the desired data symbol is shaped and directed

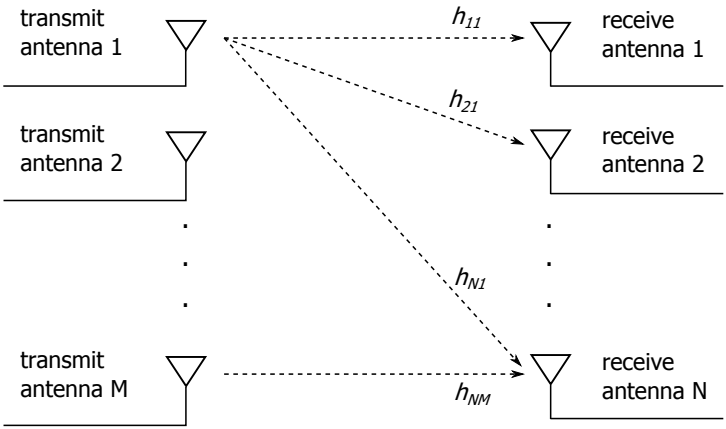

Figure 2.1.: A P2P MIMO link

towards the intended receiver, which results in an increase in received power. Thus,

$$\boldsymbol{x} = \boldsymbol{v}s$$

where

$$\boldsymbol{v} = [\exp(j\theta_1), \ldots, \exp(j\theta_M)]^{\mathrm{T}}$$

is the steering vector whose lth entry contains the phase weighting θ_l at transmit antenna l, and s is the transmitted data symbol. Forming narrow and focused beams towards a certain receiver requires the transmit antennas to be closely-spaced (namely, half a wavelength of the frequency of operation [26]), which implies that the channel responses coming from the different transmit antennas could be highly correlated. Similarly, receive beamforming can be applied to further increase the received power. The details of beamforming are skipped as it is outside the scope of the thesis.

Spatial Diversity

Provided that antenna correlation is low, the use of multiple transmit antennas provides diversity in *space*, and helps in fighting fading conditions as the channel response between each transmit-receive antenna pair will not only have a different phase but also a different amplitude. In such scenarios, the transmitted vector reads

$$\boldsymbol{x} = \mathbf{1}_M s \tag{2.3}$$

where s is the transmitted symbol, i.e., each transmit antenna sends s without any transformations. In ideal conditions, the fading experienced by different transmit and receive antennas will be independent and identically distributed (i.i.d.) with a Rayleigh distribution. Equivalently, this means $[\boldsymbol{H}]_{kl} \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $\forall k = 1,\ldots,N$, and $\forall l = 1,\ldots,M$. Additionally, it implies

$$\operatorname{rank}(\boldsymbol{H}) = \min(N, M)$$

with probability 1 as $\boldsymbol{H}$ will have independent rows or columns. Under such conditions, the receiver will see MN independent copies of s and therefore the probability that at least one of the copies doesn't experience a deep fade increases. Spatial diversity therefore constitutes a third type of diversity besides frequency and time diversity.

In practice, increasing the antenna spacing will decrease correlation and improve diversity. Rich scattering conditions decrease correlation between channel responses as well. A typically used antenna spacing to ensure decorrelation at the receiver terminals is only $\lambda/2$, where λ is the wavelength corresponding to the frequency of operation [27, 28]. This is due to the fact that the receiver is usually close to a multitude of scatterers which ensure a good scattering environment. In contrast, the required antenna spacing to ensure decorrelation at the transmitter side can go up to 10λ, as practical base stations are usually present in elevated positions with very few scatterers close by, if any [28].

The Alamouti scheme [29] is a well-known scheme which combines spatial and time diversity. It can be applied for $M = 2$ transmit antennas and any number of receive antennas. LTE standards support a transmit diversity scheme with two transmit antennas that is based on space-frequency block coding, which is very similar to Alamouti's scheme but operates in the frequency domain instead [8, Section 10.3.2.1]. Diversity schemes can substantially improve the bit error rate (BER) and are valuable in systems where low-rate, reliable communication is desired.

Spatial Multiplexing

The capacity function is logarithmic; thus, in the high SNR regime, increasing the SNR (e.g., increasing the transmit power) brings diminishing returns. For this reason, beamforming can only provide limited gains. Additionally, diversity techniques improve the BER but do not increase capacity, as these techniques do not allow the transmission of multiple symbols in a single time slot. Therefore, beamforming and diversity gains might not be sufficient, especially with the ever increasing required rates in today's communications systems. The natural extension would be therefore to support multiple symbol transmission on the same time and frequency resources, while splitting the transmit power between the different symbols. This is referred to as *spatial multiplexing* in the literature.

Given an accurate knowledge of the channel matrix $\boldsymbol{H}$ at the transmitter and/or receiver, up to $\min(N, M)$ symbols can be simultaneously transmitted on the MIMO link. This channel matrix knowledge is usually referred to *channel state information* (CSI) in the literature. Section 2.4 will discuss in detail the CSI acquisition mechanisms according to the used duplex mode. With available CSI at the transmitter (CSIT), one can perform *precoding*, i.e., a joint transformation of the data symbols prior to transmission over the air link. Precoding is thus a transmit processing operation in the spatial domain. In this thesis, we will focus on linear precoders where the output of each respective filter is a linear transformation of the corresponding input. Linear processing is employed in many practical systems due to simplicity. The performance of linear systems [whether in terms of mean-squared-error (MSE), BER, or sum-rate] is also easier to analyze and characterize in general when compared to non-linear processing.

Assuming $\boldsymbol{F} \in \mathbb{C}^{M \times r}$ is the linear precoder with rank $r \leq \text{rank}(\boldsymbol{H}) \leq \min(N, M)$, $\boldsymbol{x}$ is related to the vector of data symbols $\boldsymbol{s} \in \mathbb{C}^r$ by:

$$\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s} \tag{2.4}$$

and consequently

$$\hat{\boldsymbol{s}} = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s} + \boldsymbol{n} \tag{2.5}$$

[cf. (2.2)]. The capacity achieving precoder with perfect CSIT and CSI at the receiver (CSIR) was originally derived in [4], and is revisited in Appendix B as it will be required in later parts of the thesis. The optimal precoder is linear and has a straightforward interpretation as it aligns the transmitted symbols to the directions containing the maximum power (called the maximum channel eigenmodes). It splits the MIMO channel into up to $\min(N, M)$ parallel non-interfering SISO links. The term $\min(N, M)$ is usually referred to as the multiplexing gain as at high SNR, the capacity of the MIMO link C_{MIMO} can be approximated as (cf. [27])

$$C_{\text{MIMO}} \approx \min(N, M)\, C_{\text{AWGN}} \tag{2.6}$$

when the channel entries are i.i.d. and where C_{AWGN} is given by (2.1). In this case, $r = \min(N, M)$ and spatial multiplexing provides a $\min(N, M)$-fold increase in capacity at high SNR using the same time-frequency resources. In addition, this is obtained without any increase in transmit power.

Note that precoding was previously investigated for *code-division-multiple-access* (CDMA) systems (see, e.g., [30–32]) where it can be applied even if transmitters have single antennas and where the spreading sequences provide the necessary degrees of freedom for precoding.

Space-Division-Multiple-Access

In addition to providing spatial diversity / multiplexing for a single receiver, multiple transmit antennas combined with precoding allow the serving of multiple receivers on the same time-frequency resources. This is known as *space-division-multiple-access* (SDMA). A transmitter having M antennas can transmit up to M symbols to different receivers on the same time-frequency resources, making thus a better use of the available bandwidth. Such a model is usually referred to as the MIMO broadcast channel (BC) in the literature, as shown in Fig. 2.2. In contrast to broadcasting in the context of networking, radio, or TV where receivers acquire the same data, receivers in wireless BCs usually request and receive different data.

In a BC scenario, receivers are decentralized and cannot cooperate to jointly process or decode the obtained symbols, in contrast to a P2P scenario where the receive filter or decoder can be thought to be the result of cooperation among the different receive antennas. Therefore, the capacity of a P2P link with M transmit antennas and N receive antennas is an upper bound to the capacity of a BC scenario with M transmit antennas and a number of receivers whose sum of antennas is N. A formal proof of this statement was given by Sato in [33]. As receivers cannot cooperate, precoding and therefore CSIT knowledge become essential in providing multiplexing gains and ensuring low interference at the receiver terminals, in contrast to the P2P case where receive filtering at the receiver side can provide multiplexing gains even with no CSIT, e.g., through the use of zero-forcing (ZF) filters (cf. [27, 34]). Similarly to the P2P case, the multiplexing gain of the BC case is $\min(N, M)$.

The BC was and still is an ongoing area of research with many challenging theoretical and practical problems. The capacity-achieving precoding technique in the BC with perfect CSIT and single-antenna receivers was derived in [6, 7] and is based on dirty paper coding (DPC) [35], a non-linear strategy relying on user encoding and ordering. The capacity achieving strategy works as follows. The transmitter first chooses a codeword for receiver 1. Then he chooses a codeword for receiver 2 given the knowledge of the codeword for receiver 1 in a way such that receiver 2 does not see interference from user 1. The process is then repeated for the next users. Thus, user k only sees interference from users 1 to $k-1$. This successive interference pre-cancellation at the transmitter parallels successive interference cancellation schemes at the receiver (e.g., [3]).

The user ordering and high complexity of DPC and its higher sensitivity to CSI errors have hindered its implementation in practical systems so far. Therefore, different linear sub-optimal precoding strategies with lower complexity have been instead investigated for the BC. These strategies are well understood by now and a multitude of papers in the literature have dealt with them in different contexts. Much of these popular strategies can be split into the following

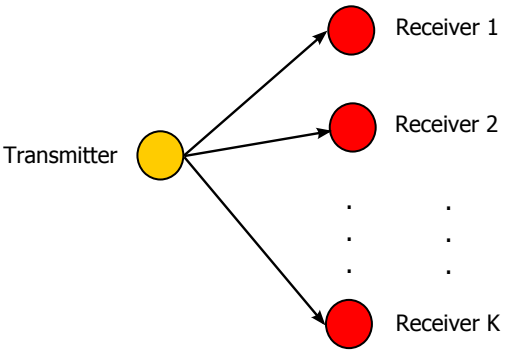

Figure 2.2.: A broadcast channel

approaches:

1. Minimum-mean-square-error (MMSE) /Wiener precoding (see, e.g, [32, 36]),
2. ZF precoding (see, e.g, [30, 36] for single-receive antennas and [37] for the general case),
3. matched filtering (see, e.g, [36, 38]), and
4. signal-to-interference ratio / signal-to-interference-plus-noise ratio (SINR) / signal-to-leakage-and-noise ratio approaches (see, e.g, [39–41]).

Much of these linear precoding techniques are similar to the previously studied receiver equalization techniques [34, 36, 42–45]. The main difference is that transmitters usually have a fixed transmit power which they can not exceed. This translates into an additional constraint in each respective optimization problem. The relatively new arising interest in linear and non-linear precoding was motivated by the development of mechanisms allowing the feedback of CSI information from the receiver to the transmitter in *frequency-division-duplex* (FDD) mode (cf. Section 2.4), the most common duplex mode used in practical wireless systems.

The performance of these strategies highly depends on the SNR level at the receivers. For instance, ZF precoders aim at canceling inter-user interference and have therefore a very good performance in the high SNR regime where inter-user interference is the performance limiting factor. However, they suffer at low SNR levels. Matched filters aim at maximizing the received signal power

and therefore perform very good in the low SNR regime. However, as they do not consider inter-user interference, they have a bad performance at high SNR as the system becomes interference limited. MRT [22] is a very similar precoding technique to matched filtering with the only exception that MRT doesn't take symbol correlations into account, which is the case for the matched filter [36]. ZF precoders and matched filters/MRT precoders have thus opposite performances at low and high SNR levels. MMSE precoders, also called Wiener filters/precoders [46], have a superior performance at all SNR ranges as by minimizing the MSE between the received and transmitted symbol, an implicit tradeoff between noise and inter-user interference is found. Joham et al. showed analytically in [36] using a BER analysis that the matched filter converges to the MMSE precoder at low SNR, and that the ZF precoder converges to the MMSE precoder at high SNR. SINR approaches have been observed to perform very close to MMSE approaches.

It should be nonetheless noted that there exists more advanced ZF schemes that operate close to capacity at all SNR ranges. For instance, authors in [47] propose a greedy ZF method, where at each step a new data stream is allocated to a certain user and the receive filter of that user is optimized in a way to maximize the sum-rate under the ZF constraint. A non-linear precoding scheme consisting of ZF and DPC was proposed in [48], and results in nearly optimum achieved sum-rates over a wide SNR range.

2.2. MIMO Communications in Interference Limited Networks

In many wireless systems such as cellular systems, a transmitter and its served receiver(s) do not constitute an isolated entity. Rather, there is a number of transmitters each serving a certain number of receivers. Furthermore, as current wireless systems are reaching their throughput limits, the whole frequency band is now available for use across the whole system (factor one frequency-reuse). This leads to another effect at the receiver side which was not considered so far: interference coming from other transmitters in the system, caused by transmission on the same time-frequency resources. This uncoordinated interference can severely damage the system performance if it is not properly accounted for. Fig. 2.3 depicts an example of this uncoordinated interference in a two-cell network with two receivers located near the cell-edge of their respective cells. In cellular networks, cell-edge users suffer from the highest level of interference. This problem has triggered research on interference limited networks in the past years and has resulted in a multitude of designs aiming at mitigating inter-cell interference and improving achievable rates near cell-edges.

Fig. 2.3 depicted a simple scenario with two interfering transmitters and receivers. A known theoretical model that captures a more general scenario

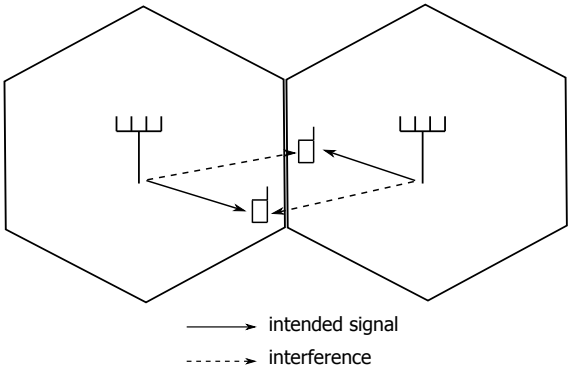

Figure 2.3.: Communications near a cell-edge with factor one frequency-reuse

can be represented by a K-user interference channel (IC), which consists of K transmitter-receiver pairs as shown in Fig. 2.4. Transmitter k transmits a message to its intended receiver k and all other transmitters cause unintended interference to receiver k. In case transmitters and receivers have multiple antennas, the channel is called a K-user MIMO IC. This thesis will deal with the K-user MIMO IC model.

Unlike the P2P MIMO and BC MIMO models which are relatively well understood and whose strategy-achieving capacity is known, much less is known about the IC. Except for the 2-user SISO case where the interference power is stronger than the desired signal and where the capacity in that case is known [49,50], the capacity of the IC is still an open problem in general. The capacity of the 2-user SISO IC has been characterized to within one bit in [51]. Similar to the BC scenario, capacity-approaching techniques in the IC rely on joint and non-linear encoding/decoding over the users. For instance, the results in [51] are based on the Han and Kobayashi scheme [52] which splits the transmitted information into two parts: a private information to be decoded only at the intended receiver and a common information to be decoded at both receivers. By decoding the common information, part of the interference can be canceled off, while the remaining private information from the other user is treated as noise.

Naturally, adding antennas to both sides of the links increases the complexity of the problem of determining the capacity-achieving strategy; therefore, many contributions in the literature dealing with the MIMO IC have focused on linear precoding and receive filtering techniques aiming at mitigating interference while treating remaining interference as noise. This is motivated by two practical facts:

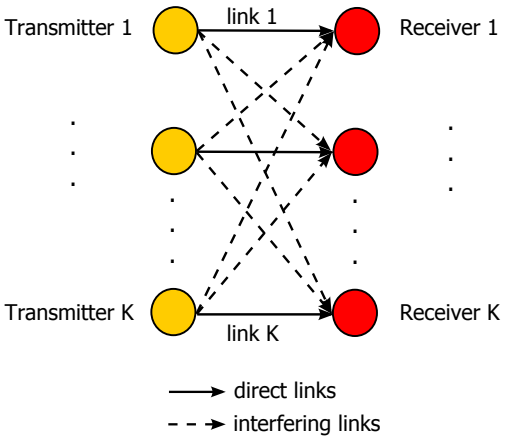

Figure 2.4.: The K-user interference channel

the first one being that receivers are usually equipped with simple decoders; the second one being that receivers do not in general know the coding and modulation schemes employed by interfering transmitters in order to decode the interference first and subtract it from the obtained signal. Many of the MIMO IC precoding techniques are a generalization of the BC linear precoding techniques which were discussed in the previous section. They require not only accurate CSIT but also accurate CSIR due to the receive filter structure. Chapter 3 will discuss in depth many of the different existing precoding schemes for the MIMO IC. Chapter 3 will show yet an additional benefit of MIMO systems; namely, the use of MIMO systems for joint precoding schemes in the context of inter-cell mitigation.

2.3. Massive MIMO Systems

Massive MIMO—also called large-scale MIMO—systems are currently investigated for future wireless systems [17,18]. The main idea behind massive MIMO systems is to equip transmitters with a large number of antennas in order to simplify precoding and receive filtering operations. This is due to the fact that in the limit of an infinite number of antennas serving a finite number of users and under ideal i.i.d. fading conditions, the effects of fast fading vanish and the channels of different users become orthogonal by the law of large numbers (cf. Appendix A.4). Therefore, the most simple precoding types such as MRT

become optimal. The only remaining rate degradation is caused by CSI errors caused by the reuse of pilot sequences from other cells, an effect known as pilot contamination[1]. Mitigating pilot contamination by designing orthogonal pilot sequences across multiple cells would cause the length of the resulting sequences to grow with the number of cells, which makes training unfeasible especially in scenarios with short coherence times. Therefore, the same pilot sequences are reused across multiple cells, leading to pilot contamination. The effects of pilot contamination have been studied in [17, 54, 55], while solutions to mitigate it have been proposed in [56–60]. Real outdoor massive MIMO measurements have been performed in the downlink using large linear and circular transmit arrays in [61], where the antenna elements are separated by only half the wavelength of the frequency of operation. The findings show the promising results that such practical arrays with simple ZF precoding or MRT can achieve up to 90% of the theoretical gains of massive MIMO and operate very close to the capacity achieving DPC scheme in non-line-of-sight conditions with rich scattering and line-of-sight conditions with well separated users.

As an example of how massive MIMO can simplify processing operations, consider the P2P link scenario of Fig. 2.1, where the transmitted symbol vector is $\boldsymbol{s} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{I}_N)$, and $\boldsymbol{H}$ has i.i.d. entries drawn from a $\mathcal{N}_{\mathbb{C}}(0,1)$ distribution. Assume the transmitter has no CSI and transmits with power $\mathrm{E_{tx}}$. In this case, the best option is to transmit equally in all directions (cf. [27]) without precoding. This can be represented by a transmit signal covariance:

$$\mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right] = \frac{\mathrm{E_{tx}}}{M}\,\mathbf{I}_{\boldsymbol{M}}.$$

In addition, the receiver applies no filter. With perfect CSIR, the achievable rate reads

$$R = \log\left|\mathbf{I}_N + \frac{\mathrm{E_{tx}}}{M}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right| \tag{2.7}$$

where without loss of generality, we assumed a noise vector with uncorrelated entries of power 1. Each element of the matrix $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}/M$ can be represented by a random variable of variance $1/M$ [cf. (C.5) and (C.8)]. As $M \to \infty$, the variance of each element goes to 0 and the elements become deterministic. Consequently, we have:

$$\lim_{M\to\infty} \frac{\mathrm{E_{tx}}}{M}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} = \mathrm{E}\left[\frac{\mathrm{E_{tx}}}{M}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right] = \mathrm{E_{tx}}\mathbf{I}_N. \tag{2.8}$$

Therefore,

$$\begin{aligned}\lim_{M\to\infty} R &= \log\left|\mathbf{I}_N + \mathrm{E_{tx}}\mathbf{I}_N\right| \\ &= N\log(1+\mathrm{E_{tx}}).\end{aligned} \tag{2.9}$$

[1] Pilot sequences are used for channel estimation, as discussed in the next section.

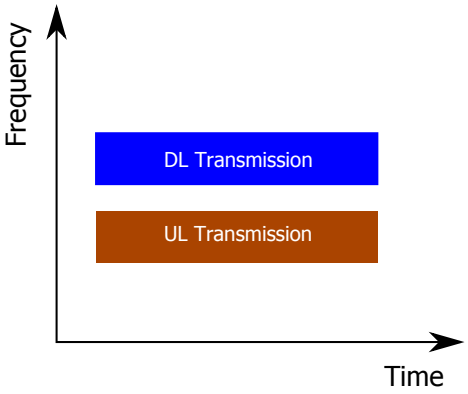

Figure 2.5.: FDD time-frequency structure

This example shows that in the limit of an infinite number of antennas, interference-free communication with full-spatial multiplexing is possible without precoding nor receive filtering. Nevertheless, it should be noted that an accurate CSIR knowledge with arbitrarily large number of transmit antennas induces arbitrarily high training overheads at the receiver, as explained next.

2.4. Duplex Modes and CSI Acquisition Mechanisms

Two commonly used duplex modes are the *frequency-division-duplex* (FDD) mode and the *time-division-duplex* (TDD) mode. In FDD mode, transmissions in the DL and UL are carried out concurrently in time but on different carrier frequencies, as depicted in Fig. 2.5. Therefore, the fading encountered in the DL is different than the one encountered in the UL. Furthermore, only the receiver can determine the DL fading. In practice, the transmitter sends pilot sequences which are known to the receiver, through which the receiver can estimate the fading coefficients. As most precoding techniques require CSIT knowledge, the receiver has to report the CSI to the transmitter in the uplink using a given number of feedback bits after CSI quantization. Many LTE systems use FDD mode.

Fig. 2.6 shows the TDD time-frequency structure. In contrast to FDD, TDD transmissions occur on the same carrier frequency but are separated in time. Therefore, under ideal *channel reciprocity* conditions, the fading experienced in the DL is similar to the one experienced in the UL, and the transmitter can estimate the CSI based on pilot sequences sent by the receiver in the uplink.

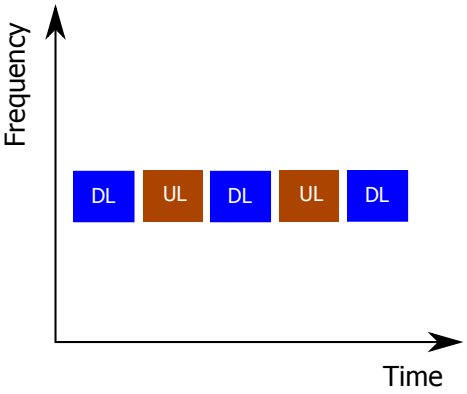

Figure 2.6.: TDD time-frequency structure

In contrast to FDD mode where a pilot symbol is required from each transmit antenna for CSIT estimation, additional transmit antennas are "overhead-free" from the CSIT estimation point of view in TDD mode, as the uplink piloting overhead scales with the number of receive antennas. Note that the time spent on data transmission in the UL and DL directions need not be the same. In the LTE standards, up to seven different configurations—six out of which are asymmetric—can be set up according to the traffic demand [8, Chapter 9.4]. One drawback of TDD is the necessity for a guard period when switching from the DL to the UL or vice versa. This is necessary to ensure that a transmission in e.g., the UL, starts after all DL data is received. This guard period, which is dictated by the fading environment and hardware, cannot be used for data transmission and thus reduces spectral efficiency.

Note: Some of the methods that this thesis deals with necessitate CSI exchange between transmitters taking place over the backhaul link, which incurs additional overheads besides piloting overheads. We mainly focus on the piloting overheads which take place via the air link. This is because such overheads can be analytically characterized as will be performed in Chapter 6, while there exists no specific measure to measure delays/overheads over backhaul links. Nonetheless, one should keep in mind that such delays and overheads limit the gains of such methods in practice.

If the transmitter(s) and/or receiver(s) have perfect or complete knowledge of the CSI matrix coefficients of the required link(s) at each time instant, this is referred to as perfect CSIT and CSIR knowledge. This assumption is not realistic in practice due to channel estimation errors and outdating in both FDD

and TDD systems, in addition to CSI quantization errors inherently present in FDD systems. Even though the performance of many systems is limited due to the inability to acquire accurate CSI, the perfect CSI assumption is necessary in understanding fundamental limits. In some cases, the performance with imperfect CSI can be deduced from the performance with perfect CSI. For example, Jindal showed in [62] for BCs in FDD mode that a number of feedback bits scaling with the number of transmit antennas and the SNR can keep the multiplexing gain constant. In that case, the performance loss due to imperfect CSIT corresponds to a right shift of the rate curve when plotted against the SNR, i.e., a power loss. Hassibi et al. showed in [63] that the CSIR error variance per antenna is inversely proportional to the number of time slots spent on downlink piloting. In TDD mode and under ideal channel reciprocity, this holds for CSIT as well. Therefore, the performance under imperfect CSIT is equivalent to the perfect CSIT performance with an additional noise term.

In case the transmitter has no CSIT, no precoding can be performed. A good option would be applying diversity techniques.

2.5. Figures of Merit

As one of the goals of future wireless systems is increasing user data rates, the achievable rate measure—which in the MIMO case is a generalization of the Shannon capacity formula in (2.1)—will be used as a figure to compare different precoding strategies in this thesis. Even though we previously mentioned that (2.1) is an upper bound, many systems are in fact operating very close to this upper bound. A result in [64] obtains the LTE achievable rates as a simple modification of (2.1). Therefore, the use of the achievable rate measure is well justified.

2.6. Summary

This chapter presented the benefits of employing multiple antennas at both the transmitter side and receiver side. Furthermore, it provided a quick overview of precoding techniques in the P2P MIMO and BC MIMO scenarios, and motivated the use of massive MIMO systems. Advantages of MIMO systems include increased received power, spatial diversity, spatial multiplexing, and space-division-multiple-access. Note that these advantages cannot be cultivated at once. For instance, closely-spaced antennas will provide high beamforming gains but not a good spatial diversity. Additionally, closely-spaced antennas might result in channels with reduced rank which do not support spatial multiplexing. Diversity techniques can substantially improve the BER but do not improve the maximum achievable rates. The converse holds for spatial multiplexing. The ma-

terial of this chapter is by no means comprehensive; rather, it only serves as a motivation for the remaining chapters of this thesis.

Ultimately, the transmission scheme and antenna spacing need to be designed according to the system requirements. In this thesis, we focus on methods aiming at maximizing the sum-rate performance; therefore, we mainly assume there is enough antenna spacing and rich scattering environments to ensure full rank channels and allow spatial multiplexing and/or interference mitigation, as will be elaborated in the next chapter.

3. Linear Precoding Methods in MIMO Interference Channels

In the previous chapter, the advantages of MIMO systems were discussed and a quick overview of precoding methods in MIMO P2P and BC scenarios was presented. In this chapter, many existing linear precoding methods in MIMO ICs are reviewed in depth. The material of this chapter provides the necessary background for later chapters of this thesis.

First, the general thesis system model is presented.

3.1. System Model

3.1.1. The K-User MIMO Interference Channel

Recall Fig. 2.4, where a user denotes a transmitter/receiver pair. Let transmitter k and receiver k have M_k and N_k antennas, respectively. Transmitter k transmits data symbol(s) $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$, $\boldsymbol{s}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{d_k}, \mathbf{I}_{d_k})$, to receiver k, where $1 \leq d_k \leq \min(M_k, N_k)$. These symbol(s) are precoded by a linear precoder $\boldsymbol{F}_k \in \mathbb{C}^{M_k \times d_k}$ before transmission. d_k is referred to as the degrees of freedom achieved by user k, as will be explained in Section 3.1.3. Transmitter k has the transmit power constraint

$$\mathrm{E}\left[||\boldsymbol{F}_k \boldsymbol{s}_k||_2^2\right] = \mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{F}_k\right) = \mathrm{E}_{\mathrm{tx},k}. \tag{3.1}$$

In case $\boldsymbol{F}_k$ has orthogonal columns, then it can be decomposed as

$$\boldsymbol{F}_k = \boldsymbol{F}_{k,\mathrm{ort}} \boldsymbol{P}_k^{\frac{1}{2}} \tag{3.2}$$

where $\boldsymbol{F}_{k,\mathrm{ort}} \in \mathbb{C}^{M_k \times d_k}$ is a matrix with orthonormal columns, i.e., $\boldsymbol{F}_{k,\mathrm{ort}}^{\mathrm{H}} \boldsymbol{F}_{k,\mathrm{ort}} = \mathbf{I}_{d_k}$. $\boldsymbol{P}_k = \mathrm{diag}(p_{k,1}, \ldots, p_{k,d_k})$ contains the power assigned to the different streams on its main diagonal, and satisfies

$$\mathrm{tr}(\boldsymbol{P}_k) = \mathrm{E}_{\mathrm{tx},k} \tag{3.3}$$

which follows directly from (3.1). At receiver k, the obtained signal is perturbed by AWGN $\boldsymbol{n}_k \in \mathbb{C}^{N_k}$, $\boldsymbol{n}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{N_k}, \sigma_n^2 \mathbf{I}_{N_k})$, in addition to unintended interference coming from other transmitters $l \neq k$. Symbols intended for different receivers are uncorrelated and additive noise is uncorrelated with the symbols as well. A linear receive filter $\boldsymbol{G}_k \in \mathbb{C}^{N_k \times d_k}$ is employed to mitigate interference

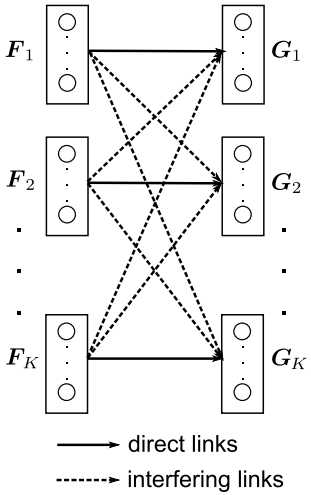

Figure 3.1.: The K-user MIMO interference channel with linear precoders and receive filters

and maximize the achievable rate. Interference is treated as additional noise at the receiver and is not decoded. The received symbol(s) $\hat{\boldsymbol{s}}_k$ at receiver k can be written as:

$$\hat{\boldsymbol{s}}_k = \boldsymbol{G}_k^{\mathrm{H}} \left(\boldsymbol{H}_{kk}\boldsymbol{F}_k\boldsymbol{s}_k + \sum_{l=1,l\neq k}^{K} \boldsymbol{H}_{kl}\boldsymbol{F}_l\boldsymbol{s}_l + \boldsymbol{n}_k \right) \tag{3.4}$$

where $\boldsymbol{H}_{kk} \in \mathbb{C}^{N_k \times M_k}$ denotes the fast-fading channel matrix of the direct link between transmitter k and receiver k while $\boldsymbol{H}_{kl} \in \mathbb{C}^{N_k \times M_l}, l \neq k$, denotes the fast-fading channel matrix of the interfering link between receiver k and transmitter l. The entries of the different channel matrices are i.i.d., drawn from a $\mathcal{N}_{\mathbb{C}}(0,1)$ distribution. Correspondingly,

$$\operatorname{rank}(\boldsymbol{H}_{km}) = \min(N_k, M_m)\ \forall k, m \in \{1, \ldots, K\}$$

with probability 1. A flat-fading model is assumed where the channels remain constant over a given coherence interval. The slow-fading coefficients of all links are normalized to 1. The chosen model represents communication near a cell-edge in a cellular system, which is a worst case scenario. Fig. 3.1 shows the K-user MIMO IC with linear precoders and receive filters.

The achievable rate of user k, R_k, with interference treated as noise is given by

$$R_k = \log \left| \mathbf{I}_{d_k} + \boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k \left(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k \right)^{\mathrm{H}} \times \left(\sum_{l=1, l \neq k}^{K} \boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \left(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \right)^{\mathrm{H}} + \sigma_n^2 \boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{G}_k \right)^{-1} \right|. \tag{3.5}$$

From the above formula it is seen that any attempt to directly maximize the achievable rates of the different users would be extremely tedious, if not intractable. This is due to the fact that the achievable rate of each user is a function of all the precoders in addition to the receive filter of that user which makes the different variables coupled, not to mention the $\log\det(\bullet)$ nature of the rate function which makes it hard to handle. Rather, the different approaches for MIMO ICs have employed proxy optimization functions which are easier to handle and optimize, as will be elaborated next. One of the cases where the rate function is directly optimized is in [65], where in the context of the two-user MISO IC a lower bound on the achievable sum-rate is maximized. However, the scheme in [65] cannot be easily generalized to more users or to the MIMO case.

Some configurations are special in the sense that transmitters and/or receivers have the same number of antennas. This gives rise to the following definitions.

Definition 3.1. *Symmetric and asymmetric channels:*
If $M_k = M, N_k = N, d_k = d,\ \forall k = 1, \ldots, K$, the channel is called symmetric and is denoted by $(M, N, d)^K$. Otherwise, it is called asymmetric.

Definition 3.2. *Square symmetric channels:*
If a symmetric channel satisfies $M = N$, it is called a square symmetric channel.

3.1.2. Number of Antennas and Precoding

The type of precoding used at the transmitters and/or receive filtering at the receivers highly depends on the spatial degrees of freedom given by the number of antennas at both ends of the communication links. There are different cases that can be distinguished:

1. No simultaneous transmission possible,
2. Simultaneous transmission without full-spatial multiplexing possible, and
3. Simultaneous transmission with full-spatial multiplexing possible.

Each case can be characterized by an inequality governing the set of parameters in the MIMO IC, i.e., the set $\{M_k, N_k, d_k\}_{k=1}^{K} \cup \{K\}$. We will mainly focus on the second case as it constitutes the most challenging case, though we discuss the other cases as well.

3.1.3. Degrees of Freedom

Definition 3.3. *If the sum-capacity of a given system can be expressed as*

$$C(\mathrm{SNR}) \approx d_{\mathrm{f}} \log(\mathrm{SNR}) + \mathrm{o}(\log(\mathrm{SNR})), \tag{3.6}$$

then d_{f} is called the degrees of freedom (DoF) of the system [66, 67]. Recalling (2.1)*, we have:*

$$d_{\mathrm{f}} = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log(\mathrm{SNR})} = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{C_{\mathrm{AWGN}}}, \tag{3.7}$$

which is the slope of the capacity curve at high SNR as shown in Fig. 3.2.

The DoF provides a capacity approximation that is accurate within $\mathrm{o}(\log(\mathrm{SNR}))$, and is therefore an important metric for characterizing achievable rates in the high SNR regime.

Example 3.1. *A P2P MIMO link with capacity achieving precoders achieves* $\min(N, M)$ *DoF, where N and M are the number of receive and transmit antennas, respectively.*

For P2P MIMO links, the DoF is equivalent to the multiplexing gain. However, the DoF is a more general term that covers ICs and BCs as well (whether MIMO, SISO etc.) as will be explained in the next section.

Having introduced the system model and necessary definitions, we start the survey of linear precoding methods in MIMO ICs with interference alignment methods.

3.2. Interference Alignment Methods

3.2.1. Concept and Conditions of Interference Alignment

Consider a K-user SISO IC where the interference power is comparable to the desired signal power. Such a setup was thought to be interference-limited for a long time, and the best solution to avoid interference was thought to be orthogonal access schemes, i.e., time-division-multiple-access (TDMA) or frequency-division-multiple-access (FDMA) schemes. Under such schemes and if resources are equally split among users, each user would get $1/K$ DoF. Cadambe et al.

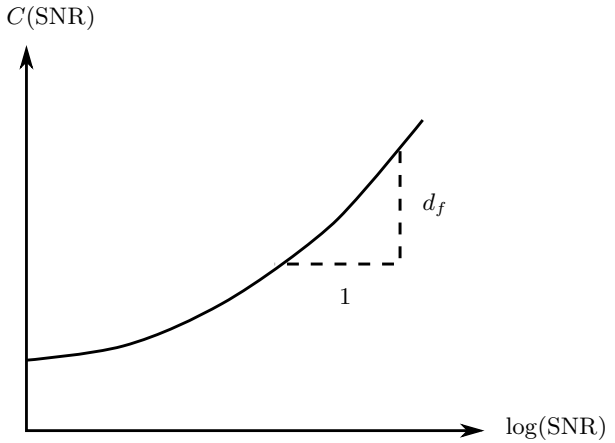

Figure 3.2.: DoF illustration

considered the time-varying K-user SISO IC in [10] and proved using the time-varying structure that in fact, each user can achieve $1/2$ degrees of freedom, i.e., the sum-capacity of the—originally thought interference-limited—SISO IC at high SNR was shown to be

$$C(\text{SNR}) \approx \frac{K}{2} \log(\text{SNR}) + \text{o}(\log(\text{SNR})) \tag{3.8}$$

which is half of the interference-free capacity. The achievable scheme is based on the idea of interference alignment (IA). The scheme jointly designs the transmit filters over multiple symbol slots such that interference occupies only a subset of the total time slots at each receiver. This leaves an interference-free number of slots which is used for desired transmission. By using longer symbol extensions, the performance of the proposed scheme gets arbitrarily close to the sum-capacity in (3.8), and each user gets arbitrarily close to achieving $1/2$ DoF. That was the concept of IA which was introduced for the SISO IC. Note that similar ideas were introduced for the MIMO X channel around the same time (see, e.g., [68]).

We now switch back to the main MIMO IC model and introduce IA in the spatial domain. The main idea is to jointly design the precoders such that interference only occupies, i.e., is aligned in, an $N_k - d_k$ dimensional subspace at receiver k, so that a ZF receive filter can subsequently null this interference.

This leaves a d_k dimensional interference-free subspace that is used for data transmission. By this choice of parameters, a receive filter of size $N_k \times d_k$ can be always found to null interference spanning an $N_k - d_k$ dimensional subspace.

The interference alignment and suppression conditions at the receivers read

$$\begin{aligned} \boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l &= \boldsymbol{0}_{d_k \times d_l} \\ \forall k \in \{1, \ldots, K\}, \forall l \in \{1, \ldots, K\} \backslash \{k\}, & \end{aligned} \tag{3.9}$$

while the full rank condition

$$\operatorname{rank}(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k) = d_k \tag{3.10}$$

ensures that d_k symbols can be transmitted between the kth transmitter/receiver pair. $\boldsymbol{G}_k$ and $\boldsymbol{F}_k$ are chosen independently of the direct channel matrix $\boldsymbol{H}_{kk}$ as will be elaborated later. Correspondingly, (3.10) is automatically satisfied as $\operatorname{rank}(\boldsymbol{H}_{kk}) = \min(N_k, M_k)$ with probability 1. An explicit proof for this statement is given in Section 4.3.1, which proves that $\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k$ has zero-mean i.i.d. entries of the same variance when the precoders and receive filters have orthogonal columns. In that case, $\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k$ has full rank with probability 1.

3.2.2. Closed-Form Solutions

We next present a scenario where closed-form precoder and receive filter solutions can be found to satisfy (3.9) for the 3-user square symmetric IC. The solution was presented in [10, Appendix IV]. For the rest of this section, we denote by N the number of antennas at each transmitter or receiver. The scheme allows each user to exactly achieve $d = N/2$ DoF when N is even. The scheme is a simpler version of the scheme presented for the SISO case. The idea is to explicitly write the IA conditions at receivers 1 to 3 as follows:

$$\operatorname{span}(\boldsymbol{H}_{12} \boldsymbol{F}_2) = \operatorname{span}(\boldsymbol{H}_{13} \boldsymbol{F}_3) \tag{3.11}$$

$$\operatorname{span}(\boldsymbol{H}_{21} \boldsymbol{F}_1) = \operatorname{span}(\boldsymbol{H}_{23} \boldsymbol{F}_3) \tag{3.12}$$

$$\operatorname{span}(\boldsymbol{H}_{31} \boldsymbol{F}_1) = \operatorname{span}(\boldsymbol{H}_{32} \boldsymbol{F}_2). \tag{3.13}$$

The conditions in (3.11)–(3.13) imply that each receiver dedicates half of its receive space to receive the overlapping interference (forced to span an $N/2$ dimensional subspace), leaving the other half interference-free. These conditions show that the IA precoder/receive filter solution set is not unique because only the span of the respective matrices need to be the same. For instance, the set

of solutions is invariant to unitary rotations. To find a possible solution, the conditions in (3.12) and (3.13) are tightened to

$$\begin{aligned} \boldsymbol{H}_{21}\boldsymbol{F}_1 &= \boldsymbol{H}_{23}\boldsymbol{F}_3 \\ \boldsymbol{H}_{31}\boldsymbol{F}_1 &= \boldsymbol{H}_{32}\boldsymbol{F}_2, \end{aligned}$$

and correspondingly

$$\begin{aligned} \boldsymbol{F}_3 &= \boldsymbol{H}_{23}^{-1}\boldsymbol{H}_{21}\boldsymbol{F}_1. \\ \boldsymbol{F}_2 &= \boldsymbol{H}_{32}^{-1}\boldsymbol{H}_{31}\boldsymbol{F}_1. \end{aligned} \tag{3.14}$$

Plugging (3.14) into (3.11) yields

$$\operatorname{span}(\boldsymbol{F}_1) = \operatorname{span}(\boldsymbol{E}\boldsymbol{F}_1) \tag{3.15}$$

where

$$\boldsymbol{E} = \boldsymbol{H}_{31}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{12}^{-1}\boldsymbol{H}_{13}\boldsymbol{H}_{23}^{-1}\boldsymbol{H}_{21}. \tag{3.16}$$

Choosing the columns of $\boldsymbol{F}_1$ as the dominant d eigenvectors of $\boldsymbol{E}$ satisfies (3.15). The chosen $\boldsymbol{F}_1$ is plugged into (3.14) to obtain $\boldsymbol{F}_2$ and $\boldsymbol{F}_3$.

After finding the precoders, they can be adjusted to satisfy the transmit power constraint without violating the alignment conditions. Choosing $\boldsymbol{G}_k \in (\boldsymbol{F}_l^{\mathrm{H}}\boldsymbol{H}_{kl}^{\mathrm{H}})^{\perp}$ for any $l \neq k$ provides a complete solution to the alignment problem that satisfies the conditions in (3.9). Fig. 3.3 depicts the conditions in (3.11)–(3.13) for the case $N = 2$. The graphs on the right correspond to the two-dimensional receive spaces of the different receivers. Here, $d = 1$ and the precoding design forces interference to lie in a one-dimensional subspace, i.e, on a line.

Besides the above scenario, there exists no closed-form IA solutions of the precoders and receive filters except for the method in [69], where the authors present a closed-form solution of the square symmetric IC in the case where

$$\begin{aligned} K &= N + 1 \\ d &= 1. \end{aligned} \tag{3.17}$$

The details are skipped as the method is heuristic and doesn't reveal any additional insights beyond the one already discussed above.

CSI requirements and implementation: There exists different ways in which the precoders can be calculated among transmitters. We list some of them here.

In a distributed implementation, transmitter 1 calculates $\boldsymbol{F}_1$ based on the CSI of *all* interfering links [cf. (3.15) and (3.16)], signals it to transmitters 2 and 3 who can then calculate $\boldsymbol{F}_2$ and $\boldsymbol{F}_3$ based on (3.14). This would then necessitate the CSI of interfering links seen by receiver 2 (resp. receiver 3) to be available at transmitter 3 to calculate $\boldsymbol{F}_3$ (resp. at transmitter 2 to calculate $\boldsymbol{F}_2$). In a centralized implementation, transmitter 1 could calculate all precoders (again

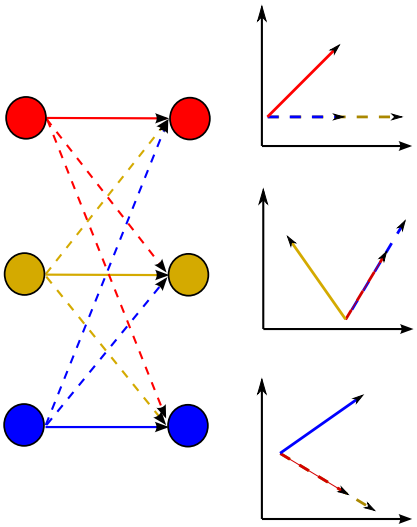

Figure 3.3.: Depicting the conditions in (3.11)–(3.13): the $(2, 2, 1)^3$ channel

based on the CSI of all interfering links) and signal $\boldsymbol{F}_2$ and $\boldsymbol{F}_3$ to transmitters 2 and 3, respectively. In both cases, the required CSI needs to be acquired and exchanged between transmitters. An alternative centralized implementation consists of a central controller who acquires the needed CSI from the respective transmitters, calculates all precoders, and then signals each precoder to the corresponding transmitter. Receiver k calculates $\boldsymbol{G}_k$ based on $\boldsymbol{H}_{kl}\boldsymbol{F}_l$ for any $l \neq k$, which needs to be estimated in the downlink. Chapter 6 will discuss in depth CSI training overheads of IA in both the downlink and uplink.

3.2.3. Iterative Solutions

The difficulty of obtaining IA solutions in MIMO ICs with arbitrary configurations triggered research on iterative IA algorithms. The first proposed iterative IA algorithm is the interference leakage (IL) algorithm [11], which aims at min-

imizing the interference power present at the different receivers. Defining

$$\begin{aligned} L_k &= \sum_{l \neq k} \mathrm{E}\left[\left|\left| \boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \boldsymbol{s}_l \right|\right|_2^2 \right] \\ &= \sum_{l \neq k} \mathrm{tr}\left(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{H}_{kl}^{\mathrm{H}} \boldsymbol{G}_k \right) \end{aligned} \tag{3.18}$$

to be at the interference power at receiver k, then the IL algorithm can be formulated as:

$$\min_{\substack{(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_K) \\ (\boldsymbol{G}_1,\ldots,\boldsymbol{G}_K)}} \sum_{k=1}^{K} L_k \quad \text{s.t. } \mathrm{rank}(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k) = d_k \;\text{ and }\; \mathrm{tr}(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{F}_k) = \mathrm{E}_{\mathrm{tx},k} \;\; \forall k. \tag{3.19}$$

The insight behind this approach is that any solution set yielding $\sum_{k=1}^{K} L_k = 0$ means IA is automatically achieved without having to explicitly solve for the IA conditions such as in (3.11)–(3.13) for an arbitrary number of users or antennas. Such a problem is non-convex in the $2K$ matrix variables; additionally, the optimum solution set cannot be found in closed-form. However, the problem can be split and solved iteratively using alternating minimization, as explained next.

With all other variables fixed, the kth receive filter $\boldsymbol{G}_k$ is the solution of the problem

$$\min_{\boldsymbol{G}_k} L_k \quad \text{s.t. } \mathrm{rank}(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k) = d_k \tag{3.20}$$

which can be directly deduced from (3.19). Enforcing $\boldsymbol{G}_k$ to have orthonormal columns, the problem in (3.20) becomes a standard trace minimization problem. Defining

$$\boldsymbol{Q}_k = \sum_{l=1, l \neq k}^{K} \boldsymbol{H}_{kl} \boldsymbol{F}_l \boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{H}_{kl}^{\mathrm{H}} \tag{3.21}$$

to be the interference covariance matrix at receiver k, then $\boldsymbol{G}_k$ is given by

$$\boldsymbol{G}_k = \boldsymbol{V}_k^{N_k - d_k + 1:N_k} \tag{3.22}$$

where

$$\boldsymbol{Q}_k = \boldsymbol{V}_k \boldsymbol{\Phi}_k \boldsymbol{V}_k^{\mathrm{H}} \tag{3.23}$$

is the eigenvalue decomposition (EVD) of $\boldsymbol{Q}_k$ (cf. Appendix A.2). In other words, the columns of $\boldsymbol{G}_k$ are the eigenvectors corresponding to the d_k smallest eigenvalues of $\boldsymbol{Q}_k$. This solution depends on the set of variables $\{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_{k-1}, \boldsymbol{F}_{k+1}, \ldots, \boldsymbol{F}_K\}$ through $\boldsymbol{Q}_k$.

In a similar fashion, the kth precoder is the solution of the problem

$$\min_{\boldsymbol{F}_k} \sum_{m\neq k} \mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{H}_{mk} \boldsymbol{G}_m \boldsymbol{G}_m^{\mathrm{H}} \boldsymbol{H}_{mk}^{\mathrm{H}} \boldsymbol{F}_k\right) \quad \text{s.t. } \mathrm{rank}(\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k) = d_k \quad \text{and} \quad \mathrm{tr}(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{F}_k) = \mathrm{E}_{\mathrm{tx},k} \tag{3.24}$$

which also follows directly from (3.19) due to the property $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate sizes. Enforcing $\boldsymbol{F}_k$ to have orthogonal columns and with equal stream power allocation, it is written as

$$\boldsymbol{F}_k = \sqrt{\frac{\mathrm{E}_{\mathrm{tx},k}}{d_k}} \boldsymbol{F}_{k,\mathrm{ort}}. \tag{3.25}$$

Consequently, the problem in (3.24) becomes a standard trace minimization problem as well, whose solution is given by

$$\boldsymbol{F}_{k,\mathrm{ort}} = \boldsymbol{X}_k^{N_k - d_k + 1:N_k} \tag{3.26}$$

with the EVD

$$\sum_{m\neq k} \frac{\mathrm{E}_{\mathrm{tx},m}}{d_m} \boldsymbol{H}_{mk} \boldsymbol{G}_m \boldsymbol{G}_m^{\mathrm{H}} \boldsymbol{H}_{mk}^{\mathrm{H}} = \boldsymbol{X}_k \hat{\boldsymbol{\Phi}}_k \boldsymbol{X}_k^{\mathrm{H}}. \tag{3.27}$$

The solution for $\boldsymbol{F}_k$ depends on the set of variables $\{\boldsymbol{G}_1, \ldots, \boldsymbol{G}_{k-1}, \boldsymbol{G}_{k+1}, \ldots, \boldsymbol{G}_K\}$.

The coupling of the different variables implies that the optimal precoders and receive filters cannot be found in closed-form. To find a solution of the initial problem in (3.19), an iterative alternating minimization procedure is used, whereby at each iteration $2K-1$ matrix variables are kept fixed and the remaining variable is solved in terms of the fixed variables. Arbitrary initial orthogonal precoders can be chosen as a starting point for the algorithm. Algorithm 3.1 summarizes the IL procedure. The IL is shown to analytically converge to a local optimum [11]. In Algorithm 3.1 and other algorithms to be later presented, convergence is reached when the global objective function [e.g., the objective in (3.19) for the IL procedure] becomes smaller than a certain defined threshold, e.g., 10^{-3} or 10^{-6}.

Other iterative IA algorithms include algorithms in [70–72]. Such algorithms rely on different objective functions to achieve IA. For instance, the algorithm in [71] takes the receive filters out of the optimization problem, and is based on the concept of *chordal distance* between subspaces, which is conveniently defined via orthogonal projectors. We review the algorithm in [71] next.

Algorithm 3.1 The Interference Leakage Algorithm

Require: Initial $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_k$ with orthogonal columns.
  repeat
    Receive filter optimization:
    for $k = 1$ to K **do**
      Obtain EVD in (3.23).
      Find $\boldsymbol{G}_k$ using (3.22).
    end for
    Precoder optimization:
    for $k = 1$ to K **do**
      Obtain EVD in (3.27).
      Find $\boldsymbol{F}_k$ using (3.26) and (3.25).
    end for
  until convergence

We start by defining orthogonal projectors. Let $\mathcal{S} \subseteq \mathbb{C}^N$ be a subspace. $\boldsymbol{P} \in \mathbb{C}^{N \times N}$ is called the orthogonal projector onto $\mathcal{S}$ if is satisfies the following properties:

$$\begin{aligned} \text{span}(\boldsymbol{P}) &= \mathcal{S}, \\ \boldsymbol{P}^{\text{H}} &= \boldsymbol{P}, \text{ and} \\ \boldsymbol{P}^2 &= \boldsymbol{P}. \end{aligned} \tag{3.28}$$

Thus, by definition, if $\boldsymbol{x} \in \mathbb{C}^N$, then $\boldsymbol{P}\boldsymbol{x} \in \mathcal{S}$ and $(\mathbf{I}_N - \boldsymbol{P})\boldsymbol{x} \in \mathcal{S}^{\perp}$. Orthogonal projectors can be used to define a chordal distance measure between subspaces. The chordal distance between two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ of the same dimensionis defined as (see, e.g., [73])

$$d(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{\sqrt{2}} \, ||\boldsymbol{P}_1 - \boldsymbol{P}_2||_{\text{F}} \tag{3.29}$$

where $\boldsymbol{P}_i$ is the orthogonal projector onto $\mathcal{S}_i$. A distance of 0 between two subspaces means that these subspaces are aligned.

Using the chordal distance between subspaces, the alignment problem can be formulated in terms of orthogonal projectors as follows. Let $\boldsymbol{P}_{kl}$ be the orthogonal projector onto the column space of $\boldsymbol{H}_{kl}\boldsymbol{F}_l$, $\forall l$ and $k \neq l$. $\boldsymbol{P}_{kl}$ uniquely defines a receive interference subspace between receiver k and transmitter l and can be written as (see, e.g., [74])

$$\boldsymbol{P}_{kl} = \boldsymbol{H}_{kl}\boldsymbol{F}_l \left(\boldsymbol{F}_l^{\text{H}} \boldsymbol{H}_{kl}^{\text{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \right)^{-1} \boldsymbol{F}_l^{\text{H}} \boldsymbol{H}_{kl}^{\text{H}} \in \mathbb{C}^{N_k \times N_k}. \tag{3.30}$$

It can be checked that $\boldsymbol{P}_{kl}$ satisfies the properties in (3.28). Defining $\alpha(\boldsymbol{F}_1, ..., \boldsymbol{F}_K)$ to be the sum of squared distances between interfering subspaces

over all receivers:

$$\alpha(\boldsymbol{F}_1, ..., \boldsymbol{F}_K) \triangleq \sum_{k=1}^{K} \sum_{l=1, l\neq k}^{K} \sum_{m=1, m>l, m\neq k}^{K} \frac{1}{2} \, ||\boldsymbol{P}_{kl} - \boldsymbol{P}_{km}||_{\mathrm{F}}^2 \,, \tag{3.31}$$

the IA problem in [71] was formulated as the problem of finding the precoders that minimize $\alpha(\boldsymbol{F}_1, ..., \boldsymbol{F}_K)$ subject to the transmit power constraint:

$$\min_{(\boldsymbol{F}_1, ..., \boldsymbol{F}_K)} \alpha(\boldsymbol{F}_1, ..., \boldsymbol{F}_K) \text{ s.t. } \mathrm{tr}\left(\boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{F}_l\right) = \mathrm{E}_{\mathrm{tx},l} \; \forall l. \tag{3.32}$$

Similar to the IL problem formulation, such a problem is non-convex in the precoder variables [due to the inverse term in (3.30)], doesn't yield closed-form solutions, and alternating optimization is needed to solve it. For fixed precoders $\boldsymbol{F}_m \forall m \neq l$, the lth precoder is the solution of the following problem

$$\min_{\boldsymbol{F}_l} \underbrace{\sum_{k\neq l} \sum_{m>l, m\neq k} ||\boldsymbol{P}_{kl} - \boldsymbol{P}_{km}||_{\mathrm{F}}^2}_{\alpha_l(\boldsymbol{F}_l)} . \tag{3.33}$$

$\alpha_l(\boldsymbol{F}_l)$ is expanded as[1]

$$\alpha_l(\boldsymbol{F}_l) = \sum_{k\neq l} \sum_{m>l, m\neq k} \left((d_l + d_m) - 2\,\mathrm{tr}\left[\boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{H}_{kl}^{\mathrm{H}} \boldsymbol{P}_{km} \boldsymbol{H}_{kl} \boldsymbol{F}_l \left(\boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{H}_{kl}^{\mathrm{H}} \boldsymbol{H}_{kl} \boldsymbol{F}_l \right)^{-1} \right] \right) . \tag{3.34}$$

This is a sum of generalized Rayleigh quotients, whose minimizer does not have a closed-form solution. However, $\alpha_l(\boldsymbol{F}_l)$ satisfies

$$\alpha_l(\boldsymbol{F}_l \boldsymbol{Q}) = \alpha_l(\boldsymbol{F}_l) \tag{3.35}$$

for any invertible or unitary $\boldsymbol{Q} \in \mathbb{C}^{d_l \times d_l}$. The invariance of $\alpha_l(\boldsymbol{F}_l)$ to unitary rotations means that the optimal precoder solution only depends on the subspace in which it lies and not on the precoder itself (as also observed in Section 3.2.2). In the context of iterative algorithms, this property is very useful because it leads to a reduction of the dimension of the optimization problem; namely, the search space is reduced to a search over subspaces. Thus, the local objective given by (3.34) can be minimized on the *complex Grassmannian manifold* $\mathrm{Gr}(M_l, d_l)$, defined next.

Definition 3.4. *Complex Grassmannian Manifold:*
The complex Grassmannian manifold $\mathrm{Gr}(M, d)$ *of* $\mathbb{C}^M$, $d < M$, *is the set of all* d*-dimensional complex subspaces of* $\mathbb{C}^M$.

[1] Note that $||\boldsymbol{X}||_2^2 = \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}})$.

The solution to (3.33) can be found using an iterative steepest descent method on the Grassmannian manifold that is especially tailored for unitary matrices [75]. Thus, the initial problem in (3.32) can be solved for unitary precoders, i.e., $\boldsymbol{F}_l^{\mathrm{H}}\boldsymbol{F}_l = \mathbf{I}_{d_l}$ where each stream is assigned unity power. The found solutions can be then scaled to satisfy the power constraint. This scaling doesn't change the value of $\alpha_l(\boldsymbol{F}_l)$, again due to (3.35). The method requires the evaluation of $\alpha_l(\boldsymbol{F}_l)$ and its gradient with respect to (w.r.t.) $\boldsymbol{F}_l$ at each iteration. The latter is given by

$$\nabla_{\boldsymbol{F}_l}(\alpha_l) = 2\frac{\partial \alpha_l}{\partial \boldsymbol{F}_l^*} = 4\sum_{k\neq l}\left[\boldsymbol{H}_{kl}^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\left(\boldsymbol{F}_l^{\mathrm{H}}\boldsymbol{H}_{kl}^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\right)^{-1}\boldsymbol{F}_l^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{F}_l - \boldsymbol{A}_k\boldsymbol{F}_l\right]\left(\boldsymbol{F}_l^{\mathrm{H}}\boldsymbol{H}_{kl}^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\right)^{-1} \tag{3.36}$$

where $\partial\alpha_l/\partial\boldsymbol{F}_l^*$ denotes the partial derivative of α_l w.r.t. $\boldsymbol{F}_l^*$ which is calculated in [71, Appendix], and

$$\boldsymbol{A}_k \triangleq \boldsymbol{H}_{kl}^{\mathrm{H}}\left(\sum_{m\notin\{l,k\}}\boldsymbol{P}_{km}\right)\boldsymbol{H}_{kl}. \tag{3.37}$$

The resulting procedure is shown in Algorithm 3.2. Steps 1 to 7 consist of the steepest descent method on the Grassmannian manifold needed to calculate a given precoder. Step 1 is similar to the first step of the steepest descent algorithm for scalar or vector variables. In step 2, the gradient is projected on the tangent space of the Grassmannian manifold to find the descent direction on the manifold $\boldsymbol{Z}$. Step 3 checks whether the steepest direction is small enough, and if so, convergence is reached. Steps 4 and 5 consist of Armijo's rule for choosing a proper step size and updating the precoder value [76]. Steps 4 and 5 result in a precoder update that lies outside $\mathrm{Gr}(M_l, d_l)$. Therefore, Steps 6 and 7 project the obtained precoder back onto $\mathrm{Gr}(M_l, d_l)$. Step 8 updates the auxiliary variables required for the following precoder calculations.

After finding the precoder solutions, they can be adjusted to satisfy the transmit power constraint. Provided the global solution is found, choosing $\boldsymbol{G}_k \in (\boldsymbol{H}_{kl}\boldsymbol{F}_l)^{\perp}$ for any $l \neq k$ provides a complete solution to the alignment problem. This algorithm was shown to analytically converge to a local minimum in [77] and numerically to converge to global optimums for different system parameters.

Note: A minor imprecision exists in the presented algorithm version of [71]. Namely, in steps 1 and 2 the gradient w.r.t. $\boldsymbol{F}_l$ is replaced with the partial derivative w.r.t. $\boldsymbol{F}_l^*$, i.e., the multiplication by 2 is missing. This is due to the formulation in [75] which treats $\nabla_{\boldsymbol{F}_l}(\alpha_l)$ and $\partial\alpha_l/\partial\boldsymbol{F}_l^*$ as equivalent. Nonetheless, this doesn't affect the correctness of the algorithm as it only corresponds to

Algorithm 3.2 Projector Based Interference Alignment via Alternating Minimization

Require: Initial $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_K$ with orthogonal columns; initial step size $\gamma = 1$.

repeat

for $l = 1$ to K **do**

Steepest descent method on the Grassmannian manifold:

1) Calculate $\nabla_{\boldsymbol{F}_l}(\alpha_l)$ [cf. (3.36) and [71]].

2) Compute the steepest direction $\boldsymbol{Z} = -(\mathbf{I}_M - \boldsymbol{F}_l \boldsymbol{F}_l^{\mathrm{H}}) \nabla_{\boldsymbol{F}_l}(\alpha_l)$.

3) Calculate $||\boldsymbol{Z}||_{\mathrm{F}}^2 = \mathrm{tr}(\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z})$. If it is sufficiently small, go to step 8.

4) If $\alpha_l(\boldsymbol{F}_l) - \alpha_l(\boldsymbol{F}_l + 2\gamma\boldsymbol{Z}) \geq \gamma\,\mathrm{tr}(\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z})$ set $\gamma \leftarrow 2\gamma$ and repeat step 4.

5) If $\alpha_l(\boldsymbol{F}_l) - \alpha_l(\boldsymbol{F}_l + \gamma\boldsymbol{Z}) < 0.5\gamma\,\mathrm{tr}(\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{Z})$ set $\gamma \leftarrow 0.5\gamma$ and repeat step 5.

6) Perform the QR decomposition of $\boldsymbol{F}_l + \gamma\boldsymbol{Z}$:
$\boldsymbol{F}_l + \gamma\boldsymbol{Z} = \boldsymbol{Q}_l \boldsymbol{R}_l$.

7) $\boldsymbol{F}_l = \boldsymbol{Q}_l^{1:d_l}$. Go back to step 1.

Update auxiliary variables:

8) Update projectors $\boldsymbol{P}_{kl}\ \forall k \neq l$ according to (3.30).

end for

until convergence

a real valued scaling operation of $\boldsymbol{Z}$ that is afterwards automatically corrected in steps 4 and 5 while choosing the step size.

In [72], an IA design based on precoders only is also presented, where the precoders are chosen to minimize the sum of the smallest eigenvalues of interference covariances matrices over all receivers, i.e., the IA problem is formulated as:

$$\min_{(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_K)} \sum_{k=1}^{K} \sum_{m=N_k-d_k+1}^{N_k} \lambda_m(\boldsymbol{Q}_k) \quad \text{s.t. } \boldsymbol{F}_l^{\mathrm{H}} \boldsymbol{F}_l = \mathbf{I}_{d_l}\ \forall l \tag{3.38}$$

where $\boldsymbol{Q}_k$ is given by (3.21). The formulation is based on matrix differentials and each precoder optimization step requires an iterative steepest descent method on the *Stiefel* Manifold [75], defined as

$$\mathrm{St}(M, d) = \{\boldsymbol{X} \in \mathbb{C}^{M \times d} : \boldsymbol{X}^{\mathrm{H}}\boldsymbol{X} = \boldsymbol{I}_d\}. \tag{3.39}$$

Note that a search over a Stiefel manifold necessitates a search over individual unitary matrices and is therefore more costly than a search over a Grassmannian manifold.

Unfortunately, the initial wish in [71,72] to simplify the iterative IA problem by removing the receive filters from the optimization problem resulted in more

complex and computationally expensive precoder solutions due to the iterative steepest descent needed to obtain each precoder, as compared to the IL algorithm where the precoders are obtained in closed-form given other variables are fixed (cf. Algorithm 3.1).

CSI requirements and implementation: There exists different ways in which the iterative IA algorithms can be implemented. We list some of them next. For all of the listed implementations, the needed CSI needs to be acquired and possibly exchanged between the concerned entities.

The IL algorithm or Algorithm 3.2 can be implemented in a centralized or distributed manner. In any case, a centralized implementation requires the processing entity to possess the CSI of all interfering links.

For the IL algorithm, there exists a decentralized implementation that includes receivers in the iterative process and works in a "ping-pong" manner as follows. Starting with initial random precoders, this implementation alternates between the following two points until convergence.

1. $\forall k$ and $l \neq k$, receiver k calculates $\boldsymbol{G}_k$ and signals it to transmitter l.
2. $\forall k$ and $l \neq k$, transmitter k calculates $\boldsymbol{F}_k$ and signals it to receiver l.

This method necessitates more signaling over the airlink than conventional methods or other IA implementations.

Algorithm 3.2 can be implemented in a distributed manner as outlined in [78]. Namely, at a given iteration, transmitter l calculates precoder $\boldsymbol{F}_l$ and updates the auxiliary variables $\boldsymbol{P}_{kl}\,\forall k \neq l$. He then signals these updated variables to other transmitters which, given this knowledge, can then update their precoder values in next iterations. Decentralized implementations necessitate the CSI of interfering links $\boldsymbol{H}_{kl}\,\forall k \neq l$ to be present at transmitter l, as seen in (3.26) or (3.33).

3.2.4. Feasibility of Interference Alignment

The initially proposed closed-form and iterative IA algorithms described in the previous section could not analytically show whether the IA conditions in (3.9) would be satisfied for a given choice of the parameters $\{d_1, \ldots, d_K\}$. Rather, this could only be checked numerically. Naturally, increasing the number of streams beyond a certain value would cause the system to become interference-limited again. This fact triggered research on the *feasibility conditions* of IA. One of the earliest works on this topic is by Yetis et al. in [79], where the following necessary condition to achieve IA was introduced:

$$\sum_{k=1}^{K} d_k(M_k + N_k - 2d_k) \geq \sum_{k=1}^{K} \sum_{l=1, l\neq k}^{K} d_k d_l, \tag{3.40}$$

which simplifies to

$$M + N \geq (K + 1)\, d \tag{3.41}$$

in the symmetric channel case. This condition was shown to be necessary but not sufficient for IA to hold in general. However, for the special case $d_k = 1\,\forall k$, the authors in [79] showed it is a sufficient condition. In that case and for i.i.d. channel entries, the conditions in (3.9) correspond to a system of linear independent equations, where the unknowns correspond to the precoders and receive filters entries. According to the number of equations and unknowns, the linear system is either overdetermined (has no solution), has exactly one solution, or is underdetermined (has infinitely many solutions). This counting of equations and unknowns can be used to obtain (3.40) (or (3.41)). In the multiple stream case, the authors noted that the equations become coupled and the coefficients are repeated. Therefore, the same framework cannot be applied. Later works on feasibility of IA were performed in [80, 81] and are based on algebraic geometry. Authors in [80] extended the results of [79] and showed that the necessary condition in (3.41) is also a sufficient condition for square symmetric channels and holds for $d \geq 1$. The authors in [81] extended the results of [79, 80] by showing that this necessary condition is also sufficient for symmetric channels with $d \geq 1$. They also provided a set of lengthy conditions in the general asymmetric channel case, which we skip for brevity and due to the fact that later chapters dealing with IA will only consider symmetric or square symmetric channels. As a summary, the condition $M + N \geq (K + 1)d$ was shown to be sufficient and necessary for symmetric channels with $d \geq 1$.

3.3. Other Approaches for MIMO Interference Channels

A number of other linear precoding methods were investigated for MIMO ICs. These methods were motivated by two facts. The first one being that IA approaches do not optimize the direct channel gains and are thus suboptimal at low SNR. The second one is that even at high but finite SNR, the rates achieved by IA might still be $o(\log(\mathrm{SNR}))$ away from capacity (cf. Section 3.1.3). For a given set of channel parameters, these methods have the same number of transmitted streams of IA methods as these correspond to the DoF of a given system. However, they have different optimization criteria and they can be split into following categories, e.g.:

1. SINR based approaches (see, e.g., [11]),
2. MSE based approaches (see, e.g., [12, 13, 82]),
3. Rank optimization approaches (see, e.g., [83, 84]), and
4. Pricing approaches (see, e.g., [85–87]).

Again, different proxy functions are optimized instead of the rate functions. We summarize some of these approaches next.

SINR approaches

The max-SINR algorithm in [11] uses the SINR metric as a proxy to the capacity function, and aims at maximizing the SINRs of the different user streams. Its formulation is based on channel reciprocity in TDD mode[2]. Under perfect channel reciprocity in TDD mode, the signaling dimensions along which a receiver sees the least interference from transmitters are also the same dimensions along which this receiver will cause the least interference to other transmitters in a reciprocal network where the roles of transmitters and receivers are reversed. This property is used to formulate the max-SINR problem. The SINR of the j-th stream of the k-th receiver in the original network equals

$$\mathrm{SINR}_{k,j} = \frac{\boldsymbol{G}_k^{j,\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k^j \boldsymbol{F}_k^{j,\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{G}_k^j}{\boldsymbol{G}_k^{j,\mathrm{H}} \boldsymbol{B}_{k,j} \boldsymbol{G}_k^j} \frac{\mathrm{E}_{tx,k}}{d_k} \tag{3.42}$$

where $\boldsymbol{B}_{k,j}$ is the interference plus noise covariance matrix of the j-th stream of the k-th receiver:

$$\begin{aligned} \boldsymbol{B}_{k,j} = & \sum_{l=1}^{K} \frac{\mathrm{E}_{tx,l}}{d_l} \sum_{u=1}^{d_l} \boldsymbol{H}_{kl} \boldsymbol{F}_l^u \boldsymbol{F}_l^{u,\mathrm{H}} \boldsymbol{H}_{kl}^{\mathrm{H}} - \\ & \frac{\mathrm{E}_{tx,k}}{d_k} \boldsymbol{H}_{kk} \boldsymbol{F}_k^j \boldsymbol{F}_k^{j,\mathrm{H}} \boldsymbol{H}_{kk}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{n}_k} \end{aligned} \tag{3.43}$$

(cf. [11]). With other variables fixed, the unit vector $\boldsymbol{G}_k^j$ that maximizes $\mathrm{SINR}_{k,j}$ is given by:

$$\boldsymbol{G}_k^j = \frac{\boldsymbol{B}_{k,j}^{-1} \boldsymbol{H}_{kk} \boldsymbol{F}_k^{:,j}}{||\boldsymbol{B}_{k,j}^{-1} \boldsymbol{H}_{kk} \boldsymbol{F}_k^{:,j}||_2}. \tag{3.44}$$

Similar SINR expressions can be obtained in the reciprocal network. After the precoders and receive filters are randomly initialized, an iterative procedure is performed between the original and reciprocal network: the receive filters optimizing the SINR in the original network are calculated for fixed precoders as in (3.44), then the precoders optimizing the SINR in the reciprocal network are calculated for fixed receive filters. This continues till convergence. The max-SINR algorithm exhibits a better performance than the IL algorithm, especially in the low and medium SNR regimes.

MSE approaches

[2]The TDD-based formulation is only needed to find the precoder and receive filter solutions. The method is also applicable to FDD systems.

In [12, 13], a sum MSE for the single-stream case is defined as

$$\alpha = \sum_{k=1}^{K} \mathrm{E}\left[|s_k - \hat{s}_k|^2\right] \tag{3.45}$$

where the system equation reads

$$\hat{s}_k = \boldsymbol{g}_k^{\mathrm{H}} \left(\boldsymbol{H}_{kk}\boldsymbol{f}_k s_k + \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_{kl}\boldsymbol{f}_l s_l + \boldsymbol{n}_k \right) \tag{3.46}$$

when $d_k = 1 \forall k$. The following optimization problem is posed:

$$\min_{\substack{(\boldsymbol{f}_1,\ldots,\boldsymbol{f}_K)\\(\boldsymbol{g}_1,\ldots,\boldsymbol{g}_K)}} \alpha \quad \text{s.t. } ||\boldsymbol{f}_k||_2^2 \leq 1 \ \forall k. \tag{3.47}$$

The Lagrangian function of the above problem is constructed as (see, e.g., [88])

$$L = \alpha + \sum_{k=1}^{K} \lambda_k (||\boldsymbol{f}_k||_2^2 - 1) \tag{3.48}$$

where $\lambda_k \geq 0$ is the Lagrangian multiplier associated with the kth transmit power constraint. With other variables fixed, the kth receive filter that minimizes α is given by

$$\begin{gathered} \frac{\partial L}{\partial \boldsymbol{g}_k^*} = 0 \implies \\ \boldsymbol{g}_k = \left(\sum_{m=1}^{K} \boldsymbol{H}_{km}\boldsymbol{f}_m\boldsymbol{f}_m^{\mathrm{H}}\boldsymbol{H}_{km}^{\mathrm{H}} + \sigma_n^2 \mathbf{I}_N \right)^{-1} \boldsymbol{H}_{kk}\boldsymbol{f}_k, \end{gathered} \tag{3.49}$$

while similarly precoder $\boldsymbol{f}_k$ is given by

$$\begin{gathered} \frac{\partial L}{\partial \boldsymbol{f}_k^*} = 0 \implies \\ \boldsymbol{f}_k = \left(\sum_{m=1}^{K} \boldsymbol{H}_{mk}^{\mathrm{H}}\boldsymbol{g}_m\boldsymbol{g}_m^{\mathrm{H}}\boldsymbol{H}_{mk} + \lambda_k \mathbf{I}_M \right)^{-1} \boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{g}_k. \end{gathered} \tag{3.50}$$

An explicit expression for λ_k cannot be given. However, the authors in [12] noted that as $||\boldsymbol{f}_k||_2^2$ is convex and decreasing in λ_k, there exists a unique solution for λ_k which can be found via Newton iterations for instance. Randomly initialized, the receive filters and precoders can be calculated according to (3.49) and (3.50) in a round robin fashion until convergence. The algorithm shows a similar

performance to the max-SINR algorithm. In addition, it can be easily adapted to the case where some users have more priorities over others, in which case the optimization problem is reformulated as a weighted MSE minimization problem, and shows a better performance than the max-SINR in low and medium SNR ranges. A later work by authors in [82] formulated a MSE minimization problem and presented a corresponding solution in the $d_k \geq 1$ case.

Even though MSE and SINR based algorithms might not perfectly cancel the interference at the different receivers, they show a better performance than the IL algorithm in general because the former approaches are better proxies of the achievable rates than the IA criterion in the finite SNR regime [11, 12, 82]. If the IL algorithm and its variants are regarded as generalized ZF algorithms operating on lower dimensional subspaces and by recalling that ZF precoding is suboptimal to SINR and MSE approaches in MIMO BCs in the finite SNR regime, a direct link between the performance of the precoding methods in MIMO ICs and precoding methods in MIMO BCs discussed in the previous chapter can be established.

Rank optimization approaches

Rank optimization approaches were motivated by the fact that when perfect IA is not possible or numerically reachable, algorithms such as the IL will result in solutions with low interference leakage/power but the resulting interference covariance matrices might possibly span multiple dimensions. Such solutions are not tightly related to the DoF, which is the pre-log capacity factor at high SNR. Indeed, the authors in [83] conjectured that solutions favoring interference covariance matrices with low rank are more representative of the DoF and would result in better sum-rate solutions than the IL algorithm or its variants when perfect IA is not possible or numerically reachable. Defining

$$\boldsymbol{J}_k \triangleq \left[\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{k1}\boldsymbol{F}_1, \ldots, \boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{k,k-1}\boldsymbol{F}_{k-1}, \boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{k,k+1}\boldsymbol{F}_{k+1}, \ldots, \boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kK}\boldsymbol{F}_K\right] \tag{3.51}$$

as the horizontal concatenation of interference matrices at receiver k, they reformulated the IA problem as the following rank constrained rank minimization problem:

$$\min_{\substack{(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_K)\\(\boldsymbol{G}_1,\ldots,\boldsymbol{G}_K)}} \sum_{k=1}^{K} \mathrm{rank}(\boldsymbol{J}_k) \quad \text{s.t. } \mathrm{rank}(\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kk}\boldsymbol{F}_k) = d_k \ \forall k \tag{3.52}$$

which is equivalent to maximizing the number of interference-free dimensions[3].

[3]The authors include the rank constraint $\mathrm{rank}(\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kk}\boldsymbol{F}_k) = d_k$—with $\boldsymbol{G}_k$ and $\boldsymbol{F}_k$ unitary—when developing a solution of the precoders and receive filters. However, this condition is satisfied with probability 1 as was discussed in Section 3.2.1. Therefore, we conjecture that this constraint can be simply dropped from the optimization problem.

Such a problem is non-convex in the precoder and receive filter variables. Even if it were, minimizing ranks of matrices is not straightforward. Therefore, they relaxed the problem and obtained a convex approximation of the objective. Such an approximation is given by

$$\begin{aligned} \text{conv}\left(\sum_{k=1}^{K}\text{rank}(\boldsymbol{J}_k)\right) &= \frac{1}{\mu}\sum_{k=1}^{K}||\boldsymbol{J}_k||_* \\ &= \frac{1}{\mu}\sum_{k=1}^{K}\sum_{t=1}^{d_k}\sigma_t(\boldsymbol{J}_k) \end{aligned} \tag{3.53}$$

where $\text{conv}(f)$ denotes the convex envelope of f and $||\boldsymbol{A}||_*$ is the nuclear norm of $\boldsymbol{A}$, i.e., the sum of the singular values of $\boldsymbol{A}$. When the maximum singular value of the interference matrices is upper bound by μ, (3.53) provides a convex approximation for $\sum_{k=1}^{K}\text{rank}(\boldsymbol{J}_k)$. They noted that the IL cost function can be written as the sum of the squared corresponding singular values, which is the main difference between the two approaches. The resulting problem was then solved numerically using convex optimization toolboxes, and numerical results showed that the proposed rank minimization algorithm results in performance improvements over the IL algorithm when perfect IA is not possible. Its performance in comparison to the max-SINR algorithm varied according to the scenario and SNR range. A possible explanation is the unpredictable convergence behavior of both algorithms.

Authors in [84] proposed a refinement of the nuclear norm approach. The reason is that the nuclear norm operator sums the singular values, while the rank operator sums the number of non-zero singular values. As a result, a nuclear norm minimization approach penalizes the larger singular values more heavily than the smaller ones, while a rank minimization approach treats them equally. Such a fine difference is not seen when perfect IA is possible (or optimal, e.g., at high SNR) as both the rank of interference covariance matrices and their nuclear norm would equal 0. However, in other cases, and to more accurately resemble a rank function, they proposed to adapt the objective in (3.53) by left multiplying $\boldsymbol{J}_k$ by a weight matrix that is iteratively updated as to contain the inverse of the obtained singular values on its diagonal after each iteration. Such a weighted nuclear norm approach shows performance improvements over the unweighted one in low and medium SNR ranges. Its performance in comparison to the max-SINR algorithm again varied according to the scenario and SNR range.

As a final word in this section, we note that MMSE, SINR, or pricing algorithms can be applied to the MISO IC as well, unlike the IA algorithms which require at least 2 receive antennas to align the interference. Similarly, rank optimization approaches necessitate 2 receive antennas at least.

CSI requirements and implementation: For the algorithms of this section and besides the CSI of the interfering links, the CSI set of direct links $\{\boldsymbol{H}_{11}, \ldots, \boldsymbol{H}_{KK}\}$ is also involved in the optimization process, as can be seen from (3.44), (3.49), or (3.50). Besides this fact, the implementation of any of the algorithms of this section follows similarly to the one of the previous section.

3.4. Other Channel Configurations

3.4.1. Configurations Where Interference Cannot be Overcome

As the number of streams that can be simultaneously transmitted depends on the number of transmit antennas, receive antennas, and users, this implies that there are some configurations where no simultaneous transmission is possible across the K transmitter/receiver links. This case can be represented by the inequality:

$$\sum_{k=1}^{K} d_k(M_k + N_k - 2d_k) < \sum_{k=1}^{K} \sum_{l=1, l\neq k}^{K} d_k d_l \tag{3.54}$$

which simplifies to

$$M + N < (K+1)\, d \tag{3.55}$$

in the symmetric channel case. As the conditions in (3.40) and (3.41) are necessary, any system not satisfying them will be interference limited. This gives (3.54) and (3.55).

Example 3.2. *The scenario* $(2,1,1)^3$ *is interference limited.*

A possible solution to mitigate interference here is through orthogonal access schemes. For instance, TDMA can be used to mute a subset of the transmitters at each transmission interval. Muting one transmitter at each transmission interval in the original $(2,1,1)^3$ scenario would lead to a $(2,1,1)^2$ scenario satisfying (3.58) (cf. Section 3.4.2) which is not interference limited anymore. Another solution would be moving a subset of transmissions to other frequency bands, i.e, FDMA.

3.4.2. Configurations With Full-Spatial Multiplexing

So far, we have discussed configurations where either simultaneous transmissions across the K links without full-spatial multiplexing are possible, or configurations where simultaneous transmissions are not possible at all. We next discuss configurations where enough antennas are available at the transmitters and/or receivers to support full-spatial multiplexing, i.e., the transmission of up to $d_k = \min(N_k, M_k)$ data streams to receiver k. We focus on the case where

transmitters are equipped with at least as many antennas as receivers. This case yields $d_k = N_k$.

It is first important to point out that all the iterative algorithms from Section 3.2.1 can still be used here. However, the additional number of antennas present in this case allows for non-iterative algorithms to be implemented. One example is the BD algorithm [37], which is a multi-user ZF scheme. It is mainly intended for users with multiple antennas, and supports full-spatial multiplexing. Even though originally presented for the MIMO BC, the BD can be implemented in the MIMO IC as well. The BD uses the available degrees of freedom to construct ZF precoders directly, without the need for an iterative design nor a joint precoder-receive filter design. In contrast to (3.9) and with enough transmit antennas, ZF precoders satisfy

$$\begin{aligned} \boldsymbol{H}_{kl}\boldsymbol{F}_l &= \mathbf{0}_{N_k \times N_l} \\ \operatorname{rank}(\boldsymbol{H}_{kk}\boldsymbol{F}_k) &= N_k \\ k \in \{1,\ldots,K\}, l &\in \{1,\ldots,K\}\backslash\{k\}. \end{aligned} \tag{3.56}$$

Thus, all the processing is done at the transmitter side in this case and no processing is required at the receiver side in contrast to Sections 3.2 and 3.3. The BD is therefore a good option in fast-varying systems with short coherence times. The conditions in (3.56) correspond again to a linear system consisting of a number of equations and a number of unknowns. It yields a solution if and only if:

$$M_k \geq \sum_{l=1}^{K} N_l \tag{3.57}$$

which simplifies to

$$M \geq KN \tag{3.58}$$

in the symmetric channel case. The inequality in (3.57) was originally derived in [37] based on a singular value decomposition analysis of the aggregate interference matrix at each receiver. They can also be obtained from (3.40) by setting $d_k = N_k$.

Example 3.3. *The scenario* $(8,2,2)^3$ *satisfies condition* (3.58).

If the above conditions are satisfied, interference can be simply zero-forced by choosing $\boldsymbol{F}_k$ to lie in the nullspace of

$$\left[\boldsymbol{H}_{1k}^{\mathrm{T}},\ldots,\boldsymbol{H}_{k-1,k}^{\mathrm{T}},\boldsymbol{H}_{k+1,k}^{\mathrm{T}},\ldots,\boldsymbol{H}_{Kk}^{\mathrm{T}}\right]^{\mathrm{T}}.$$

The last step consists in maximizing the achievable rate of user k given this choice of $\boldsymbol{F}_k$. Writing the final precoder as $\boldsymbol{F}_k\tilde{\boldsymbol{F}}_k$ where $\tilde{\boldsymbol{F}}_k \in \mathbb{C}^{N\times N}$, this is

equivalent to solving the problem:

$$\max_{\tilde{\boldsymbol{F}}_k:\mathrm{tr}\left(\boldsymbol{F}_k\tilde{\boldsymbol{F}}_k\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{F}_k^{\mathrm{H}}\right)\leq \mathrm{E}_{\mathrm{tx},k}} \log\left|\boldsymbol{I}_N+\frac{1}{\sigma_n^2}\boldsymbol{H}_{kk}\boldsymbol{F}_k\tilde{\boldsymbol{F}}_k\tilde{\boldsymbol{F}}_k^{\mathrm{H}}\boldsymbol{F}_k^{\mathrm{H}}\boldsymbol{H}_{kk}^{\mathrm{H}}\right|. \tag{3.59}$$

This is a standard P2P rate maximization problem whose solution is given in Appendix B. The receive filters are square matrices that do not perform any rank reductions. Therefore, they can be simply set to $\boldsymbol{G}_k = \mathbf{I}_{N_k}$ as any other choice would not improve the system performance from an achievable rate point of view.

A modified version of BD, the regularized BD, was proposed in [89]. It aims at improving the performance of BD at low SNR. The regularized BD solution introduces an SNR dependent regularization term to the BD solution and can be thus considered as a generalized Wiener filter design for receivers with multiple antennas.

Yet one final case can be distinguished from the general case covered in this section. It corresponds to the case

$$M \gg KN. \tag{3.60}$$

All algorithms covered in this chapter so far can be applied here. However, this case deserves a special attention. To get a better understanding, we first introduce MRT precoding. MRT is a simple and well known precoding strategy which aims at maximizing the received power of a given user without considering the interference that might be caused to other users. The precoder is a Hermitian version of the CSI of the direct link, i.e.:

$$\boldsymbol{F}_k = \alpha_k \boldsymbol{H}_{kk}^{\mathrm{H}} \tag{3.61}$$

where α_k is a term used to satisfy the transmit power constraint. For simple scenarios such as P2P MISO links where inter-user and inter-stream interference are non-existent, MRT is optimal. However, in MIMO scenarios where M is comparable to N and in the high SNR regime, MRT is suboptimal as it results in strong interference to unintended users. Similar observations hold for eigenmode precoding which is optimal in P2P links. As discussed in the previous chapter, MRT becomes optimal in MIMO BCs as $M \to \infty$. We can also generalize this statement and say that all precoding strategies which only depend on the direct channels, e.g., eigenmode precoding, are asymptotically optimal in MIMO ICs. Therefore, in systems where (3.60) is satisfied, the simple MRT or eigenmode precoding methods might constitute an attractive choice.

3.5. Classification From a Game Theory Point of View

In game theory, it is usual to define a set of players, a set of actions by each player, and a utility function for each player [90]. Furthermore, actions are split into non-cooperative/selfish and cooperative strategies. An essential requirement for cooperation among players is achieving a benefit for all of them. In the context of MIMO ICs, players correspond to users, actions correspond to precoding strategies, and utility functions correspond to achievable rate functions. All the algorithms discussed in this chapter except for MRT and eigenmode precoding are considered to be cooperative strategies. This is because transmitters design their precoders to minimize the interference to other receivers in the system, instead of each transmitter trying to maximize the rate of its dedicated receiver only. The result is increased achievable rates for all users. A non-cooperative strategy such as eigenmode precoding and MRT where any transmitter aims at maximizing the rate of its dedicated receiver would result in strong interference to other receivers in general. Consequently, only a low network sum-rate can be achieved.

3.6. Common Drawbacks of Cooperative Methods

One common problem with the cooperative methods of this chapter (except for BD and the closed-form IA solutions) is the convergence behavior due to the non-convex nature of the different optimization problems and the coupling of the different variables. At best, iterative IA algorithms are analytically guaranteed to converge to a local minimum. Even so, the alternating optimization procedure is computationally expensive and might require hundreds or thousands of iterations for medium-sized systems to converge. For this reason, they might not be feasible in practical systems as by the time a solution is found, the users' channels would have changed and the obtained solutions would be outdated. The methods of Section 3.3 are not guaranteed to converge and their performance can be only checked via simulations. The performance of many of these methods was observed to be unpredictable in many scenarios (see, e.g., [83]).

Another issue that should be considered is the higher CSI requirements of these methods compared to non-cooperative methods, as discussed in the previous sections. As elaborated in Section 2.4, CSI is acquired via pilot symbols sent in the downlink and/or uplink. The knowledge of the CSI of additional links necessitates more pilot symbols, which occupy more time slots that would be otherwise used for data transmission. The effects of these piloting overheads on spectral efficiency are often neglected and the required CSI is assumed to be available for free. Finally, the dependency of such methods on CSI from multiple links makes them highly non-robust to practical channel estimation errors,

which lead to mismatched precoders as well as receive filters and consequently limit the gains of cooperative methods in practice.

MRT or eigenmode precoding combined with a large number of transmit antennas deserve a special attention because this combination allows climbing out of the "interference-limited" regime merely because of the size of the transmit antenna array. In addition, MRT and eigenmode precoding exhibit the following properties which are desirable in practical systems:

1. the precoders can be obtained in closed-form for arbitrary system parameters,[4]

2. these methods are more robust to CSI errors than cooperative methods.

The second point is merely due to the inherent nature of non-cooperative methods in which receivers treat interference as noise and do not perform any processing based on interfering links. Similarly, transmitters do not perform any processing based on interfering links.

One of the goals of future cellular systems is creating noise-limited systems. Naturally, this can be achieved through either cooperative precoding strategies or through non-cooperative precoding strategies combined with large transmit arrays. In the second case, however, the physical required number of antennas has not been investigated so far. Keeping in mind that cooperative and non-cooperative strategies translate to multi-cell and single-cell strategies, respectively, such a number is important for designing the overall system architecture. Initial massive MIMO architectures envisioning hundreds of antennas might not be feasible due to high costs, lack of space in urban areas, etc. In this thesis, we attempt to find out how much transmit antennas a non-cooperative strategy requires so that it emulates the performance of a noise-limited system, i.e., so that its resulting performance is similar to the one of a system where transmitters with a fixed number of antennas employ a cooperative strategy.

3.7. Relation to LTE Cooperative Transmission Schemes

In the LTE standards, cooperation between transmitters is referred to as coordinated multipoint transmission (CoMP) and is split into two categories [91]:

1. coordinated scheduling / beamforming (CS, CB), and

2. joint processing (JP), also called Network MIMO.

[4] MRT is especially attractive because the transmitter simply constructs its precoder as a Hermitian version of the channel estimate without any decompositions nor inversions as is the case for other designs.

CS corresponds to the simplest cooperation type as transmitters only cooperate to find the best scheduling decisions. An example of CS was already covered in Section 3.4.1 which corresponds to a muting decision by one or more transmitters to reduce interference. All the algorithms presented in this chapter except for MRT and eigenmode precoding fall into the CB class.

When JP is employed, the same data symbol(s) are sent from multiple transmitters to a particular receiver in contrast to CS and CB. JP is the most demanding class of cooperation as it requires CSI exchange as well as data symbol(s) exchange among the transmitters. Information exchange occurs over the backhaul link connecting the transmitters which is called the X2 interface in LTE systems; therefore, JP requires a higher backhaul capacity than CS and CB. Theoretically, JP can convert interference power from dominant interferers into additional useful signal power coming from the latter, assuming transmission and reception procedures are properly synchronized. In order to keep the backhaul requirements moderate, JP is not considered in this thesis and the part of the thesis related to CoMP transmission will only deal with CB.

3.8. Summary

In this chapter, we presented different linear precoding methods for MIMO ICs. We differentiated three main regimes according to the network configuration in terms of the number of transmit antennas, receive antennas, desired streams, and users. Two of these regimes allow simultaneous transmission over the K links while at the same time mitigating interference. In one of these two regimes, mainly iterative methods such as IA need to be used. We reviewed different iterative IA algorithms in that regime and a closed-form IA solution in one special case. We then summarized other approaches such as MSE, SINR, or rank based approaches. The other interference-mitigating regime allows additionally for non-iterative methods such as BD to be used. In that regime, we identified the massive MIMO special case and discussed possible precoding options when transmitters are equipped with a large and non-conventional numbers of antennas. In the last regime, there are too little antennas and/or too many users to allow simultaneous transmission on the same time-frequency resources. Consequently, transmissions have to be orthogonalized in time or frequency. We also discussed the different methods' CSI requirements and convergence behavior, presented different possible centralized and decentralized implementations of cooperative methods, and discussed some drawbacks of cooperative methods in the context of practical systems. We finally showed how these methods relate to current LTE cooperative transmission schemes.

This chapter didn't include designs with limited CSI as this is outside the scope of this thesis. Nonetheless, this is an important topic with ongoing re-

search, most of which is concerned with IA. Recent publications include works in [92–96]. Authors in [92] deal with single stream transmission and show that for some channel configurations, interference can be first aligned in a subset of the K interfering links. A $K_s < K$ sub-IC is considered where the IA problem is feasible and K_s precoders and K_s receive filters are designed given only the CSI of interfering links of this sub-IC, resulting in a reduction of CSI requirements for the given K_s users. Afterwards, the remaining $K - K_s$ precoders/receive filters can be aligned to the already constructed ones. This is done using the CSI of all interfering links seen by the last $K - K_s$ users, and doesn't result in a CSI reduction for these users. The works in [93–96] deal with limited feedback and codebook design for IA. They build on the fact that optimal IA precoders lie on Grassmannian manifolds to construct codebooks entries belonging to such manifolds. They are therefore different from many conventional codebook construction techniques.

4. Large System Performance of Interference Alignment

In this chapter, we will derive rate expressions of cooperative IA methods when applied to transmitters and receivers with large numbers of antennas. RMT tools, large system analysis (LSA), and the law of large numbers (LLN) constitute an integral part of this chapter. In general, analytical rate expressions of MIMO algorithms are hard to obtain. However, LSA is a very useful tool to overcome this problem. In an LSA, when two of system parameters, e.g., number of transmit and receive antennas, go to infinity at a finite fixed ratio, algorithms' performance metrics such as achievable rates or BERs become possible to obtain. This is due to the fact that the eigenvalue or singular value distributions of large random matrices often converge to fixed asymptotic distributions. This important property can be used to derive analytical rate expressions in the large system limit. Examples of such distributions include the Marčenko-Pastur distribution, which is the asymptotic eigenvalue distribution of large Wishart matrices. Another example is the quarter circle distribution, which is the asymptotic singular value distribution of large square matrices.

Even though the resulting expressions are only exact asymptotically, they usually provide a very good estimate for a finite number of system parameters as well. It has been applied to CDMA systems [97–102] as well as to multi-user MIMO systems. It has been used in [103] to derive rate lower bound expressions of the successive-encoding-successive-allocation method [48] in the MIMO BC. In [104], an LSA is performed in the MIMO BC and rate expressions of the BD algorithm and rate lower bound expressions of BD with dirty paper coding are derived. LSA has been applied to IA in [105,106]. In [105], the optimal rate curve offset (the y-axis intercept) of IA in the single stream square symmetric MIMO IC is analyzed using LSA. In [106], an uplink two-cell scenario with Ricean channels is considered and the impact of the Ricean factor on IA performance is analyzed using LSA. LSA for MRT and ZF has been performed in [54] for non-cooperative multi-cell MISO systems. In [107], a two cooperating MISO cell network is considered and an LSA is performed to obtain asymptotic values of the SINRs and bit rates.

In this chapter, large system rate expressions will be obtained in closed-form under equal power allocation, and in quasi closed-form for the optimal water-filling (WF) power allocation. The resulting expressions in the latter case depend

on the asymptotic WF level which has to be found numerically. An intermediary needed result on the asymptotic eigenvalue distribution of the direct channel gains resulting from IA will be obtained as well. As explained in Chapter 3, different IA variants can be grouped together in one class as they aim at aligning and subsequently canceling interference from other transmitters through a proper precoder/receive filter design. Afterwards, the rate of the different users can be maximized through WF. Thus, we expect the obtained large system expressions to be valid for different IA variants. A similar conjecture applies for the obtained asymptotic eigenvalue distribution. The simulation results at the end of the chapter will validate these conjectures.

We start this chapter by revisiting some of the main RMT tools which upcoming derivations will be based on. These tools will be needed to characterize the performance of eigenmode precoding in Chapter 5 as well. The material of this chapter has been published in [108].

4.1. Preliminaries

4.1.1. The Marčenko-Pastur Distribution

Let $\boldsymbol{A}$ be a real or complex random $s \times t$ matrix with i.i.d. zero mean elements of variance $1/s$. As $s, t \to \infty$ at a fixed ratio $t/s = \alpha$, the distribution of the eigenvalues of $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ converges almost surely to a fixed distribution with density [19–21]

$$f_{\mathrm{MP}}(\lambda) = \left(1 - \frac{1}{\alpha}\right)^{+} \delta(\lambda) + \frac{\sqrt{(\lambda - a)^{+}(b - \lambda)^{+}}}{2\pi\alpha\lambda} \tag{4.1}$$

where $a = (1 - \sqrt{\alpha})^2$, $b = (1 + \sqrt{\alpha})^2$, and $y^{+} = \max(0, y)$.

4.1.2. The Shannon Transform

The Shannon transform of a non-negative random variable x is defined as

$$\mathcal{V}_x(\gamma) = \mathrm{E}\left[\log(1 + \gamma x)\right] \tag{4.2}$$

where $\gamma \in \mathbb{R}^{+}$. If x follows a Marčenko-Pastur distribution, its Shannon transform is additionally a function of α which is given in Section 4.1.1 and is found in closed form as [21]

$$\begin{aligned} \mathcal{V}(\gamma, \alpha) = &\log\left(1 + \gamma - \frac{1}{4}\mathcal{F}(\gamma, \alpha)\right) + \\ &\frac{1}{\alpha}\log\left(1 + \gamma\alpha - \frac{1}{4}\mathcal{F}(\gamma, \alpha)\right) - \frac{\log(e)}{4\gamma\alpha}\mathcal{F}(\gamma, \alpha), \end{aligned} \tag{4.3}$$

where e is Euler's number and

$$\mathcal{F}(\gamma,\alpha)=\left(\sqrt{\gamma(1+\sqrt{\alpha})^2+1}-\sqrt{\gamma(1-\sqrt{\alpha})^2+1}\right)^2.$$

4.1.3. The Quarter Circle Distribution

Let $\boldsymbol{A}$ be a real or complex square random $s\times s$ matrix with i.i.d. zero mean elements of variance $1/s$. As $s\to\infty$, the distribution of the singular values of $\boldsymbol{A}$ converges almost surely to a fixed distribution with density [20, 21]

$$f_{\mathrm{QC}}(\sigma)=\frac{1}{\pi}\sqrt{4-\sigma^2},\quad \sigma\in[0,2). \tag{4.4}$$

4.2. Equivalent Modified Channel and Transmit Power Models

Without loss of generality and for brevity of expressions, we focus on the symmetric channel $(M,N,d)^K$ (cf. Def. 3.1). We later show that the analysis is easily extendable to asymmetric ones. The analysis is performed for precoders with orthogonal columns. In case the obtained precoders do not satisfy this (e.g., closed-form solution in Section 3.2.2), they can be right-multiplied by a rotation matrix to restore the orthogonality condition without violating the IA conditions.

Recall the system model of Section 3.1.1 which will be used throughout this chapter. However, for the purpose of derivations, we use a modified but equivalent channel and transmit power model. Namely, we divide the channel matrices by $\sqrt{d}$ and multiply the precoders by $\sqrt{d}$, i.e.,

$$\begin{aligned}\tilde{\boldsymbol{H}}&=\frac{1}{\sqrt{d}}\,\boldsymbol{H}\\ \tilde{\boldsymbol{F}}&=\sqrt{d}\,\boldsymbol{F}=\boldsymbol{F}_{\mathrm{uni}}\underbrace{\left(\sqrt{d}\boldsymbol{P}^{\frac{1}{2}}\right)}_{\tilde{\boldsymbol{P}}^{1/2}}\end{aligned} \tag{4.5}$$

where the user indices have been dropped. This modified system model will be mainly needed for the rate derivation in the case of WF. As $\tilde{\boldsymbol{H}}\tilde{\boldsymbol{F}}=\boldsymbol{H}\boldsymbol{F}$, this model is equivalent to the one of Section 3.1.1. The equivalent model results in modified i.i.d. channel entries having variance $1/d$, i.e., $[\tilde{\boldsymbol{H}}]_{i,j}\sim\mathcal{N}_{\mathbb{C}}(0,1/d)$ $\forall i,j$. The modified transmit power constraint of transmitter k equals

$$Q_k=\mathrm{tr}(\tilde{\boldsymbol{P}}_k)=\mathrm{E}_{\mathrm{tx},k}\,d \tag{4.6}$$

which grows with d. The fading power and transmit power in the equivalent model are modified; however, the received power is still the same as the one in

model of Section 3.1.1. The equivalent model will be used subsequently used in this chapter as it substantially facilitates the performed derivations. However, the simulations will be run using the model of Section 3.1.1 in order to check the validity of the obtained expressions.

After the IA precoders and receive filters are found and supposing they lead to a global optimum of the IA problem, the K-user MIMO IC is decomposed into K interference-free P2P MIMO links. The rate of user k is expressed as

$$\begin{aligned} R_k &= \log\det\left(\mathbf{I}_d + \frac{1}{\sigma_n^2}\,\boldsymbol{F}_{k,\mathrm{uni}}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\boldsymbol{G}_k\boldsymbol{G}_k^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}\boldsymbol{F}_{k,\mathrm{uni}}\tilde{\boldsymbol{P}}_k\right) \\ &= \log\det\left(\mathbf{I}_d + \frac{1}{\sigma_n^2}\,\boldsymbol{H}_{kk,\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{kk,\mathrm{eff}}\tilde{\boldsymbol{P}}_k\right) \end{aligned}$$

where

$$\boldsymbol{H}_{kk,\mathrm{eff}} = \boldsymbol{G}_k^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}\boldsymbol{F}_{k,\mathrm{uni}} \in \mathbb{C}^{d\times d} \tag{4.7}$$

is the effective channel of the kth pair and $\tilde{\boldsymbol{P}}_k = \mathrm{diag}(p_{k,1}, p_{k,2}, \ldots, p_{k,d})$. R_k can be rewritten as

$$R_k = \sum_{l=1}^{d} \log\left[1 + \frac{1}{\sigma_n^2}\,p_{k,l}\,\lambda_l\left(\boldsymbol{H}_{kk,\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{kk,\mathrm{eff}}\right)\right] \tag{4.8}$$

where the set of powers $\{p_{k,l}\}$ is chosen according to equal power allocation or according to the optimal WF rule. The set of powers satisfies $\sum_{l=1}^{d} p_{k,l} = Q_k$ as explained in (4.6). The achievable rate of user k depends on the eigenvalues of the matrix $\boldsymbol{H}_{kk,\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{kk,\mathrm{eff}}$.

4.3. Large System Rate Analysis

To perform an LSA of the achievable rates, we first assume the system is fully loaded, i.e., the maximum number of streams is transmitted (this condition will be relaxed later). In this case, the IA feasibility condition is achieved with equality:

$$M + N = (K+1)\,d \tag{4.9}$$

[cf. (3.41)]. In an LSA, it is common to let two system parameters go to infinity at a finite fixed ratio, while keeping the remaining one(s) fixed. The system parameters at hand are K, M, N (and implicitly d). As we would like to obtain expressions for an arbitrary large number of streams, the LSA is performed under the assumption that $d \to \infty$. As $d \leq \min(N, M)$, this furthermore necessitates that both N and M go to infinity. By fixing K and letting M and N go

to infinity at a finite fixed ratio β, d will go to infinity as well at a finite fixed ratio r w.r.t. the sum $M+N$:

$$\begin{aligned} &M \to \infty, N \to \infty \;\text{ with }\; \beta = \frac{M}{N} \Longrightarrow \\ &d \to \infty \;\text{ with }\; r = \frac{M+N}{d} = K+1. \end{aligned} \tag{4.10}$$

No conditions are imposed on the ratio β which can take any finite value. For the rest of this chapter, we abbreviate the conditions in (4.10) by simply $d \to \infty$. The following Lemma presents the asymptotic distribution of the eigenvalues of $\boldsymbol{H}_{kk,\text{eff}}^{\text{H}} \boldsymbol{H}_{kk,\text{eff}}$.

4.3.1. Direct Channels' Asymptotic Eigenvalue Distribution

Lemma 4.1. *As $d \to \infty$, the distribution of the eigenvalues of $\boldsymbol{H}_{kk,\text{eff}}^{\text{H}} \boldsymbol{H}_{kk,\text{eff}}$ converges almost surely to a Marčenko-Pastur distribution with $\alpha = 1$ and corresponding density*

$$f(\lambda) = \frac{\sqrt{\lambda(4-\lambda)}}{2\pi\lambda}, \quad \lambda \in (0,4), \tag{4.11}$$

irrespective of the ratios r and β.

Proof. In the IL algorithm or its variants, the pairs of matrices $\{\boldsymbol{G}_k, \boldsymbol{F}_{\text{uni},k}\}$ are chosen based on the interfering links—independently of the direct links—to align interference in lower dimensional subspaces at the receivers before subsequently canceling it. $\{\boldsymbol{F}_{\text{uni},k}\}$ are unitary matrices and $\{\boldsymbol{G}_k\}$ are chosen as unitary zero-forcing filters. Left and right multiplying the matrix $\tilde{\boldsymbol{H}}_{kk}$—containing i.i.d. Gaussian entries—by the independent, unitary, and known matrices $\boldsymbol{G}_k$ and $\boldsymbol{F}_{k,\text{uni}}$ respectively leads to a new matrix with i.i.d. Gaussian entries of the same variance [21]. Note that the matrices $\boldsymbol{G}_k$ and $\boldsymbol{F}_{k,\text{uni}}$ need not to be constant for this to hold. This is the case here as these matrices change for every set of channel realizations. As $\tilde{\boldsymbol{H}}_{kk}$ has i.i.d. entries drawn from a $\mathcal{N}_{\mathbb{C}}\left(0, 1/d\right)$ distribution (cf. Section 4.2), the matrix $\boldsymbol{H}_{kk,\text{eff}} = \boldsymbol{G}_k^{\text{H}} \tilde{\boldsymbol{H}}_{kk} \boldsymbol{F}_{k,\text{uni}} \in \mathbb{C}^{d \times d}$ has i.i.d. elements drawn from a $\mathcal{N}_{\mathbb{C}}\left(0, 1/d\right)$ distribution[1] as well. The result of the above Lemma can then be directly obtained by replacing $\alpha = d/d = 1$ in (4.1). □

Note: In some cases, the parameters of a symmetric channel do not allow for a fully loaded scenario, i.e., the spatial degrees of freedom are not totally utilized and we have $M + N > (K+1)\,d$ (e.g., $K = 3, M = 6, N = 4$ results

[1] The distribution is conditioned on $\boldsymbol{G}_k$ and $\boldsymbol{F}_{k,\text{uni}}$.

in $d = 2$ and a non-fully loaded scenario). Here, no explicit ratio r can be found relating $M + N$ and d. However, the result obtained in (4.11) still holds as long as d grows with $\min(N, M)$ to infinity as the asymptotic eigenvalue distribution is independent of r or any other system parameter. Therefore, it holds for fully loaded as well as non-fully loaded systems. Consequently, the derived large system rates per stream in the next sections hold for fully loaded as well non-fully loaded systems.

Alternatively, the degrees of freedom can be fully utilized by assigning a subset of users more streams than the remaining users. In this case, we will have a fully loaded non-symmetric channel. Again, the derived asymptotic eigenvalue distribution and the derived large system rates per stream hold. However, the total achievable rate in this case is higher than the non-fully loaded scenario because more streams are transmitted.
In what follows, we drop the index k for brevity.

4.3.2. Large System Analysis and The Law of Large Numbers

Convergence in distribution is the weakest type of convergence of random variables: it does not imply that the eigenvalues $(\lambda_1, \ldots, \lambda_d)$ themselves converge to fixed values. Nonetheless, an important consequence of the convergence in distribution of the eigenvalues is the following. Let $g(\lambda)$ be an arbitrary function, well defined on the interval $\lambda \in (0, 4)$. Then, the averaged sum over l of $g(\lambda_l)$ converges as $d \to \infty$ according to:

$$\frac{\sum_{l=1}^{d} g(\lambda_l)}{d} \underset{d\to\infty}{\longrightarrow} \mathrm{E}_\lambda[g(\lambda)] = \int_0^4 g(\lambda) f(\lambda) d\lambda \tag{4.12}$$

as a consequence of the LLN (cf. Appendix A.4). The left-hand side corresponds to a single realization of a large matrix with eigenvalues $(\lambda_1, \ldots, \lambda_d)$ while the right-hand side corresponds to an expectation over a random variable λ whose distribution is $f(\lambda)$, the Marčenko-Pastur distribution given by (4.11). Eq. (4.12) combines the large system assumption in (4.10), the Marčenko-Pastur distribution, and the LLN. It constitutes a very useful identity as $g(\lambda)$ can correspond to the rate function as well as any other needed function as will be observed in Section 4.3.4.

4.3.3. Achievable Rates Under Equal Power Allocation

When equal power is allocated to each stream, we have

$$p_l = \frac{Q}{d} = \mathrm{E}_{\mathrm{tx}} \tag{4.13}$$

[cf. (4.6)] and the rate formula reads as

$$R_{\mathrm{EP}} = \sum_{l=1}^{d} \log\left[1 + \frac{\mathrm{E_{tx}}}{\sigma_n^2}\,\lambda_l\Big(\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff}}\Big)\right].$$

A direct application of (4.12) implies that the achievable rate per stream R_{EP}/d converges to the large system rate per stream $R_{\mathrm{LS,EP}}$ as $d \to \infty$:

$$\frac{R_{\mathrm{EP}}}{d} \underset{d\to\infty}{\longrightarrow} R_{\mathrm{LS,EP}}, \tag{4.14}$$

where

$$R_{\mathrm{LS,EP}} = \mathrm{E}_{\lambda(\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff}})}\left[\log\left[1 + \frac{\mathrm{E_{tx}}}{\sigma_n^2}\,\lambda\Big(\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff}}\Big)\right]\right] \tag{4.15}$$

which is the Shannon transform of $\lambda(\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{eff}})$ evaluated at $\mathrm{E_{tx}}/\sigma_n^2$ [cf. (4.2)]. As the eigenvalues follow a Marčenko-Pastur distribution with $\alpha = 1$, $R_{\mathrm{LS,EP}}$ can be expressed in closed form as

$$\begin{aligned} R_{\mathrm{LS,EP}} &= \mathcal{V}\left(\frac{\mathrm{E_{tx}}}{\sigma_n^2}, 1\right) \\ &= 2\log\left(1 + \frac{\mathrm{E_{tx}}}{\sigma_n^2} - \frac{1}{4}\mathcal{F}\left(\frac{\mathrm{E_{tx}}}{\sigma_n^2}, 1\right)\right) \\ &\quad - \frac{1}{4}\log(e)\mathcal{F}\left(\frac{\mathrm{E_{tx}}}{\sigma_n^2}, 1\right)\frac{\sigma_n^2}{\mathrm{E_{tx}}} \end{aligned} \tag{4.16}$$

[cf. (4.3)]. The obtained expression is a function of $\mathrm{E_{tx}}$ and σ_n^2 only.

Under equal power allocation, all streams have the same received signal-to-noise ratio (SNR), averaged over all variables. Furthermore, the SNR, denoted $\mathrm{SNR_{IA}}$, equals $\mathrm{E_{tx}}/\sigma_n^2$:

$$\begin{aligned} \mathrm{SNR_{IA}} &= \frac{\mathrm{E}\left[\left|\boldsymbol{e}_l^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff}}\tilde{\boldsymbol{P}}^{1/2}\boldsymbol{s}\right|^2\right]}{\mathrm{E}\left[\left|\boldsymbol{e}_l^{\mathrm{T}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{n}\right|^2\right]} = \frac{\mathrm{E}\left[\boldsymbol{e}_l^{\mathrm{T}}\boldsymbol{H}_{\mathrm{eff}}\tilde{\boldsymbol{P}}\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}\boldsymbol{e}_l\right]}{\mathrm{E}\left[\boldsymbol{e}_l^{\mathrm{T}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\boldsymbol{G}\boldsymbol{e}_l\right]} \\ &= \frac{\boldsymbol{e}_l^{\mathrm{T}}\,\mathrm{E_{tx}}\mathbf{I}_d\,\boldsymbol{e}_l}{\boldsymbol{e}_l^{\mathrm{T}}\,\sigma_n^2\mathbf{I}_d\,\boldsymbol{e}_l} = \frac{\mathrm{E_{tx}}}{\sigma_n^2} \end{aligned} \tag{4.17}$$

with the stream index $l = 1, \ldots, d$ and the used identities $\tilde{\boldsymbol{P}} = \mathrm{E_{tx}}\mathbf{I}_d$ and $\mathrm{E}[\boldsymbol{H}_{\mathrm{eff}}\boldsymbol{H}_{\mathrm{eff}}^{\mathrm{H}}] = \mathbf{I}_d$. Therefore, the large system rate expression under equal power allocation is a function of the average received SNR of each stream.

4.3.4. Achievable Rates Under Water-Filling

Recall (4.8):

$$R = \sum_{l=1}^{d} \log\left(1 + \frac{1}{\sigma_n^2} p_l \, \lambda_l\right) \tag{4.18}$$

and remember that the eigenvalues are arranged in a decreasing order

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d.$$

In the finite case, the solution to p_l under WF is given by (cf. Appendix B.2)

$$p_l = \max\left(0, \eta - \frac{\sigma_n^2}{\lambda_l}\right), \tag{4.19}$$

where η is the WF level given by

$$\eta = \frac{1}{t}\left(Q + \sum_{j=1}^{t} \frac{\sigma_n^2}{\lambda_j}\right) \tag{4.20}$$

and where $1 \leq t \leq d$ is the number of served streams, i.e., streams with strictly positive power. From (4.19) it is seen that a stream with $\lambda_l \leq \sigma_n^2/\eta$ is not served.

The main challenge in deriving the large system WF rates consists in finding the asymptotic WF level η^∞. This is because there is no explicit relationship between t and d, and in addition (4.20) doesn't hold asymptotically. The following Lemma shows how η^∞ can be calculated.

Lemma 4.2. *The asymptotic WF level η^∞ is the solution of the fixed-point equation:*

$$\eta^\infty = \frac{\mathrm{E_{tx}} + \int_{\lambda_{\min}^\infty}^{4} \frac{\sigma_n^2}{\lambda} f(\lambda)\mathrm{d}\lambda}{\int_{\lambda_{\min}^\infty}^{4} f(\lambda)\mathrm{d}\lambda}, \tag{4.21}$$

where

$$\lambda_{\min}^\infty = \frac{\sigma_n^2}{\eta^\infty} \tag{4.22}$$

is the asymptotic smallest served eigenvalue and $f(\lambda)$ is given by (4.11)*. η^∞ is constant over different channel realizations.*

Proof. To find an expression of η^∞, we first proceed by writing t, the number of served streams, as a sum of indicator functions:

$$t = \sum_{l=1}^{d} I(\lambda_l) \tag{4.23}$$

where

$$I(\lambda_l) = \begin{cases} 1, & \lambda_l > \frac{\sigma_n^2}{\eta}, \\ 0, & \lambda_l \leq \frac{\sigma_n^2}{\eta}. \end{cases}$$

Dividing both sides of (4.23) by d, we get

$$\frac{t}{d} = \frac{\sum_{l=1}^{d} I(\lambda_l)}{d},$$

which is the sample average of the function $I(\lambda)$. Applying (4.12) with $g(\lambda) = I(\lambda)$ yields

$$\begin{aligned} &\lim_{d\to\infty} \frac{t}{d} = \lim_{d\to\infty} \frac{\sum_{l=1}^{d} I(\lambda_l)}{d} = \\ &\mathrm{E}[I(\lambda)] = \int_0^4 I(\lambda) f(\lambda)\,\mathrm{d}\lambda = \\ &\int_0^{\lambda_{\min}^{\infty}} I(\lambda) f(\lambda)\,\mathrm{d}\lambda + \int_{\lambda_{\min}^{\infty}}^4 I(\lambda) f(\lambda)\,\mathrm{d}\lambda = \\ &\int_{\lambda_{\min}^{\infty}}^4 f(\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{4.24}$$

where $0 < \lambda_{\min}^{\infty} < 4$ is the asymptotic smallest served eigenvalue [cf. (4.22)], and $I(\lambda) = 0$ for $\lambda < \lambda_{\min}^{\infty}$. Thus, as $d \to \infty$, the limit of the ratio t/d is independent of instantaneous channel realizations and their corresponding eigenvalues and is only a function of the asymptotic eigenvalue distribution. This follows directly from the LLN.

Next, consider the term $\left(\sum_{j=1}^{t} \frac{\sigma_n^2}{\lambda_j}\right)/t$. Considering (4.12) again, it can be observed that this term converges to the conditional expectation[2]

$$\mathrm{E}\left[\frac{\sigma_n^2}{\lambda} \;\middle|\; \lambda > \lambda_{\min}^{\infty}\right]$$

as $d \to \infty$, and therefore [109]

$$\lim_{d\to\infty} \left(\frac{1}{t}\sum_{j=1}^{t} \frac{\sigma_n^2}{\lambda_j}\right) = \frac{\int_{\lambda_{\min}^{\infty}}^4 \frac{\sigma_n^2}{\lambda} f(\lambda)\mathrm{d}\lambda}{\int_{\lambda_{\min}^{\infty}}^4 f(\lambda)\mathrm{d}\lambda}. \tag{4.25}$$

Similarly, the limit of the above ratio is independent of the instantaneous eigenvalues and is only a function of the asymptotic eigenvalue distribution.

[2] The considered term only contains the t served eigenvalues. Thus, as $d \to \infty$ the condition $\lambda > \lambda_{\min}^{\infty}$ has to be included.

The asymptotic WF level can be then written using (4.20), (4.24), and (4.25) as

$$\begin{aligned}\eta^{\infty} &= \lim_{d\to\infty} \eta \\ &= \lim_{d\to\infty} \frac{1}{t}\left(Q + \sum_{j=1}^{t} \frac{\sigma_n^2}{\lambda_j}\right) \\ &= \frac{\lim_{d\to\infty} \frac{Q}{d} + \int_{\lambda_{\min}^{\infty}}^{4} \frac{\sigma_n^2}{\lambda} f(\lambda)\mathrm{d}\lambda}{\int_{\lambda_{\min}^{\infty}}^{4} f(\lambda)\mathrm{d}\lambda} \\ &= \frac{\mathrm{E_{tx}} + \int_{\lambda_{\min}^{\infty}}^{4} \frac{\sigma_n^2}{\lambda} f(\lambda)\mathrm{d}\lambda}{\int_{\lambda_{\min}^{\infty}}^{4} f(\lambda)\mathrm{d}\lambda}.\end{aligned} \tag{4.26}$$

Evaluating the integrals in (4.26) and substituting $\lambda_{\min}^{\infty} = \frac{\sigma_n^2}{\eta^{\infty}}$ leads to [110]

$$\begin{aligned}&\eta^{\infty}\left[\sqrt{\lambda(4-\lambda)} + 2\arcsin\left(\frac{\lambda-2}{2}\right)\right]\Bigg|_{\frac{\sigma_n^2}{\eta^{\infty}}}^{4} - \\ &\sigma_n^2\left(-2\sqrt{\frac{4-\lambda}{\lambda}} - \arcsin\left(\frac{\lambda-2}{2}\right)\right)\Bigg|_{\frac{\sigma_n^2}{\eta^{\infty}}}^{4} = \\ &2\pi\,\mathrm{E_{tx}}\end{aligned} \tag{4.27}$$

which can be further arranged as a fixed-point equation in η^{∞} and solved for η^{∞}. Furthermore, $\lambda_{\min}^{\infty}$ is obtained using (4.22). This concludes the proof. □

The main difference between WF in the finite and infinite case is that in the finite case there is a different WF level for each channel realization. In contrast, the asymptotic WF level is constant across different channel realizations and only depends on the asymptotic eigenvalue distribution as observed in (4.26) and (4.27). Applying (4.12) again allows us to state the following Theorem.

Theorem 4.1. *As $d \to \infty$, the achievable WF rate per stream converges to the large system WF rate per stream $R_{\mathrm{LS,WF}}$:*

$$\frac{R_{\mathrm{WF}}}{d} \underset{d\to\infty}{\longrightarrow} R_{\mathrm{LS,WF}}, \tag{4.28}$$

where

$$\begin{aligned} R_{\mathrm{LS,WF}} = & \\ & \frac{2}{\pi \ln(2)} \left[\left(\ln(\sigma) - \frac{1}{2} \right) \left(\frac{\sigma}{2} \sqrt{4 - \sigma^2} + 2 \arcsin\left(\frac{\sigma}{2} \right) \right) \right] \Bigg|_{\sigma_{\min}}^{2} \\ & + \log\left(\frac{\eta^{\infty}}{\sigma_n^2} \right) \frac{1}{\pi} \left[\frac{\sigma}{2} \sqrt{4 - \sigma^2} + 2 \arcsin\left(\frac{\sigma}{2} \right) \right] \Bigg|_{\sigma_{\min}}^{2} \\ & - \frac{2}{\pi \ln(2)} A(\sigma) \Big|_{\sigma_{\min}}^{2}, \end{aligned} \tag{4.29}$$

$\sigma_{\min} = \sqrt{\lambda_{\min}^{\infty}}$, *and*

$$A(\sigma) = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n (2n+1)^2} \left(\frac{\sigma}{2} \right)^{2n} \sigma \tag{4.30}$$

is an infinite series which converges for $0 \leq \sigma \leq 2$.

Proof. Looking at (4.12) and (4.18), the stream rate function in the WF case reads

$$g(\lambda) = \log\left(1 + \frac{1}{\sigma_n^2} p\, \lambda \right). \tag{4.31}$$

As p_l is implicitly a function of λ_l, the index l of p_l is dropped. $R_{\mathrm{LS,WF}}$ can be then written as:

$$\begin{aligned} R_{\mathrm{LS,WF}} &= \int_{\lambda=0}^{4} \log\left(1 + \frac{1}{\sigma_n^2} p\, \lambda \right) f(\lambda)\, \mathrm{d}\lambda \\ &= \int_{\lambda=0}^{4} \log\left(\max\left(1, \frac{\eta^{\infty}}{\sigma_n^2} \lambda \right) \right) f(\lambda)\, \mathrm{d}\lambda \\ &= \int_{\lambda=0}^{\lambda_{\min}^{\infty}} \log(1) f(\lambda)\, \mathrm{d}\lambda + \int_{\lambda_{\min}^{\infty}}^{4} \log\left(\frac{\eta^{\infty}}{\sigma_n^2} \lambda \right) f(\lambda)\, \mathrm{d}\lambda \\ &= \int_{\lambda_{\min}^{\infty}}^{4} \log\left(\frac{\eta^{\infty}}{\sigma_n^2} \lambda \right) \frac{\sqrt{\lambda(4-\lambda)}}{2\pi\lambda}\, \mathrm{d}\lambda. \end{aligned} \tag{4.32}$$

We evaluate this integral using the quarter circle distribution of the singular values instead of the Marčenko-Pastur eigenvalue distribution as it results in tractable expressions. As

$$\lambda\left(\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A} \right) = \sigma^2\left(\boldsymbol{A} \right)$$

for any matrix $\boldsymbol{A}$, using the change of variable $\sigma^2 = \lambda$ in (4.32) results in

$$R_{\mathrm{LS,WF}} = \frac{1}{\pi} \int_{\sigma_{\min}}^{2} \log\left(\frac{\eta^{\infty}}{\sigma_n^2}\,\sigma^2\right) \sqrt{4-\sigma^2}\,\mathrm{d}\sigma \tag{4.33}$$

which is the asymptotic WF rate expressed in terms of σ^2 ($\boldsymbol{H}_{kk,\mathrm{eff}}$). In addition, $\sigma_{\min} = \sqrt{\lambda_{\min}^{\infty}}$ and

$$f_{\mathrm{QC}}(\sigma) = \frac{1}{\pi}\sqrt{4-\sigma^2}$$

is the quarter-circle distribution of the singular values of the square matrix $\boldsymbol{H}_{kk,\mathrm{eff}}$ (cf. Section 4.1.3). The integral in (4.33) can be expressed as

$$\begin{aligned} R_{\mathrm{LS,WF}} = \log\left(\frac{\eta^{\infty}}{\sigma_n^2}\right) \frac{1}{\pi} \int_{\sigma_{\min}}^{2} \sqrt{4-\sigma^2}\,\mathrm{d}\sigma \\ + \frac{2}{\pi} \int_{\sigma_{\min}}^{2} \log(\sigma)\sqrt{4-\sigma^2}\,\mathrm{d}\sigma. \end{aligned} \tag{4.34}$$

Evaluating the first operand yields [110]

$$\log\left(\frac{\eta^{\infty}}{\sigma_n^2}\right) \frac{1}{\pi} \left[\frac{\sigma}{2}\sqrt{4-\sigma^2} + 2\arcsin\left(\frac{\sigma}{2}\right)\right]\Bigg|_{\sigma_{\min}}^{2}, \tag{4.35}$$

and the second operand is integrated by parts to give

$$\begin{aligned} &\frac{2}{\pi\ln(2)} \left[\left(\ln(\sigma) - \frac{1}{2}\right)\left(\frac{\sigma}{2}\sqrt{4-\sigma^2} + 2\arcsin\left(\frac{\sigma}{2}\right)\right)\right]\Bigg|_{\sigma_{\min}}^{2} \\ &- \frac{2}{\pi\ln(2)} \int_{\sigma_{\min}}^{2} \frac{2}{\sigma}\arcsin\left(\frac{\sigma}{2}\right)\mathrm{d}\sigma. \end{aligned} \tag{4.36}$$

The last integral doesn't have a solution in closed form. To solve this problem, we use the Taylor/Maclaurin series expansion of $\arcsin\left(\frac{\sigma}{2}\right)$ [110]:

$$\begin{aligned} \arcsin\left(\frac{\sigma}{2}\right) &= \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n\,(2n+1)} \left(\frac{\sigma}{2}\right)^{2n+1} \\ &= \sum_{n=0}^{\infty} b_n(\sigma), \end{aligned} \tag{4.37}$$

where $b_n(\sigma)$ is the implicitly defined n-th term of the infinite series. This representation of the arcsin($\bullet$) function is valid, i.e., the infinite series converges,

Table 4.1.: Finding the Large System Water-Filling Rate

1) Calculate the asymptotic WF level η^{∞}, the solution of a fixed-point equation [cf. (4.27)].

2) Calculate the asymptotic smallest served eigenvalue $\lambda_{\min}^{\infty}$ [cf. (4.22)].

3) Calculate $\sigma_{\min} = \sqrt{\lambda_{\min}^{\infty}}$.

4) Calculate the large system WF rate per stream [cf. (4.29)].

when $|\sigma/2| \leq 1$. As $0 \leq \sigma < 2$ [cf. (4.4)], the convergence condition is always satisfied. Then,

$$\frac{2}{\sigma} \arcsin\left(\frac{\sigma}{2}\right) = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n\,(2n+1)} \left(\frac{\sigma}{2}\right)^{2n}, \quad \sigma \neq 0. \tag{4.38}$$

Therefore,

$$\begin{aligned} &\int_{\sigma_{\min}}^{2} \frac{2}{\sigma} \arcsin\left(\frac{\sigma}{2}\right) \mathrm{d}\sigma = \\ &\int_{\sigma_{\min}}^{2} \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n\,(2n+1)} \left(\frac{\sigma}{2}\right)^{2n} \mathrm{d}\sigma = \\ &\sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n\,(2n+1)^2} \left(\frac{\sigma}{2}\right)^{2n} \sigma \Big|_{\sigma_{\min}}^{2} = A(2) - A(\sigma_{\min}). \end{aligned} \tag{4.39}$$

Summing (4.35), (4.36) and (4.39) gives the desired result in (4.29). $R_{\mathrm{LS,WF}}$ is given in a quasi closed-form expression due to the dependency on η^{∞} which has to be obtained numerically via a fixed-point equation. $A(2)$ and $A(\sigma_{\min})$ have to calculated numerically as well. □

What yet remains is proving that the infinite series $A(\sigma)$ is convergent on the interval $0 \leq \sigma \leq 2$. The proof is presented in Appendix C.3. A summary for finding the large system WF rate per stream is presented in Table 4.1.

The large system expressions usually provide accurate approximations of the ergodic (average) rates for finite system parameters (this will be checked in the simulations section). Therefore, the average rates achieved with equal power allocation and WF can be approximated as:

$$\begin{aligned} E_{\boldsymbol{H}_{kk}}[R_{\mathrm{EP}}] &\approx d\, R_{\mathrm{LS,EP}} \\ E_{\boldsymbol{H}_{kk}}[R_{\mathrm{WF}}] &\approx d\, R_{\mathrm{LS,WF}} \end{aligned} \tag{4.40}$$

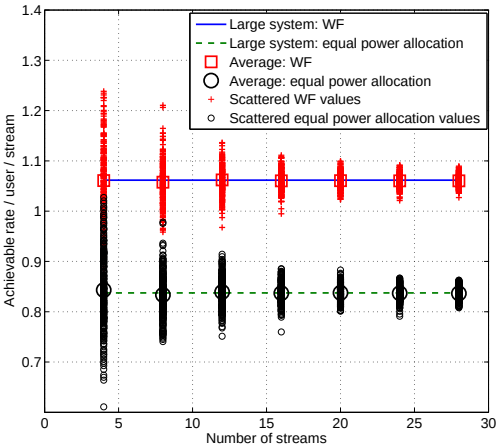

Figure 4.1.: IA performance: Scenario $(N, N, N/2)^3$ with $\mathrm{SNR}_{\mathrm{IA}} = 0$ dB

for finite system parameters.

4.4. Simulation Results and Discussion

Simulation results are averaged over 250 i.i.d. channel realizations, where the individual entries of all links are drawn from a $\mathcal{N}_{\mathbb{C}}(0, 1)$ distribution. Throughout the simulations, we set $\mathrm{E}_{\mathrm{tx},k} = \mathrm{E}_{\mathrm{tx}} = 1\ \forall k$. The fixed-point iteration in (4.27) is allowed to run for up to 1000 iterations. First, the fully loaded scenario $(N, N, N/2)^3$ is considered. In this case, closed-form solutions of precoders and receive filters can be obtained according to the method described in Section 3.2.2. Figure 4.1 shows the derived large system rates, the average of the simulated rates, in addition to the 250 scattered individual rate values around their average for an increasing number of streams with $\mathrm{SNR}_{\mathrm{IA}} = 0$ dB, where $\mathrm{SNR}_{\mathrm{IA}}$ is given by (4.17). Figure 4.2 shows the same results with $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB. First, it can be seen that the derived large system rates provide accurate approximations of the average achievable rates, even for a small number of streams $d = 4$ (which corresponds to $M = N = 8$). This is in line with numerous previous works on LSA, which show via numerical results that LSA expressions provide accurate approximations of the average rates for finite system parameters. Second, the variance of the simulated rate values decreases as the number

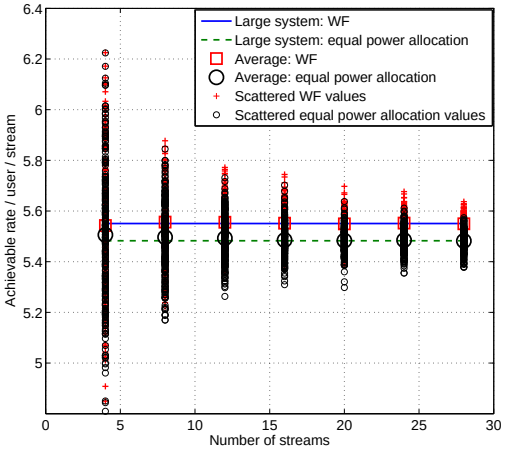

Figure 4.2.: IA performance: Scenario $(N, N, N/2)^3$ with $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB

of streams increases (this can be observed by the cloud of scattered rate values around the large system rates getting smaller and denser); correspondingly, with an increasing number of streams, the large system rates provide increasingly better estimates of the achievable rates of any channel realization and not only of the average. This matches the theoretical results in (4.14) and (4.28). As expected, WF provides a noticeable performance improvement over equal power allocation when $\mathrm{SNR}_{\mathrm{IA}} = 0$ dB and a minor improvement when $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB.

Figures 4.3 and 4.4 correspond to the scenarios $(4, 4, 2)^3$ and $(8, 8, 4)^3$, where the average rates and the large system rates are plotted against the SNR. Figure 4.3 shows that the large system expressions provide good approximations of the average achievable rates even with a very small parameter $d = 2$, especially at low and medium SNR. Figure 4.4 with $d = 4$ confirms this as well for all SNR ranges and it is observed that all curves converge at high SNR.

Next, the fully loaded scenario $(3N/2, N, N/2)^4$ is considered. In this case, no closed-form solutions exist and iterative algorithms are required to find the precoders and receive filters. The IL algorithm from Section 3.2.3 is run. Figures 4.5 and 4.6 show the large system rates, the average of the simulated rates, and the scattered individual rate values with $\mathrm{SNR}_{\mathrm{IA}} = 0$ dB and $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB, respectively. Similar observations to the ones made in Figures 4.1 and 4.2

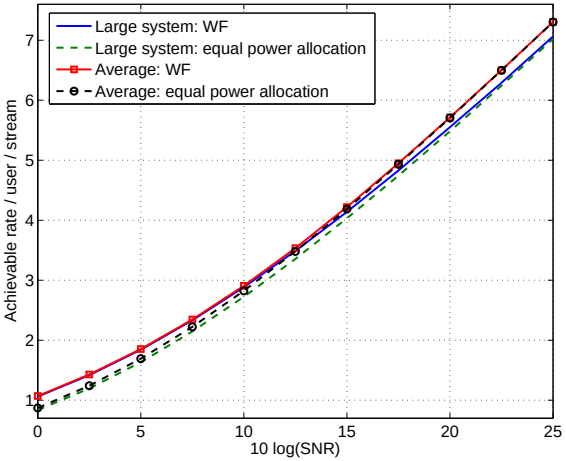

Figure 4.3.: IA performance: Scenario $(4, 4, 2)^3$

can be made. The large system rates provide an accurate approximation of the average rates for a finite number of streams. Moreover, as the number of streams increases, the large system rates provide increasingly better estimates of the achievable rates of any channel realization. In addition, the simulated rates of the two different algorithms, i.e., the closed-form solution and IL, converge to the large system rates with an increasing number of streams. This asserts the hypothesis posed in the beginning of this chapter; namely, that different IA variants belong to the same class and that the obtained large system expressions are valid for different IA variants. Finally, for the same SNR, the simulated rates per stream are almost identical for two different scenarios having different system parameters β and K (compare Figure 4.1 and Figure 4.5, or Figure 4.2 and Figure 4.6). This shows that the achievable rates per stream are independent of the system parameters, as concluded from the analytical rate expressions.

4.5. Summary

In this chapter, large system rate expressions of IA were derived for both equal power allocation and the optimal WF power allocation. The key for deriving the rate expressions was finding the asymptotic eigenvalue distribution of the

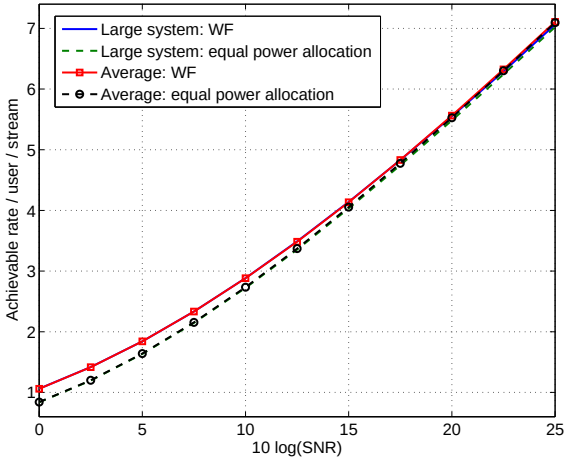

Figure 4.4.: IA performance: Scenario $(8, 8, 4)^3$

resulting channel gains using RMT tools in order to calculate the large system rate per stream in each case. The derived expressions were found to be functions of the transmit power and noise power and are independent of any other system parameters. Simulation results showed that the achievable rates of different IA algorithms converge to the large system rates as the number of transmit and receive antennas increases, thereby showing that the large system expressions are valid for different IA variants. Moreover, simulation results showed that large system expressions provide accurate estimates of the average achievable rates for small and finite system parameters. The asymptotic eigenvalue distribution—also accurate for finite system parameters—is a useful result by itself as it can be used in other applications of IA.

Even though we assumed the channel entries to be Gaussian distributed, such an assumption didn't play any role in the derivations as the Marčenko-Pastur distribution only depends on the mean and variance of the channel entries. Consequently, the derived large system expressions are valid for channel entries which are non-Gaussian distributed as well, given their means and variances equal 0 and 1, respectively. We also note that the derived expressions do not hold in the case where a further precoder optimization is performed after obtaining the IA solution to improve the achievable rates as done in, e.g., [111].

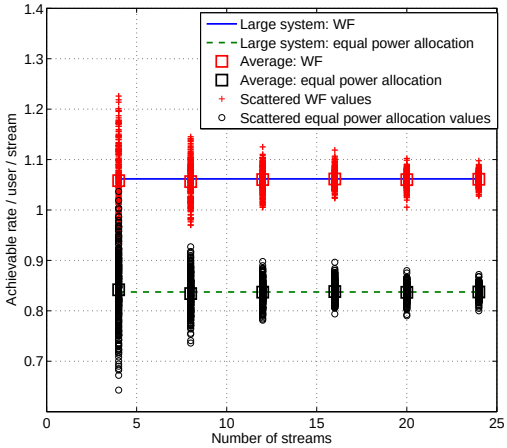

Figure 4.5.: IA performance: Scenario $(3N/2, N, N/2)^4$ with $\mathrm{SNR}_{\mathrm{IA}} = 0$ dB

We focused on IA as it results in an attractive and tractable interference-free model. The large system characterization of other cooperative methods, e.g., the ones reviewed in Section 3.3, is not a straightforward task. This is due to the fact that in these methods, the precoder/receive filter set of user k, $\{\boldsymbol{F}_k, \boldsymbol{G}_k\}$, depends on the direct channel $\boldsymbol{H}_{kk}$. This creates an effective channel $\boldsymbol{G}_k^{\mathrm{H}} \boldsymbol{H}_{kk} \boldsymbol{F}_k$ whose statistics are harder to evaluate due to the dependency between the different terms. In addition, such methods result in non-zero residual interference at the receivers in general. Characterizing the statistics of the remaining interference is a non-trivial task when the precoders and/or receive filters are not unitary, which is the case for many of the methods of Section 3.3.

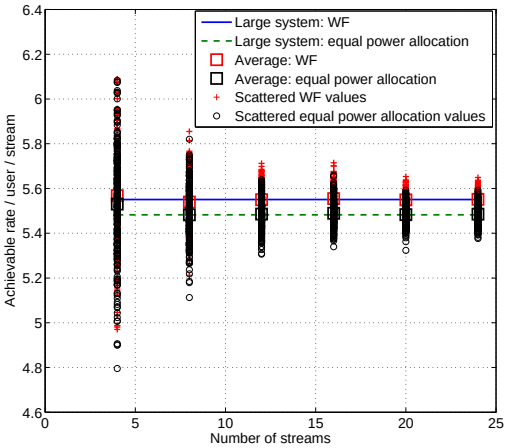

Figure 4.6.: IA performance: Scenario $(3N/2, N, N/2)^4$ with $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB

5. Performance of Non-Cooperative Methods

In this chapter, we characterize the performance of two non-cooperative methods in the MIMO IC. These are the eigenmode precoding and MRT which were discussed in Section 3.4.2. These methods result in non-zero interference at the receiver side, which makes obtaining analytical expressions harder than in the previous chapter. Nonetheless, due to the properties of each precoding type this interference can be characterized and lower bounds of achievable rates based on separate stream decoding can be obtained. Similar to IA, the performance of eigenmode precoding will be studied asymptotically. For MRT, the analysis is non-asymptotic, i.e., the expressions hold for arbitrary system parameters.

We focus on the symmetric MIMO IC where $M \geq N$. Furthermore, we consider full spatial multiplexing with each transmitter transmitting $d = N$ symbols to its corresponding receiver. $\boldsymbol{F}_k \in \mathbb{C}^{M \times N}$ and $\boldsymbol{G}_k \in \mathbb{C}^{N \times N}$ are chosen based on the direct link between the kth transmitter/receiver pair only. This models single-cell precoding in a cellular system.

5.1. Large System Performance of Eigenmode Precoding

The eigenmode precoder is the capacity achieving precoder in a P2P scenario, but is suboptimal in the MIMO IC as it causes interference to other users. Defining

$$\boldsymbol{H}_{kk} = \boldsymbol{U}_k \boldsymbol{\Phi}_k \boldsymbol{V}_k^{\mathrm{H}}$$

to be the singular value decomposition (SVD) of $\boldsymbol{H}_{kk}$, precoder $\boldsymbol{F}_k$ is given by (cf. Appendix B)

$$\boldsymbol{F}_k = \boldsymbol{V}_k^{1:N} \boldsymbol{P}_k^{1/2} \tag{5.1}$$

where $\boldsymbol{P}_k = \mathrm{diag}(p_{k,1}, \ldots, p_{k,N})$. The entries of $\boldsymbol{P}_k$ satisfy

$$\sum_{t=1}^{N} p_{k,t} = E_{\mathrm{tx},k}$$

and are either given by the optimal WF rule or are equally allocated over the N streams. We first focus on the latter case and later discuss the former one. The receive filter $\boldsymbol{G}_k$ is given by

$$\boldsymbol{G}_k = \boldsymbol{U}_k. \tag{5.2}$$

Under the assumption that the entries of $\boldsymbol{H}_{kk}$ are i.i.d., $\boldsymbol{H}_{kk}$ and $\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}$ have rank N with probability 1. Thus, the diagonal elements of

$$\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathrm{T}} = \mathrm{diag}\left(\lambda_1\left(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}\right),\ldots,\lambda_N\left(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}\right)\right) \tag{5.3}$$

are non-zero.

With equal power allocation, (5.1) reads

$$\boldsymbol{F}_k = \sqrt{\frac{\mathrm{E}_{\mathrm{tx},k}}{N}}\,\boldsymbol{V}_k^{1:N}. \tag{5.4}$$

Using (5.2) and (5.4), the obtained symbol vector $\hat{\boldsymbol{s}}_k$ at receiver k reads

$$\hat{\boldsymbol{s}}_k = \sqrt{\frac{\mathrm{E}_{\mathrm{tx},k}}{N}}\,\boldsymbol{\Phi}_k^{1:N}\boldsymbol{s}_k + \sum_{l=1,l\neq k}^{K}\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\boldsymbol{s}_l + \boldsymbol{n}_k \tag{5.5}$$

where we used the fact that $\boldsymbol{\Phi}_k\boldsymbol{V}_k^{\mathrm{H}}\boldsymbol{V}_k^{1:N} = \boldsymbol{\Phi}_k^{1:N}$. Noting that

$$\boldsymbol{\Phi}_k^{1:N}\left(\boldsymbol{\Phi}_k^{1:N}\right)^{\mathrm{T}} = \boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathrm{T}}, \tag{5.6}$$

the achievable rate with interference treated as noise reads

$$\begin{aligned} R_k &= \log\left|\mathbf{I}_N + \frac{\mathrm{E}_{\mathrm{tx},k}}{N}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathrm{T}}\left(\sum_{l\neq k}\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\left(\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{F}_l\right)^{\mathrm{H}} + \sigma_n^2\mathbf{I}_N\right)^{-1}\right| \\ &= \log\left|\mathbf{I}_N + \frac{\mathrm{E}_{\mathrm{tx},k}}{N}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathrm{T}}\left(\sum_{l\neq k}\frac{\mathrm{E}_{\mathrm{tx},l}}{N}\,\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\left(\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\right)^{\mathrm{H}} + \sigma_n^2\mathbf{I}_N\right)^{-1}\right|. \end{aligned} \tag{5.7}$$

5.1.1. Achievable Lower Bounds

Eq. (5.7) is hard to evaluate analytically due to the interference caused by transmitters $l \neq k$. Therefore, we first present a lower bound of (5.7) obtained by assuming receivers perform separate stream decoding. With separate decoding, the off-diagonal entries of the interference covariance matrix

$$\boldsymbol{H}_{\mathrm{int},k} = \sum_{l\neq k}\frac{\mathrm{E}_{\mathrm{tx},l}}{N}\,\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\left(\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\right)^{\mathrm{H}} \tag{5.8}$$

are not taken into account, which is suboptimal unless the off-diagonal entries of $\boldsymbol{H}_{\mathrm{int},k}$ equal 0. The separate decoding assumption is merely used here to obtain tractable analytical expressions. As Lemma 5.1 shows, the bound is an ergodic one as the statistics of the interfering links are used to replace the K-user MIMO IC with K proxy P2P MIMO channels.

Lemma 5.1.

$$\mathrm{E}_{\{\boldsymbol{H}_{kl}\}_{\forall l\neq k}}\,[R_k|\boldsymbol{H}_{kk}] \geq \sum_{t=1}^{N} \log\left(1+\frac{\mathrm{E}_{\mathrm{tx},k}}{N}\frac{\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})}{\sum_{l\neq k}\mathrm{E}_{\mathrm{tx},l}+\sigma_n^2}\right). \tag{5.9}$$

Proof. With separate decoding and using (5.3) and (5.6), the decoded tth stream equals [cf. (5.5)]:

$$\begin{aligned}\hat{s}_{k,t} &= \boldsymbol{e}_t^{\mathrm{T}}\hat{\boldsymbol{s}}_k \\ &= \boldsymbol{e}_t^{\mathrm{T}}\sqrt{\frac{\mathrm{E}_{\mathrm{tx},k}}{N}}\boldsymbol{\Phi}_k^{1:N}\boldsymbol{s}_k + \boldsymbol{e}_t^{\mathrm{T}}\left(\sum_{l\neq k}\sqrt{\frac{\mathrm{E}_{\mathrm{tx},l}}{N}}\,\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\boldsymbol{s}_l+\boldsymbol{n}_k\right) \\ &= \sqrt{\frac{\mathrm{E}_{\mathrm{tx},k}}{N}}\sqrt{\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})}[\boldsymbol{s}_k]_t + \boldsymbol{e}_t^{\mathrm{T}}\left(\sum_{l\neq k}\sqrt{\frac{\mathrm{E}_{\mathrm{tx},l}}{N}}\,\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\boldsymbol{s}_l+\boldsymbol{n}_k\right)\end{aligned} \tag{5.10}$$

$t=1,\ldots,N$. The achievable rate $\hat{R}_k \leq R_k$ obtained with separate decoding reads

$$\hat{R}_k = \sum_{t=1}^{N}\hat{R}_{k,t} \tag{5.11}$$

where

$$\hat{R}_{k,t} = \log\left(1+\frac{(\mathrm{E}_{\mathrm{tx},k}/N)\,\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})}{\sum_{l\neq k}(\mathrm{E}_{\mathrm{tx},l}/N)\left|\left|\boldsymbol{e}_t^{\mathrm{T}}\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\right|\right|_2^2+\sigma_n^2}\right) \tag{5.12}$$

is the rate of the tth stream.

Remember that $\boldsymbol{H}_{kl}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries. With eigenmode precoding, $\boldsymbol{V}_l^{1:N}$ is a unitary matrix that is chosen based on $\boldsymbol{H}_{ll}$, and is therefore independent of $\boldsymbol{H}_{kl}$. A similar argument holds for $\boldsymbol{G}_k$ that is chosen based on $\boldsymbol{H}_{kk}$. Using the fact that left and right multiplying a matrix with i.i.d. entries by independent matrices with orthonormal columns results in a new matrix with i.i.d. entries having the same variance as those of the initial matrix, we conclude that $\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N} \in \mathbb{C}^{N\times N}$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries [21]. Therefore, the following holds by definition:

$$\mathrm{E}\left[\left|\left|\boldsymbol{e}_t^{\mathrm{T}}\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\right|\right|_2^2\right] = N, \quad t=1,\ldots,N \text{ and } l\neq k. \tag{5.13}$$

Finally, we note that with $\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})$ fixed, $\hat{R}_{k,t}(\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}))$ is a convex function in

$$u = \sum_{l\neq k}\frac{\mathrm{E}_{\mathrm{tx},l}}{N}\left|\left|\boldsymbol{e}_t^{\mathrm{T}}\boldsymbol{G}_k^{\mathrm{H}}\boldsymbol{H}_{kl}\boldsymbol{V}_l^{1:N}\right|\right|_2^2+\sigma_n^2. \tag{5.14}$$

Consequently, using Jensen's inequality and performing an expectation over the set of independent variables $\{\boldsymbol{H}_{kl}\}_{\forall l \neq k}$ yields [112]

$$\begin{aligned} E[\hat{R}_{k,t}|\boldsymbol{H}_{kk}] &\geq \log\left(1 + \frac{\mathrm{E}_{\mathrm{tx},k}}{N} \frac{\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})}{\mathrm{E}[u]}\right) \\ &= \log\left(1 + \frac{\mathrm{E}_{\mathrm{tx},k}}{N} \frac{\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})}{\sum_{l\neq k}\mathrm{E}_{\mathrm{tx},l} + \sigma_n^2}\right). \end{aligned} \tag{5.15}$$

Using (5.15), (5.11), and $R_k \geq \hat{R}_k$ gives the result of Lemma 5.1. □

The proof and result of Lemma 5.1 show that the interference power present at a given receiver is bounded by the transmit power coming from interfering transmitters, but is independent of the number of antennas of interfering transmitters. Consequently, increasing the number of antennas at all transmitters simultaneously increases the array gain and achievable rates for all receivers without causing any additional interference, as elaborated next.

5.1.2. Large System Analysis

In this section, we perform an LSA to obtain closed-form expressions of the obtained bound. Similarly to Section 4.2 and to facilitate the LSA, we rewrite $\lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk})$ as:

$$\begin{aligned} \lambda_t(\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{H}_{kk}) &= \lambda_t\left[\left(\frac{1}{\sqrt{N}}\boldsymbol{H}_{kk}^{\mathrm{H}}\right)\left(\frac{1}{\sqrt{N}}\boldsymbol{H}_{kk}\right)N\right] \\ &= N\lambda_t\left(\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}\right) \end{aligned} \tag{5.16}$$

where

$$\tilde{\boldsymbol{H}}_{kk} = \frac{1}{\sqrt{N}}\boldsymbol{H}_{kk} \tag{5.17}$$

has i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1/N)$ entries. Using (5.16), we rewrite the bound in Lemma 5.1 as

$$R_{\mathrm{LB},k} = \sum_{t=1}^{N} \log\left(1 + c_k\,\lambda_t\left(\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}\right)\right) \tag{5.18}$$

where

$$c_k = \frac{\mathrm{E}_{\mathrm{tx},k}}{\sum_{l\neq k}\mathrm{E}_{\mathrm{tx},l} + \sigma_n^2}. \tag{5.19}$$

Similarly to the previous chapter and for our scenario of interest, we fix K and assume that

$$M \to \infty, N \to \infty \;\text{ with }\; \beta = \frac{M}{N} \geq 1. \tag{5.20}$$

Next, we present the asymptotic eigenvalue distribution of $\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}$.

Direct Channels' Asymptotic Eigenvalue Distribution

With the conditions in (5.20) and $\tilde{\boldsymbol{H}}_{kk}$ having i.i.d. entries of variance $1/N$, the distribution of the eigenvalues of $\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}$ converges almost surely to a Marčenko-Pastur distribution with density

$$f_{\mathrm{MP}}(\lambda) = \left(1 - \frac{1}{\beta}\right)\delta(\lambda) + \frac{\sqrt{(\lambda - a)^+(b - \lambda)^+}}{2\pi\beta\lambda} \tag{5.21}$$

where $a = (1 - \sqrt{\beta})^2$, $b = (1 + \sqrt{\beta})^2$, and $y^+ = \max(0, y)$.

The result in (5.21) is a direct application of Section 4.1.1 to the scaled channel matrices.

Large System Bounds

Recall (4.12), which we apply to (5.18). Note that as rank $(\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{H}}_{kk}) = N$ with probability 1, there exists $M - N$ zero eigenvalues that constitute part of the distribution in (5.21) and have to be accounted for. Therefore, the summation in (4.12) needs to be taken over M and not over N. Then:

$$\begin{aligned} &\frac{1}{M}\sum_{t=1}^{M}\log(1 + c_k\lambda_t) = \\ &\frac{1}{M}\sum_{t=1}^{N}\log(1 + c_k\lambda_t) \underset{\substack{N,M\to\infty \\ \beta = M/N}}{\longrightarrow} \mathrm{E}_\lambda\left[\log(1 + c_k\lambda)\right]. \end{aligned} \tag{5.22}$$

The right-hand side of (5.22) is the Shannon transform $\mathcal{V}(c_k, \beta)$ of λ. As λ follows a Marčenko-Pastur distribution, its Shannon transform can be obtained in closed-form as

$$\begin{aligned} \mathcal{V}(c_k, \beta) = &\log\left(1 + c_k - \frac{1}{4}\mathcal{F}(c_k, \beta)\right) + \\ &\frac{1}{\beta}\log\left(1 + c_k\beta - \frac{1}{4}\mathcal{F}(c_k, \beta)\right) - \frac{\log(e)}{4c_k\beta}\mathcal{F}(c_k, \beta) \end{aligned} \tag{5.23}$$

as explained in Section 4.1.2.

It is usually desired to define an achievable rate per stream, or in our case a lower bound of the achievable rate per stream. Multiplying (5.22) by β leads to the following theorem.

Theorem 5.1. *As $N, M \to \infty$ at a fixed ratio $\beta = M/N$, the lower bound of the achievable rate per stream of the kth user using eigenmode precoding in a*

MIMO IC converges under equal power allocation as follows:

$$\frac{1}{N} R_{\mathrm{LB},k} = \frac{1}{N} \sum_{t=1}^{N} \log(1 + c_k \lambda_t) \underset{\substack{N,M\to\infty \\ \beta = M/N}}{\longrightarrow} \beta\, \mathcal{V}(c_k, \beta) \tag{5.24}$$

where c_k and $\mathcal{V}(c_k, \beta)$ are given by (5.19) *and* (4.3)*, respectively.*

Theorem 5.1 shows that asymptotically, the bound of any channel realization [left-hand side of (5.24)] converges to a quantity that only depends on the parameters β and c_k, and is independent of the instantaneous eigenvalues [right-hand side of (5.24)].

The large system expressions usually provide accurate approximations of the ergodic rates for finite system parameters (this will be checked in the numerical results section). Therefore, Therorem 5.1 can be written as:

$$\mathrm{E}_{\boldsymbol{H}_{kk}}[R_{\mathrm{LB},k}] \approx M\, \mathcal{V}(c_k, \beta), \tag{5.25}$$

and consequently

$$\mathrm{E}_{\{\boldsymbol{H}_{km}\}_{\forall m}}[R_k | \boldsymbol{H}_{kk}] \geq M\, \mathcal{V}(c_k, \beta) \tag{5.26}$$

for finite M and N. The last inequality holds because under the assumption that the channel entries are i.i.d., the channel matrices are independent; thus, the expectation over the matrices $\{\boldsymbol{H}_{km}\}_{\forall m}$ can be performed in any random order. An expectation over the interfering links first gives the bound in Lemma 5.1, then an expectation over the direct link $\boldsymbol{H}_{kk}$ gives the large system approximation of the obtained bound and consequently (5.26).

5.1.3. Comparison To Interference Alignment Large System Properties

It is useful at this point to compare the large system properties of eigenmode precoding to the large system properties of IA. The first observation is the asymptotic eigenvalue distribution. IA algorithms consist of rank reducing precoders *and* receive filters which lead to square effective channel matrices of size $d \times d$. The result is that the parameter β doesn't appear in the asymptotic eigenvalue distribution of the direct channels in (4.11). In contrast, in the case of eigenmode precoding, the asymptotic eigenvalue distribution of the effective channel gains in (5.21) depends on β. Similarly, the large system lower bound expressions are a function of β in contrast to the IA obtained expressions in (4.16) and (4.29) which are independent of β.

The second observation follows closely from the first one. So far, the analysis in this chapter assumed equal power allocation. Unlike the previous chapter, obtaining quasi closed-form expressions in the case of WF power allocation is

only possible in the case $\beta = 1$. Recall that the rate function in the case of IA depends on the terms

$$\lambda_1 \left(\boldsymbol{H}_{kk,\mathrm{eff}}^{\mathrm{H}} \boldsymbol{H}_{kk,\mathrm{eff}} \right), \ldots, \lambda_d \left(\boldsymbol{H}_{kk,\mathrm{eff}}^{\mathrm{H}} \boldsymbol{H}_{kk,\mathrm{eff}} \right)$$

[cf. (4.8)]. Evaluating the rate function using the Marčenko-Pastur distribution is intractable. However, this is circumvented by noting that IA methods result in a square $\boldsymbol{H}_{kk,\mathrm{eff}}$ regardless of β. Consequently, the quarter circle distribution of the singular values of $\boldsymbol{H}_{kk,\mathrm{eff}}$ can be used instead to obtain quasi closed-form expressions. In the case of eigenmode precoding, the rate function depends on the terms

$$\lambda_1 \left(\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk} \right), \ldots, \lambda_N \left(\boldsymbol{H}_{kk}^{\mathrm{H}} \boldsymbol{H}_{kk} \right),$$

and $\boldsymbol{H}_{kk}$ is not square in general. Thus, the quarter circle distribution cannot be used to obtain closed-form expressions. $\boldsymbol{H}_{kk}$ is only square in the case $\beta = 1$. In this case, a WF analysis follows exactly as in Section 4.3.4. Achievable lower bounds with WF can be obtained by replacing σ_n^2 by $\sigma_n^2 + \sum_{l \neq k} \mathrm{E}_{\mathrm{tx},l}$ in the fixed-point equation (4.27) and the WF achievable rate per stream (4.29).

5.2. Performance of Maximum Ratio Transmission

So far, we analyzed the performance of two precoding schemes in this thesis using LSA. Nonetheless, there exists some other cases where performance can be characterized without the need for LSA. MRT with limited CSIR constitutes one of these cases, and is covered next. The material of this section has been published in [113]. In order to keep a unified framework in this chapter, a different proof for the main result in (5.36) than the one in [113] is provided here, and is similar to the one developed in Section 5.1.1.

The MRT precoder is a Hermitian version of $\boldsymbol{H}_{kk}$, which needs to be properly scaled to satisfy the transmit power constraint. Here, we relax the instantaneous transmit power constraint to an average one[1], i.e.:

$$\mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{F}_k^{\mathrm{H}} \boldsymbol{F}_k \right) \right] = \mathrm{E}\left[||\boldsymbol{F}_k||_{\mathrm{F}}^2 \right] = E_{\mathrm{tx},k}. \tag{5.27}$$

With equal power allocated to the N streams, this results in

$$\boldsymbol{F}_k = \sqrt{\frac{E_{\mathrm{tx},k}}{MN}}\, \boldsymbol{H}_{kk}^{\mathrm{H}} \tag{5.28}$$

as $\mathrm{E}\left[||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2 \right] = MN$. In the $M \gg 1$ regime, there is little difference between an instantaneous and an average power constraint. The reason is that $||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2$

[1] Averaging is done over the symbols and over the desired channel as well.

can be expanded as

$$||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2 = \sum_{i=1}^{N}\sum_{j=1}^{M} |[\boldsymbol{H}_{kk}]_{i,j}|^2$$

which is a sum of $MN \gg 1$ variables. As by definition $\{[\boldsymbol{H}_{kk}]_{i,j}\}$ is a set of i.i.d. variables with mean 0 and variance 1, $\{|[\boldsymbol{H}_{kk}]_{i,j}|^2\}$ is a set of i.i.d. variables with mean 1. Therefore, we have:

$$\frac{||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2}{MN} \approx 1 \implies ||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2 \approx \mathrm{E}\left[||\boldsymbol{H}_{kk}||_{\mathrm{F}}^2\right] \tag{5.29}$$

for any channel realization by the LLN. As the instantaneous Frobenius norm of any channel realization is close to its average, there is little difference between an instantaneous and an average power constraint in the $M \gg 1$ regime. As it suffices to normalize the MRT precoders by a quantity depending on the physical parameters instead of computing the norm of each channel realization, this implies a simpler implementation of MRT in practice. It also makes obtaining analytical expressions easier, especially for the $N > 1$ case. The system equation then reads:

$$\begin{aligned}\hat{\boldsymbol{s}}_k &= \sqrt{\frac{E_{\mathrm{tx},k}}{MN}}\,\boldsymbol{H}_{kk}\boldsymbol{H}_{kk}^{\mathrm{H}}\boldsymbol{s}_k + \sum_{l\neq k}\sqrt{\frac{E_{\mathrm{tx},l}}{MN}}\,\boldsymbol{H}_{kl}\boldsymbol{H}_{ll}^{\mathrm{H}}\boldsymbol{s}_l + \boldsymbol{n}_k \\ &= \tilde{\boldsymbol{H}}_{kk}\boldsymbol{s}_k + \sum_{l\neq k}\tilde{\boldsymbol{H}}_{kl}\boldsymbol{s}_l + \boldsymbol{n}_k\end{aligned} \tag{5.30}$$

where we defined the effective channels

$$\tilde{\boldsymbol{H}}_{km} = \sqrt{\frac{E_{\mathrm{tx},m}}{MN}}\,\boldsymbol{H}_{km}\boldsymbol{H}_{mm}^{\mathrm{H}} \in \mathbb{C}^{N\times N}, \quad m = 1,\ldots,K \tag{5.31}$$

seen at receiver k for brevity. Furthermore, we have set $\boldsymbol{G}_k = \mathbf{I}_N$ in order to find a lower bound on the achievable rate as explained next.

5.2.1. Limited CSIR Model

We assume receiver k has no CSI except for the mean of the effective channel gain $\mathrm{E}[\tilde{\boldsymbol{H}}_{kk}]$, which can be estimated on a long-term basis for instance. This facilitates obtaining analytical expressions and has been used previously in, e.g., [54, 55]. Furthermore, this might be desirable in scenarios with very short coherence times in TDD mode, as estimating the CSIR requires downlink pilots sent by the transmitter. In scenarios with very short coherence times, it might be desired to free up the time slots occupied by these downlink pilots and use

them for data transmission instead[2]. The CSI of interfering links are neither known to a given transmitter nor receiver. The received signal at receiver k is decomposed as:

$$\begin{aligned}\hat{\boldsymbol{s}}_k &= \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\right]\boldsymbol{s}_k \\ &\quad + \underbrace{\left(\tilde{\boldsymbol{H}}_{kk} - \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\right]\right)\boldsymbol{s}_k + \sum_{l=1,l\neq k}^{K} \tilde{\boldsymbol{H}}_{kl}\boldsymbol{s}_l + \boldsymbol{n}_k}_{\hat{\boldsymbol{n}}_k}.\end{aligned} \tag{5.32}$$

$\mathrm{E}[\tilde{\boldsymbol{H}}_{kk}]$ is the channel gain estimated by the receiver and $(\tilde{\boldsymbol{H}}_{kk} - \mathrm{E}[\tilde{\boldsymbol{H}}_{kk}])\boldsymbol{s}_k$ is an additional noise term resulting from the mismatch between the instantaneous channel gain and its mean. The CSIR error vector and the interfering symbol vectors are collected into $\hat{\boldsymbol{n}}_k$. Eq. (5.32) represents a MIMO additive noise channel model, where the channel gain consists of statistical information only.

5.2.2. Ergodic Lower Bounds

In order to obtain lower bound expressions, expectations of different matrix products are required. In Appendix C.2, we show that if $\{\boldsymbol{H}, \boldsymbol{A}\} \in \mathbb{C}^{N\times M}$ are two uncorrelated matrices with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, we have:

$$\begin{aligned}\mathrm{E}\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}/\sqrt{M}\right] &= \sqrt{M}\,\mathbf{I}_N \\ \mathrm{E}\left[\left(\frac{1}{\sqrt{M}}\,\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^2\right] &= (M+N)\,\mathbf{I}_N \\ \mathrm{E}\left[\frac{1}{M}\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\right)^{\mathrm{H}}\right] &= N\,\mathbf{I}_N.\end{aligned} \tag{5.33}$$

As the channel matrices are modeled to be uncorrelated with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, using (5.33) with an appropriate scaling yields:

$$\begin{aligned}\mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\right] &= \sqrt{\frac{M\,E_{\mathrm{tx},k}}{N}}\,\mathbf{I}_N \\ \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\tilde{\boldsymbol{H}}_{kk}^{\mathrm{H}}\right] &= \frac{(M+N)\,E_{\mathrm{tx},k}}{N}\,\mathbf{I}_N \\ \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kl}\tilde{\boldsymbol{H}}_{kl}^{\mathrm{H}}\right] &= E_{\mathrm{tx},l}\,\mathbf{I}_N, \quad l \neq k.\end{aligned} \tag{5.34}$$

[2]This discussion is not related to the uplink channel estimation which is needed in any case to obtain the CSI at the transmitter.

As $\tilde{\boldsymbol{H}}_{kk}$ is estimated by its mean which is a scaled identity matrix, inter-stream interference is neglected at the receiver. Using (5.34) and the statistical properties of the data symbols, the covariance matrix of $\hat{\boldsymbol{n}}_k$—averaged over the data symbols and channels statistics—can be calculated as:

$$\begin{aligned}\boldsymbol{C}_{\hat{\boldsymbol{n}}_k} &= \mathrm{E}\left[\left(\tilde{\boldsymbol{H}}_{kk} - \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\right]\right)\left(\tilde{\boldsymbol{H}}_{kk} - \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kk}\right]\right)^{\mathrm{H}}\right] + \sum_{l=1,l\neq k}^{K} \mathrm{E}\left[\tilde{\boldsymbol{H}}_{kl}\tilde{\boldsymbol{H}}_{kl}^{\mathrm{H}}\right] \\ &= E_{\mathrm{tx},k}\,\mathbf{I}_N + \sum_{l=1,l\neq k}^{K} E_{\mathrm{tx},l}\,\mathbf{I}_N = \sum_{m=1}^{K} E_{\mathrm{tx},m}\,\mathbf{I}_N.\end{aligned} \tag{5.35}$$

Using (5.35), we can state the main result of this section.

Theorem 5.2. *The ergodic rate of receiver k achieved with perfect CSIT, limited CSIR, and full multiplexing is lower bounded by:*

$$\overline{R}_{\mathrm{MRT,LB}} = N \log\left(1 + \frac{M\,E_{\mathrm{tx},k}}{N\left(\sum_{i=1}^{K} E_{\mathrm{tx},i} + \sigma_n^2\right)}\right). \tag{5.36}$$

Proof. For the model of (5.32) and with separate stream decoding, the decoded tth stream equals

$$\begin{aligned}\hat{s}_{k,t} &= \boldsymbol{e}_t^{\mathrm{T}}\hat{\boldsymbol{s}}_k \\ &= \sqrt{\frac{M\,E_{\mathrm{tx},k}}{N}}\,[\boldsymbol{s}_k]_t + [\hat{\boldsymbol{n}}_k]_t + [\boldsymbol{n}_k]_t.\end{aligned} \tag{5.37}$$

The achievable rate $\hat{R}_{k,t}$ of the tth stream obtained with separate stream decoding equals

$$\hat{R}_{k,t} = \log\left(1 + \frac{M\,E_{\mathrm{tx},k}}{N}\,\frac{1}{\mathrm{E}_s[|[\hat{\boldsymbol{n}}_k]_t|^2] + \sigma_n^2}\right) \tag{5.38}$$

where $\mathrm{E}_s[\bullet]$ refers to an expectation over the symbols. Similarly to Section 5.1.1, we can use Jensen's inequality in combination with (5.35) to conclude that

$$\mathrm{E}[\hat{R}_{k,t}] \geq \log\left(1 + \frac{M\,E_{\mathrm{tx},k}}{N\left(\sum_{m=1}^{K} E_{\mathrm{tx},m} + \sigma_n^2\right)}\right) \tag{5.39}$$

where the averaging in (5.39) is additionally done over the channels.

Due to the i.i.d. assumption, all streams have the same ergodic lower rate bound and therefore the total achievable ergodic rate under separate stream decoding $E[R_k]$ can be bounded by

$$\begin{aligned} \mathrm{E}[R_k] &= \sum_{t=1}^{N} \mathrm{E}[R_{k,t}] \\ &\geq \sum_{t=1}^{N} \mathrm{E}[\hat{R}_{k,t}] \\ &\geq N \log\left(1 + \frac{M\, E_{\mathrm{tx},k}}{N\left(\sum_{m=1}^{K} E_{\mathrm{tx},m} + \sigma_n^2\right)}\right) \\ &= \overline{R}_{\mathrm{MRT,LB}}. \end{aligned} \tag{5.40}$$

□

Looking at (5.36), it is observed that the denominator of the SINR term—which includes the CSIR error and interference—scales with the number of receive antennas while its numerator scales with the number of transmit antennas. Thus, increasing the number of antennas at all transmitters simultaneously with the use of MRT doesn't increase the interference level at the receivers. Rather, it only increases the numerator and resulting SINR given a fixed transmit power. This also confirms why MRT is considered an attractive option in massive MIMO scenarios.

5.3. Simulation Results and Discussion

We first check the validity of the large system expressions obtained for eigenmode precoding by illustrating Theorem 5.1. To that purpose, consider first a $(2N, N, N)^3$ IC. For 250 channel realizations and $E_{\mathrm{tx}}/\sigma_n^2 = 10$ dB, Fig. 5.1 plots the average of the simulated bound values in Theorem 5.1, the separate scattered bound values around their average, and the large system derived expressions. It can be seen that the derived large system rates provide accurate approximations of the average achievable rates, even for the case $M = 8, N = 4$. This is in line with numerous previous works on LSA, which show via numerical results that LSA expressions provide accurate approximations of the average rates for finite system parameters. Second, the variance of the simulated bound values decreases as the number of streams increases; this can be observed by the cloud of scattered rate values around the large system rates getting smaller and denser. Correspondingly, as the number of streams N increases, the large system expressions provide increasingly better estimates of the achievable rates of any

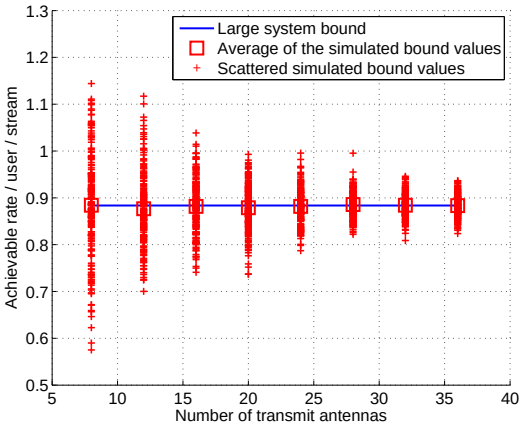

Figure 5.1.: Validating Theorem 5.1: Scenario $(2N, N, N)^3$ at $E_{\text{tx}}/\sigma_n^2 = 10$ dB

channel realization and not only of the average. This matches the theoretical results of Theorem 5.1 which predict that the achievable rate (bound) of any channel realization converges to the large system rate (bound) as $M, N \to \infty$. These conclusions are similar to the ones of Section 4.4.

Having validated Theorem 5.1 holds, we proceed to check whether (5.26) holds. Here, the results are averaged over 1000 channel realizations. Given $N = 4$, $K = 3$, and $E_{\text{tx}}/\sigma_n^2 = 10$ dB, Fig. 5.2 plots the derived bound and the average of the simulated rates with separate and joint decoding against M. The bound holds but there exists a performance loss between separate and joint decoding which is not negligible with 4 receive antennas. A horizontal curve shift can be approximated to correspond to $5-7$ transmit antennas between the derived bound and simulated rates with joint decoding. Nonetheless, increasing the number of interferers will make the obtained bound tighter, as elaborated next. Fig. 5.3 shows the achievable rate curves and derived rate bound for the scenario where the number of interferers is double the one of the scenario of Fig. 5.2, i.e., $K = 5$, with all other parameters kept fixed. This is due to the fact that with more interferers which according to the system model are independent, the interference covariance matrix in (5.8) tends to become white—a diagonal matrix—containing less information that can be exploited when joint decoding is performed. Consequently, the performance gap between separate and joint decoding decreases.

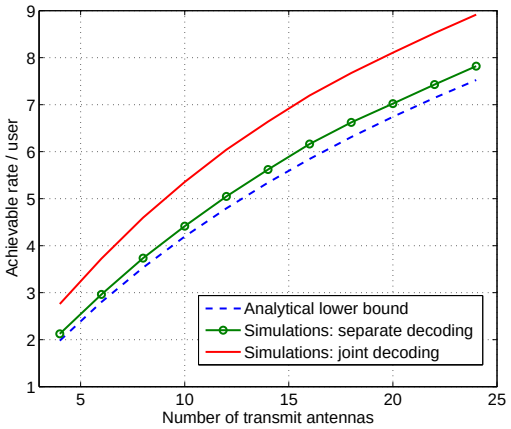

Figure 5.2.: Eigenmode precoding performance: Scenario $(M, 4, 4)^3$ at $E_{\text{tx}}/\sigma_n^2 = 10$ dB

We next check the validity of the derived expressions for MRT, where the results are averaged over 1000 channel realizations. We consider the scenario $N = 2$, $K = 3$ and plot the derived bound and the average of the simulated rates with separate and joint decoding against M in Fig. 5.4. The simulated rates with joint and separate decoding are based on perfect CSIR. In Fig. 5.5, a similar scenario is simulated but the number of interferers is doubled ($K = 5$). The figures show that the derived expression in (5.36) holds. Additionally, comparing both figures leads to similar conclusions to the ones made in the case of eigenmode precoding; namely, that the performance gap between separate and joint decoding decreases as the number of independent interferers increases. As we are mainly interested in $K \geq 3$ scenarios, the expressions derived in this chapter will provide a good basis for comparing eigenmode precoding or MRT to IA in the next chapter.

5.4. Summary

In this chapter, we obtained analytical lower bound rate expressions for eigenmode precoding and MRT assuming separate stream decoding. In the first case and similarly to Chapter 4, the bounds were obtained using RMT tools such as the Marčenko-Pastur distribution and the Shannon transform. In the second

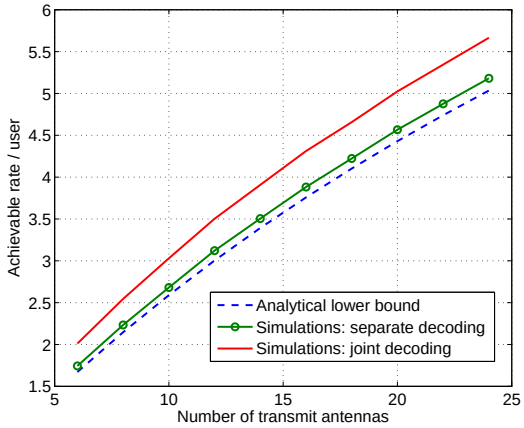

Figure 5.3.: Eigenmode precoding performance: Scenario $(M, 4, 4)^5$ at $E_{\mathrm{tx}}/\sigma_n^2 = 10$ dB

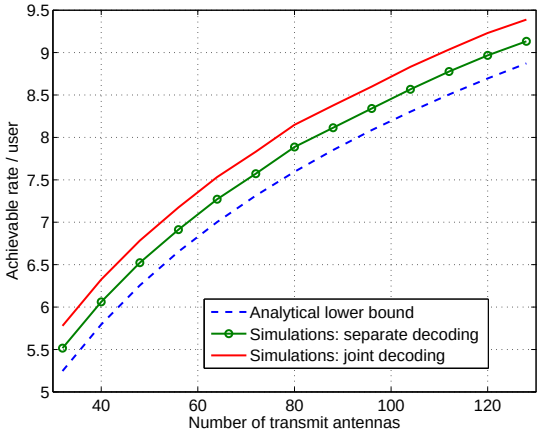

Figure 5.4.: MRT performance: Scenario $(M, 2, 2)^3$ at $E_{\mathrm{tx}}/\sigma_n^2 = 10$ dB

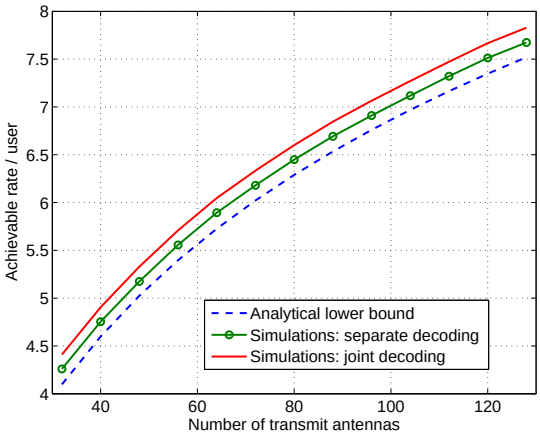

Figure 5.5.: MRT performance: Scenario $(M, 2, 2)^5$ at $E_{\mathrm{tx}}/\sigma_n^2 = 10$ dB

case, we made use of a special model based on a limited CSIR assumption. Along with the fact that precoders which are scaled versions of the users' channels facilitates characterizing the interference power, MRT rate lower bounds were obtained. Naturally, the obtained bounds with limited CSIR served as bounds for the perfect CSIR case. It is observed that the bounds are around 0.5 (resp. 0.3) bits/sec/Hz away from the perfect CSIR case with joint decoding for the case $K = 3$ (resp. $K = 5$) for the given choice of $E_{\mathrm{tx}}/\sigma_n^2$. Alternatively, the limited CSIR assumption can be used to reduce the piloting overheads which need to be included in the spectral efficiency analysis, as will be performed in Chapter 6.

For both precoding types, it was observed that increasing the number of interferers makes the obtained bounds tighter. This is due to the fact that with more interferers, the interference covariance matrix of each receiver tends to become a diagonal matrix containing less information that can be exploited when joint decoding is performed. As our main interest consists in $K \geq 3$ scenarios, the expressions derived in this chapter will provide a good basis for comparing the performance of eigenmode precoding or MRT to the one of IA in Chapter 6.

6. How Much Antennas is "Massive"?

In Chapters 4 and 5, we obtained analytical rate expressions—or lower bounds thereof—of different cooperative and non-cooperative linear precoding methods. Even though these constitute standalone results, our ultimate goal lies in finding how many transmit antennas do non-cooperative methods need to emulate the performance of a noise-limited system, i.e., so that their resulting performance is similar to the one of a system where transmitters have a fixed number of antennas and employ the IA strategy. One could use the obtained expressions in Chapters 4 and 5 to answer that question. However, this does not constitute a fair comparison in real systems, as CSI acquisition overheads vary from method to method. IA and other cooperative methods require additional training overheads compared to non-cooperative methods due to their higher CSI requirements at both the transmitter and receiver side. These overheads take up a possibly large number of time slots, and result in a smaller portion left for data transmission for a given coherence interval. This can have a big effect on the spectral efficiency in scenarios with short coherence times. Therefore, a direct application of the expressions in Chapters 4 and 5 is not valid, except in special cases discussed at the end of the chapter.

The literature dealing with the tradeoff between training and/or feedback overheads on one side and data transmission on the other include works in [63, 114–117], most of which deal with FDD mode. Such works mainly aim at optimizing the training interval length and/or number of feedback bits such that the spectral efficiency is maximized for a given coherence interval length. A P2P unprecoded MIMO link is considered in [63] and the optimal number of downlink training pilots maximizing a lower bound to capacity is found. The analysis is valid for both FDD and TDD modes. The optimal training and feedback resource allocation in FDD mode maximizing spectral efficiencies or lower bounds thereof is treated in [114] for a two-way P2P link MISO link, in [115] for a two-way P2P MIMO link, in [116] for a MIMO BC with ZF precoding, and in [117] for the MIMO IC with IA precoding and analog feedback.

Literature dealing with cooperative and non-cooperative strategies in a single framework include [118], where MRT and the cooperative ZF strategy were studied from a game theoretic point of view for the two-user MISO IC. The authors in [119] show that for the latter system model, any achievable rate pairs lying on the Pareto boundary correspond to linear combinations of ZF and MRT precoders.

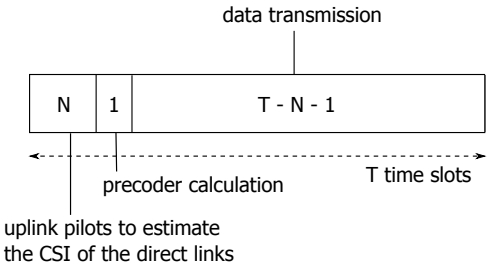

Figure 6.1.: Coherence interval structure of MRT

The framework of this chapter is different from the above mentioned references as the tradeoff between training overheads and spectral efficiency in the context of cooperative and non-cooperative strategies will be given by a required number of transmit antennas. We focus on square symmetric ICs where the IA solution can be found in closed-form as explained in Section 3.2.2 for the following reasons. Scenarios which require iterative alternating minimization make it hard to estimate the number of time slots required for the precoders/receive filters calculation. Furthermore, they might not be feasible in practical systems as by the time a solution is found, the users' channels might have changed and the obtained solutions would be outdated. Thus, the resulting performance in these scenarios would be worse than in scenarios where the IA solution can be obtained in closed-form. Parts of this chapter have been published in [120].

6.1. Coherence Interval Structures

In this section, we calculate the training overheads associated with CSI acquisition for each precoding type, assuming the channels remain constant for a coherence interval of length T time slots. The analysis is done for a TDD system assuming ideal channel reciprocity. The number of time slots spent on piloting is not optimized in this analysis; rather, we assume the minimum number of slots is spent. Furthermore, we assume that the used pilot sequences by receivers and transmitters are orthogonal. Thus, the analysis does not take pilot contamination into account for the simple reason that in case it occurs, the performance of IA would heavily deteriorate. This would defeat the original purpose of comparing both precoding methods assuming IA reaches its best theoretical performance. For the same reason, even though the rate expressions used here are based on perfect CSIT and/or CSIR, we note that in practice CSI estimation errors would cause higher rate degradations for IA as it relies on the

CSI of more links. Therefore, CSI errors would reduce the number of antennas required for MRT or eigenmode precoding to perform similarly to IA.

6.1.1. Maximum Ratio Transmission

For transmitter k to estimate $\boldsymbol{H}_{kk} \in \mathbb{C}^{N \times M}$, receiver k has to send a sequence of orthogonal pilots in the uplink for a duration of at least N time slots. The overhead of channel estimation in the uplink is independent of the number of transmit antennas in TDD mode. Transmitter k then calculates $\boldsymbol{F}_k$, which we heuristically assume to take one time slot. Even if this assumption is not completely true, it would not affect the main message and conclusions of this chapter. The lower bound on the achievable rate is based on the assumption that receiver k only needs the term $\mathrm{E}[\boldsymbol{H}_{kk}\boldsymbol{F}_k]$ to decode the data symbols. With this assumption, no instantaneous downlink pilots are necessary, and an overhead of $N+1$ time slots is spent on CSIT estimation and precoder calculation before data transmission, as shown in Fig. 6.1. The coherence interval structure here is similar to existing ones in the literature (see, e.g., [53]).

6.1.2. Eigenmode Based Precoding

Similarly to MRT, $N+1$ time slots are first necessary to obtain the CSIT of each direct link and perform the precoder calculation. Additionally, the receiver needs to estimate the resulting effective channel $\boldsymbol{H}_{kk}\boldsymbol{F}_k \in \mathbb{C}^{N \times N}$ to obtain $\boldsymbol{G}_k$ and apply it to decode the received symbols. This is done by sending downlink orthogonal pilot symbols precoded by $\boldsymbol{F}_k$ during N time slots at least. Therefore, a minimum of of $2N+2$ time slots are spent before data transmission as shown in Fig. 6.2. This leaves $T-2N-2$ time slots for data transmission.

6.1.3. Interference Alignment

Here, each transmitter sends $d = d_{\mathrm{IA}}$ symbols to its corresponding receiver. The following overheads are required:

1. The CSI of interfering links $\boldsymbol{H}_{mk} \in \mathbb{C}^{N \times M}\ \forall m \neq k$ needs to be known at transmitter k for the precoder to be calculated
2. The CSI of the effective channels $\boldsymbol{H}_{km}\boldsymbol{F}_m \in \mathbb{C}^{N \times d_{\mathrm{IA}}}\ \forall m$ needs to be known at receiver k:
   - The terms $\boldsymbol{H}_{kl}\boldsymbol{F}_l$, $\forall l \neq k$ are required for the kth receive filter to be calculated according to the obtained precoders and zero-force the interference
   - The term $\boldsymbol{H}_{kk}\boldsymbol{F}_k$ is required to decode the received symbols.

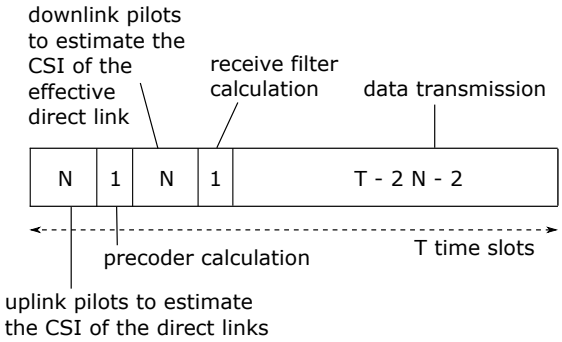

Figure 6.2.: Coherence interval structure of eigenmode precoding

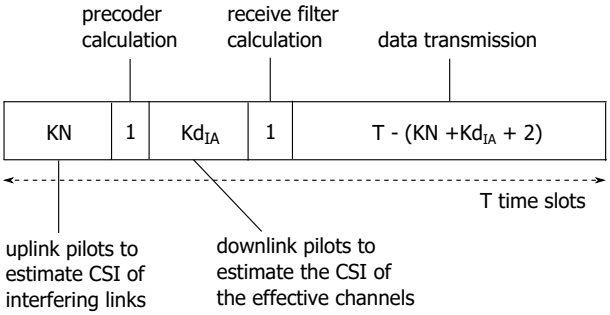

Figure 6.3.: Coherence interval structure of IA

Table 6.1.: Summary of CSIR and CSIT requirements

	MRT	eigenmode precoding	IA
CSIT of direct links	instantaneous	instantaneous	none
CSIR of direct links	statistical	instantaneous	instantaneous
CSIT of interfering links	none	none	instantaneous
CSIR of interfering links	none	none	instantaneous

For the transmitters to be able to properly separate and estimate the CSI corresponding to K receivers with N antennas each, the receivers have to send orthogonal pilot sequences for a minimum duration of KN symbol slots. This corresponds to the overhead of step 1. For step 2, estimating the effective channels is done by sending pilot symbols precoded by $\boldsymbol{F}_m$ at transmitter m. These channels have size $N \times d_{\mathrm{IA}}$. In a P2P link, a transmitter would have to send precoded pilots over a period of at least d_{IA} time slots to the receiver. In an IC, if K transmitters simultaneously send orthogonal pilot sequences, then the receivers can estimate the required effective channels in the downlink after a minimum of Kd_{IA} time slots. In addition, precoder and receive filter calculations at the transmitters and receivers respectively last 1 time slot each. In total, a minimum of

$$T_o = KN + Kd_{\mathrm{IA}} + 2 \tag{6.1}$$

time slots are spent on piloting and calculation before data transmission starts, as shown in Fig. 6.3. Table 6.1 summarizes the CSI requirements of the different precoding types given their system model assumptions.

6.2. Spectral Efficiency Analysis

For the rest of the chapter, we assume without loss of generality that all transmitters transmit with power E_{tx}. This results in $\mathrm{SNR}_{\mathrm{IA},k} = \mathrm{SNR}_{\mathrm{IA}}$ [cf. (4.17)] and $c_k = c$ [cf. (5.19)] $\forall k$ where c is calculated in terms of $\mathrm{SNR}_{\mathrm{IA}}$ as

$$c = \frac{\mathrm{SNR}_{\mathrm{IA}}}{(K-1)\,\mathrm{SNR}_{\mathrm{IA}} + 1}. \tag{6.2}$$

6.2.1. Eigenmode Precoding vs. Interference Alignment

We combine the rate results of Chapters 4 and 5 with the coherence interval structures of Section 6.1. Data transmission using eigenmode precoding oc-

curs over a portion of $(T-2N-2)/T$ of the coherence interval with a minimum ergodic rate $M\mathcal{V}(c,\beta)$ [cf. (5.26)], while data transmission using IA occurs over a portion of $(T-T_o)/T$ of the coherence interval with an ergodic rate $d_{\mathrm{IA}}\mathcal{V}(\mathrm{SNR}_{\mathrm{IA}},1)$ [cf. (4.40) and (4.16)]. To find the minimum number of antennas $M_{\min}$ required for transmitters employing eigenmode precoding to match the performance of transmitters with N antennas employing IA, the following equality is posed:

$$\begin{aligned}&\left(\frac{T-2N-2}{T}\right) M\mathcal{V}(c,M/N) = \\ &\left(\frac{T-T_o}{T}\right) d_{\mathrm{IA}}\mathcal{V}(\mathrm{SNR}_{\mathrm{IA}},1).\end{aligned} \tag{6.3}$$

The left-hand side is the minimum eigenmode precoding spectral efficiency while the right-hand side is the IA spectral efficiency. Eq. (6.3) is then solved in M for a given fixed set of parameters T, N, K, and $\mathrm{SNR}_{\mathrm{IA}}$ (d_{IA} is implicitly defined by K and N). The smallest integer satisfying (6.3) gives $M_{\min}$. $M_{\min}$ cannot be found in closed form; nonetheless, noting that the left-hand side of (6.3) is increasing in M, a bisection method can be used to get a solution for $M_{\min}$. The results for the K-user IC with constant channel coefficients can be obtained from (6.3) by setting $T\to\infty$. In that case, both fractions appearing before the rate terms in (6.3) equal 1 and all discussed overheads in Section 6.1 have no effect on the spectral efficiency.

6.2.2. Maximum Ratio Transmission vs. Interference Alignment

Similarly to (6.3), we can set up an equality to find how much antennas are required for MRT to perform similarly to IA:

$$\begin{aligned}&\left(\frac{T-N-1}{T}\right) \overline{R}_{\mathrm{MRT,LB}} = \\ &\left(\frac{T-T_o}{T}\right) d_{\mathrm{IA}}\mathcal{V}(\mathrm{SNR}_{\mathrm{IA}},1).\end{aligned} \tag{6.4}$$

Defining

$$f = \frac{d_{\mathrm{IA}}\mathcal{V}(\mathrm{SNR}_{\mathrm{IA}},1)(T-T_o)}{N(T-N-1)}, \tag{6.5}$$

then $M_{\min}$ can be found in closed-form as

$$\begin{aligned}&M_{\min} = \\ &\left\lceil N(2^f-1)\frac{K\,\mathrm{SNR}_{\mathrm{IA}}+1}{\mathrm{SNR}_{\mathrm{IA}}}\right\rceil\end{aligned} \tag{6.6}$$

which again is a function of the set $\{T,N,K,\mathrm{SNR}_{\mathrm{IA}}\}$.

6.2.3. Closing Remarks

The found $M_{\min}$ corresponds in fact to an upper bound on the required number of antennas as

1. the left-hand sides of (6.3) and (6.4) constitute lower bound of the spectral efficiencies obtained with eigenmode precoding and MRT, and

2. CSI errors would cause higher spectral efficiency degradations for IA.

In the analysis, we assumed that a transmission of N data streams with eigenmode precoding and MRT but only a transmission of $N/2$ data streams with IA. This models the most naive non-cooperative strategy vs. the smartest cooperative strategy. Note that a transmission of more than $N/2$ streams while attempting IA would cause the system to be interference limited. Thus, the choice of $N/2$ streams in the latter case is optimal. In contrast, the considered non-cooperative strategies simply transmit the maximum number of streams despite possibly causing large inter-symbol interference. From the result of Lemma 5.1, it can be observed that streams suffer from a high level of interference. Consequently, a higher achievable rate for eigenmode precoding could be achieved via WF; namely, shutting off the weak streams and transmitting less than N streams. However, this is not considered to model the most naive type of a given transmitter which has no information about the interference caused by other transmitters.

Finally, the left and right-hand sides of (6.3) and (6.4) constitute standalone results as they can used to calculate the spectral efficiency of any of the considered precoding methods for a given set of parameters, taking into account the piloting overheads.

6.3. Numerical Examples

We solve (6.3) using a bisection method and evaluate (6.6) for different values of the set $\{T, N, K, \mathrm{SNR}_{\mathrm{IA}}\}$. We are not interested in the low SNR regime, as cooperation would only bring small benefits in this regime. We first consider the IA $(4, 4, 2)^3$ scenario where each stream has a receive $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB, corresponding to an average spectral efficiency of approximately 11 bits/s/Hz/user when $T \to \infty$. Fig. 6.4 plots the resulting required number of antennas $M_{\min}$ against T for both eigenmode precoding and MRT to perform similarly to IA. Fig. 6.5 plots the same scenario for a receive $\mathrm{SNR}_{\mathrm{IA}} = 15$ dB, corresponding to an average spectral efficiency of approximately 8 bits/s/Hz/user when $T \to \infty$. The results of the constant IC are also plotted for comparison. Fig. 6.4 shows that in the worst-case corresponding to the constant IC, transmitters with less than 70 and 50 antennas employing MRT and eigenmode precoding

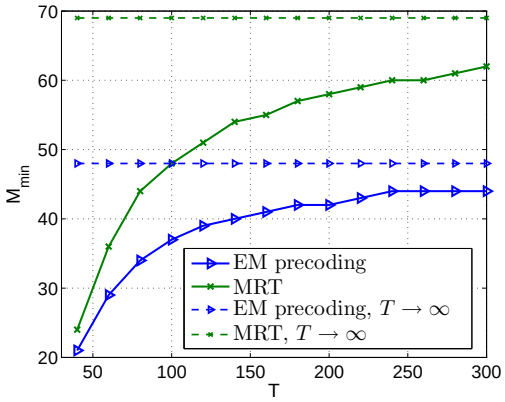

Figure 6.4.: $(M_{\min}, 4, 4)^3$ vs. IA $(4, 4, 2)^3$: $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB

respectively perform similarly to transmitters with 4 antennas employing IA. In Fig. 6.5, this number drops to less than 40 and 30 for MRT and eigenmode precoding, respectively. This is due to the fact that increasing the noise level reduces the performance gap between IA and MRT or eigenmode precoding[1]; consequently, $M_{\min}$ decreases with decreasing SNR. Most importantly, Figs. 6.4 and 6.5 show the relationship between T and $M_{\min}$: the smaller T is, the smaller the required $M_{\min}$. Note, for instance, the big difference between $M_{\min}$ at $T = 40$ and $T \to \infty$. This is an expected result as the training overheads of IA take a much larger portion of the coherence interval for small values of T compared to the overheads of MRT or eigenmode precoding (cf. Figs. 6.1, 6.2, and 6.3). As a result, MRT or eigenmode precoding have a much larger portion of the coherence time left for data transmission and a relatively small $M_{\min}$—corresponding to small transmit array gains—is already enough to give spectral efficiencies comparable to IA.

Fig. 6.6 corresponds to the $(4, 4, 1)^5$ scenario with $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB. In this case and for a given value of T, $M_{\min}$ is smaller than its corresponding value in Fig. 6.4 because 3 receive antennas are reserved for IA and cancellation and only 1 data stream can be transmitted. As MRT and eigenmode precoding allow full spatial multiplexing with 4 transmitted streams, it doesn't take much more additional antennas to perform similarly to IA, despite possibly large existent

[1] Remember that IA is a desired strategy at high SNRs while MRT or eigenmode precoding are desired strategies at low SNRs.

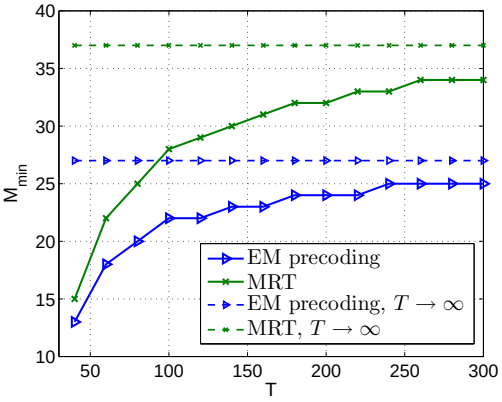

Figure 6.5.: $(M_{\min}, 4, 4)^3$ vs. IA $(4, 4, 2)^3$: $\mathrm{SNR}_{\mathrm{IA}} = 15$ dB

inter-symbol interference. The relationship between T and $M_{\min}$ is seen in this figure as well. Figs. 6.4-6.6 emphasize the importance of including the training overheads in the spectral efficiency calculations to allow a fair performance comparison between cooperative and non-cooperative precoding techniques. Figs. 6.4-6.6 show that even the simple MRT strategy does not require massive configurations to emulate noise-limited systems. Additionally, the required number of antennas for eigenmode precoding is smaller, as the latter has a better performance than MRT.

Finally, we consider the $(5, 5, 2)^4$ IC scenario. This configuration doesn't yield a closed-form IA solution; nonetheless, we focus on the constant coefficients case and plot $M_{\min}$ against $\mathrm{SNR}_{\mathrm{IA}}$ in Fig. 6.7, covering average spectral efficiencies up to 14 bits/s/Hz/user. The conclusions of the previous figures hold here as well; namely, it is observed that $M_{\min}$ increases with increasing SNR. Figs. 6.4, 6.5, 6.6, and 6.7 show that for practical systems with short coherence times and finite SNRs and in the context of inter-cell interference cancellation, non-cooperative (single-cell) precoding strategies do not necessitate massive numbers of antennas to emulate the performance of interference-free cooperative systems operating with traditional numbers of antennas.

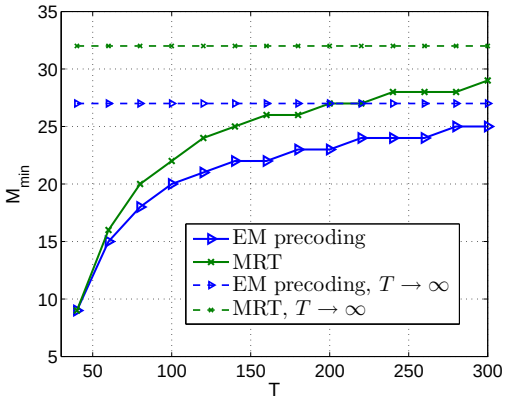

Figure 6.6.: $(M_{\min}, 4, 4)^5$ vs. IA $(4, 4, 1)^5$: $\mathrm{SNR}_{\mathrm{IA}} = 20$ dB

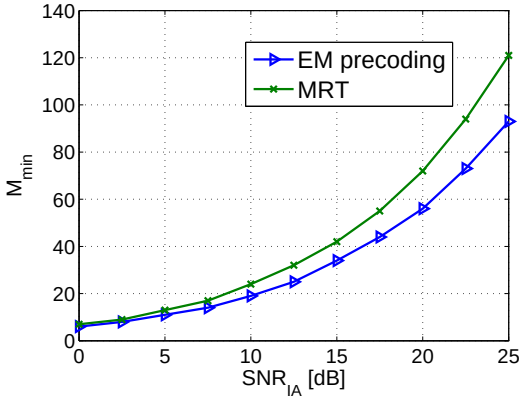

Figure 6.7.: $(M_{\min}, 5, 5)^4$ vs. IA $(5, 5, 2)^4$ for the constant coefficients IC

6.4. Summary

This chapter provided a framework for comparing the spectral efficiencies of IA, MRT, and eigenmode precoding schemes in a MIMO IC. This was achieved through the following two points:

- considering scenarios where the IA solution can be found in closed-form, and
- studying the coherence interval structures of each precoding type and including the training overheads in the spectral efficiency analysis.

Using the rate expressions from Chapters 4 and 5 in combination with the training overhead analysis in TDD mode allowed obtaining the minimum number of transmit antennas $M_{\min}$ required by naive transmitters employing MRT or eigenmode precoding to result in good enough achievable rates—occurring over a larger portion of the coherence interval—to match the performance of a system where smart cooperating transmitters have a fixed number of antennas and employ the IA scheme. $M_{\min}$ was observed to vary significantly according the coherence interval length and the SNR. In practice, CSI errors—whether related to estimation errors or pilot contamination—would reduce $M_{\min}$ for a given transmit to noise power ratio as IA relies on the CSI of more links and is therefore more sensible to CSI errors. Along with the fact that the MRT and eigenmode precoding analytical expressions correspond to lower bounds, the found $M_{\min}$ in each case corresponds to an upper bound on the number of antennas for practical systems. Whereas massive MIMO literature envisions base stations with hundreds or even thousands of antennas to facilitate precoding operations and create noise-limited systems, the conclusions of this chapter are that in the context of inter-cell interference cancellation, even the simple MRT strategy does not require massive configurations to emulate the performance of noise-limited systems in practical scenarios with finite SNRs and short coherence times. This is mainly due to two reasons. The first one being the larger CSIT acquisition overhead for cooperative systems such as IA, which takes a larger portion of the coherence interval and reduces the resulting spectral efficiency. The second reason is the limited benefits of IA at finite SNRs.

The investigations of this chapter lead to interesting unanswered questions, as we did not claim that cooperation among transmitters having $M_{\min}$ antennas (e.g., through ZF precoders) would not lead to any benefits at all compared to non-cooperative methods. The answer to that question cannot be deduced from the performed analysis. A possible line of work consists in investigating whether there exists a threshold given again, by, a number of transmit antennas, where cooperative methods would not provide performance gains compared to non-cooperative methods at medium and high SNRs. The framework of this chapter could be a starting point for such an analysis.

7. Conclusions

In this thesis, the rate performance of different linear precoding methods was characterized. These methods were split into cooperative methods (e.g., multi-cell precoding) and non-cooperative ones (e.g., single-cell precoding). These methods were compared and an upper bound on the required number of transmit antennas of the single-cell methods to perform similarly to multi-cell methods operating with a fixed number of antennas was found. The i.i.d. channel entries assumption used throughout this thesis was used as it results in a superior performance of cooperative methods, compared to scenarios where individual channels' entries and/or channels of different users are correlated, as explained in Chapter 6. Next, we provide a brief summary of the thesis chapters and follow up by a final discussion on cooperative methods.

Chapter 2 of this thesis motivated the need for MIMO systems by discussing several advantages of MIMO systems including capacity increase, spatial diversity, spatial multiplexing, etc. Additionally, it briefly reviewed the linear precoding designs in P2P and BC MIMO scenarios.

Chapter 3 introduced the thesis system model, the K-user MIMO IC, and reviewed many of the existing linear precoding methods for this model. Three main communication modes were distinguished, according to the system configuration in terms of the number of transmit antennas, receive antennas, desired streams, and users. We focused on the most challenging mode, and reviewed different linear precoding methods in that mode such as IA. We then discussed the possible transmission options in the other two modes. We additionally differentiated between *cooperative* and *non-cooperative* methods, listing the advantages, disadvantages, and requirements of each method.

In Chapter 4, we focused on the class of IA methods due to its analytical tractability and derived large system closed-form expressions of its achievable rates using RMT tools. Even though the derived expressions are only exact asymptotically, simulation results showed that these expressions provide accurate estimates for small and finite system parameters as well. IA methods always result in square effective channel matrices with the result that—contrary to other methods—achievable rates under WF power allocation can always be found in quasi closed-form expressions. Chapter 4 also revealed some interesting observations about the asymptotic WF level which, contrary to the finite case, is independent of the instantaneous eigenvalues and is completely determined by the asymptotic eigenvalue distribution.

In Chapter 5, we focused on the non-cooperative eigenmode precoding and MRT techniques and derived achievable rate lower bounds using a combination of separate stream decoding assumption and Jensen's inequality. The bounds for eigenmode precoding were derived using RMT tools as well, while the special nature of MRT combined with the limited CSIR model allowed deriving lower bounds without the need for RMT. Simulation results showed that the bounds become tighter as the number of interferers increases, because the interference covariance matrix of each receiver in that case would approach a white (diagonal) matrix, containing less information that can be exploited when joint decoding is performed.

In Chapter 6, we used the derived expressions of Chapters 4 and 5 in the context of a spectral efficiency analysis, which included the training overheads of the different precoding methods in a TDD system. We investigated the required number of transmit antennas for the non-cooperative strategies of Chapter 5 to perform similarly to IA. We derived a relatively modest upper bound $M_{\min}$ on that required number when compared to the literature assumptions, and concluded that massive configurations are not necessary to emulate the performance of noise-limited ICs using IA. We showed how the training overheads of IA can significantly reduce its spectral efficiency in scenarios with short coherence times, and lead to a considerable drop of the value of this upper bound. The findings emphasize the importance of including the training overheads in the spectral efficiency calculations to allow a fair performance comparison between cooperative and non-cooperative precoding techniques. Such overheads are predominantly neglected in the literature which focuses on channels with constant coefficients. We did not claim that cooperation among transmitters having $M_{\min}$ antennas would not lead to any performance improvements over non-cooperative methods. Such an answer cannot be deduced from the performed analysis and necessitates further investigations using a similar framework to the one of Chapter 6.

The conclusions of Chapter 6 do not hold for FDD systems, as acquiring CSIT requires downlink piloting and feedback from the receiver side; consequently, each additional transmit antenna increases piloting and feedback overheads. Nonetheless, we focused on TDD mode as it is the most suitable mode of operation for massive MIMO systems from the point of view of CSI acquisition and precoding, given hardware aspects are properly handled (e.g., the calibration of the transmit and receive hardware chains at the transmitter to ensure channel reciprocity and mitigate any CSI mismatch between the uplink and downlink directions).

It is yet unclear whether IA or other reviewed cooperative methods in this thesis will be implemented in practice, and whether the theoretical gains of such methods can be retained in a real implementation. Recent papers on IA have contradicting views regarding this aspect [121–123]. In any case, even though

IA is well understood by now, it still has open research problems. As explained in Chapter 3, closed-form solutions have been only obtained in two special cases and the problem of obtaining closed-form solutions is still open in general. This problem is interesting from a theoretical as well as practical point of view, as closed-form solutions facilitate practical implementations. One interesting line of work consists in investigating whether similar ideas to the ones in [92] can be applied to find closed-form solutions for new cases. That would involve first reducing the MIMO IC to a sub-IC where IA is feasible and where closed-form solutions can be found (e.g, a three-user square symmetric channel), then trying to construct the final precoders and decoders as a function of the precoders and decoders of the initial sub-IC without the need for iterative algorithms. It is unclear at this point whether such a problem has a tractable solution.

Even though a lot of questions on the MIMO IC have been answered, it still has its share of interesting unsolved problems and will continue to attract a lot of interest in the future.

A. Mathematical Basics

A.1. Linear Transformations

A transformation $T : \mathbb{C}^n \to \mathbb{C}^m$ is called linear if and only if the following relationships hold for all vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{C}^n$ and all scalars $c \in \mathbb{R}$ [124]:

1. $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$
2. $T(c\,\boldsymbol{u}) = c\,T(\boldsymbol{u})$.

It can be easily checked that all the precoding and receive filtering operations considered in this thesis correspond to linear transformations of the input of the corresponding filters.

A.2. The Eigenvalue Decomposition

A widely used matrix decomposition is the EVD. Let $\boldsymbol{S} \in \mathbb{C}^{N \times N}$. By the theory of eigenvalues and eigenvectors of matrices, the eigenvalues $(\phi_1, \ldots, \phi_N)$ and eigenvectors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ of $\boldsymbol{S}$ are the solutions of the following equations [74]:

$$\boldsymbol{S}\,\boldsymbol{x}_l = \phi_l\,\boldsymbol{x}_l$$

for $l = 1, \ldots, N$. If $\boldsymbol{S}$ is full rank, its eigenvalues are non-zero and its eigenvectors are distinct. The EVD of $\boldsymbol{S}$ is then defined as [74]

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{\Phi}\boldsymbol{X}^{-1}$$

where $\boldsymbol{\Phi} = \operatorname{diag}(\phi_1, \ldots, \phi_N)$, $\phi_1 \geq \phi_2 \geq \ldots \geq \phi_N$, and $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$.

If $\boldsymbol{S}$ is Hermitian (full rank or rank deficient), then the EVD can be written as

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{\Phi}\boldsymbol{X}^{\mathrm{H}}.$$

A.3. The Central Limit Theorem

Let $\{x_1, x_2, \ldots, x_N\}$ be $N \gg 1$ independent complex random variables. Let the mean and variance of x_l equal μ_l and σ_l^2, $l = 1, 2, \ldots, N$. Then, the distribution

of the sum $x_1+x_2+\ldots+x_N$ approaches a Gaussian distribution, whose mean and variance equal the sum of the individual means and variances, respectively [109]:

$$x_1 + x_2 + \ldots + x_N \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{C}} \left(\sum_{l=1}^{N} \mu_l, \sum_{l=1}^{N} \sigma_l^2 \right).$$

A.4. The Law of Large Numbers

The LLN describes the outcome of an experiment when repeated a large number of times. Let x be a random variable with mean μ, and let the set $\{x_{\mathrm{s},1}, x_{\mathrm{s},2}, \ldots, x_{\mathrm{s},N}\}$ correspond to N samples or realizations of x. The LLN states that as the number of realizations grows without bounds, the sample average converges to the mean, i.e.:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{N} x_{\mathrm{s},l} = \mu$$

(cf. [109]).

B. Capacity of Point-to-Point MIMO Links

B.1. Capacity-Achieving Precoders

The link between the N receive antennas and the M transmit antennas is modeled by $\boldsymbol{H} \in \mathbb{C}^{N \times M}$. The system equation of the P2P link reads

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{B.1}$$

where $\boldsymbol{y} \in \mathbb{C}^N, \boldsymbol{x} \in \mathbb{C}^M$, and $\boldsymbol{n} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}_N, \sigma_n^2 \mathbf{I}_N) \in \mathbb{C}^N$ denote the received signal vector, the transmitted signal vector, and the noise vector at the receiver, respectively. The transmit power constraint reads

$$\mathrm{E}\left[||\boldsymbol{x}||_2^2\right] = \mathrm{tr}(\boldsymbol{Q}) \leq \mathrm{E}_{\mathrm{tx}} \tag{B.2}$$

where $\boldsymbol{Q} \in \mathbb{C}^{M \times M}$ is the covariance matrix of $\boldsymbol{x}$.

Besides the additive noise, inter-symbol interference is another bottleneck in such a scenario. The capacity of this link is given by [27]

$$C_{\mathrm{MIMO}} = \max_{f(\boldsymbol{x}):\mathrm{E}\left[||\boldsymbol{x}||_2^2\right] \leq \mathrm{E}_{\mathrm{tx}}} \mathcal{I}(\boldsymbol{y}; \boldsymbol{x}) \tag{B.3}$$

where $\mathcal{I}(\boldsymbol{y}; \boldsymbol{x})$ is the mutual information between the output and input of the channel [112], and $f(\boldsymbol{x})$ is the probability density function of $\boldsymbol{x}$. It can be shown that $\mathcal{I}(\boldsymbol{y}; \boldsymbol{x})$ is maximized when $\boldsymbol{x}$ is circularly symmetric complex Gaussian (CSCG). In that case, $\mathcal{I}(\boldsymbol{y}; \boldsymbol{x})$ can be found in closed-form as

$$\mathcal{I}(\boldsymbol{y}; \boldsymbol{x}) = \log \left| \mathbf{I}_N + \frac{1}{\sigma_n^2} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}} \right|. \tag{B.4}$$

What remains is choosing $\boldsymbol{Q}$ to maximize (B.4). The argument of the determinant is non-negative definite. By Hadamard's inequality, (B.4) is maximized when the non-negative definite argument is a diagonal matrix [125, Theorem 7.8.1]. Defining

$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{H}} \tag{B.5}$$

to be the EVD of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$, choosing

$$\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{P}\boldsymbol{V}^{\mathrm{H}} \tag{B.6}$$

where $\boldsymbol{P} = \mathrm{diag}(p_1, \ldots, p_M)$ maximizes the mutual information and results in

$$\mathcal{I}(\boldsymbol{y}; \boldsymbol{x}) = \log \left| \mathbf{I}_M + \boldsymbol{\Phi}\boldsymbol{P} \right| . \tag{B.7}$$

The real entries of $\boldsymbol{P}$ correspond to the power loading coefficients and are chosen according to the WF rule, as explained in Appendix B.2. Defining $r = \mathrm{rank}(\boldsymbol{H})$, then $\mathrm{rank}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}) \leq r$ and consequently

$$C_{\mathrm{MIMO}} = \sum_{l=1}^{r} \log(1 + \phi_l p_l) \tag{B.8}$$

as no more than r eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ are non-zero. The last $M - r$ entries of $\boldsymbol{P}$ are thus set to 0.

To obtain the optimal precoder $\boldsymbol{F} \in \mathbb{C}^{M \times r}$ from (B.6), we write $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s}$ where $\boldsymbol{s} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_r, \mathbf{I}_r)$ is the input symbol vector. The input covariance matrix in this case is $\boldsymbol{Q} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}$ due to the statistical properties of $\boldsymbol{s}$. Consequently, one solution for $\boldsymbol{F}$ can be obtained from (B.6) as

$$\boldsymbol{F} = \boldsymbol{V}^{1:r} \hat{\boldsymbol{P}}^{1/2} \tag{B.9}$$

where $\hat{\boldsymbol{P}} = \mathrm{diag}(p_1, \ldots, p_r)$. This choice keeps the desired CSCG properties of $\boldsymbol{x}$. This is because left multiplying the CSCG $\boldsymbol{s}$ by $\hat{\boldsymbol{P}}^{1/2}$ only corresponds to a scaling operation and keeps $\hat{\boldsymbol{P}}^{1/2}\boldsymbol{s}$ CSCG. Similarly, it can be checked that left multiplying the CSCG $\hat{\boldsymbol{P}}^{1/2}\boldsymbol{s}$ by the unitary $\boldsymbol{V}^{1:r}$ keeps the resulting product CSCG.

The chosen precoder aligns the data symbols to the strongest eigenmodes of the channel, i.e., it directs the transmission to the subspace containing the maximum signal power, and splits the MIMO link into r parallel and non-interfering SISO links with the end result that inter-symbol interference is cancelled. The obtained precoder is *linear*.

B.2. Water-Filling Power Allocation

The entries of $\hat{\boldsymbol{P}}$ are given by the WF rule [27], i.e.,

$$p_l = \max\left(0, \mu - \frac{\sigma_n^2}{\phi_l}\right), \quad \forall l = 1, \ldots, r \tag{B.10}$$

where μ is the WF level. Additionally,

$$\sum_{l=1}^{r} p_l = \mathrm{E}_{\mathrm{tx}}. \tag{B.11}$$

WF assigns more transmit power to directions with strong power and less to directions with weak power to reach capacity. If $t \leq r$ eigenvalues are served, then μ can be calculated from (B.10) and (B.11) as

$$\mu = \frac{1}{t}\left(\mathrm{E}_{\mathrm{tx}} + \sum_{l=1}^{t} \frac{\sigma_n^2}{\phi_l}\right). \tag{B.12}$$

μ and $\{p_1, \ldots, p_t\}$ are coupled. By noting that (B.12) holds if $t = r$ or if $\mu < 1/\phi_{t+1}$ with $t < r$ as seen from (B.10), the following procedure can be used to find a solution for μ:

1. Start with $t = 1$.
2. Find μ using (B.12).
3. If $\mu < 1/\phi_{t+1}$ or $t = r$, then the solution is found. Else, set $t \leftarrow t + 1$ and go back to step 2.

At low SNRs, WF can shut off transmissions in certain directions and even only allow single-stream transmission. In this case, the multiple receive antennas serve as to increase the received signal power without spatial multiplexing [27]. In contrast, μ converges to $\mathrm{E}_{\mathrm{tx}}/r$ at high SNR and all streams with non-zero eigenvalues are approximately allocated equal power.

C. Derivations and Proofs

C.1. Distributions of Inner Vector Products

Let $\{\boldsymbol{h}, \boldsymbol{a}\} \in \mathbb{C}^M$ be two uncorrelated Gaussian vectors with i.i.d. entries of mean 0 and variance 1. We wish to find the distributions of the vector products $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}/\sqrt{M}$ and $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a}/\sqrt{M}$. To find the distribution of the first term, we expand $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}$ as

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h} = \sum_{l=1}^{M} |h_l|^2 = \sum_{l=1}^{M} \left(\Re(h_l)^2 + \Im(h_l)^2\right)$$

which is a sum of $2M \gg 1$ real Gaussian i.i.d. variables. The expectation and variance of each of the variables are given by:

$$\begin{aligned} \mathrm{E}\left[\Re(h_l)^2\right] &= \mathrm{E}\left[\Im(h_l)^2\right] = \frac{1}{2}, \\ \mathrm{E}\left[\Re(h_l)^4\right] &= 3\left(\frac{1}{2}\right)^2, \\ \mathrm{var}\left(\Re(h_l)^2\right) &= \mathrm{E}\left[\Re(h_l)^4\right] - \left(\mathrm{E}\left[\Re(h_l)^2\right]\right)^2 \\ &= \frac{1}{2} = \mathrm{var}\left(\Im(h_l)^2\right), \end{aligned} \tag{C.1}$$

$l \in \{1, \ldots, M\}$ (cf. [109]). Therefore,

$$\begin{aligned} \mathrm{var}(\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}) &= \mathrm{var}\left(\sum_{l=1}^{M} \left(\Re(h_l)^2 + \Im(h_l)^2\right)\right) \\ &= \sum_{l=1}^{M} \mathrm{var}\left(\left(\Re(h_l)^2\right) + \mathrm{var}\left(\Im(h_l)^2\right)\right) = M. \end{aligned} \tag{C.2}$$

Furthermore, $\mathrm{E}[\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}] = M$ by definition. Based on the central limit theorem (cf. Appendix A.3), we have:

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h} \xrightarrow{\mathrm{d}} \mathcal{N}(M, M) \tag{C.3}$$

for $M \gg 1$. With a scaling of $1/\sqrt{M}$ and $1/M$, we obtain:

$$\frac{1}{\sqrt{M}}\, \boldsymbol{h}^{\mathrm{H}}\boldsymbol{h} \xrightarrow{\mathrm{d}} \mathcal{N}(\sqrt{M}, 1) \tag{C.4}$$

and

$$\frac{1}{M}\,\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h} \xrightarrow{\mathrm{d}} \mathcal{N}\left(1, \frac{1}{M}\right). \tag{C.5}$$

Similarly to the above, the inner product $\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a}$ is expanded as

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a} = \sum_{l=1}^{M} h_l^* a_l$$

which is a sum of $M \gg 1$ i.i.d. complex variables. Similar to (C.3), it can be shown that:

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a} \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{C}}(0, M). \tag{C.6}$$

Consequently:

$$\frac{1}{\sqrt{M}}\,\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a} \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{C}}(0, 1) \tag{C.7}$$

and

$$\frac{1}{M}\,\boldsymbol{h}^{\mathrm{H}}\boldsymbol{a} \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{C}}\left(0, \frac{1}{M}\right). \tag{C.8}$$

C.2. Moments of Matrix Products

Let $\{\boldsymbol{H}, \boldsymbol{A}\} \in \mathbb{C}^{N\times M}$ be two uncorrelated matrices with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries. We wish to calculate the following quantities:

$$\boldsymbol{R} = \mathrm{E}\left[\frac{1}{\sqrt{M}}\,\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right]$$

$$\boldsymbol{P} = \mathrm{E}\left[\left(\frac{1}{\sqrt{M}}\,\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^2\right]$$

$$\boldsymbol{Q} = \mathrm{E}\left[\frac{1}{M}\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\right)^{\mathrm{H}}\right].$$

The lth row of $\boldsymbol{H}$ is $\boldsymbol{h}_l^{\mathrm{T}}$, i.e., $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N]^{\mathrm{T}}$. Similarly, $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]^{\mathrm{T}}$. To obtain $\boldsymbol{R}$, the term $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}/\sqrt{M}$ is first expanded as

$$\frac{1}{\sqrt{M}}\,\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} = \frac{1}{\sqrt{M}}\begin{pmatrix} \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_1^* & \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_2^* & \ldots & \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_N^* \\ \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_1^* & \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_2^* & \ldots & \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_N^* \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{h}_N^{\mathrm{T}}\boldsymbol{h}_1^* & \boldsymbol{h}_N^{\mathrm{T}}\boldsymbol{h}_2^* & \ldots & \boldsymbol{h}_N^{\mathrm{T}}\boldsymbol{h}_N^* \end{pmatrix}. \tag{C.9}$$

Noting that the matrix entries of (C.9) are the complex conjugates of the quantities in (C.4) and (C.7), we obtain $\boldsymbol{R} = \sqrt{M}\,\mathbf{I}_N$.
To calculate $\boldsymbol{P}$, we focus on its (i,j)th entry p_{ij} which equals:

$$\begin{aligned} p_{ij} &= \frac{1}{M}\ \mathrm{E}\left[\boldsymbol{e}_i^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)\boldsymbol{e}_j\right] \\ &= \frac{1}{M}\ \mathrm{E}\left(\left[\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_1^*, \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_2^*, \ldots, \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_N^*\right]\begin{bmatrix} \boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_j^* \\ \boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_j^* \\ \vdots \\ \boldsymbol{h}_N^{\mathrm{T}}\boldsymbol{h}_j^* \end{bmatrix}\right) \\ &= \frac{1}{M}\ \mathrm{E}\Big[\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_1^*\boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_j^* + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_2^*\boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_j^* + \ldots \\ &\quad + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_i^*\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_j^* + \ldots + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_N^*\boldsymbol{h}_N^{\mathrm{T}}\boldsymbol{h}_j^*\Big]. \end{aligned} \tag{C.10}$$

Correspondingly, using (C.4) and (C.7) yields:

$$\begin{aligned} p_{ij} &= \frac{1}{M}\begin{cases} 0, & i \neq j, \\ \mathrm{E}\left[\left|\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_i^*\right|^2\right] + \sum_{l\neq i}\mathrm{E}\left[\left|\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{h}_l^*\right|^2\right], & i = j. \end{cases} \\ &= \begin{cases} 0, & i \neq j, \\ \mathrm{E}\left[\left|\frac{1}{\sqrt{M}}\,\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{h}_i\right|^2\right] + \sum_{l\neq i}\mathrm{E}\left[\left|\frac{1}{\sqrt{M}}\,\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{h}_l\right|^2\right], & i = j. \end{cases} \\ &= \begin{cases} 0, & i \neq j, \\ M+N, & i = j. \end{cases} \end{aligned} \tag{C.11}$$

Thus, $\boldsymbol{P} = (M+N)\mathbf{I}_N$.

Using a similar procedure, the (i,j)th term q_{ij} of $\boldsymbol{Q}$ equals:

$$\begin{aligned} q_{ij} &= \frac{1}{M}\ \mathrm{E}\left[\boldsymbol{e}_i^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\right)\left(\boldsymbol{H}\boldsymbol{A}^{\mathrm{H}}\right)^{\mathrm{H}}\boldsymbol{e}_j\right] \\ &= \frac{1}{M}\ \mathrm{E}\Big[\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{a}_1^*\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{h}_j^* + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{a}_2^*\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{h}_j^* + \ldots \\ &\quad + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{a}_i^*\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{h}_j^* + \ldots + \boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{a}_N^*\boldsymbol{a}_N^{\mathrm{T}}\boldsymbol{h}_j^*\Big] \\ &= \frac{1}{M}\begin{cases} 0, & i \neq j, \\ \mathrm{E}\left[\sum_{l=1}^{N}\left|\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{a}_l^*\right|^2\right], & i = j. \end{cases} \\ &= \begin{cases} 0, & i \neq j, \\ N, & i = j. \end{cases} \end{aligned} \tag{C.12}$$

Therefore, $\boldsymbol{Q} = N\,\mathbf{I}_N$.

C.3. Convergence Proof of $A(\sigma)$

Rewriting $a_n(\sigma)$, the n-th term of the infinite series $A(\sigma)$ as

$$a_n(\sigma) = \frac{\binom{2n}{n}}{4^n\,(2n+1)} \left(\frac{\sigma}{2}\right)^{2n+1} \frac{2}{2n+1}, \tag{C.13}$$

and comparing it with $b_n(\sigma)$ [cf. (4.37)], it can be observed that

$$a_n(\sigma) = b_n(\sigma)\,\frac{2}{2n+1}, \tag{C.14}$$

and therefore $a_n(\sigma) < b_n(\sigma)$ for $n \geq 1$. As $\sum_{n=1}^{\infty} b_n(\sigma)$ is a convergent series for $0 \leq \sigma \leq 2$, this implies that $\sum_{n=1}^{\infty} a_n(\sigma)$ is a convergent series for $0 \leq \sigma \leq 2$. Furthermore, this implies that the infinite series

$$A(\sigma) = \sigma + \sum_{n=1}^{\infty} a_n(\sigma)$$

converges for $0 \leq \sigma \leq 2$.

D. Abbreviations and Acronyms

AWGN	additive white Gaussian noise
BC	broadcast channel
BD	block diagonalization
BER	bit error rate
CB	coordinated beamforming
CDMA	code-division-multiple-access
CLT	central limit theorem
CoMP	coordinated multipoint transmission
CS	coordinated scheduling
CSCG	circularly symmetric complex Gaussian
CSI	channel state information
CSIR	channel state information at the receiver
CSIT	channel state information at the transmitter
DoF	degrees of freedom
DPC	dirty paper coding
EVD	eigenvalue decomposition
FDD	frequency-division-duplex
FDMA	frequency-division-multiple-access
IA	interference alignment
IC	interference channel
IL	interference leakage (algorithm)
i.i.d.	independent and identically distributed
JP	joint processing
LLN	law of large numbers
LSA	large system analysis
LTE	long term evolution (systems)

MIMO	multiple-input-multiple-output
MISO	multiple-input-single-output
MSE	mean-squared-error
MMSE	minimum-mean-square-error
MRT	maximum ratio transmission
MU-MIMO	multi-user MIMO
OFDM	orthogonal frequency division multiplexing
P2P	point-to-point
RMT	random matrix theory
SINR	signal-to-interference-plus-noise ratio
SISO	single-input-single-output
SNR	signal-to-noise ratio
SDMA	space-division-multiple-access
TDD	time-division-duplex
TDMA	time-division-multiple-access
WF	water-filling
w.r.t.	with respect to
ZF	zero-forcing

E. List of Author's Publications

Note: the publications marked with an asterisk are not related to the topic of the thesis.

1. * S. Bazzi and W. Xu, "Mitigating inter-cell pilot interference via network-based greedy sequence selection and exchange," *IEEE Wireless Communications Letters*, to appear.

2. S. Bazzi, G. Dietl, and W. Utschick, "How many transmit antennas emulate the performance of noise-limited systems?," *19th International ITG Workshop on Smart Antennas (WSA)*, March 2015.

3. S. Bazzi, G. Dietl, and W. Utschick, "Large system analysis of interference alignment achievable rates for the MIMO interference channel," *IEEE Transactions on Signal Processing,* vol. 63, no. 6, pp. 1490-1499, March 2015.

4. S. Bazzi, G. Dietl and W. Utschick, "Maximum ratio transmission for massive MIMO: non-asymptotic analysis with limited CSIR," *10th International ITG conference on Systems, Communications, and Coding (SCC)*, Feb. 2015.

5. * S. Bazzi, G. Dietl, and W. Utschick, "Subspace precoding with limited feedback for the massive MIMO interference channel," *8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM),* June 2014.

6. * H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing base station power consumption," *IEEE Journal on Selected Areas in Communications,* vol. 32, no. 2, pp. 297-306, Feb. 2014.

7. S. Bazzi, G. Dietl, and W. Utschick, "Interference alignment with imperfect channel state information at the transmitter," *9th International Symposium on Wireless Communication Systems (ISWCS),* invited paper, August 2012.

8. S. Bazzi, G. Dietl, and W. Utschick, "Interference alignment via minimizing projector distances of interfering subspaces," *13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* June 2012.

9. S. Bazzi, G. Dietl, and W. Utschick, "Power signaling based cooperative precoding for the MISO interference channel," *13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* June 2012.

List of Figures

Bibliography

[1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for the 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26-35, May 2014.

[2] NTT DOCOMO, Inc., "DOCOMO 5G white paper. 5G radio access: requirements, concept, and technologies," July 2014.

[3] G.J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell Systems Technical Journal*, 1996.

[4] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[5] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691-1706, July 2003.

[6] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912-19, Aug. 2003.

[7] H. Weingarten, Y. Steinberg and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936-3964, Sep. 2006.

[8] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband.* Academic Press, 2nd edition, 2014.

[9] 3GPP, TR 36.819 V 11.2.0, "Coordinated multi-point operation for LTE physical layer aspects," Rel. 11.

[10] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425-2441, Aug. 2008.

[11] K. Gomadam, V. Cadambe, and S. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2008.

[12] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp.1106-1110, 1-4 Nov. 2009.

[13] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Comparison of distributed beamforming algorithms for MIMO interference networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 461-471, July 2013.

[14] S. Shamai Shitz and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. of the 53rd IEEE Vehicular Technology Conference (VTC), VTC Spring*, April 2001.

[15] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1910-1921, April 2009.

[16] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380-1408, December 2010.

[17] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas,"' *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[18] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol.30, no.1, pp.40-60, Jan. 2013.

[19] V. Marčenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math USSR Sbornik*, vol. 1, pp. 457-483, 1967.

[20] R. R. Müller, "Applications of large random matrices in communications engineering," Habilitation Thesis [Online]. Internet: http://www.iet.ntnu.no/ ralf/papers/ipsi03.pdf.

[21] A. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications.* Foundations and Trends in Communications and Information Theory, Now Publishers, 2004, vol. 1.

[22] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, no. 10, pp. 1458-1461, Oct. 1999.

[23] C. E. Shannon, "A mathematical theory of communications," *Bell Systems Technical Journal*, vol. 27, pp. 379-423, 623-656, July, October, 1948.

[24] R. V. L. Hartley, "Transmission of information," *Bell Systems Technical Journal*, vol. 7, pp. 535-563, July 1928.

[25] Andrea Goldsmith, *Wireless Communications.* Cambridge University Press, 1st edition, 2005.

[26] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge University Press, 1st edition, 2005.

[27] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. Vincent Poor, *MIMO Wireless Communications.* Cambridge University Press, 2007.

[28] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281-302, April 2003.

[29] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451-1458, Oct. 1998.

[30] B. R. Vojčić and W. M. Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Transactions on Communications*, vol. 46, no. 10, October 1998.

[31] M. Joham and W. Utschick, "Downlink processing for mitigation of intracell interference in DS-CDMA systems," in *Proc. ISSSTA*, vol. 1, Sep. 2000, pp. 15-19.

[32] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener filter for the downlink of TDD DS-CDMA systems," in *Proc. ISSSTA*, vol. 1, Sep. 2002, pp. 9-13.

[33] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-24, pp. 374-377, May 1978.

[34] M. Joham, *MIMO Systems.* Associate Institute for Signal Processing, Technical University of Munich, 2010.

[35] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[36] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, August 2005.

[37] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461-471, Feb. 2004.

[38] R. Esmailzadeh and M. Nakagawa, "Pre-RAKE diversity combination for direct sequence spread spectrum mobile communications systems," in *Proc. of the IEEE Conference on Communications (ICC)*, May 1993.

[39] H. Boche and M. Schubert, "Optimal multi-user interference balancing using transmit beamforming," *Wireless Personal Communications (WPC)*, September 2003.

[40] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18-28, August 2004.

[41] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711-1721, May 2007.

[42] S. Verdú, *Multiuser Detection.* Cambridge University Press, 1998.

[43] R. Price and P. E. Green, "A communication technique for multipath channels," *Proc. IRE*, vol. 46, no. 3, pp. 555-570, Mar. 1958.

[44] D. A. Shnidman, "A generalized Nyquist criterion and an optimum linear receiver for a pulse modulation system," *Bell Systems Technical Journal*, vol. 46, no. 9, pp. 2163-2177, Nov. 1967.

[45] D. A. George, "Matched filters for interfering signals," *IEEE Transactions on Information Theory,* vol. 11, no. 1, pp. 153-154, Jan. 1965.

[46] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* MIT Press, 1949.

[47] C. Guthy, W. Utschick, and G. Dietl, "Low-complexity linear zero-forcing for the MIMO broadcast channel," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 6, pp. 1106-1117, Dec. 2009.

[48] P. Tejera, W. Utschick, G. Bauch, and J.A. Nossek, “Subchannel allocation in multiuser multiple-input multiple-output systems," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4721-4733, Oct. 2006.

[49] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 569-570, Sep. 1975.

[50] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Transactions on Information Theory*, vol. 27, pp. 786-788, Nov. 1981.

[51] R. H. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534-5562, Dec. 2008.

[52] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49-60, Jan. 1981.

[53] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, USA, Oct./Nov. 2006.

[54] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[55] J. Jose, A. Ashkhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640-2651, Aug. 2011.

[56] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, issue 2, pp. 264-273, Feb. 2013.

[57] D. Neumann, A. Gründinger, M. Joham, and W. Utschick, "Pilot coordination for large-scale multi-cell TDD systems," in *Proc. of the International ITG Workshop on Smart Antennas (WSA)*, Mar. 2014.

[58] D. Neumann, M. Joham, and W. Utschick, "Suppression of pilot-contamination in massive MIMO systems," in *Proc. of the 15th IEEE International Symposium on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014.

[59] D. Neumann, A. Gründinger, M. Joham, and W. Utschick, "On the amount of training in coordinated massive MIMO systems," in *Proc. of the 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014.

[60] S. Bazzi and W. Xu, "Mitigating inter-cell pilot interference via network-based greedy sequence selection and exchange," *IEEE Wireless Communications Letters*, to appear.

[61] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. of the 46th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, USA, Nov. 2012.

[62] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[63] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, no. 4, April 2003.

[64] P. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K.I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the Shannon bound," in *Proc. of the 65th IEEE Vehicular Technology Conference (VTC), VTC Spring*, April 2007.

[65] S. Bazzi, G. Dietl, and W. Utschick, "Power signaling based cooperative precoding for the MISO interference channel," in *Proc. of the 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012.

[66] S. Jafar and M. Fakhreddin, "Degrees of freedom for the MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2637-2642, July 2007.

[67] T. Gou and S. Jafar, "Degrees of freedom of the K user $M \times N$ MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6040-6057, Dec. 2010.

[68] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: interference alignment, decomposition, and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, pp. 3457-3470, Aug. 2008.

[69] R. Tresch, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," *15th IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 277-280, Sep. 2009.

[70] S. W. Peters and R. W. Heath, Jr., "Interference alignment via alternating minimization," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2445-2448.

[71] S. Bazzi, G. Dietl, and W. Utschick, "Interference alignment via minimizing projector distances of interfering subspaces," in *Proc. of the 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012.

[72] H. G. Ghauch and C. B. Papadias, “Interference alignment: a one-sided approach," *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011.

[73] A. Barg and D. Y. Nogin, “Bounds on packings of spheres in the Grassmann manifold”, *IEEE Transactions on Information Theory*, vol. 48, pp. 2450-2454, Sep. 2002.

[74] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Johns Hopkins, 1996.

[75] J. H. Manton, “Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 635-650, Mar. 2002.

[76] L. Armijo, “Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1-4, 1966.

[77] S. Bazzi, G. Dietl, and W. Utschick, "Interference alignment with imperfect channel state information at the transmitter," *9th International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2012.

[78] S. Bazzi and G. Dietl, "Method for reducing interference at a terminal of a wireless cellular network, wireless cellular network, node and central node of a wireless network," *European Patent Application EP 12155149.3,* pending.

[79] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4771-4782, Sep. 2010.

[80] G. Bresler, D. Cartwright, and D. Tse, "Feasibility of interference alignment for the MIMO interference channel: the symmetric square case," *IEEE Information Theory Workshop (ITW)*, 2011.

[81] L. Ruan, V. K. N. Lau, and M. Z. Win, "The feasibility conditions for interference alignment in MIMO networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2066-2077, April 2013.

[82] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Transactions on Wireless Communications*, vol.9, no.11, pp.3480-3489, Nov. 2010.

[83] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization problem," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4278-4288, August 2012.

[84] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias, "Reweighted nuclear norm approach for interference alignment," *IEEE Transactions on Wireless Communications*, vol. 61, no. 9, pp. 3754-3765, Sep. 2013.

[85] C. Shi, R. A. Berry, and M. L. Honig, "Distributed interference pricing with MISO Channels," in Proc. *46th Annual Allerton Conference 2008*, Urbana-Champaign, IL, pp. 539-546, Sept. 2008.

[86] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, "Distributed interference pricing for the MIMO interference channel," *IEEE International Conference on Communications (ICC)*, June 2009.

[87] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Distributed resource allocation schemes," *IEEE Signal Processing Magazine*, pp. 53-63, Sept. 2009.

[88] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, Mar. 2004.

[89] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, Mar. 2008.

[90] G. Owen, *Game Theory.* Academic press, 1995.

[91] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission / reception techniques for LTE-advanced," *IEEE Wireless Communications Magazine,* June 2010.

[92] P. de Kerrert and D. Gesbert, "Interference alignment with incomplete CSIT sharing," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2563-2573, May 2014.

[93] M. Rezaee and M. Guillaud, "Limited feedback for interference alignment in the K-user MIMO interference channel," *IEEE Information Theory Workshop (ITW)*, 2012.

[94] R. T. Krishnamachari and M. K. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," *International Symposium on Information Theory (ISIT)*, pp. 619-623, June 2010.

[95] J. Kim, S. Moon, S. Lee, and I. Lee, "A new channel quantization strategy for MIMO interference alignment with limited feedback," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 358-366, Jan. 2012.

[96] O. El Ayash and R. W. Heath, "Grassmannian differential limited feedback for interference alignment," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6481-6494, Dec. 2012.

[97] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Asymptotic analysis of LMMSE multiuser receivers for multi-signature multicarrier CDMA in Rayleigh fading," *IEEE Transactions on Communications*, vol. 52, no. 6, pp. 964-972, June 2004.

[98] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, pp. 622-640, March 1999.

[99] R. R. Müller, "Multiuser receivers for randomly spread signals: fundamental limits with and without decision-feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 268-283, Jan. 2001.

[100] L. Cottatellucci, R. R. Müller, and M. Debbah, "Asynchronous CDMA systems with random spreading-part I: fundamental limits," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1477 - 1497, April 2010.

[101] L. Cottatellucci, R. R. Müller, and M. Debbah, "Asymptotic design and analysis of linear detectors for CDMA systems," *International Symposium on Information Theory (ISIT)*, 2004.

[102] K. Takeuchi, M. Vehkapera, T. Tanaka, and R. R. Müller, "Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1385 - 1412, March 2012.

[103] C. Guthy, W. Utschick, and M. L. Honig, "Large system analysis of sum capacity in the Gaussian MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 149-159, February 2013.

[104] C. Guthy, W. Utschick, and M. L. Honig, "Large system analysis of projection based algorithms for the MIMO broadcast channel," *IEEE International Symposium on Information Theory (ISIT)*, pp. 2128-2132, June 2010.

[105] D. A. Schmidt, W. Utschick, and M. L. Honig, "Large system performance of interference alignment in single-beam MIMO networks," *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2010.

[106] A. M. Masucci, A. M. Tulino, and M. Debbah, "Asymptotic analysis of uplink interference alignment in Ricean small cells," [Online]. Internet: http://www.supelec.fr/offres/file_inline_src/342/342_P_11867_129.pdf.

[107] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: large system analysis," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2079-2106, April 2012.

[108] S. Bazzi, G. Dietl, and W. Utschick, "Large system analysis of interference alignment achievable rates for the MIMO interference channel," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1490-1499, March 2015.

[109] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill series in electrical engineering, 3rd edition, 1991.

[110] G. B. Thomas and R. L. Finney, *Calculus and Analytical Geometry.* Addison-Wesley, 1995.

[111] I. Santamaria, O. Gonzalez, R. W. Heath, and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2010.

[112] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley Series in Telecommunications and Signal Processing, 2nd edition, 2006.

[113] S. Bazzi, G. Dietl, and W. Utschick, "Maximum ratio transmission for massive MIMO: non-asymptotic analysis with limited CSIR," *10th International ITG Conference on Systems, Communications, and Coding (SCC)*, February 2015.

[114] M. Castañeda, A. Mezghani, and J. A. Nossek, "Optimal resource allocation in the downlink/uplink of single-user MISO/SIMO FDD systems with limited feedback," *10th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2009.

[115] C. Au-Yeung and D. Love, “Design and analysis of two-way limited feedback beamforming systems," *Forty-First Asilomar Conference on Signals, Systems and Computers*, 4-7 Nov. 2007.

[116] M. Kobayashi, G. Caire, and N. Jindal, “How much training and feedback are needed in MIMO broadcast channels?," *IEEE International Symposium on Information Theory (ISIT)*, July 2008.

[117] O. El Ayash, A. Lozano, and R. W. Heath, “On the overhead of interference alignment: training, feedback, and cooperation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 4192-4203, Nov. 2012.

[118] E. A. Jorswieck and E. G. Larsson, “Competition versus cooperation on the MISO interference channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, Sep. 2008.

[119] E. A. Jorswieck, E. G. Larsson, and D. Danev, “Complete characterization of the Pareto boundary for the MISO interference channel,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292-5296, Oct. 2008.

[120] S. Bazzi, G. Dietl, and W. Utschick, "How many transmit antennas emulate the performance of noise-limited systems?," *19th International ITG Workshop on Smart Antennas (WSA)*, March 2015.

[121] O. El Ayash, S. Peters, and R. Heath, "The feasibility of interference alignment over measured MIMO-OFDM channels," *IEEE Transactions on Vehicular Communications*, vol. 59, no. 9, pp. 4309-4321, Nov. 2010.

[122] S. Lee, A. Gerstlauer, and R. W. Heath, "Distributed real-time implementation of interference alignment with analog feedback," *IEEE Transactions on Vehicular Technology*, to be published.

[123] R. Mungara, D. Morales-Jimenez, and A. Lozano, "System-level performance of interference alignment," *IEEE Transactions on Wireless Communications*, to be published.

[124] H. Anton and C. Rorres, *Elementary Linear Algebra, Applications Version.* John Wiley and Sons, Inc., 9th edition, 2005.

[125] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1st edition, 1985.