THE UNIVERSITY *of* EDINBURGH

Advanced Interference Management Techniques for Future Wireless Networks

SEYED MORTEZA RAZAVI

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
2014

Abstract

In this thesis, we design advanced interference management techniques for future wireless networks under the availability of perfect and imperfect channel state information (CSI). We do so by considering a generalized imperfect CSI model where the variance of the channel estimation error depends on the signal-to-noise ratio (SNR).

First, we analyze the performance of standard linear precoders, namely channel inversion (CI) and regularized CI (RCI), in downlink of cellular networks by deriving the received signal-to-interference-plus-noise ratio (SINR) of each user subject to both perfect and imperfect CSI. In this case, novel bounds on the asymptotic performance of linear precoders are derived, which determine how much accurate CSI should be to achieve a certain quality of service (QoS). By relying on the knowledge of error variance in advance, we propose an adaptive RCI technique to further improve the performance of standard RCI subject to CSI mismatch.

We further consider transmit-power efficient design of wireless cellular networks. We propose two novel linear precoding techniques which can notably decrease the deployed power at transmit side in order to secure the same average output SINR at each user compared to standard linear precoders like CI and RCI.

We also address a more sophisticated interference scenario, i.e., wireless interference networks, wherein each of the K transmitters communicates with its corresponding receiver while causing interference to the others. The most representative interference management technique in this case is interference alignment (IA). Unlike standard techniques like time division multiple access (TDMA) and frequency division multiple access (FDMA) where the achievable degrees of freedom (DoF) is one, with IA, the achievable DoF scales up with the number of users. Therefore, in this thesis, we quantify the asymptotic performance of IA under a generalized CSI mismatch model by deriving novel bounds on asymptotic mean loss in sum rate and the achievable DoF. We also propose novel least squares (LS) and minimum mean square error (MMSE) based IA techniques which are able to outperform standard IA schemes under perfect and imperfect CSI. Furthermore, we consider the implementation of IA in coordinated networks which enable us to decrease the number of deployed antennas in order to secure the same achievable DoF compared to standard IA techniques.

Lay Summary

The next generation of wireless cellular networks will provide access to high-speed data applications such as video streaming and internet browsing, together with relatively low-rate applications such as voice and therefore enable the mobile users to experience much higher quality of service (QoS). While there is an increasingly high demand for wireless services, radio resources (e.g., bandwidth spectrum and transmit power) are often scarce and resource allocation is very conservative. Therefore, spectrum and energy efficient communications are major design goals for future wireless networks. In this thesis, we address these concerns by proposing advanced interference management techniques to increase the QoS in wireless cellular networks while keeping the usage of radio resources as low as possible.

We do so by first analyzing and improving the performance of standard linear precoders in downlink of cellular networks when imperfect channel state information (CSI) is available at base station (BS). This becomes particularly important, since from the practical point of view, the availability of perfect CSI at BS is not pragmatic. We also consider energy-efficient design of wireless networks. By considering a fixed transmit power at BS, we introduce novel techniques which can deliver more signal-to-interference-plus-noise ratio (SINR) to each mobile user compared to standard linear precoders. Equivalently, we propose two novel transmit-power efficient linear precoding schemes which enable us to decrease the deployed power at BS to meet an averagely constant output SINR at each mobile user in comparison with two well-known linear precoding techniques, namely channel inversion (CI) and regularized CI (RCI), in downlink of cellular networks.

Recently, an interference mitigation technique, named interference alignment (IA), has been proposed which can significantly increase the throughput in wireless cellular networks compared to standard techniques, like time division multiple access (TDMA) and frequency division multiple access (FDMA), and consequently achieves much better spectral efficiency. Therefore, in this thesis, we further consider performance analysis and improvement of IA under both perfect and imperfect CSI by deriving novel bounds on the asymptotic performance of IA and proposing new IA algorithms which outperform standard IA schemes.

Declaration of Originality

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. The presented materials in this thesis are entirely my own and have been done in institute for digital communications, school of engineering, the University of Edinburgh.

Seyed Morteza Razavi

Acknowledgements

Pursuing a PhD is a long journey with many unprecedented encounters and it can not be ended up without help. First, I would like to deeply thank my family to support me before and throughout this journey. Especially, I would like to highlight the infinite impact factor of my parents who literally devoted themselves to provide a better future for me.

My deepest gratitude goes to my supervisor, Dr Tharmalingam Ratnarajh, who gave me this opportunity to pursue my aspirations and gave me enough degrees of freedom in research. Whilst I was not confident, he believed in me and continuously pushed me to improvement not only in my research but also in my professional attitude. I greatly appreciate his kind consideration in removing the obstacles that I came across throughout my research.

I likewise owe the significant progress during my research to my former colleague Dr Christos Masouros.

Finally, I would like to acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number: HíATUS-265578.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

Alt-Min	Alternating Minimization
AP	Access Point
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BS	Base Station
CAGR	Compound Annual Growth Rate
CCDF	Complementary Cumulative Distribution Function
CI	Channel Inversion
CSI	Channel State Information
CSIT	Channel State Information at Transmitter
dB	Decibel
DoF	Degrees of Freedom
DPC	Dirty Paper Coding
ESPAR	Electronically Steerable Passive Array Radiator
FDMA	Frequency Division Multiple Access
GSM	Global System for Mobile
IA	Interference Alignment
IBC	Interference Broadcast Channel
IC	Interference Channel
i.i.d.	independent and identically distributed
LS	Least Squares
LZF	Linear Zero Forcing
Max-SINR	Maximum Signal-to-Interference-plus-Noise Ratio
MIMO	Multiple-Input Multiple-Output
Min-WLI	Minimum Weighted Leakage Interference
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Square Error
MT	Mobile Terminal
PA	Phase Alignment

PDF	Probability Density Function
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
QR	Orthogonal-triangular decomposition
RCI	Regularized Channel Inversion
RCRM	Rank Constrained Rank Minimization
RHS	Right Hand Side
RPA	Regularized Phase Alignment
RRH	Remote Radio Head
SER	Symbol Error Rate
SIC	Successive Interference Cancelation
SINR	Signal-to-Interference-plus-Noise Ratio
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
TDMA	Time Division Multiple Access
V-BLAST	Vertical-Bell Laboratories Layered Space-Time Architecture
ZF	Zero Forcing

List of Symbols

$\lvert\cdot\rvert$	Element-wise absolute value
$\lVert\cdot\rVert_2$	Vector 2-norm
$\lVert\cdot\rVert_*$	Nuclear norm
$[\cdot]_{\ell\star}$	ℓth row of a matrix
$[\cdot]_{\ell,x}$	(ℓ, x)th element of a matrix
$\approx$	Approximation
$\Longleftarrow$	Assignment through an in-place manner
$\Longrightarrow$	Results in
$\succeq$	Positive semi-definiteness
∂	Differentiation
∞	Infinity
$\propto$	Linear proportionality
$\odot$	Hadamard product
a	Scalar
$\mathbf{a}$	Vector
$\mathbf{A}$	Matrix
blkerr	Block error rate
$\mathrm{blkdiag}$	Block diagonal operator
$\mathbf{c}$	Transmitted symbols chosen from a desired constellation
C	A constant
$\mathcal{C}$	Capacity
$\mathbb{C}$	Complex numbers set
χ	Chi-squared random variable
d	Degrees of freedom
$\det$	Determinant of a square matrix
diag	Diagonal operator
ε	Multiplicative factor of the identity matrix
ϵ	Arbitrarily small number
η	Output SINR

$\mathbb{E}$	Expectation operator
$\mathbf{E}$	Error matrix
eig	Eigenvalue decomposition
$\mathrm{eig}\,(\cdot)_{1:d}$	Eigenvectors related to the d smallest eigenvalues
$\mathrm{eig}\,(\cdot)_{d+1:M}$	Eigenvectors related to the $M-d$ largest eigenvalues
erfc	Complementary error function
f	Multiplicative factor
$\mathcal{F}$	Demodulating operator
γ	Nominal SNR
$\Delta\gamma$	Power offset
Γ	Gamma function
g	Scaling factor in cellular downlink
$(\cdot)^{\mathsf{H}}$	Hermitian transpose
$h_{k,j}$	(k,j)th element of channel matrix
$\mathbf{H}_{k,j}$	Channel from transmitter j to receiver k
$\widehat{\mathbf{H}}_{k,j}$	Estimated channel from transmitter j to receiver k
$\breve{\mathbf{H}}_{k,j}$	Auxiliary error matrix
$\mathbf{I}$	Identity matrix
$\Im$	Imaginary part of a complex number
K	Number of transceiver pairs (users)
λ	Eigenvalue
$\log_2$	Logarithm to the base 2
$\lim$	Limit
$\mathcal{M}$	Number of constellation point
M	Number of receive antennas
$\min$	Minimize
N	Number of transmit antennas
$\mathcal{N}$	Normal random variable
$\mathcal{N}_{\mathbb{C}}$	Complex normal random variable
N_e	Number of equations
N_v	Number of variables
orth	Orthonormal basis of a (sub)space
$\mathbf{\Psi}$	Precoding matrix
P	Transmit power

pinv	pseudo-inverse
$\mathbf{Q}$	Covariance matrix
$\rho_{k,j}$	(k,j)th element of covariance matrix of channel
R	Sum rate
$\mathbf{R}$	Covariance matrix
$\Re$	Real part of a complex number
ΔR	Mean loss in sum rate
rank	Rank of a matrix
σ^2	Noise variance
$\mathbf{s}$	Transmitted signal in single-cell scenario
span	Range of a (sub)space
$\mathrm{s.\,t.}$	Subject to
$(\cdot)^\top$	Transpose
τ	Channel estimation error variance
Tr	Trace operator
$\mathbf{T}$	Covariance matrix
$\mathbf{U}$	Unitary matrix
$\mathbf{V}$	Unitary matrix
var	Variance
vec	Vector form of a matrix
ξ	Power efficiency
$\mathbf{x}$	Transmitted signal in interference channels
$\mathbf{y}$	Received signal
$\mathbf{z}$	Noise vector

Chapter 1

Introduction

1.1 Background

Wireless communications have been going through many advances to become what it is today; from its primitive shapes in the form of the transmission of simple codes like Morse code to much more complicated shapes like satellite and cellular communications. Guglielmo Marconi was the pioneer in this filed who brought the attention of the world to this spectacular technology by carrying out series of experiments and demonstrations during 1897. Since then, wireless communications have been significantly evolved through various efforts from many researches across the globe.

Perhaps, it is fair to say that the most astonishing advancements of wireless communications have been appeared within the area of mobile and cellular communications which directly affect our daily lives. This becomes more prominent since instead of the traditional form of cellular communications meant solely for voice, nowadays we can easily have access to high-speed data applications such as video streaming and internet browsing on our smart phones or tablets.

To manifest this exponential growth of data traffic within cellular networks, it suffices to mention that the overall mobile data traffic will grow at a compound annual growth rate (CAGR) of 66 percent from 2012 to 2017, and it is expected to grow to 11.2 exabytes per month by 2017, a 13-fold increase over 2012 [1] (see Fig. 1.1). This exponential growth of demand for data requires advanced interference management techniques which are going to be addressed in this thesis.

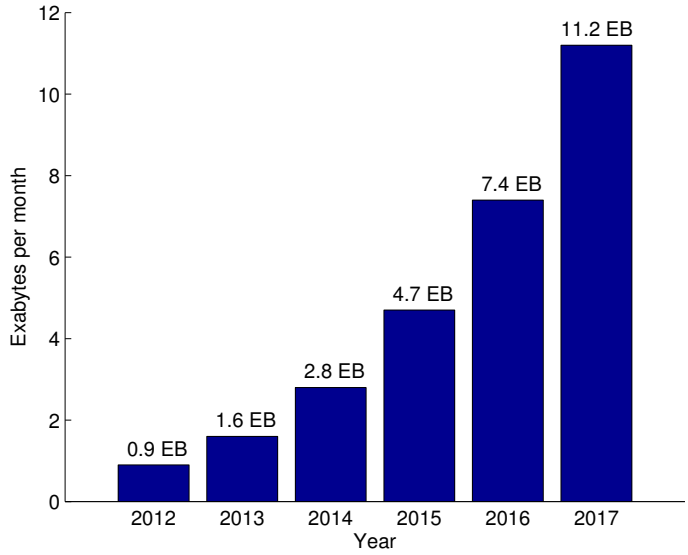

Figure 1.1: *Global Mobile Data Traffic, 2012 to 2017 [1].*

1.2 Contributions

In all communication scenarios, the availability of perfect channel state information (CSI) at transmit and/or receive side is vital to achieve the desired performance. For example, to achieve the sum capacity in downlink of cellular networks (and consequently achieve the maximum data rate), perfect CSI must be available at the base station (BS). From the practical point of view, however, having access to perfect CSI is not readily possible and this compromises the intrinsic capabilities of a communication system. Therefore, it is essential to evaluate and enhance the performance of communication techniques under CSI mismatch. In this thesis, we aim to do so by first introducing a generalized imperfect CSI model which accommodates a variety of distinct scenarios (like CSI feedback and reciprocal channels). We then focus on the performance analysis of prevalent and promising interference management techniques in future wireless networks under this imperfect CSI model. The most important research contributions of this thesis can be summarized as follows:

- We derive novel bounds on asymptotic mean loss in sum rate and achievable degrees of freedom (DoF) as a function of channel estimation error variance for multiuser multiple-input multiple-output (MIMO) downlink. We then propose an adaptive regularized channel inversion (RCI) precoding technique, which with the knowledge of error variance in advance, can significantly improve the performance of standard

RCI subject to the availability of imperfect CSI at BS.

- Similar to the case of RCI which improves the performance of channel inversion (CI), we propose the idea of regularized phase alignment (RPA) by deriving optimum regularization parameters. It is shown that the proposed RPA precoding outperforms standard linear precoders (i.e., CI, RCI, and phase alignment (PA)) under both perfect and imperfect CSI. Also it is shown that PA and RPA precoding techniques are power-efficient alternatives for CI and RCI precoding, respectively.
- We derive novel bounds on asymptotic mean loss in sum rate and achievable DoF as a function of channel estimation error variance for wireless interference networks with interference alignment (IA). We then propose an adaptive maximum signal-to-interference-plus-noise-ratio (Max-SINR) algorithm, which with the knowledge of error variance in advance, can significantly improve the performance of original Max-SINR under the availability of imperfect CSI. Moreover, two novel IA schemes based on least squares (LS) and minimum mean square error (MMSE) criteria are also proposed. It is shown that the proposed LS and MMSE based IA schemes outperform standard IA techniques under both perfect and imperfect CSI.
- We propose novel IA techniques for wireless interference networks with partially coordinated receivers in order to decrease the number of deployed antennas. In this case, on average, half of the total decoded data is needed to be shared by receive nodes. It is shown that even with this reduced number of deployed antennas, the proposed schemes are able to achieve notable performance compared to standard IA techniques under both perfect and imperfect CSI.

1.3 Thesis Organization

Chapter 2 provides an overview of wireless communications and the state of the art interference management techniques. In Chapter 3, we consider performance analysis of CI and RCI precoding under imperfect CSI by deriving novel bounds regarding their asymptotic performance. We also propose an adaptive RCI technique to improve the performance of standard RCI under CSI mismatch. In Chapter 4, we propose two novel transmit-power efficient linear precoding schemes which enable us to decrease the power at BS to secure the same average output SINR at each user compared to standard linear precoders like CI and RCI. In Chapter 5, we consider performance analysis of constant MIMO IA under CSI mismatch by deriving asymptotic bounds on mean loss in sum rate and the achievable

DoF, and also proposing an adaptive Max-SINR algorithm. By relying on the concept of partial coordination in Chapter 6, we propose novel IA techniques which enable us to decrease the deployed number of transmit and/or receive antennas compared to standard IA techniques to secure the same DoF. Section 7 contains novel LS and MMSE based IA techniques which are shown to achieve better performance than standard IA methods under both perfect and imperfect CSI. Finally, Section 8 contains concluding remarks and further notes to expand the presented materials within this thesis for future work.

Chapter 2

An Overview of Wireless Communications

2.1 Introduction

In this chapter, we reiterate some basic concepts of wireless communications along with some key metrics for evaluating their performance. First, we consider the point-to-point communications and we revisit the advantages of MIMO signaling over single-input single-output (SISO) systems by emphasizing on the achievable DoF. Next, we reintroduce the idea of transmit beamforming (or precoding) for single-cell broadcast channels. We then discuss why single-cell communication techniques are not spectrally efficient and we therefore overview the state of the art interference management techniques in multi-cell (or wireless interference) networks along with their limitations. We finally place our focus on one of the promising interference management techniques, namely IA, by pointing out some of its key features.

2.2 MIMO Communications

The simplest form of wireless communications includes one single-antenna transmitter and one single-antenna receiver. Affected by the surrounding environment, the transmitted signal goes through various paths from the transmitter to the receiver and consequently may undergo a change in amplitude, phase and frequency. In this case, the received multipath components may add up destructively and this results in a degraded quality of the transmitted signal at the receiver. The amplitude variations of the received signals are

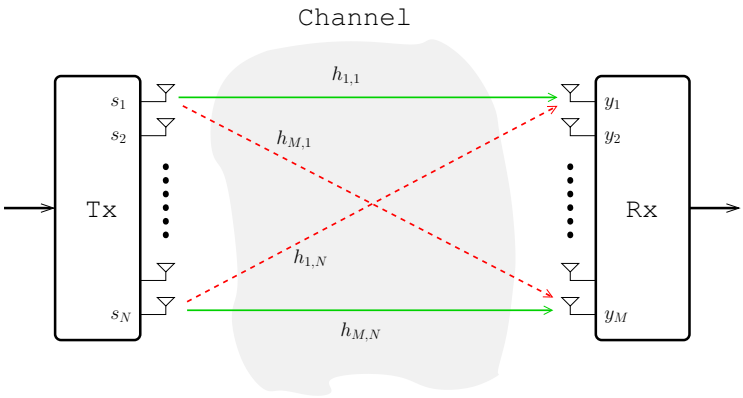

Figure 2.1: $M \times N$ *point-to-point MIMO communications.*

referred as *fading* [2]. Therefore, by relying on the knowledge of CSI, the transmit beamforming tries to steer the transmitted signal into specific paths so that the received multipath components at the intended receiver add up constructively and better sinal reception can be experienced.

Over the past years, the idea of MIMO communications has attracted lots of attention due to its inherent capability to overcome the destructive fading and to deliver higher quality of service (QoS) to the receiver [2]. In this case, multiple antennas at both transmit and receive side of a communication link can be deployed. Fig. 2.1 illustrates a point-to-point MIMO system where the transmitter and receiver have N and M antennas, respectively. The received signal can then be shown as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M,1} & h_{M,2} & \cdots & h_{M,N} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix} + \begin{pmatrix} z_1 \\ \vdots \\ z_M \end{pmatrix}$$
$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{z} \tag{2.1}$$

where $s_j,\ j = 1, \ldots, N$, is the transmitted signal from the jth antenna, $h_{k,j}$ is the time-variant channel response between the jth transmit antenna and the kth receive antenna, and $z_k,\ k = 1, \ldots, M$, is the noise at the kth receive antenna. In a rich scattering environment, the entries of $\mathbf{H}$, i.e., $h_{k,j},\ k = 1, \ldots, M,\ j = 1, \ldots, N$, are considered to be independent and identically distributed (i.i.d.) zero mean unit variance circularly symmetric Gaussian

random variables which denote Rayleigh fading channel model, i.e.,

$$h_{k,j} = \mathcal{N}\left(0, \frac{1}{2}\right) + \sqrt{-1} \cdot \mathcal{N}\left(0, \frac{1}{2}\right) \tag{2.2}$$

where $\mathcal{N}\left(0, \frac{1}{2}\right)$ designates a real Gaussian random variable with mean zero and variance $\frac{1}{2}$. Moreover, $\mathbf{z}$ is the additive white Gaussian noise (AWGN) vector which its entries are considered to be independent zero mean Gaussian random variables with variance σ^2, that is

$$z_k = \mathcal{N}\left(0, \frac{\sigma^2}{2}\right) + \sqrt{-1} \cdot \mathcal{N}\left(0, \frac{\sigma^2}{2}\right) \tag{2.3}$$

which results to the fact that $\mathbb{E}\left\{\mathbf{z}\mathbf{z}^{\mathsf{H}}\right\} = \sigma^2 \mathbf{I}$, where $\mathbb{E}\left\{\cdot\right\}$ denotes the expectation operator, and the superscript $(\cdot)^{\mathsf{H}}$ designates the Hermitian transpose.

Compared to SISO, MIMO systems provide remarkable performance gains as follows [2]:

- *Power gain*: by judiciously steering the transmit and/or receive signals, we can increase the signal-to-noise ratio (SNR) at the receiver. This can be done by coherently combining the received signals and/or by allocating more power to the transmit antenna with the better gain which further enhances the quality of the reception. It is worthwhile to note that the power gain can be gleaned in single-input multiple-output (SIMO) systems as well. In the case of the availability of CSI at the transmitter, multiple-input single-output (MISO) systems are also able to deliver power gain.
- *Diversity gain*: multiple replicas of the transmitted signal are received by the sink node. Consequently, it becomes more likely that at least one of these copies does not undergo a deep fading. This results in a more reliable recovery of the transmitted signal. The diversity gain can be gleaned in both MISO and SIMO systems as well.
- *Interference suppression*: In multiuser systems and under the same usage of time and frequency resources, while one transmitted data stream is just meant for one specific user, it causes interference to the other unintended users. In this case, deploying multiple antennae enables us to efficiently null out this interference. In the case of downlink communications, this interference gets precanceled by precoding, which will be discussed in detail within section 2.4.
- *Multiplexing gain (or DoF)*: the transmitter is now able to simultaneously send parallel independent data streams to the receiver using the same radio frequency bands

for all data streams. The DoF can be further interpreted as the pre-log factor of the sum rate, i.e.,

$$R = d \log_2 (\text{SNR}) + o (\log_2 (\text{SNR})) \tag{2.4}$$

wherein R is the achievable sum rate and d denotes DoF. In other words, at high SNRs, the achievable sum rate linearly scales with DoF. Unlike power gain and diversity gain which can be gleaned for MISO and SIMO systems, the multiplexing gain is only achievable in MIMO systems. In the following section, we provide more discussions regarding the role of the multiplexing gain.

2.3 Performance Metrics

Sum capacity (or the achievable sum rate) is one of the most important key metrics to evaluate the performance of communication systems. In the remainder of this section, we provide an overview of the deterministic and ergodic sum capacity at low and high SNR regimes.

2.3.1 Deterministic Capacity

For a simple SISO system, the capacity can be shown as

$$\mathcal{C}_{\texttt{SISO}} = \log_2 (1 + \gamma |h|) \quad \text{bits per channel use} \tag{2.5}$$

where $\gamma = \frac{P}{\sigma^2}$ is the operational (or nominal) SNR, P is the transmit power, σ^2 is the noise power at the receiver, h is the channel response between the transmit antenna and the receive antenna and $|\cdot|$ represents element-wise absolute value. For the MISO and SIMO channels, the sum capacity can be represented by

$$\mathcal{C}_{\texttt{SIMO}|\texttt{MISO}} = \log_2 (1 + \gamma \|\mathbf{h}\|_2) \quad \text{bits per channel use} \tag{2.6}$$

where $\|\cdot\|_2$ is the vector 2-norm. In the case of the MISO channel, $\mathbf{h}$ is a $1 \times N$ vector and for the case of the SIMO channel, $\mathbf{h}$ is a $M \times 1$ vector.

The sum capacity of the MIMO systems by considering the availability of CSI at the receiver can be shown as

$$\mathcal{C}_{\texttt{MIMO}} = \max_{\mathbf{Q}:\ \text{Tr}[\mathbf{Q}] \leq P} \log_2 \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{H} \mathbf{Q} \mathbf{H}^{\mathsf{H}} \right) \quad \text{bits per channel use} \tag{2.7}$$

where $\det(\cdot)$ denotes the determinant of a square matrix and $\mathbf{Q}$ is the covariance matrix of the transmitted data stream. Note that the above optimization depends on the knowledge of the transmitter regarding the channel matrix $\mathbf{H}$.

If full CSI is available at transmitter, the optimum transmission scheme can be obtained by assigning more power to the subchannels with more gains. This strategy is also referred as waterfilling [2]. If the transmitter has no knowledge about the CSI, the optimum transmission scheme is equal power allocation among all transmit antennas. In this case the sum capacity is equal to

$$\mathcal{C}_{\text{Eq.}} = \log_2 \det\left(\mathbf{I} + \frac{\gamma}{N}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right) \quad \text{bits per channel use} \tag{2.8}$$

2.3.2 Ergodic Capacity

Apart from deterministic capacity which deals with one specific realization of $\mathbf{H}$, another important performance metric is ergodic capacity which considers the effects of channel fading. In the case of time-variant channels, $\mathbf{H}$ becomes a random matrix, and consequently the ergodic capacity is defined as the maximum achievable rate averaged over all channel realizations as follows:

$$\mathcal{C}_{\text{Erg.}} = \max_{\mathbf{Q}:\ \text{Tr}[\mathbf{Q}]\le P} \mathbb{E}\left\{\log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathsf{H}}\right)\right\} \quad \text{bits per channel use} \tag{2.9}$$

It has been shown that in the case of the i.i.d. Rayleigh fading model, $\mathbf{Q} = \frac{P}{N}\mathbf{I}$ is the optimal covariance matrix, i.e., equal power allocation among all transmit antennas is the optimal transmission scheme. With equal powers, the resulting sum capacity is

$$\begin{aligned}
\mathcal{C} &= \mathbb{E}\left\{\log_2 \det\left(\mathbf{I} + \frac{\gamma}{N}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right\} \\
&= \mathbb{E}\left\{\sum_{\ell=1}^{d} \log_2\left(1 + \frac{\gamma}{N}\lambda_\ell\right)\right\} \\
&= \sum_{\ell=1}^{d} \mathbb{E}\left\{\log_2\left(1 + \frac{\gamma}{N}\lambda_\ell\right)\right\} \quad \text{bits per channel use}
\end{aligned} \tag{2.10}$$

where $d = \text{rank}(\mathbf{H})$ is the achievable DoF, and λ_ℓ is the ℓth largest eigenvalue of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$. If $\mathbf{H}$ has i.i.d. elements chosen from a continuous distribution (like Rayleigh), we have $d = \min(M, N)$.

It is worthwhile to point out that by considering i.i.d. Gaussian input signaling, the

achievable sum rate can be similarly shown as

$$R \approx \sum_{\ell=1}^{d} \mathbb{E}\left\{\log_2\left(1+\eta_\ell\right)\right\} \qquad \text{bits per channel use} \tag{2.11}$$

where η_ℓ is the output (or received) SNR per DoF and it heavily depends on the channel statistics and also the employed communication technique. The value of η_ℓ for some well-known communication scenarios will be determined in the upcoming chapters.

2.3.3 High SNR Regime

By considering the ergodic capacity in (2.10), we have

$$\sum_{\ell=1}^{d} \log_2\left(1+\frac{\gamma}{N}\lambda_\ell\right) \overset{①}{\leq} d\,\log_2\left(1+\frac{\gamma}{N}\left[\frac{1}{d}\sum_{\ell=1}^{d}\lambda_\ell\right]\right) \tag{2.12}$$

where ① is due to the Jensen's inequality. Therefore at high SNRs, the capacity for the i.i.d. Rayleigh channel can be given by

$$\mathcal{C} \approx d\,\log_2\left(\frac{\gamma}{N}\right) + \sum_{\ell=1}^{d}\mathbb{E}\left\{\log_2\lambda_\ell\right\} \tag{2.13}$$

Since $\mathbb{E}\left\{\log_2\lambda_\ell\right\} > -\infty$, the achievability of d DoF is guaranteed [2]. In fact, we have

$$\sum_{\ell=1}^{d}\mathbb{E}\left\{\log_2\lambda_\ell\right\} = \sum_{\ell=|N-M|+1}^{\max(M,N)}\mathbb{E}\left\{\log_2\chi^2_{2\ell}\right\} \tag{2.14}$$

where $\chi^2_{2\ell}$ is a χ-square distributed random variable with 2ℓ degrees of freedom. Therefore, the sum capacity can be represented as

$$\mathcal{C} = d\log_2\gamma + o\left(\log_2\gamma\right) \tag{2.15}$$

which has the same exact for as the achievable sum rate in (2.4). In other words, the sum capacity and the achievable sum rates are linearly increasing with the DoF at high SNRs. Note that the number of DoF is restricted to the minimum of the number of transmit and receive antennas. This further implies that to achieve the multiplexing gain of more than one, we need multiple transmit and multiple receive antennas.

2.3.4 Low SNR Regime

At low SNRs, the sum capacity in (2.10) can be given as

$$\begin{aligned} \mathcal{C} &= \sum_{\ell=1}^{d} \mathbb{E}\left\{\log_2\left(1+\frac{\gamma}{N}\lambda_\ell\right)\right\} \\ &\overset{②}{\approx} \sum_{\ell=1}^{d} \frac{\gamma}{N}\mathbb{E}\left\{\lambda_\ell\right\}\log_2 e \\ &= \frac{\gamma}{N}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}}\right]\right\}\log_2 e \\ &\overset{③}{=} M\gamma\log_2 e \quad \text{bits per channel use} \end{aligned} \tag{2.16}$$

where ② follows the fact that for sufficiently small a we have $\log_2(1+a) \approx a\,\log_2 e$, and ③ is due to the fact that for i.i.d. Rayleigh fading channel we have $\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}}\right]\right\} = MN$ [3] where $\mathrm{Tr}\left[\cdot\right]$ denotes trace operator.

Thus at low SNRs, an $M \times N$ MIMO system yields a power gain of M over a single antenna system. Interestingly, it can be seen that by increasing the number of transmit antennas, the power gain is not boosted, except for the case where the channel state is available at the transmitter and the waterfilling strategy is applicable to attain power gain.

By considering (2.13) and (2.16), we can conclude that:

1. At high SNRs, the capacity is approximately equal (up to an additive constant) to $d\,\log_2\gamma$ bits per channel use where $d = \min(M,N)$ is the DoF.
2. At low SNRs, the capacity is approximately equal to $M\gamma\log_2 e$ bits per channel use, so only a receive power gain can be achieved.
3. For the case of $M = N = d$, the sum capacity (or the achievable sum rate) increases linearly with d over the entire SNR range.
4. When CSI is available at the transmitter, an additional power gain can be gleaned by using waterfilling at low SNR ranges. In other words, at low SNR, the capacity with full CSI at transmitter is significantly larger than when CSI is only available at the receiver. However, at high SNR, the difference between the two goes to zero.

From the above discussions we can conclude that the total DoF, i.e., d, is one of the most important factors to achieve higher sum rates, i.e., the larger the d , the higher data rates can be achieved for all SNR ranges.

2.4 Precoding

One of the applications of point-to-point communications is in cellular networks where a serviced area is partitioned into several cells, each typically has one BS serving some mobile terminals (MTs). The very earlier forms of cellular communications consist of single-antenna BSs and single-antenna MTs. For this scenario, since each BS has one antenna, it can only transmit to one user at each time. Since there are several users expecting to be serviced by the BS, they need to be spread across time or frequency. In this case, time division multiple access (TDMA) and frequency division multiple access (FDMA) are prevalent techniques which provide fair access to all MTs within a cell. For example, one of the earliest standards for cellular communications is Global System for Mobile (GSM) where users within each cell are serviced via TDMA [4]. However even for this case, if there are two or more cells and the adjacent cells share the same set of radio frequencies, the transmission of each BS can cause interference to the other active MTs in nearby cells. Therefore, one approach to suppress this *inter-cell* interference is allocating distinct radio frequency bands to adjacent cells.

In the following, we first consider the scenario where different sets of frequencies have been assigned to nearby cells. In this case, each cell can be analyzed separately and we therefore turn our focus on the transmission schemes which provide reasonable performance in downlink of single-cell scenario. We will next consider the scenarios wherein the nearby cells share the same radio frequency bands and we introduce the relevant state of the art interference management techniques to overcome the inter-cell interference in this case.

Note that in the cellular downlink, since BS can have access to partial or perfect CSI, it is more appropriate to shift the major signal processing enhancements to transmit side to keep MTs simple and low-cost.

In downlink of cellular networks, it has been shown that if instead of merely one antenna, multiple antennas are deployed at BS, significant throughput gains can be achieved [5]. Now, all the aforementioned performance gains of MIMO systems can be gleaned. However, instead of assuming one user with multiple antennas, it is more reasonable to consider several MTs with a single antenna. This is due to the fact that deploying more than one antenna at each MT results in larger handsets and also leads to more power consumption, which is not practical from the user's perspective. Therefore, in the case of multi-antenna BS and single-antenna MTs, each transmit antenna can be used to serve one single-antenna MT, and consequently several MTs can be simultaneously serviced within

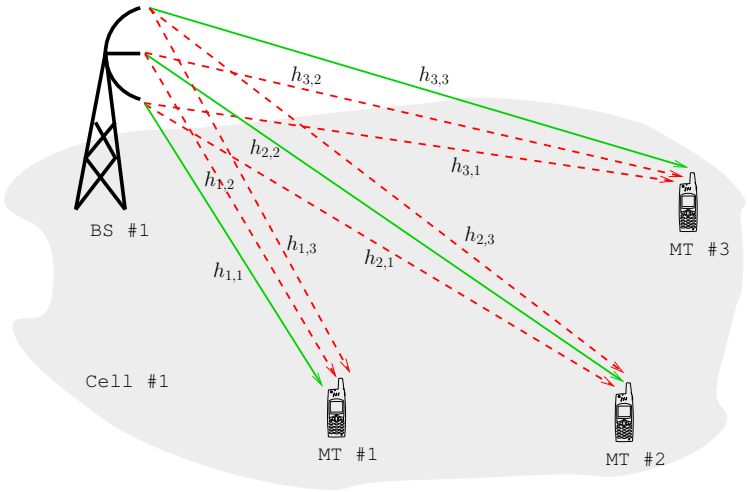

Figure 2.2: *Single-cell broadcast channel where dash red arrows represent intra-cell interference while solid green arrows denote desired links.*

each cell. This is the multiplexing gain of MIMO systems, discussed in the previous section, and results in higher sum rates, i.e., multiple MTs can now be provided with higher data rates.

Although using multiple antennas at BS results in higher multiplexing gains, it causes *intra-cell* interference, i.e., while each transmitted signal from one specific antenna at the BS is intended for just one specific single-antenna MT, it causes interference to the other receiving MTs within the same cell. Consequently, the downlink transmission strategy tries to design the beam patterns such that each MT receives its intended signal interference free. The more judicious transmission schemes try to increase the received SNR by having the received signals from the various transmit antennas add up in-phase (coherently) and/or by allocating more power to the transmit antenna with the better gain. This strategy, i.e., aligning the transmit signal in the direction of the transmit antenna array pattern, is called transmit beamforming (or hereafter we call it "precoding").

There are various precoding schemes, each of which has been designed to meet a certain criterion. Based on how the transmitted signals are related to the input data streams, the precoding techniques are categorized as *linear* and *nonlinear*. For example, it has been shown that the dirty paper coding (DPC), which is a nonlinear precoding, is capable of achieving the downlink capacity [6]. Nevertheless, due to its very high complexity, some less complex nonlinear precoders like vector perturbation [7–9] and Tomlinson-Harashima [10,11] are also of particular interest. The simplest transmission scheme for multi-antenna downlink is CI [12], which is linear, such that the intra-cell interferences are precanceled at

BS in order to enable each MT to receive its intended signal interference free. For example, as illustrated in Fig. 2.2, it has been assumed that BS has 3 antennas and is therefore able to send 3 independent data streams simultaneously to 3 single-antenna MTs. Therefore, each transmitted data stream acts as an intra-cell interference to the other 2 unintended MTs. In this case, the channel coefficient form antenna j to the kth MT can be modeled by $h_{k,j},\ k,j=1,\ldots,3$. We can represent the whole CSI as $\mathbf{H}\in\mathbb{C}^{M\times N}$ which collectively denotes the channel coefficients from the N-antenna BS to M single-antenna MTs, and for the special case of $M=N=3$ as denoted in Fig. 2.2, we have

$$\mathbf{H}=\begin{pmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & h_{2,2} & h_{2,3} \\ h_{3,1} & h_{3,2} & h_{3,3} \end{pmatrix} \tag{2.17}$$

It is worthwhile to note that the total number of MTs is typically thought to be equal to or less than the total number of antennas at BS, i.e., $M\leq N$, otherwise the users must be scheduled since one BS is only able to service up to N MTs simultaneously. We further denote $\mathbf{c}\in\mathbb{C}^{M\times 1}$ as the data vector containing the M symbols chosen from an arbitrary constellation that are intended for the M MTs. By considering the fact that the perfect CSI, i.e., $\mathbf{H}$, is known at BS, the transmitted signal based on CI precoding can be represented by [13]

$$\mathbf{s}_{\mathtt{CI}}=g_{\mathtt{CI}}\mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\mathbf{c} \tag{2.18}$$

where $g_{\mathtt{CI}}$ is the scaling factor which its value will be given in subsection 3.3.1. Now as can be seen in (2.18), the transmitted signal from BS, i.e., $\mathbf{s}_{\mathtt{CI}}$, is linearly dependent on the input data vector $\mathbf{c}$. In this case, the transmitted signals are received interference free as follows:

$$\begin{aligned} \mathbf{y}_{\mathtt{CI}} &= \sqrt{P}\mathbf{H}\mathbf{s}_{\mathtt{CI}}+\mathbf{z} \\ &= \sqrt{P}g_{\mathtt{CI}}\underbrace{\left[\mathbf{H}\mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\right]}_{\texttt{canceled interference}}\mathbf{c}+\mathbf{z} \\ &= \sqrt{P}g_{\mathtt{CI}}\mathbf{c}+\mathbf{z} \end{aligned} \tag{2.19}$$

where $\mathbf{z}\in\mathbb{C}^{M\times 1}$ is the noise vector, and P is the transmit power. As seen in (2.19), for each MT, all intra-cell interferences get suppressed and eventually each MT receives a scaled version of its transmitted data which can be readily recovered.

Although CI precoding achieves reasonable performance in downlink of cellular net-

works with multi-antenna BSs, it may deteriorate the performance of users at low-to-intermediate SNRs. This is due to the fact that some of the received signals may undergo a deep fading and this can further lead to an ill-conditioned channel matrix such that taking the inverse of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$, i.e., $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$, becomes problematic. In this case and as will be shown in detail within section 3.4, regularizing this inverse by adding an appropriate multiple of the identity matrix can notably improve the performance of CI. This precoding is called RCI [13] such that the transmitted signal from BS can now be shown as

$$\mathbf{s}_{\mathtt{RCI}} = g_{\mathtt{RCI}}\mathbf{H}^{\mathsf{H}}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \varepsilon\,\mathbf{I}\right)^{-1}\mathbf{c} \tag{2.20}$$

where the optimum values of ε will be determined in chapter 3. Now as can be seen in (2.20), the transmitted signal from BS, i.e., $\mathbf{s}_{\mathtt{RCI}}$, is linearly dependent on the input data vector $\mathbf{c}$. This further implies that similar to CI, RCI precoding is linear as well. However, unlike the CI precoding where all the intra-cell interferences are completely canceled, RCI precoding allows some constructive interference from unwanted MTs spill over the received signal of each user. In other words, the regularization parameter ε controls the amount of interference introduced to each user which is not zero at low-to-intermediate SNRs. Therefore, RCI precoding tries to constructively add up the interference with the desired signals such that the received SNR gets increased.

As stated earlier, even by deploying multiple antennas at BS and using one of the aforementioned linear or nonlinear precoding and in order to avoid inter-cell interference, we still have to assign different radio frequency bands to adjacent cells. Since spectrum allocation is extremely conservative and expensive, in order to increase the spectral efficiency of the entire network, it is more desirable to use the same set of radio frequencies for two or more nearby cells. To meet this demand, several advanced precoding techniques of increased complexity have recently emerged for wireless access, which are inherently cooperative schemes. One of the promising techniques is *network MIMO* [14] which enables BSs to share the same frequency bands by the combined use of multiple antennas in several neighboring cell sites. However, the BSs further need to share all the transmitted data streams through, for example, low-latency high-capacity backhaul links like the optical fibers, as illustrated in Fig. 2.3. This way, the inter-cell interference gets canceled and each user receives its intended data interference free. However, these backhaul links request for additional infrastructures which may not be readily implementable due to the excessive needs of data sharing between different sites and the need for additional antennas, which is impractical due to many hardware and cost constraints.

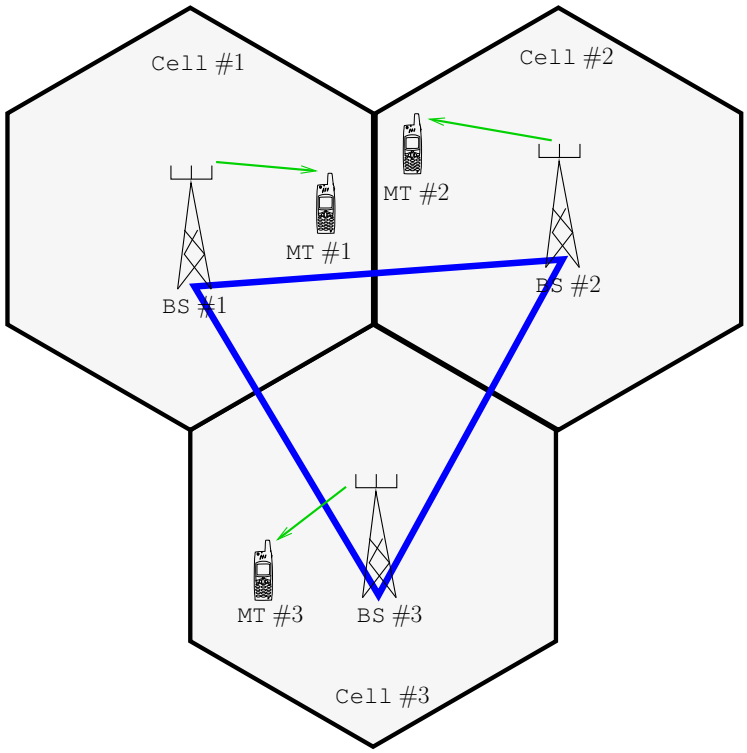

Figure 2.3: *Network MIMO for cellular communications where all the BSs share the transmitted data streams via backhaul links which are denoted as thick blue lines. This way, the inter-cell interference is canceled and each MT received its intended data stream interference free.*

Apart from network MIMO, another interesting approach towards higher spectral efficiency is *massive MIMO* where unlike network MIMO, there is no need to share the data streams between BSs [15]. However as illustrated in Fig. 2.4, large number of antennas are needed to be deployed. This way, the transmitted signals are beamformed towards the intended MTs without causing interference to the unintended MTs in nearby cells. Massive MIMO communications heavily depend on the two emerging technologies [16]:

1. Remote radio heads (RRHs) which allow for geographically distributed access via radio-over-fibre connections to a BS.
2. Electronically steerable passive array radiators (ESPARs) which provide multi-antenna-like functionality with a single active radio frequency chain only.

Although massive MIMO seems to be a part of the future wireless networks, it also needs extra infrastructure like fiber connections between each BS and the RRHs, and the spread-out installation of large number of antennas across a wide area which may result in huge implementation costs. Therefore, there is a need for a different interference management

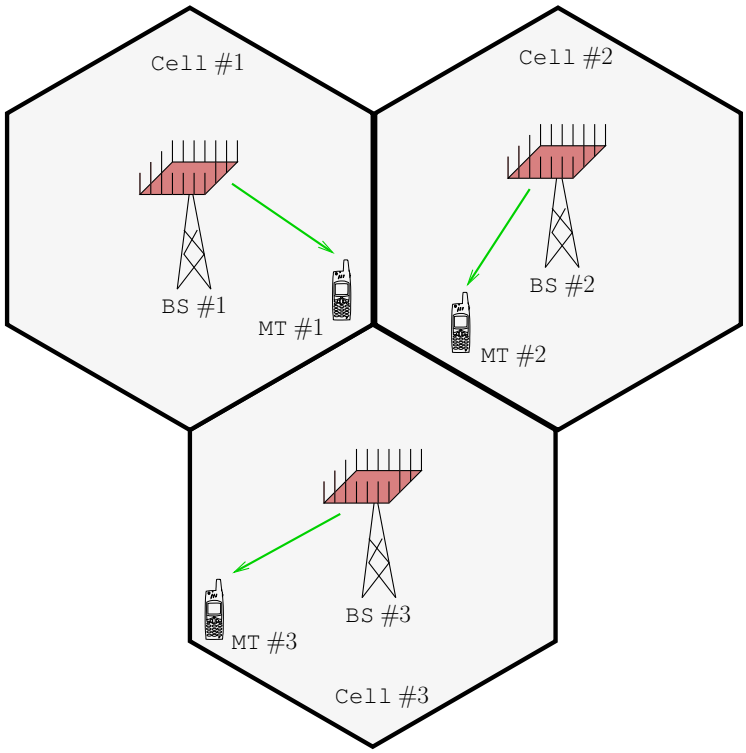

Figure 2.4: *Massive MIMO for cellular communications. Having large number of antennas, each BS is now able to beamform the transmitted data streams to its intended MT without causing inter-cell interference.*

technique which removes the extra overhead of network MIMO and massive MIMO systems.

2.5 Interference Alignment

As mentioned in the previous section, even by using CI and RCI precoding, the MTs in nearby cells can still experience inter-cell interference if the assigned radio frequency bands of two or more adjacent cells would be the same. This implies that even for the case of multi-antenna BSs, to avoid inter-cell interference, different sets of frequencies are still needed to be used. With respect to the fact that in contemporary urban environments, there are several BSs, if each cell requests a unique set of frequencies, this leads to a huge spectrum reservation for the entire network. However, since resource allocation (like energy and spectrum) is extremely conservative and valuable, it is more desirable to design advanced interference management techniques such that more adjacent cells can share the same set of frequencies. We also pointed out the two emerging and promising

wireless communication techniques, namely network MIMO and massive MIMO, which enable us to use the same radio frequency bands for the nearby cells by precoding the data streams towards the intended users while suppressing the inter-cell interference. However, both network MIMO and massive MIMO communications need extra infrastructure to be available. This includes radio-over-fiber connections which may not be readily implementable.

Therefore in this section we represent another potential and alternative approach towards interference management in future wireless networks. To do so, we first reintroduce the concept of wireless interference networks (or interference channels (IC)) where it has been shown that IA will be a dominant technique since it enables several BSs to transmit to their corresponding MTs simultaneously on the same radio frequencies. Now in this case and unlike network MIMO or massive MIMO communications, there is no need for backhaul links or excessive number of antennas. Therefore while each BS communicates with its intended MT, it causes inter-cell interference. This has been depicted in Fig. 2.5 where dash red arrows represent inter-cell interference while solid green arrows denote desired links. In this case, although each BS makes interference to the unintended MTs, IA enables us to design appropriate transmit filters ("precoders") for BSs and receive filters ("combiners") for MTs such that each MT is now able to cancel the inter-cell interferences received from nearby BSs [17].

Unlike point-to-point, multiple access (uplink) and broadcast (downlink) communications, the capacity characterization of interference channels is still an open problem. This prompted the information theorists to propound the concept of DoF to outline the performance of IC. It is worthwhile to reiterate that the DoF is the pre-log factor of the sum rate or the multiplexing gain, as discussed in section 2.3.

To highlight the spectrum-efficiency of IA compared to standard techniques, it suffices to mention that while for standard cellular communications like GSM and IS-136, the typical frequency reuse factor[1] is $\frac{1}{3}$ and $\frac{1}{7}$, respectively, for IA, the frequency reuse factor is one.

Example 2.1: In this example, we demonstrate the difference between conventional downlink precoding techniques (as discussed in section 2.4) and IA in terms of the achievable DoF. We consider the same scenario as the one in Fig. 2.5, i.e., there are three cells in the network such that each cell has one BS and each BS possesses three antennas and communicates with one single-antenna MT. In this case and by using IA, all BSs are able to

[1]The frequency reuse factor of $\frac{1}{n}$ means that there are n cells in one cluster and each and every cell has its own set of frequencies.

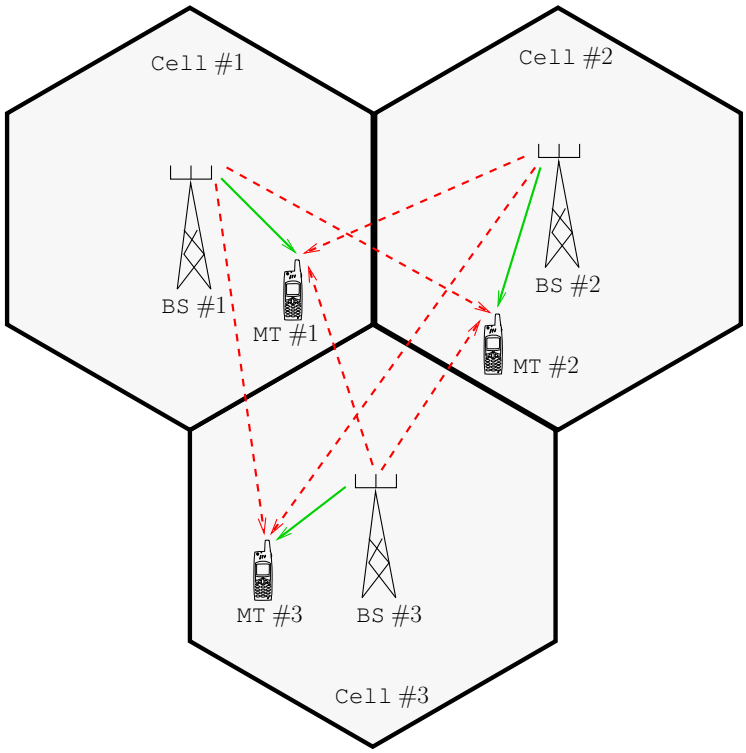

Figure 2.5: *3-cell wireless interference network where all BSs transmit simultaneously on the same radio frequencies. Each BS communicates to its corresponding MT while causing interference to the other MTs in adjacent cells. In this case, dash red arrows represent inter-cell interference while solid green arrows denote desired links.*

communicate with their corresponding MTs simultaneously on the same radio frequency bands. Therefore, the achievable DoF is now equal to three, i.e., all three MTs can now be serviced by using the same frequency and time resources. However by deploying downlink precoding techniques, the BSs can not communicate to their intended MTs at the same time on the same radio frequencies. In this case and in order to avoid inter-cell interference, each cell should be assigned to unique radio frequencies (using OFDM) or the MTs need to be scheduled in different time slots (using TDMA). Therefore by using downlink precoding, the achieve DoF is now equal to one, i.e., only one MT can be serviced by relying on the same frequency and time resources. Consequently, for the depicted scenario in Fig. 2.5, IA achieves three times as many DoF as downlink precoding techniques and is thus more spectral efficient.

With respect to the above example and the depicted scenario in Fig. 2.5, IA is distinguishable from standard single-cell downlink precoding as follows:

1. Using IA, each individual BS performs precoding to its user of interest through pre-

multiplying the transmitted signals by a precoding matrix. In this case all the interference is aligned within a reduced subspace of the received signal space. However for single-cell downlink precoding and under the same usage of frequency and time resources, only one BS can perform precoding and therefore no alignment of the interference takes place.

2. In standard downlink precoding, the single-antenna MTs do not need to use an interference suppression filter. However, in the case of IA, each user is able to null out the inter-cell interference through postmultiplying its received signal by an interference suppression filter.
3. In downlink precoding each BS only needs to know the CSI for its own MTs, however in the case of IA, each BS should have access to the CSI of all MTs in nearby cells as well.
4. In single-cell downlink precoding each BS uses different frequency or time resources and therefore there is no need for BSs to perform coordinated beamforming. However in the case of the interference channel, IA relies on coordinated precoding transmissions from all BSs.

Now the importance of IA in wireless interference networks can be more pronounced with the help of DoF. By assuming that there are K cells each possessing a single-antenna BS, while TDMA and FDMA achieve only one DoF, with IA, the achievable DoF can now be linearly scaled up with K. More precisely, in this case, IA is able to achieve $d = \frac{K}{2}$ DoF [17]. In other words, in this K-cell scenario, while with TDMA or FDMA we can only send one data stream to one specific user, with IA, $\frac{K}{2}$ independent data streams can be send throughout the network. Therefore, IA enables us to more efficiently use the available spectrum by transmitting higher sum rates while using the same time and frequency resources compared to standard interference management techniques like TDMA and FDMA.

IA is a broad area of research and it mainly falls within two separate fields. The first one includes information theoretic studies like the achievable DoF [17–19], and the second one includes the signal processing aspects like assessing the feasibility conditions [20, 21], or designing different algorithms meant for particular scenarios based on different optimization criteria. Designing IA algorithms can split into distinct directions. For example, while some IA schemes have been designed based on symbol extension across time or frequency [17, 19, 22], some other IA techniques have been proposed by relying on the signal space alignment using multiple antennas [23–34] to cancel the inter-

ference. These IA schemes, which have been set upon the spatial dimensions, are mostly iterative but dissolve the need of symbol extension across time or frequency, and therefore they are more practical and appealing. Of such schemes, minimum weighted leakage interference (Min-WLI) [23], alternating minimization (Alt-Min) [34], Max-SINR [23], weighted MMSE [24], and rank constrained rank minimization (RCRM) [26] are the most representative IA algorithms. Consequently, in this thesis, we place our focus on both information theoretic and signal prossing aspects of IA by addressing the achievable sum rates and DoF under imperfect CSI, feasibility conditions of IA in partially coordinated networks and also designing new iterative IA techniques based on the concept of signal space alignment using multiple antennas.

2.6 Limitations of the State of the Art

With respect to the above discussions, we can conclude that due to the scarcity of radio resources and incessant requests for higher QoS, new interference management techniques are needed to be addressed in order to support the excessively huge amount of data demand in future cellular networks. Note that all the aforementioned standard and future techniques including CI, RCI and IA have been mainly designed and analyzed based on the availability of perfect CSI. However, from the practical point of view and also due to deployment challenges, only partial CSI may be accessible, and this can adversely affect the achievable bounds which are predicted by the information theoretic studies. For example, it has been shown that CI can achieve the same asymptotic sum capacity as that of DPC, as the number of users goes to infinity [35], provided that perfect CSI is available. Also as mentioned earlier, although IA can achieve $\frac{K}{2}$ DoF, this can be attained under the availability of perfect CSI [17]. Consequently, performance analysis of these standard and future wireless communication techniques under imperfect CSI is vital and is thus going to be addressed in this thesis.

More specifically, first we introduce a generalized CSI mismatch model where the variance of the channel estimation error is a function of SNR and covers the CSI feedback and reciprocal channels as special scenarios. Then, unlike most of the literature wherein the proposed schemes have been set upon the knowledge of perfect CSI, we aim to design adaptive precoders and interference alignment techniques such that the design criteria are based on the availability of imperfect CSI and the knowledge of error variance in advance. This provides a versatile design since these adaptive schemes can accommodate the case

of perfect CSI as a spacial scenario when the variance of the channel estimation error is set to zero. We will show that these adaptive designs outperform standard techniques in more realistic scenarios where due to the channel dynamics and channel estimation schemes only imperfect CSI is accessible. Apart from designing adaptive interference management techniques, we also consider the power-efficient design of wireless networks towards the aims of "green communications" [36].

Chapter 3

Channel Inversion and Regularization

3.1 Introduction

Deploying multiple antennas at BS can significantly improve the achievable throughput in broadcast channels [5] where DPC is capable of achieving the downlink capacity [6]. Nevertheless, due to its high complexity, some less complex nonlinear precoders like vector perturbation [7–9] and Tomlinson-Harashima [10, 11] are also of particular interest.

Although achieving less throughput, linear precoders are more practical due to reduced possessing complexity compared to their nonlinear counterparts. The least complex of the available techniques named CI [12] is a linear precoding technique which yields reasonable performance in downlink; in other words, although CI achieves less sum rates than DPC, the former is linear whereas the latter is nonlinear with high complexity. Nonetheless, it has been shown that CI precoding, while generally suboptimal, can achieve the same asymptotic sum capacity as that of DPC, as the number of users goes to infinity [35]. However, in a downlink scenario where the number of antennas at BS is equal to the total number of single-antenna users and both are finite, it has been shown that with increasing the number of antennas at BS (and increasing the number of users accordingly), the symbol error rate (SER) of each user, due to deploying CI, deteriorates. Also in this case, while the sum capacity linearly increases with the number of transmit antennas, the sum rate of CI does not. This inferior performance of CI is related to the erratic behavior of the largest eigenvalue of the inverse of the covariance matrix of the channel. One approach to alleviate this malfunction is to employ the concept of regularization by adding a multiple of the identity matrix to the covariance matrix of the channel before inverting. This precoding, which has been dubbed RCI in [13], improves the performance of CI, such that

with increasing the number of antennas, the SER of each user remains fixed at low-to-intermediate SNRs, and slightly improves at high SNRs. Also, by using RCI, the sum rate has now a linear growth with the number of transmit antennas. Plus, for a fixed number of antennas at BS, RCI achieves higher throughput than CI at low-to-intermediate SNRs.

However, this comparative improvements of standard RCI to CI can mainly be gleaned when perfect CSI is available at BS, which is a very stringent requirement in practice. Hence, performance analysis of CI and RCI under CSI mismatch becomes more important, and is thus going to be addressed in this chapter.

First, we quantify the performance of CI and RCI under a generalized CSI mismatch model where the variance of the channel estimation error depends on the SNR. In this case, we derive novel bounds regarding the asymptotic mean loss in sum rate and the achievable DoF as a function of error variance. For instance, it will be shown that when the variance of the channel estimation error scales with the inverse of SNR, full DoF is achievable and therefore the asymptotic mean loss in sum rate is a nonzero bounded value.

Although the performance analysis of standard RCI under imperfect CSI has been addressed in [37], it has been assumed that the actual channel matrix (i.e., the perfect CSI) depends on the channel estimation error and the variance of the channel estimation error is constant. Meanwhile, only the bit error rate (BER) performance analysis of RCI under this imperfect CSI model has been considered in [37].

However, in this chapter (as well as the subsequent chapters), we assume that the actual channel matrix is independent of the channel estimation error, which is more pragmatic. Also as mentioned earlier, we consider an SNR-dependent model for error variance which automatically accommodates the case of constant variance as a special scenario. Then under this imperfect CSI model, we analyze the achievable sum rates of CI and RCI by deriving the output SINR of each user. Moreover, based on the knowledge of the error variance in advance, we propose an adaptive RCI precoding by deriving an appropriate regularization parameter. Simulation results verify that with the presence of CSI mismatch, while the performance improvement of standard RCI compared to CI becomes negligible, the proposed adaptive RCI compensates this degraded performance of the standard RCI.

We present the system model under perfect and imperfect CSI in Section 3.2. In Section 3.3, we quantify novel bounds on the asymptotic performance of CI. Section 3.4 deals with the performance analysis of standard RCI under imperfect CSI by deriving the output SINR of each user. In Section 3.5, an adaptive RCI technique is proposed. In Section 3.6, we use simulations to corroborate the undergone analyses in this chapter, and finally Section 3.7 contains a summary of the presented materials within this chapter.

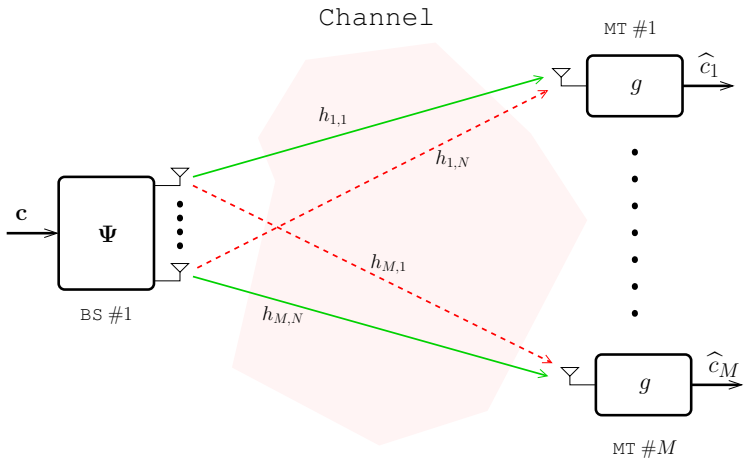

Figure 3.1: *Single-cell broadcast channel where dash red arrows represent intra-cell interference while solid green arrows denote desired links. $\mathbf{\Psi}$ designates the precoding matrix, g is the scaling factor, and $h_{k,j}$ denotes the time-variant channel response between the jth transmit antenna of BS and the kth MT. While $\mathbf{c} = (c_1, \ldots, c_M)^\top$ is the transmitted data stream from BS, $\widehat{c}_k$ is the decoded data at the kth user.*

3.2 Preliminaries

3.2.1 System Model

We consider a multiuser downlink scenario where an N-antenna transmitter communicates with MTs with M receive antennas in total. More specifically, we assume that the total number of receive antennas is equal to or less than the total number of transmit antennas, i.e., $M \leq N$. Since no signal processing treatment is going to be considered at each MT, the system configuration is irrespective to whether the receive antennas cooperate or not, therefore the total number of receive antennas can belong to one user or be shared by several users; however, as purely transmitter-based precoders are most useful with single antenna receivers, we consider single-antenna MTs, i.e., MISO downlink. A simple downlink communication scenario has been illustrated in Fig. 3.1.

Without loss of generality and analogous to [13,37], we assume that all single-antenna users are homogeneous and experience independent fading. The received signals of all users can be expressed by

$$\mathbf{y} = \sqrt{P}\mathbf{H}\mathbf{s} + \mathbf{z} \tag{3.1}$$

where $\mathbf{y} \in \mathbb{C}^{M\times 1}$, $\mathbf{H} \in \mathbb{C}^{M\times N}$ denotes the channel from N-antenna transmitter to M single-antenna users such that the magnitude of channel coefficients is bounded away from zero and infinity. We also consider block fading model where channel coefficients

are static for the duration of a transmission but may change between successive transmission. We further assume that elements of $\mathbf{H}$ can be modeled by independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance, i.e., $\mathrm{vec}\,(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \mathbf{I})$, $\mathbf{s} \in \mathbb{C}^{N\times 1}$ is the transmitted signal from BS, and $\mathbf{z} \in \mathbb{C}^{M\times 1}$ is the circularly symmetric additive white Gaussian noise with zero mean and variance σ^2, i.e., $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \sigma^2\mathbf{I})$. We further assume that the transmitted signal $\mathbf{s}$ in (3.1) can be expressed as $\mathbf{s} = g\mathbf{\Psi}\mathbf{c}$. Similar to [13, 38], we consider g as the scaling factor that ensures transmit power constraint i.e., $\mathbb{E}\left\{\|\mathbf{s}\|_2^2\right\} = 1$. $\mathbf{\Psi}$ is the precoding matrix and $\mathbf{c}$ represents the vector containing the symbols chosen from a desired constellation and since we assume i.i.d. Gaussian input signaling, we have $\mathbb{E}\left\{\mathbf{c}\mathbf{c}^{\mathsf{H}}\right\} = \mathbf{I}$. We also define the nominal SNR as $\gamma = P/\sigma^2$ where P is the transmit power. Note that although the concept of regularization is most beneficial for the case of equal number of transmit and receive antennas [13, 37], without loss of generality, we assume $M \leq N$.

3.2.2 Imperfect CSI Model

Unlike [37], where the perfect CSI, viz. $\mathbf{H}$, depends on the channel estimation error, here we model the imperfect CSI as

$$\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{E} \tag{3.2}$$

where the actual channel matrix $\mathbf{H}$ is thought to be independent of channel measurement error $\mathbf{E}$. Similar to [8], we further consider $\mathbf{E}$ as a Gaussian matrix consisting of i.i.d. elements with mean zero and variance τ, i.e.,

$$\mathrm{vec}\,(\mathbf{E}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \tau\mathbf{I}) \text{ with } \tau \triangleq \beta\gamma^{-\alpha}, \quad \beta > 0, \quad \alpha \geq 0 \tag{3.3}$$

In this case, the error variance can depend on SNR $(\alpha \neq 0)$ or be independent of that $(\alpha = 0)$. Notice the variance model in (3.3) is versatile since it is potentially able to accommodate a variety of distinct scenarios. In particular, perfect CSI is regained when $\alpha \to \infty$. Three representative cases of (3.3), which will be widely referred throughout this thesis, are

- *Reciprocal Channels:* This case represents the reciprocal systems like time division duplex where uplink and downlink channel are identical. The downlink channel can thus be estimated through pilots sent over the uplink channel and the channel measurement error $\mathbf{E}$ depends on the noise level at BS as well as the pilot power. If

the pilot power proportionally increases with P, the channel estimation error scales inversely with increasing γ. This case is modeled by (3.3) with $\alpha = 1$.

- *Mismatched Reciprocal Channels:* This may be the case where the mobile and BS powers are not in the same range (or of the same order). This scenario can be represented by $0 < \alpha < 1$ where there is a significant power imbalance between the transmitters and receivers, i.e., the feedback power is much smaller than the feedforward power.
- *CSI Feedback:* Here, the channel matrix can be estimated by pilot transmissions in the downlink. Then, a quantized version of this channel estimate is sent back to BS through a dedicated feedback link. This way, the imperfect CSI will be mostly dominated by the errors caused through quantization and feedback delay, which can eventually result in outdated CSI at BS if the channel coherence time is smaller than the feedback delay. Since channel coherence time and the resolution of quantizer do not depend on γ, the channel estimation error variance τ becomes independent of γ as well. This case is captured by (3.3) with $\alpha = 0$.

More specifically, τ can be interpreted as a parameter that captures the quality of the channel estimation which is possible to be known a priori, depending on the channel dynamics and channel estimation schemes (see [39] and references therein). For example:

- For the CSI feedback scenario where each single-antenna MT sends the index of the estimated channel matrix back to BS by using a codebook of size 2^b, it has been shown that $\tau = 2^{-\frac{b}{N}}$ [37].
- In a block Rayleigh fading channel of coherence time T and by using orthogonal training signals (which are optimum for spatially white inputs) with training interval length of T', $\tau = \dfrac{1}{1 + \frac{T'}{N}P'}$ where P' is the transmit power of the training symbols [40].
- In a continuously time-varying Rayleigh-fading channel with a bandlimited low-pass spectrum with the cutoff frequency f_c , using pilot symbols with sampling rate $1/L \geq 2f_c$ yields $\tau = \dfrac{1}{1 + \frac{1}{2f_cL}P'}$ [41].

Fig 3.2 depicts the performance trend of τ as a function of γ for different values of α and β.

To facilitate the performance analysis of CI and RCI under CSI mismatch model in (3.2), it is more appropriate to have the statistical properties of $\mathbf{H}$ conditioned on $\widehat{\mathbf{H}}$ by using following lemma [42]:

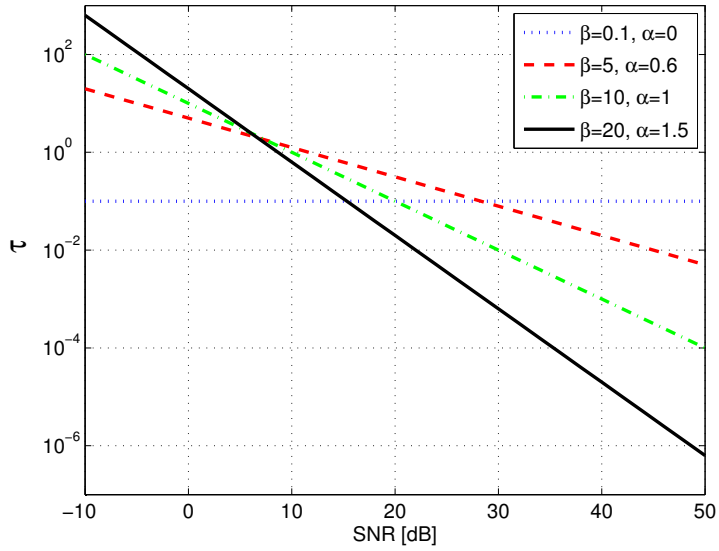

Figure 3.2: *Channel estimation error variance as a function of SNR for different values of* α *and* β.

Lemma 3.1: Conditioned on $\widehat{\mathbf{H}}$, $\mathbf{H}$ has a Gaussian distribution with mean $\widehat{\mathbf{H}}/(1+\tau)$ and statistically independent elements of variance $\tau/(1+\tau)$, i.e.,

$$\mathbf{H} = \frac{1}{1+\tau}\widehat{\mathbf{H}} + \breve{\mathbf{H}} \tag{3.4}$$

where $\mathrm{vec}\left(\breve{\mathbf{H}}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{\tau}{1+\tau}\mathbf{I}\right)$ is statistically independent of $\widehat{\mathbf{H}}$.

3.3 Asymptotic Performance of Channel Inversion

In this section, we first derive the output SINR of each user when CI is deployed at BS. We then derive novel bounds on the asymptotic mean loss in sum rate and the achievable DoF when BS is in possession of imperfect CSI. Consequently, we assume that the channel estimate $\widehat{\mathbf{H}}$ is only available at BS, and the signal preprocessing at BS is thus going to be done upon the knowledge of $\widehat{\mathbf{H}}$.

3.3.1 Channel Inversion under Perfect CSI

When the perfect channel state information is available at BS, the transmitted signal can be represented as

$$\mathbf{s}_{\mathtt{CI}} = g_{\mathtt{CI}} \mathbf{\Psi}_{\mathtt{CI}} \mathbf{c} \tag{3.5}$$

where the precoding matrix is $\mathbf{\Psi}_{\mathtt{CI}} = \mathbf{H}^{\mathsf{H}} \left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ and the scaling factor can be defined as [13]

$$g_{\mathtt{CI}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\right]}} \tag{3.6}$$

In this case, the received signal can be shown as

$$\mathbf{y}_{\mathtt{CI}} = \sqrt{P} g_{\mathtt{CI}} \mathbf{H}\mathbf{H}^{\mathsf{H}} \left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1} \mathbf{c} + \mathbf{z} = \sqrt{P} g_{\mathtt{CI}} \mathbf{c} + \mathbf{z} \tag{3.7}$$

and consequently the unified output SINR of each user is equal to

$$\eta_{\mathtt{CI}} = \frac{P g_{\mathtt{CI}}^2}{\sigma^2} \tag{3.8}$$

With respect to the fact that $\mathrm{Tr}\left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\right] = \sum_{\ell=1}^{M} \left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\right]_{\ell,\ell}$, the output SINR of the ℓth user can be shown as [35]

$$\dot{\eta}_{\mathtt{CI}} = \frac{P}{M\sigma^2 \left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}\right]_{\ell,\ell}} \tag{3.9}$$

where $[\cdot]_{\ell,\ell}$ denotes the ℓth diagonal element.

Without loss of generality and to avoid cumbersome formulation, and also to simplify the analysis within this subsection and also the next one, we employ the unified output SINR in lieu of the output SINR of the ℓth user, since this interchangeability does not compromise the validity of the asymptotic performance analysis as SNR goes to infinity. The achievable sum rate under perfect CSI by considering the unified output SINR can thus be expressed as [13]

$$R_{\mathtt{Perfect\ CSI}} = M \log_2 \left(1 + \eta_{\mathtt{CI}}\right) = M \log_2 \left(1 + \frac{P g_{\mathtt{CI}}^2}{\sigma^2}\right) \tag{3.10}$$

and the achievable DoF is defined as [43]

$$D_{\texttt{Perfect CSI}} = \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}}\left\{R_{\texttt{Perfect CSI}}\right\}}{\log_2 P} = \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}}\left\{M\log_2\left(1+\frac{Pg_{\texttt{CI}}^2}{\sigma^2}\right)\right\}}{\log_2 P} = M \tag{3.11}$$

Note that in (3.11), both $g_{\texttt{CI}}^2$ and σ^2 are finite while the transmit power P goes to infinity.

3.3.2 Channel Inversion under Imperfect CSI

With the presence of imperfect CSI at BS, the precoding matrix can now be defined as $\widehat{\mathbf{\Psi}}_{\texttt{CI}} = \widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}$. Consequently, in this case, the transmitted signal in (3.1) can be shown as

$$\widehat{\mathbf{s}}_{\texttt{CI}} = \widehat{g}_{\texttt{CI}}\widehat{\mathbf{\Psi}}_{\texttt{CI}}\mathbf{c} \tag{3.12}$$

where the scaling factor is equal to

$$\widehat{g}_{\texttt{CI}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\right]}} \tag{3.13}$$

Therefore the received signal in (3.1) can be represented by

$$\begin{aligned}
\widehat{\mathbf{y}}_{\texttt{CI}} &= \sqrt{P}\widehat{g}_{\texttt{CI}}\mathbf{H}\widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\mathbf{c}+\mathbf{z} \\
&\overset{①}{=} \sqrt{P}\widehat{g}_{\texttt{CI}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}+\breve{\mathbf{H}}\right)\widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\mathbf{c}+\mathbf{z} \\
&= \underbrace{\frac{\sqrt{P}\widehat{g}_{\texttt{CI}}}{1+\tau}\mathbf{c}}_{\texttt{desired term}} + \underbrace{\sqrt{P}\widehat{g}_{\texttt{CI}}\breve{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\mathbf{c}+\mathbf{z}}_{\texttt{interference plus noise term}}
\end{aligned} \tag{3.14}$$

where ① follows from (3.4). To further proceed, we consider the following lemmas:

Lemma 3.2: If $\mathbf{A}\in\mathbb{C}^{M\times N}$ represents a Gaussian matrix with i.i.d. elements of mean zero and variance a, then $\mathbb{E}\left\{\mathbf{A}\mathbf{A}^{\mathsf{H}}\right\} = aN\,\mathbf{I}$.

Proof: Since $\mathbf{A}$ is a Gaussian matrix, we have $\mathrm{vec}\left(\mathbf{A}\right)\sim\mathcal{N}_{\mathbb{C}}\left(\mathbf{0},a\mathbf{I}\right)$. In other words, if $\mathbf{a}$ represents an arbitrary column of $\mathbf{A}$, then $\mathbb{E}\left\{\mathbf{a}\mathbf{a}^{\mathsf{H}}\right\} = a\,\mathbf{I}$ [3]. However, since $\mathbf{A}$ has N independent columns, the claim follows. ■

Lemma 3.3: Throughout this chapter (as well as the subsequent chapters), we assume

that the noise and data vectors are independent of each other and are also independent of the actual channel matrix $\mathbf{H}$ which is consistent with [13]. Since due to lemma 3.1, $\mathbf{H}$ depends on both $\widehat{\mathbf{H}}$ and $\breve{\mathbf{H}}$, the data and noise are likewise considered to be independent of $\widehat{\mathbf{H}}$ and $\breve{\mathbf{H}}$. However, to make the SINR of each user dependent only on the channel estimate $\widehat{\mathbf{H}}$, we additionally take the expectation over $\breve{\mathbf{H}}$. This is also consistent with [37, 44] wherein the expectation was taken over the redundant channel measurement error.

Therefore based on lemma 3.2 and lemma 3.3 and by considering (3.13), we have $\mathbb{E}_{\breve{\mathbf{H}},\mathbf{c}}\left\{\left\|\sqrt{P}\widehat{g}_{\mathtt{CI}}\breve{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\mathbf{c}\right\|_2^2\right\}=\frac{PM\tau}{1+\tau}$, and consequently for a given realization of $\widehat{\mathbf{H}}$, the unified output SINR of each user can be given by

$$\widehat{\eta}_{\mathtt{CI}}=\frac{P\widehat{g}_{\mathtt{CI}}^2}{P\tau\left(1+\tau\right)+\sigma^2\left(1+\tau\right)^2} \tag{3.15}$$

With respect to the fact that $\mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\right]=\sum_{\ell=1}^{M}\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\right]_{\ell,\ell}$ and by considering (3.8)–(3.9), the output SINR of the ℓth user can be shown as

$$\ddot{\eta}_{\mathtt{CI}}=\frac{P}{M\left(P\tau\left(1+\tau\right)+\sigma^2\left(1+\tau\right)^2\right)\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\right)^{-1}\right]_{\ell,\ell}} \tag{3.16}$$

Note that by setting $\tau=0$, $\ddot{\eta}_{\mathtt{CI}}$ boils down to $\dot{\eta}_{\mathtt{CI}}$ in (3.9) which is the output SINR of the ℓth user under perfect CSI.

Consequently and by considering the unified output SINR in (3.15), the achievable sum rate of CI under imperfect CSI can be represented by

$$R_{\texttt{Imperfect CSI}}=M\log_2\left(1+\widehat{\eta}_{\mathtt{CI}}\right)=M\log_2\left(1+\frac{P\widehat{g}_{\mathtt{CI}}^2}{P\tau\left(1+\tau\right)+\sigma^2\left(1+\tau\right)^2}\right) \tag{3.17}$$

3.3.3 Mean Loss in Sum Rate and Achievable DoF

In this subsection, we derive novel bounds regarding the asymptotic mean loss in sum rate and the achievable DoF of CI. We do so with respect to the unified output SINRs of each user when CI is deployed at BS with the presence of imperfect CSI.

With respect to (3.10) and (3.17), the mean loss in sum rate can be defined as

$$\Delta R=\mathbb{E}_{\mathbf{H}}\left\{R_{\texttt{Perfect CSI}}\right\}-\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{R_{\texttt{Imperfect CSI}}\right\}$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathbf{H}} \left\{ M \log_2 \left(1 + \frac{P g_{\mathtt{CI}}^2}{\sigma^2} \right) \right\} - \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}} \left\{ M \log_2 \left(1 + \frac{P \widehat{g}_{\mathtt{CI}}^2}{P\tau\left(1+\tau\right) + \sigma^2 \left(1+\tau\right)^2} \right) \right\} \\
&= \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left\{ M \log_2 \left(\frac{\left(P g_{\mathtt{CI}}^2 + \sigma^2\right) \left(P\tau\left(1+\tau\right) + \sigma^2 \left(1+\tau\right)^2\right)}{\sigma^2 \left(P \widehat{g}_{\mathtt{CI}}^2 + P\tau\left(1+\tau\right) + \sigma^2 \left(1+\tau\right)^2\right)} \right) \right\} \\
&\overset{②}{=} \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left\{ M \log_2 \left(\frac{\left(P g_{\mathtt{CI}}^2 + \sigma^2\right) \left(\beta P^{1-\alpha}\sigma^{2\alpha}\left(1+\beta P^{-\alpha}\sigma^{2\alpha}\right) + \sigma^2 \left(1+\beta P^{-\alpha}\sigma^{2\alpha}\right)^2\right)}{\sigma^2 \left(P \widehat{g}_{\mathtt{CI}}^2 + \beta P^{1-\alpha}\sigma^{2\alpha}\left(1+\beta P^{-\alpha}\sigma^{2\alpha}\right) + \sigma^2 \left(1+\beta P^{-\alpha}\sigma^{2\alpha}\right)^2\right)} \right) \right\}
\end{aligned}
\tag{3.18}
$$

where ② is due to the fact that in (3.3), $\tau = \beta\gamma^{-\alpha}$. Thus, with respect to the fact that

$$
\lim_{P\to\infty} \beta P^{-\alpha}\sigma^{2\alpha} = \begin{cases} 0 & \alpha > 0 \\ \beta\sigma^{2\alpha} & \alpha = 0 \end{cases} \tag{3.19}
$$

the asymptotic mean loss in sum rate can be evaluated as follows

$$
\lim_{P\to\infty} \Delta R = \begin{cases} \infty & 0 \le \alpha < 1 \\ C & \alpha = 1 \\ 0 & 1 < \alpha \end{cases} \tag{3.20}
$$

where $0 < C < \infty$ is a constant which its value is given in the following theorem:

Theorem 3.1: When the error variance scales with the inverse of SNR, i.e., when $\alpha = 1$, the asymptotic mean loss in sum rate is equal to $C = M \log_2 \left(1+\beta\right)$.

Proof: By considering (3.18), we have

$$
\begin{aligned}
\lim_{\substack{\alpha=1 \\ P\to\infty}} \Delta R &= \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left\{ M \log_2 \left(\frac{\left(1+\beta\right) g_{\mathtt{CI}}^2}{\widehat{g}_{\mathtt{CI}}^2} \right) \right\} \\
&= \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left\{ M \log_2 \left(\frac{\left(1+\beta\right) \mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} \right)^{-1} \right]}{\mathrm{Tr}\left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} \right)^{-1} \right]} \right) \right\}
\end{aligned}
\tag{3.21}
$$

Due to the fact that $\lim\limits_{\substack{\alpha=1 \\ P\to\infty}} \tau = \lim\limits_{P\to\infty} \beta \left(P/\sigma^2\right)^{-1} = 0$, and with respect to (3.2), we have

$$
\lim_{\substack{\alpha=1 \\ P\to\infty}} \mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} \right)^{-1} \right] = \mathrm{Tr}\left[\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} \right)^{-1} \right] \tag{3.22}
$$

and consequently

$$C = \lim_{\substack{\alpha=1 \\ P\to\infty}} \Delta R = M \log_2 (1+\beta) \quad \text{bits per channel use} \tag{3.23}$$

■

Since each of the sum-rate curves has a slope of $M/3$ in units of bits per channel use per dB, this rate offset C (i.e., the vertical offset between the curve representing the perfect CSI and the one denoting the imperfect CSI case of $\alpha = 1$) can be translated into a power offset (i.e., a horizontal offset) as follows:

$$\Delta\gamma\Big|_{\substack{\alpha=1 \\ P\to\infty}} = \frac{3}{M}\Delta R\Big|_{\substack{\alpha=1 \\ P\to\infty}} = 3\log_2(1+\beta) \quad \text{dB} \tag{3.24}$$

In other words, the power offset in (3.24) implies that for the case of $\alpha = 1$, we should increase the transmit power by $\Delta\gamma$ dB to achieve the same sum rate as the perfect CSI.

Now that we established bounds on asymptotic mean loss in sum rate, in the sequel, it is revealed that when $0 \le \alpha < 1$, an α fraction of the total DoF, i.e., $\alpha D_{\texttt{Perfect CSI}}$ DoF, is achievable, where $D_{\texttt{Perfect CSI}}$ is defined in (3.11).

The total DoF subject to imperfect CSI can be quantified as follows:

$$\begin{aligned}
D_{\texttt{Imperfect CSI}} &= \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{R_{\texttt{Imperfect CSI}}\}}{\log_2 P} \\
&= \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{M\log_2\left(1+\dfrac{\widehat{g}_{\texttt{CI}}^2 P}{P\tau(1+\tau)+\sigma^2(1+\tau)^2}\right)\right\}}{\log_2 P} \\
&= \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{M\log_2\left(\widehat{g}_{\texttt{CI}}^2 P + P\tau(1+\tau)+\sigma^2(1+\tau)^2\right)\right\}}{\log_2 P} \\
&- \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{M\log_2\left(P\tau(1+\tau)+\sigma^2(1+\tau)^2\right)\right\}}{\log_2 P} \\
&\ge \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{M\log_2(\widehat{g}_{\texttt{CI}}^2 P)\}}{\log_2 P} \\
&- \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{M\log_2\left(P\tau(1+\tau)+\sigma^2(1+\tau)^2\right)\right\}}{\log_2 P} \\
&= M - \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{M\log_2\left(P\tau(1+\tau)+\sigma^2(1+\tau)^2\right)\right\}}{\log_2 P}
\end{aligned}$$

$$= \begin{cases} M & 1 \leq \alpha \\ \alpha\, M & 0 \leq \alpha < 1 \end{cases} \tag{3.25}$$

Notice $0 < \alpha < 1$ reflects the scenario in which feedback power is much smaller than feedforward power. This may be the case where the mobile and BS powers are not in the same range. Therefore, the BS can reciprocally learn the forward link, but instead of full DoF, i.e., M, only an α fraction of that, i.e, $\alpha\, M$, is achievable. Note that the results of (3.25) are inherently related to those in (3.20). For example, for the case of $\alpha = 0$, i.e., finite-rate feedback, while (3.25) implies that the achievable DoF is equal to zero, (3.20) indicates that for this case and by increasing SNR, the asymptotic mean loss in sum rate is unboundedly increasing. Also (3.25) implies that when $\alpha \geq 1$, full DoF is achievable, that is, the asymptotic mean loss in sum rate is constant. In this case, (3.20) implies that when $\alpha = 1$, this asymptotic mean loss converges to a non-zero constant whereas for $\alpha > 1$, it tends to zero.

3.4 Standard RCI

In section 3.3, we evaluated the performance of CI under CSI mismatch. However, as it is well-known, RCI precoding achieves better performance than CI precoding under perfect CSI by adding a multiple of the identity matrix to $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ before inverting. This is due to the fact that CI precoding inherently utilizes $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$, which is problematic when $\mathbf{H}$ becomes ill-conditioned, and this becomes more pronounced when $N = M$. An explanation for this poor behavior comes from looking at the eigenvalues of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ for the case $M = N$. As shown in [45], the smallest eigenvalue of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ has an exponential distribution as follows:

$$p\left(\lambda_s\right) = Me^{-M\lambda_s} \tag{3.26}$$

Therefore, the largest eigenvalue of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ has the distribution

$$p\left(\lambda_L\right) = \left(\frac{M}{\lambda_L^2}\right) e^{-\frac{M}{\lambda_L}} \tag{3.27}$$

which is also referred as inverse-gamma distribution with parameter one. This density is zero at $\lambda_L = 0$, but decays as $\frac{1}{\lambda_L^2}$ as $\lambda_L \to \infty$ for any values of M. It is thus a long-tailed distribution with infinite mean. Figs. 3.3–3.4 illustrate a numerical comparison of the four largest eigenvalues of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ for cases $M = N - 1$ and $M = N$, respec-

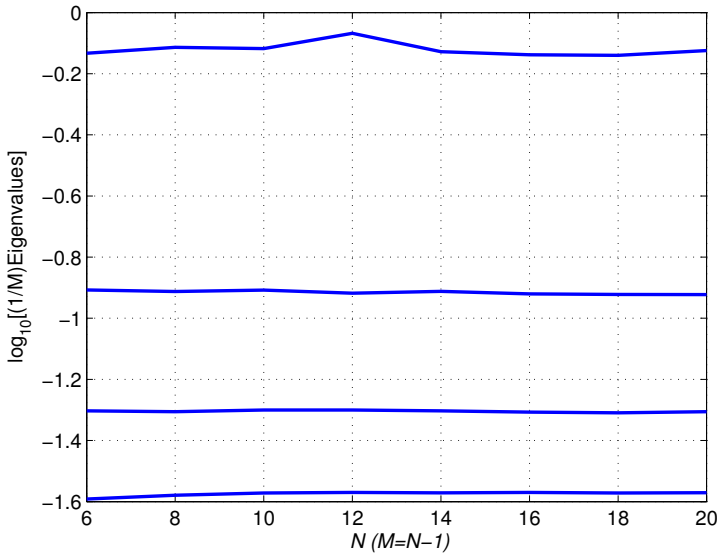

Figure 3.3: *Numerical results for the mean behavior of the four largest eigenvalues of* $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ *as a function of* N *when* $M = N - 1$.

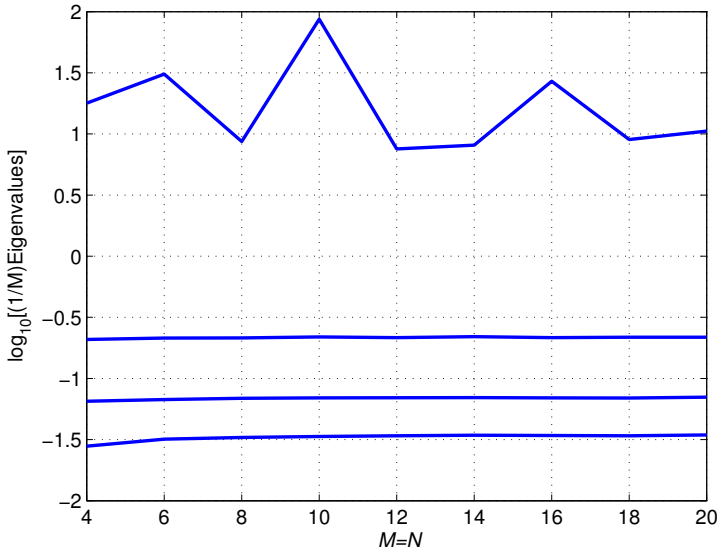

Figure 3.4: *Numerical results for the mean behavior of the four largest eigenvalues of* $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ *as a function of* N *when* $M = N$.

tively. As seen in Fig. 3.3, when $M = N - 1$, all four largest eigenvalues (including the largest one) have a relatively smooth behavior. However, for the case $M = N$ and as revealed in Fig. 3.4, except the largest eigenvalue, the others have much smoother plots. Plus, the magnitude of the largest eigenvalue is much larger than that of the remaining

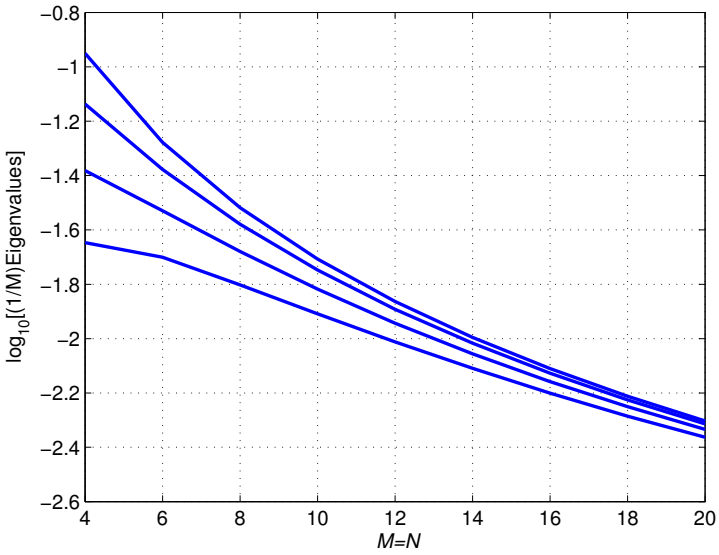

Figure 3.5: *Numerical results for the mean behavior of the four largest eigenvalues of* $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{M}{2}\mathbf{I}\right)^{-1}$ *as a function of* N *when* $M = N$.

eigenvalues. Consequently, the case of $M = N$ is much more problematic than $N > M$ when the channel is ill-conditioned. With respect to the fact that the smallest eigenvalue of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ probabilistically concentrates around $\frac{1}{4M}$ as $M \to \infty$, any approach to improve the performance of CI must seek to reduce the effects of the largest eigenvalue.

It has been shown that one very effective and simple way to alleviate this erratic behavior of the largest eigenvalue of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ is the regularization via adding a multiple of the identity matrix to $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ before inverting [46, 47]. RCI precoding aims to do so such that this multiplicative factor of the identity matrix is a function of M and γ [13]. It is also worthwhile to mention that, although the benefits of regularization extend to the case $N > M$, it is most beneficial when $M = N$. As shown in Fig. 3.5, by adding a multiple of the identity matrix, e.g., $\frac{M}{2}\mathbf{I}$, to $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ before inverting, the behavior of the largest eigenvalue of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{M}{2}\mathbf{I}\right)^{-1}$ becomes regularized. In other words, now all the four largest eigenvalues have the similarly monotonic behavior. This is opposed to the sans regularization case (as denoted in Fig. 3.4), where the performance trend of the largest eigenvalue is nonmonotonic.

In the remainder of this section, we evaluate the performance of standard RCI precoding by deriving the output SINR of each user when RCI is deployed at BS with the knowledge of imperfect CSI. Note that the SINR analysis of RCI has been addressed in [13] where the derived formula is just meant for the perfect CSI and is also dependent

on the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$. However, in this section, we derive the output SINR of each user with a different approach, which makes the RCI precoding especially amenable to the performance analysis subject to imperfect CSI.

Under the assumption of perfect CSI, the RCI precoder is defined as

$$\mathbf{\Psi}_{\texttt{RCI}} = \mathbf{H}^{\mathsf{H}} \left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \varepsilon \mathbf{I} \right)^{-1} \tag{3.28}$$

where $\varepsilon = M\gamma^{-1}$ is the regularization parameter which is optimal for large M but works well for small M too [13]. The scaling factor is thus equal to:

$$g_{\texttt{RCI}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\mathbf{H}\mathbf{H}^{\mathsf{H}} \left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \varepsilon \mathbf{I} \right)^{-2} \right]}} \tag{3.29}$$

By considering the fact that only imperfect channel estimate $\widehat{\mathbf{H}}$ is available at BS, the transmitted signal can be shown as

$$\widehat{\mathbf{s}}_{\texttt{RCI}} = \widehat{g}_{\texttt{RCI}} \widehat{\mathbf{\Psi}}_{\texttt{RCI}} \mathbf{c} \tag{3.30}$$

such that the precoding matrix is defined as

$$\widehat{\mathbf{\Psi}}_{\texttt{RCI}} = \widehat{\mathbf{H}}^{\mathsf{H}} \left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \varepsilon \mathbf{I} \right)^{-1} \tag{3.31}$$

and the scaling factor is equal to

$$\widehat{g}_{\texttt{RCI}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} \left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \varepsilon \mathbf{I} \right)^{-2} \right]}} \tag{3.32}$$

Accordingly, the received signal at all users can be represented as

$$\begin{aligned} \widehat{\mathbf{y}}_{\texttt{RCI}} &= \sqrt{P}\widehat{g}_{\texttt{RCI}} \mathbf{H}\widehat{\mathbf{H}}^{\mathsf{H}} \left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \varepsilon \mathbf{I} \right)^{-1} \mathbf{c} + \mathbf{z} \\ &= \sqrt{P}\widehat{g}_{\texttt{RCI}} \mathbf{H} \left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}} + \varepsilon \mathbf{I} \right)^{-1} \widehat{\mathbf{H}}^{\mathsf{H}} \mathbf{c} + \mathbf{z} \end{aligned} \tag{3.33}$$

Let $\widehat{\mathbf{h}}_{\ell}^{\mathsf{H}} \in \mathbb{C}^{1\times N}$ denote the ℓth row of $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{H}}_{\ell} \in \mathbb{C}^{(M-1)\times N}$ designate the submatrix

obtained by striking $\widehat{\mathbf{h}}_\ell^{\mathsf{H}}$ out of $\widehat{\mathbf{H}}$. The received signal at the ℓth user is then given by

$$\begin{aligned}
\widehat{y}_{\texttt{RCI}}^{[\ell]} &= \sqrt{P}\widehat{g}_{\texttt{RCI}}\mathbf{h}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}^{\mathsf{H}}\mathbf{c}+z_\ell \\
&= \sqrt{P}\widehat{g}_{\texttt{RCI}}\mathbf{h}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell c_\ell+\sqrt{P}\widehat{g}_{\texttt{RCI}}\mathbf{h}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\mathbf{c}_\ell+z_\ell \\
&\overset{③}{=} \underbrace{\sqrt{P}\frac{\widehat{g}_{\texttt{RCI}}}{1+\tau}\widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell c_\ell}_{\texttt{desired term}} \\
&+ \underbrace{\sqrt{P}\frac{\widehat{g}_{\texttt{RCI}}}{1+\tau}\widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\mathbf{c}_\ell+\sqrt{P}\widehat{g}_{\texttt{RCI}}\breve{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}^{\mathsf{H}}\mathbf{c}+z_\ell}_{\texttt{interference plus noise term}}
\end{aligned} \tag{3.34}$$

where ③ follows from (3.4) and we further considered $\mathbf{c}_\ell$ as a subvector obtained by removing c_ℓ from $\mathbf{c}$. To further continue, we consider the following lemma [48]:

Lemma 3.4: If $\mathbf{a}$ is a column vector of an appropriate size, then

$$\left(\mathbf{A}+\mathbf{a}\mathbf{a}^{\mathsf{H}}\right)^{-1}\mathbf{a}=\frac{\mathbf{A}^{-1}\mathbf{a}}{1+\mathbf{a}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{a}}$$

Based on lemma 3.4, we have

$$\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell=\frac{\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell}{1+\widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell} \tag{3.35}$$

In this case, the desired signal energy is equal to

$$E_{\texttt{desired signal}}=P\left[\frac{\widehat{g}_{\texttt{RCI}}A_\ell}{(1+\tau)(1+A_\ell)}\right]^2 \tag{3.36}$$

where $A_\ell=\widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell$. With respect to lemmas 3.1, 3.2 and 3.3, and by taking the expectation over $\mathbf{c}$ and $\breve{\mathbf{H}}$, the energy of the interference term can be written as

$$E_{\texttt{interference}}=P\left[\frac{\widehat{g}_{\texttt{RCI}}}{(1+\tau)(1+A_\ell)}\right]^2 B_\ell+G \tag{3.37}$$

where

$$B_\ell=\widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell$$

and by considering (3.32), we have

$$G = P\widehat{g}_{\mathtt{RCI}}^2\,\mathbb{E}_{\breve{\mathbf{H}},\mathbf{c}}\left\{\breve{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}^{\mathsf{H}}\mathbf{c}\mathbf{c}^{\mathsf{H}}\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^{\mathsf{H}}\widehat{\mathbf{H}}+\varepsilon\mathbf{I}\right)^{-1}\breve{\mathbf{h}}_\ell\right\} = \frac{P\tau}{1+\tau} \tag{3.38}$$

Consequently, the output SINR of the ℓth user can be shown as

$$\begin{aligned}\widehat{\eta}_{\mathtt{RCI}} &= \frac{E_{\mathtt{desired\ signal}}}{E_{\mathtt{interference}}+\sigma^2} \\ &= \frac{\widehat{g}_{\mathtt{RCI}}^2 A_\ell^2 P}{\widehat{g}_{\mathtt{RCI}}^2 B_\ell P + P\tau\left(1+\tau\right)\left(1+A_\ell\right)^2 + \sigma^2\left(1+\tau\right)^2\left(1+A_\ell\right)^2}\end{aligned} \tag{3.39}$$

Note that the output SINR of each user due to the standard RCI with perfect CSI can be easily obtained by setting $\tau = 0$ and replacing $\widehat{\mathbf{H}}$ with $\mathbf{H}$.

Remark 3.1: Note that although the derived bounds in (3.20)–(3.25) are based on the output SINR of CI precoding, they likewise hold for RCI. This is due to the fact that the output SINR of MMSE equalizers (conditioned on the channel realization) is asymptotically equal to that of zero forcing (ZF) equalizers plus a gap [49]. Furthermore, these aforementioned bounds will be numerically corroborated for standard RCI precoding in section 3.6.

3.5 Adaptive RCI

Subject to perfect CSI, standard RCI outperforms CI; however with the presence of imperfect CSI, its comparative improvement to CI deteriorates. Therefore, in this section and by deriving an appropriate regularization parameter, we propose an adaptive RCI which outperforms the standard RCI under CSI mismatch. To do so, we further assume that the BS knows the variance of the channel estimation error, i.e., τ, which is possible to be known a priori, as discussed in subsection 3.2.2.

We obtain the adaptive RCI precoder by using the following optimization criterion:

$$\arg\min_{\widehat{\mathbf{\Psi}}}\ \mathbb{E}\left\{\left\|\sqrt{P}\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}+f\mathbf{z}-\sqrt{P}\mathbf{c}\right\|_2^2\right\} \tag{3.40}$$

where

$$f = \frac{1+\tau}{g} = (1+\tau)\sqrt{\mathrm{Tr}\left[\widehat{\mathbf{\Psi}}^{\mathsf{H}}\widehat{\mathbf{\Psi}}\right]} \tag{3.41}$$

and g is the scaling factor. The inclusion of f in (3.40) is due to the fact that in all precoding schemes like CI and RCI, the power of noise is affected by the precoding matrix, and consequently, this effect can be reflected through a multiplicative factor like f. This can be perceived with respect to the fact that at transmit side, the transmitted signals get scaled by g to meet the power constraints; consequently at receive side and by considering (3.34), to have an unbiased detection, the received signals should be scaled back by $(1+\tau)/g$ which further appears as a multiplicative factor for the noise vector.

The objective function in (3.41) can be shown as

$$\begin{aligned} F &= \mathbb{E}\left\{\mathrm{Tr}\left[\left(\sqrt{P}\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}+f\mathbf{z}-\sqrt{P}\mathbf{c}\right)\left(\sqrt{P}\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}+f\mathbf{z}-\sqrt{P}\mathbf{c}\right)^{\mathsf{H}}\right]\right\} \\ &= \mathbb{E}\Big\{\mathrm{Tr}\Big[P\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}\mathbf{c}^{\mathsf{H}}\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}+f^2\mathbf{z}\mathbf{z}^{\mathsf{H}}-P\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}\mathbf{c}^{\mathsf{H}}-P\mathbf{c}\mathbf{c}^{\mathsf{H}}\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}+P\mathbf{c}\mathbf{c}^{\mathsf{H}} \\ &+ \sqrt{P}\,f\mathbf{H}\widehat{\mathbf{\Psi}}\mathbf{c}\mathbf{z}^{\mathsf{H}}+\sqrt{P}f\mathbf{z}\mathbf{c}^{\mathsf{H}}\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}-\sqrt{P}f\mathbf{c}\mathbf{z}^{\mathsf{H}}-\sqrt{P}f\mathbf{z}\mathbf{c}^{\mathsf{H}}\Big]\Big\} \\ &\overset{④}{=} P\,\mathrm{Tr}\left[\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\widehat{\mathbf{\Psi}}\right]+M\sigma^2(1+\tau)^2\,\mathrm{Tr}\left[\widehat{\mathbf{\Psi}}^{\mathsf{H}}\widehat{\mathbf{\Psi}}\right]-P\,\mathrm{Tr}\left[\mathbf{H}\widehat{\mathbf{\Psi}}\right] \\ &- P\,\mathrm{Tr}\left[\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right]+PM \end{aligned} \tag{3.42}$$

where ④ follows lemma 3.3. To obtain the sought precoder, we can differentiate F with respect to $\widehat{\mathbf{\Psi}}$ by first considering the following assumptions [50, 51]:

1. $\widehat{\mathbf{\Psi}}$ and $\widehat{\mathbf{\Psi}}^{\mathsf{H}}$ are treated as independent variables.
2. $\dfrac{\partial\,\mathrm{Tr}\left[\mathbf{A}\widehat{\mathbf{\Psi}}\right]}{\partial\widehat{\mathbf{\Psi}}}=\dfrac{\partial\,\mathrm{Tr}\left[\widehat{\mathbf{\Psi}}\mathbf{A}\right]}{\partial\widehat{\mathbf{\Psi}}}=\mathbf{A}.$

Following the preceding assumptions, the differentiation of F with respect to $\widehat{\mathbf{\Psi}}$ gives

$$\begin{aligned} \frac{\partial F}{\partial\widehat{\mathbf{\Psi}}} &= P\widehat{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}+M\sigma^2(1+\tau)^2\,\widehat{\mathbf{\Psi}}^{\mathsf{H}}-P\mathbf{H} \\ &\overset{⑤}{=} P\widehat{\mathbf{\Psi}}^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right)^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right)+M\sigma^2(1+\tau)^2\,\widehat{\mathbf{\Psi}}^{\mathsf{H}} \\ &- P\left(\frac{\widehat{\mathbf{H}}}{1+\tau}+\breve{\mathbf{H}}\right) \end{aligned} \tag{3.43}$$

where ⑤ follows from (3.4). The adaptive precoder can then be found by setting $\dfrac{\partial F}{\partial\widehat{\mathbf{\Psi}}}$ equal to zero and taking the expectation over $\breve{\mathbf{H}}$. Therefore with respect to the fact that due to

lemma 3.1, $\widehat{\mathbf{H}}$ and $\breve{\mathbf{H}}$ are independent, we can represent the precoding matrix as

$$\mathbb{E}_{\breve{\mathbf{H}}}\left\{\frac{\partial F}{\partial \widehat{\mathbf{\Psi}}}\right\} = \mathbf{0} \implies \widehat{\mathbf{\Psi}} = \widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \widehat{\varepsilon}\,\mathbf{I}\right)^{-1} \tag{3.44}$$

where the regularization parameter $\widehat{\varepsilon}$ can now be expressed as

$$\widehat{\varepsilon} = M\left(1+\tau\right)\left(\tau + \gamma^{-1}\left(1+\tau\right)^{3}\right) \tag{3.45}$$

Note that by setting $\tau = 0$, $\widehat{\varepsilon}$ boils down to $\varepsilon = M\gamma^{-1}$ which is the appropriate regularization parameter under perfect CSI.

Remark 3.2: From (3.43), it can be concluded that since the Hessian matrix (second order derivative) of the objective function is positive definite, the expression in (3.44) is a global minimizer, and therefore the regularization parameter $\widehat{\varepsilon}$ is optimum.

Therefore, for the proposed adaptive RCI, the transmitted signal from BS can be shown as

$$\widehat{\mathbf{s}}_{\texttt{adaptive RCI}} = \widetilde{g}\widehat{\mathbf{\Psi}}\mathbf{c} \tag{3.46}$$

where $\widehat{\mathbf{\Psi}}$ is defined in (3.44), and $\widetilde{g}$ is the scaling factor which can be represented as

$$\widetilde{g} = \frac{1}{\sqrt{\mathrm{Tr}\left[\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \widehat{\varepsilon}\,\mathbf{I}\right)^{-2}\right]}} \tag{3.47}$$

Similar to the standard RCI, it is straightforward to show that the output SINR of the ℓth user based on the proposed adaptive RCI can now be expressed as

$$\widetilde{\eta}_{\texttt{RCI}} = \frac{\widetilde{g}^2\widehat{A}_\ell^2 P}{\widetilde{g}^2\widehat{B}_\ell P + P\tau\left(1+\tau\right)\left(1+\widehat{A}_\ell\right)^2 + \sigma^2\left(1+\tau\right)^2\left(1+\widehat{A}_\ell\right)^2} \tag{3.48}$$

where $\widehat{A}_\ell = \widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell + \widehat{\varepsilon}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell$ and

$$\widehat{B}_\ell = \widehat{\mathbf{h}}_\ell^{\mathsf{H}}\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell + \widehat{\varepsilon}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell\left(\widehat{\mathbf{H}}_\ell^{\mathsf{H}}\widehat{\mathbf{H}}_\ell + \widehat{\varepsilon}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{h}}_\ell$$

such that $\widehat{\varepsilon}$ is defined in (3.45).

Note that with respect to remark 3.1, the derived bounds in (3.20)–(3.25) are also valid for the proposed adaptive RCI technique.

3.6 Numerical Results

In this section, by using simulation results, we substantiate the analytically derived bounds in (3.20)–(3.25). We also demonstrate the superior performance achieved by the proposed adaptive RCI compared to the standard RCI. For the case of imperfect CSI, we assume that the channel estimation error variance obeys (3.3), i.e., $\tau = \beta\gamma^{-\alpha}$ where γ is the SNR, α is the SNR exponent and β is the SNR scaling factor.

We evaluate the sum capacity as [13]

$$\log_2 \det\left(\mathbf{I} + \frac{\gamma}{N}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right) \tag{3.49}$$

and the achievable sum rates as [13, 52]

$$\sum_{\ell=1}^{M} \log_2\left(1 + \mathrm{SINR}_\ell\right) \tag{3.50}$$

where SINR_ℓ denotes the output SINR of the ℓth user. For instance, for the proposed adaptive RCI, $\mathrm{SINR}_\ell = \widetilde{\eta}_{\mathtt{RCI}}$ where $\widetilde{\eta}_{\mathtt{RCI}}$ is defined in (3.48).

Note that similar to standard RCI and CI, adaptive RCI is also independent of the constellation, i.e., it is based on Gaussian input signaling. However, without loss of generality and in the interest of verifying the accuracy of the derived SINRs, we utilize the following formula within Fig. 3.6, which is a good criterion to analytically evaluate the SER of each user when QPSK constellation is used [53]

$$\mathrm{SER}_{\mathtt{QPSK}} = \mathrm{erfc}\left(\sqrt{\frac{\mathrm{SINR}_\ell}{2}}\right)\left[1 - \frac{1}{4}\,\mathrm{erfc}\left(\sqrt{\frac{\mathrm{SINR}_\ell}{2}}\right)\right] \tag{3.51}$$

wherein

$$\mathrm{erfc}\left(x\right) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\mathrm{d}t \tag{3.52}$$

is the complementary error function. In Fig. 3.6, we assume $M = N = 10$, and simulated results are due to counting the number of occurred errors in received signals when the transmitted signals are based on what is expressed in (3.46). As revealed, both analytical and simulated results are in close agreement, which verify the validity of the derived output SINR in (3.48). This can be similarly used to verify the validity of the derived SINRs of CI and standard RCI as well, though the corresponding curves are omitted for the sake of compactness.

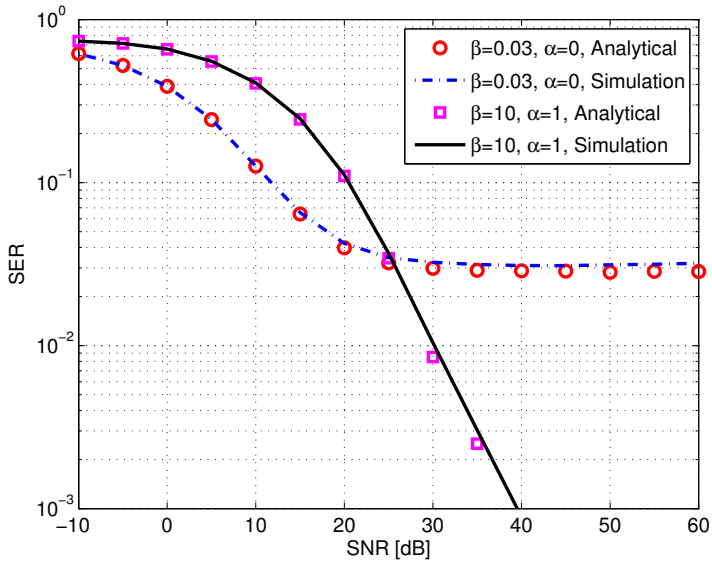

Figure 3.6: *Comparison between analytical and simulated SER of adaptive RCI precoding under QPSK signaling for* $M = N = 10$ *and for the cases* $\beta = 0.03,\ \alpha = 0$*, and* $\beta = 10,\ \alpha = 1$.

As mentioned earlier, although the derived bounds in (3.20)–(3.25) are based on the output SINR of CI precoding, they are likewise valid for both standard and adaptive RCI. However, without loss of generality and to avoid congestion in Fig. 3.7, we just certify the aforementioned bounds using standard RCI subject to different CSI qualities and for the case $M = N = 10$, where the following performance trends are observed:

1. $\alpha > 1$: While (3.20) indicates that the asymptotic mean loss in sum rate should be equal to zero, (3.25) denotes that full DoF, i.e., 10 DoF, is achievable. All these bounds are certified where the corresponding curve overlaps with the one representing the perfect CSI at high SNRs.

2. $\alpha = 1$: While (3.20) indicates that the asymptotic mean loss in sum rate should be equal to a nonzero finite constant, (3.25) denotes that full DoF, i.e., 10 DoF, should be achievable. These bounds are also certified where the corresponding curve has the same slope with that of the perfect CSI, and consequently there is a nonzero constant gap between them.

3. For $\beta = 10,\ \alpha = 1$ and based on (3.23), we expect that the asymptotic mean loss in sum rate should be 35 bits per channel use which is verified by the depicted results in Fig. 3.7. Plus, based on (3.24), we expect that in the case of $\beta = 10,\ \alpha = 1$

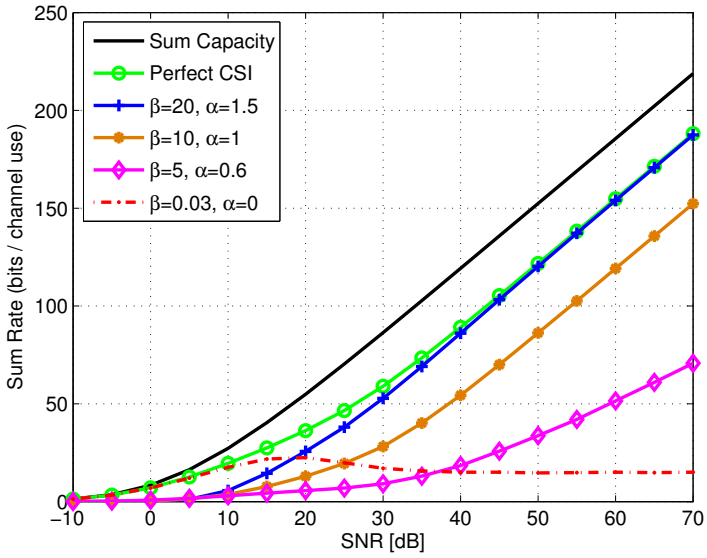

Figure 3.7: *Sum rate of standard RCI precoding for $M = N = 10$ and under different CSI qualities.*

and to achieve the same sum rate as the perfect CSI, we should increase the transmit power by 10.4 dB, which is again certified in Fig. 3.7.

4. $0 \leq \alpha < 1$: In this case, equation (3.20) indicates that the asymptotic mean loss in sum rate is unboundedly increasing with SNR. This is again certified in Fig. 3.7, such that when SNR gets larger, the gap between the corresponding curves and that of the perfect CSI becomes wider. Also based on (3.25), we expect that an α fraction of the total DoF should be achievable. By considering the slope of the curves in the same figure, while for $\alpha = 0.6$ the achievable DoF is now 6, for the case $\alpha = 0$, this is equal to zero.

Note that the preceding performance trends can also be regarded as an approval for the validity of the derived SINR of standard RCI in (3.39), since the depicted results in Fig. 3.7 are in line with the analytically derived bounds in equations (3.20)–(3.25).

Although the promised improvement of adaptive RCI over standard RCI and CI can be gleaned for various values of α, in Figs. 3.8–3.13, we just focus on two representative cases: $\alpha = 0$ (which imitates the CSI feedback scenario), and $\alpha = 1$ (which imitates the reciprocal channels). More specifically and with respect to the error variance τ defined in (3.3), we consider two cases: $\beta = 10,\ \alpha = 1$ and $\beta = 0.03,\ \alpha = 0$. We also assume that $M = N = 10$ unless stated otherwise.

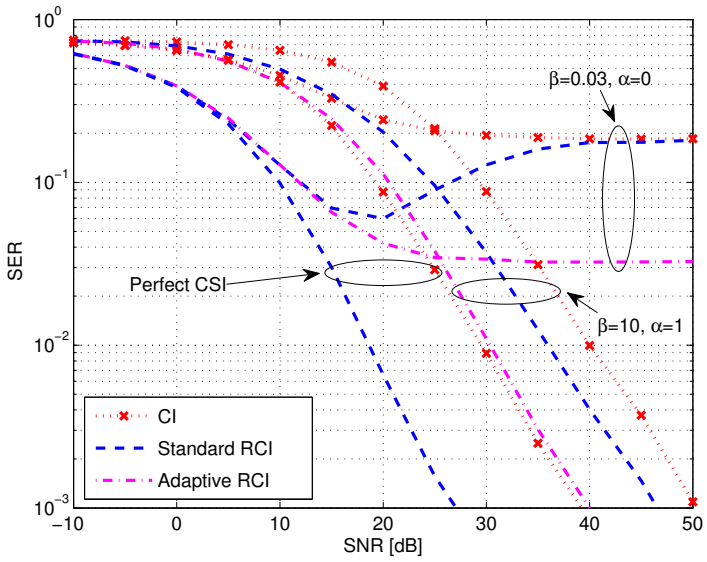

Figure 3.8: *SER under QPSK signaling for* $M = N = 10$ *and for the cases* $\beta = 0.03,\ \alpha = 0$*, and* $\beta = 10,\ \alpha = 1$.

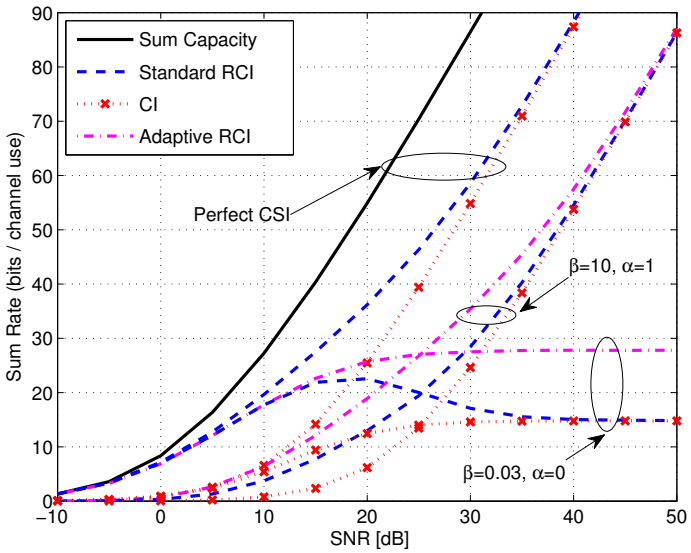

Figure 3.9: *Sum rate for* $M = N = 10$ *and for the cases* $\beta = 0.03,\ \alpha = 0$*, and* $\beta = 10,\ \alpha = 1$.

Fig. 3.8 illustrates the SER of CI and RCI under QPSK signaling. As demonstrated, the proposed adaptive RCI achieves better SER than standard RCI. For example, when $\alpha = 1$, adaptive RCI achieves nearly 6 dB gain compared to standard RCI to achieve the SER of 10^{-2}.

In Fig. 3.9, sum rates of CI and RCI under perfect and imperfect CSI are depicted.

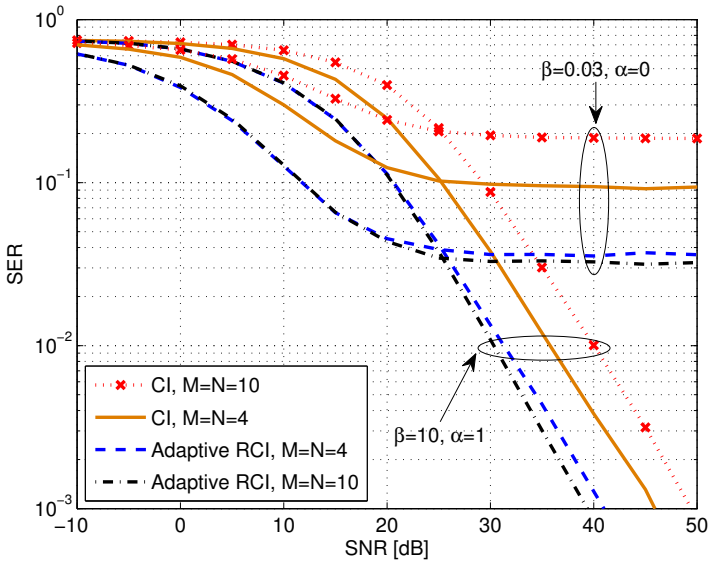

Figure 3.10: *SER of CI and adaptive RCI under QPSK signaling for* $M = N = 4,\ M = N = 10$ *and for the cases* $\beta = 0.03,\ \alpha = 0$*, and* $\beta = 10,\ \alpha = 1$.

As shown, for different values of α, the proposed adaptive RCI achieves higher sum rates than standard RCI, e.g., for the case $\alpha = 0$, adaptive RCI achieves 13 bits per channel use gain in sum rate compared to standard RCI at high SNRs. Note that for the case of $\alpha = 1$, we expect that the achievable DoF should be the same as that of the perfect CSI. This is confirmed in Fig. 3.9 where it can be seen that the slope of the curves related to the case $\alpha = 1$ is the same as that of the perfect CSI at high SNRs.

By considering Figs. 3.8–3.9, one interesting observation is that when $\alpha = 0$, while the performance trend of the standard RCI is nonmonotonic, that of the proposed adaptive RCI is monotonic.

Fig. 3.10 illustrates the SER of CI and adaptive RCI under QPSK signaling and for $M = N = 4$ and $M = N = 10$ under $\beta = 0.03,\ \alpha = 0$, and $\beta = 10,\ \alpha = 1$. As revealed, under different CSI qualities and by increasing M, while the SER due to CI precoding deteriorates, that of the adaptive RCI remains fixed at low-to-intermediate SNRs, and improves slightly at high SNRs.

Fig. 3.11 depicts the sum rate of linear precoders as a function of M and N, at $\gamma = 10$ dB and $\gamma = 25$ dB and in the case of $\alpha = 1$. As revealed, with increasing M, while the sum rate of CI does not linearly increase, those of standard and adaptive RCI do. Also adaptive RCI outperforms standard RCI at both low and high SNRs such that the larger the M and N, the more gain in sum rate can be gleaned by deploying adaptive RCI.

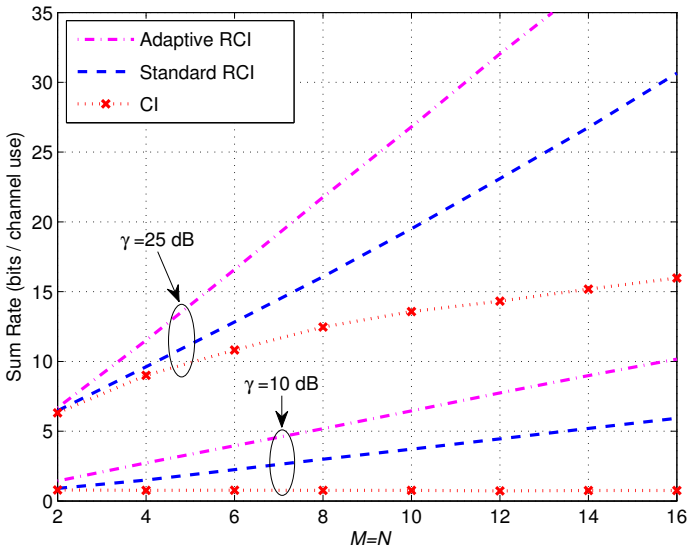

Figure 3.11: *Sum rate of adaptive and standard RCI compared to that of CI as a function of* M *and* N*, at SNRs of* 10 *dB and* 25 *dB and for the case of* $\beta = 10, \alpha = 1$.

Fig. 3.12 demonstrates the average sum rate of linear precoders as a function of M and N, at $\gamma = 10$ dB and $\gamma = 25$ dB when $\alpha = 0$. In this case, at low SNRs and with increasing M, the sum rate of adaptive RCI does not improve in comparison with that of standard RCI. However, as SNR gets increased, the performance of adaptive RCI becomes more prominent than that of standard RCI.

As mentioned earlier, the concept of regularization is most beneficial when there are equal number of transmit and receive antennas. Nevertheless, in Fig. 3.13, we compare the SER of adaptive RCI with standard RCI and CI when the number of antennas at BS is more than the number of receive antennas in total. The results are depicted under QPSK signaling, $\beta = 10,\ \alpha = 1$ and $\beta = 0.03,\ \alpha = 0$, when $M = 8$ and $N = 10$. As revealed, even for nonsquare channels, the proposed adaptive RCI achieves better performance that standard RCI and CI. For instance, when $\alpha = 1$, while the SER of standard RCI is almost the same as that of CI, adaptive RCI achieves 2dB gain to reach the same SER. It is also worthwhile to point out that for the case of $\alpha = 0$, SER curves get saturated at high SNRs due to the fact that in this case, the system becomes interference limited.

Finally, it is worthwhile to point out why we reckon that the proposed adaptive RCI compensates the degraded performance of the standard RCI compared to CI subject to CSI mismatch. This can be clearly observed in Figs. 3.8, 3.9, and 3.13. For example, as illustrated in Fig. 3.8, while under perfect CSI, standard RCI achieves nearly 10 dB gain

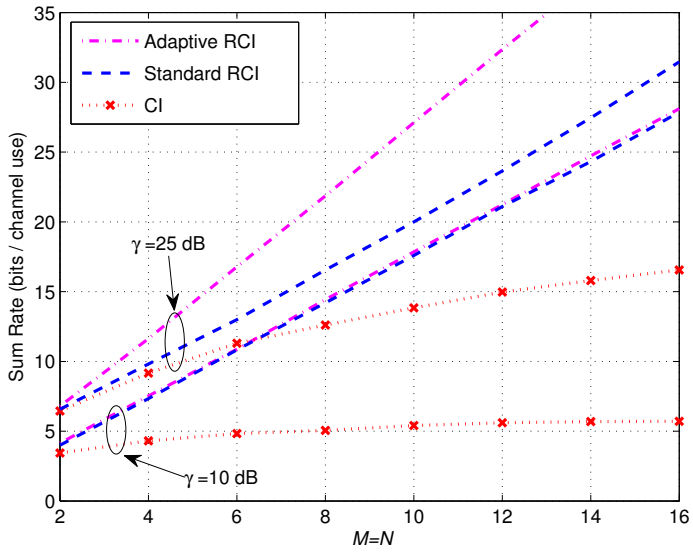

Figure 3.12: *Sum rate of adaptive and standard RCI compared to that of CI as a function of* M *and* N*, at SNRs of* 10 *dB and* 25 *dB and for the case of* $\beta = 0.03, \alpha = 0$*.*

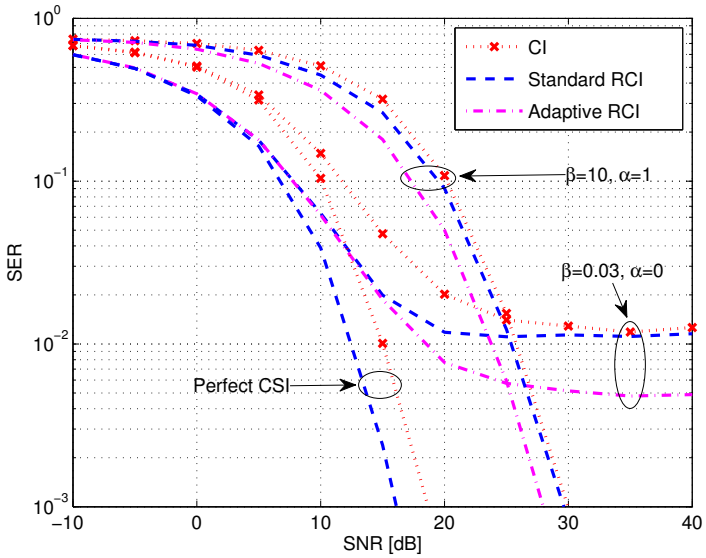

Figure 3.13: *SER under QPSK signaling for* $M = 8,\ N = 10$ *and for the cases* $\beta = 0.03,\ \alpha = 0$*, and* $\beta = 10,\ \alpha = 1$*.*

compared to CI to reach the SER of 10^{-2}, under imperfect CSI, say $\alpha = 1$, this gain is nearly 4 dB. However, adaptive RCI compensates this 6 dB loss in performance such that the achieved gain is now equal to 10 dB.

3.7 Summary

In this chapter, we quantified the performance of linear precoders, namely CI and RCI, in the MISO downlink under an imperfect CSI model where the variance of the CSI measurement error depends on the SNR. We first analyzed the performance of CI precoding and showed that when $M = N$, the largest eigenvalue of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$ has a highly erratic behavior which results in a degraded performance for CI. In this case and by employing the concept of regularization, better performance can be gleaned. This way, the RCI precoder now consists of $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \varepsilon\,\mathbf{I}\right)^{-1}$ as opposed to the CI precoder which involves $\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)^{-1}$. We then derived novel bounds regarding the asymptotic mean loss in sum rate and the achievable DoF of CI and RCI. For example, we showed that when this error variance scales with the inverse of SNR, full DoF is achievable, and the asymptotic mean loss in sum rate is therefore a nonzero finite value. More precisely, we showed that if the intention is to keep the asymptotic mean loss in sum rate bounded, the error variance must at least scales with the inverse of SNR. It was also shown that subject to imperfect CSI, the comparative improvement of standard RCI to CI becomes negligible. Accordingly, we proposed an adaptive RCI by deriving an appropriate regularization parameter as a function of error variance. Simulation results showed that the proposed adaptive RCI outperforms standard RCI under CSI mismatch such that it compensates the degraded performance of standard RCI compared to CI.

Chapter 4

Transmit-Power Efficient Linear Precoding

4.1 Introduction

In chapter 3, we analyzed the performance of linear precoders, namely CI and RCI, with the presence of perfect and imperfect CSI at BS. In this chapter, we propose alternative precoders which enable us to glean more benefits compared to CI and RCI in downlink cellular communications. These enhanced precoders are likewise linear which can be readily implemented alongside CI and RCI.

In line with the aim of linear precoders to achieve reduced complexity compared to nonlinear precoders, [54] proposed a precoding technique based on phase rotation (which hereafter we call it PA) for multiantenna downlink communications, where instead of removing the harmful symbol-to-symbol interference, it rotates the phases of the transmitted symbols such that the destructive interference becomes constructive, and eventually leading to more output SINRs for a fixed transmit power at BS. Further, the superior performance of PA precoding compared to standard linear precoders has been investigated in [55, 56] for cognitive radio networks.

As discussed in chapter 3, the aim of standard and adaptive RCI is to improve the SER and the achievable sum rates compared to CI. Aside from improving the SER or throughput in downlink cellular networks, designing power-efficient precoders has become important in recent years. The idea is to minimize the transmit power while securing the same QoS for each user. Accordingly, we focus on designing such precoders that enable us to decrease the transmit power in order to achieve the same average output SINR for each

user. Due to their practical complexity, we focus on linear precoders.

First, we reformulate and enhance the performance analysis of PA precoding in [54] so that, in line with the aims of green communications [36], the power efficiency (as opposed to output SINR) can be optimized. In addition, we complete the performance analysis of PA by analytically calculating the scaling factors, where in [54] only empirical scaling factors were used for the theoretical results.

We also propose an enhanced PA technique, namely RPA, where based on the performance analysis of PA, we derive the required regularization factor for RPA. We analytically derive the output SINR of the proposed RPA scheme. We also show that to achieve the same average output SINR for each user, the transmit power reduction achieved by RPA compared to RCI precoding is the same as that of PA to CI precoding.

It is also shown that the power gains of RPA compared to its counterparts PA, CI and RCI magnify as the number of transmit antennas increases, which aligns the proposed scheme with the aims of massive MIMO [57]. In particular, we observe up to more than 50 times savings in the transmit power for RPA (PA) compared to RCI (CI) for systems with up to a hundred transmit antennas.

Since from the practical point of view, it is more pragmatic to assume that only partial CSI is available at BS, we also consider the effect of channel estimation errors on the performance of the proposed schemes. We show that with imperfect transmit-side channel state information (CSIT), the performance trend of PA and RPA follows the one of the standard CI and RCI precoders, which further implies that PA and RPA are as sensitive as the others to erroneous CSIT. Moreover, we consider an adaptive design for the proposed RPA precoding which achieves significantly better performance under the availability of imperfect CSI. We do so by deriving an optimum regularization parameter which is a function of error variance.

In Section 4.2, we reformulate and enhance the performance analysis of PA precoding. In Section 4.3, we propose RPA precoding which improves the performance of PA. In this case, we derive optimum regularization parameters under both perfect and imperfect CSI. In Section 4.4, power efficiency of PA and RPA precoding is evaluated. In Section 4.5 and by using numerical simulations, we show that RPA precoding outperforms CI, RCI and PA precoding and enables us to save more power at transmit side for an averagely constant output SINR at each user. Finally Section 4.6 contains a summary of the presented materials within this chapter.

4.2 Phase Alignment

In this section, we represent the basic idea of PA precoding in downlink cellular communications. Analogous to the system model discussed in subsection 3.2.1, we consider MISO downlink communications with block Rayleigh fading such that $\mathbf{H} \in \mathbb{C}^{M\times N}$ has zero-mean unit-variance elements, i.e., $\mathrm{vec}\,(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \mathbf{I})$. Without loss of generality, we further assume that the total number of transmit antennas is equal to or more than the total number of receive antennas, i.e., $N \geq M$. We also define $\mathbf{R}$ as the covariance matrix of the channel, i.e., $\mathbf{R} = \mathbf{H}\mathbf{H}^{\mathsf{H}}$. To distinguish the nominal SNR from the output SNR at each MT, throughout this chapter, we use the term "input SNR" in lieu of the nominal SNR, i.e., $\gamma = P/\sigma^2$.

Note that although the PA precoding was defined in [54], it needs to be redefined in a relatively different way. This is due to the fact that there is no closed-form expression for average output SNR of PA precoding. However, in this chapter, we derive a closed-form expression for this average output SNR which eventually enables us to calculate the amount of transmit-power reduction of PA precoding compared to CI precoding for a constant output SNR at each user. This also facilitates the selection of an optimized regularization parameter for the proposed RPA precoding.

We note that the concept of phase alignment is most beneficial in high interference scenarios where more gains are to be gleaned by exploiting interference. In these scenarios typically low order modulation is employed to secure low error rates. Therefore, while the benefits of the proposed scheme extend to high order QAM modulations, here we focus on low order PSK.

With PA precoding, instead of nulling out the destructive symbol-to-symbol (or co-channel) interference (as being done by using $\mathbf{R}^{-1}$ for CI precoding in (3.7)), the knowledge of the data's and channels' covariance matrices at transmit side can be used to make the harmful interference constructive. Fig. 4.1 shows how PA precoding works under QPSK signaling. If we consider the signal of interest as $c_\ell = (1+j)/\sqrt{2}$ and the interfering symbol as $c_x = (-1+j)/\sqrt{2}$, the symbol-to-symbol co-channel interference $\rho_{\ell,x}c_x$ resulted from c_x to c_ℓ through the (ℓ, x)th element of channel's covariance matrix $\mathbf{R}$, i.e., $\rho_{\ell,x}$, is denoted by the dashed red arrow in the figure. The phase of the interference $\rho_{\ell,x}c_x$ with respect to the signal of interest c_ℓ is denoted by $\theta_{\ell,x}$. For QPSK constellation the real and imaginary axes are decision thresholds. It is clear that for the interfering symbol c_x the resulting interference $\rho_{\ell,x}c_x$ is harmful since its accumulation with the signal of interest move the received symbol $\ddot{y}_\ell$ closer to the QPSK decision thresholds. The goal of the

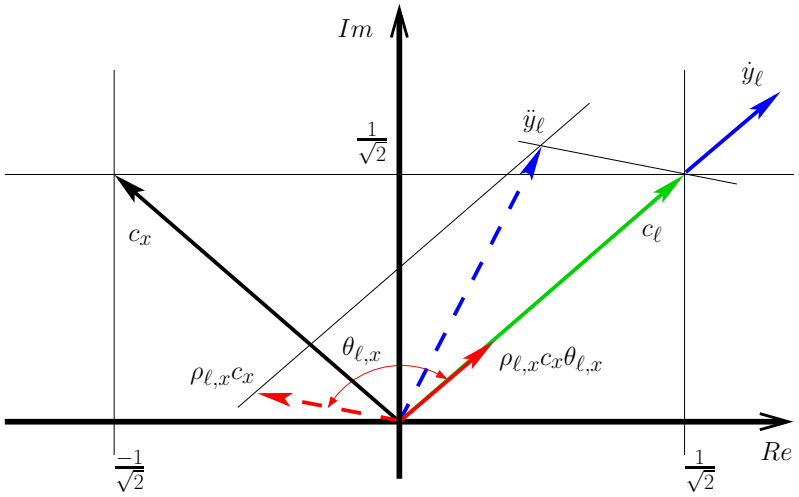

Figure 4.1: *Phase alignment for QPSK constellation. $\ddot{y}_\ell$ is the received symbol without phase alignment while $\dot{y}_\ell$ is the received symbol with phase alignment.*

phase alignment precoding is to correct the phase of all transmitted symbols and rotating the angle of correlation between them such that the resulting symbols after precoding are aligned to the signal of interest c_ℓ. The desired symbol c_ℓ and the aligned interference $\rho_{\ell,x}c_x\theta_{\ell,x}$, respectively, are shown by the solid green and red arrows in Fig. 4.1, which add up to $\dot{y}_\ell$ denoted by solid blue one. With respect to the fact that the magnitudes of c_ℓ and c_x are equal to one, i.e., $|c_\ell| = |c_x| = 1$, the relative phase $\theta_{\ell,x}$ can be expressed as

$$\rho_{\ell,x}c_x\theta_{\ell,x} \propto c_\ell \Longrightarrow \theta_{\ell,x} = \frac{(\rho_{\ell,x}c_x)^{\mathsf{H}}}{|\rho_{\ell,x}|}c_\ell \tag{4.1}$$

where $\propto$ means linear proportionality. From (4.1) it is evident that $|\theta_{\ell,x}| = 1$ and therefore the amplitude of the rotated correlations remains unchanged. Now the matrix $\mathbf{R}_\theta$ which contains the phase rotated correlation elements can be shown as

$$[\mathbf{R}_\theta]_{\ell,x} = \rho_{\ell,x}\theta_{\ell,x} = \rho_{\ell,x}\frac{(\rho_{\ell,x}c_x)^{\mathsf{H}}}{|\rho_{\ell,x}|}c_\ell = |\rho_{\ell,x}|c_\ell c_x^{\mathsf{H}} \tag{4.2}$$

From the matrix algebra perspective, (4.2) is equivalent to

$$\mathbf{R}_\theta = |\mathbf{R}| \odot \mathbf{C} \tag{4.3}$$

where $\mathbf{C} = \mathbf{c}\mathbf{c}^{\mathsf{H}}$ is the covariance matrix of the input data vector $\mathbf{c}$ and $\odot$ denotes Hadamard (element-wise) matrix product.

Remark 4.1: By considering (4.3) and with respect to the fact that the Hadamard product is a linear operator, it can be seen that PA precoding (and consequently the proposed RPA precoding as being discussed later) is linear, as also stated in [54].

Remark 4.2: As denoted in (4.3), PA precoding is only dependant on the amplitudes of the elements of the channel's covariance matrix as well as the covariance matrix of the transmitted data which are known at BS prior to downlink transmission. Therefore similar to CI and RCI precoding, the signal processing enhancement of PA precoding is going to be done at BS and consequently no overhead is introduced to MTs.

With respect to (3.1), the receive signal can now be shown as

$$\mathbf{y}_{\mathtt{PA}} = \sqrt{P}\mathbf{H}\mathbf{s}_{\mathtt{PA}} + \mathbf{z} \tag{4.4}$$

where

$$\mathbf{s}_{\mathtt{PA}} = g_{\mathtt{PA}}\mathbf{\Psi}_{\mathtt{PA}}\mathbf{c} \tag{4.5}$$

is the transmitted signal, and the precoding matrix can be defined as

$$\mathbf{\Psi}_{\mathtt{PA}} = \mathbf{H}^{\mathsf{H}}\mathbf{R}^{-1}\mathbf{R}_{\theta} \tag{4.6}$$

and the scaling factor is equal to [54]

$$g_{\mathtt{PA}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\mathbf{R}_{\theta}^{2}\mathbf{R}^{-1}\right]}} \tag{4.7}$$

4.2.1 Instantaneous Output SNR

In this subsection and by assuming PA precoding, the instantaneous output SNR of each user is derived. After going through the channel, the received signal related to the ℓth user ($\ell = 1, \ldots, M$) can be shown as

$$\begin{aligned} y_{\mathtt{PA}}^{[\ell]} &= \sqrt{P}g_{\mathtt{PA}}\left[\mathbf{R}_{\theta}\right]_{\ell\star}\mathbf{c} + z_{\ell} = \sqrt{P}g_{\mathtt{PA}}\sum_{x=1}^{M}|\rho_{\ell,x}|c_{\ell}c_{x}^{\mathsf{H}}c_{x} + z_{\ell} \\ &= \sqrt{P}g_{\mathtt{PA}}c_{\ell}\sum_{x=1}^{M}|\rho_{\ell,x}| + z_{\ell} \\ &= \underbrace{\sqrt{P}g_{\mathtt{PA}}c_{\ell}|\rho_{\ell,\ell}|}_{\mathtt{desired\ signal}} + \underbrace{\sqrt{P}g_{\mathtt{PA}}c_{\ell}\sum_{\substack{x=1\\x\neq\ell}}^{M}|\rho_{\ell,x}|}_{\mathtt{constructive\ interference}} + z_{\ell} \end{aligned} \tag{4.8}$$

	Type of Random Variable	$\mathbb{E}\{\lvert\rho_{\ell,x}\rvert\}$	$\mathbb{E}\{\lvert\rho_{\ell,x}\rvert^2\}$
$\ell \neq x$	Rayleigh	$\sigma_h^2\sqrt{N\pi}/2$	$\sigma_h^4 N$
$\ell = x$	χ-square	$\sigma_h^2 N$	$\sigma_h^4 N\left(N+1\right)$

Table 4.1: *Statistical properties of $|\rho_{\ell,x}|$, where $\rho_{\ell,x}$ is the (ℓ,x)th element of $\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^{\mathsf{H}}$ and $\mathrm{vec}\left(\boldsymbol{\mathcal{H}}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma_h^2\mathbf{I}\right)$ (Appendix A).*

where $[\mathbf{R}_\theta]_{\ell\star}$ denotes the ℓth row of matrix $\mathbf{R}_\theta$ in (4.3), and z_ℓ is the ℓ-th element of the noise vector $\mathbf{z}$ which is the circularly symmetric additive white Gaussian noise with zero mean and variance σ^2, i.e., $z_\ell \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma^2\right)$. From (4.8), it can be seen that the received signal, due to the phase alignment of the co-stream interference, is a factor of only the desired symbol c_ℓ and not the interfering symbol c_x, as depicted in Fig. 4.1. Since this interference contributes to the signal power, the effective SINR instead of the standard form ($\check{\eta} = \frac{S}{I+\sigma^2}$) can be expressed as $\hat{\eta} = \frac{S+I}{\sigma^2}$ where S denotes the desired signal's power, I is the additional signal power due to the constructive interference and σ^2 denotes the noise variance of each user. Therefore, it is basically a case of signal-plus-noise at the receiver and consequently for the case of PA precoding, we present an SNR calculation as opposed to SINR. Hence, similar to CI precoding, the output SNR of ℓth user based on PA precoding can be shown as

$$\eta_{\mathrm{PA}} = \frac{P g_{\mathrm{PA}}^2}{\sigma^2}\left(\sum_{x=1}^{M}\lvert\rho_{\ell,x}\rvert\right)^2 \tag{4.9}$$

4.2.2 Average Output SNR

Since no closed-form expression for average output SNR of PA precoding is presented in [54], in this subsection, we calculate this value. In order to compare the power efficiency of PA precoding to that of CI precoding and also to find an optimized regularization parameter for our proposed scheme in the following section, we should represent the output SNR of PA precoding similar to that of CI precoding in (3.9). To do so, by considering some simplifying steps and by taking the expectation over η_{PA} in (4.9), we can represent the average output SNR of each user as

$$\eta'_{\mathrm{PA}} = \frac{P\,\mathbb{E}\left\{\left(g_{\mathrm{PA}}\sum_{x=1}^{M}|\rho_{\ell,x}|\right)^2\right\}}{\sigma^2} = \frac{P\,\mathbb{E}\left\{g_{\mathrm{PA}}^2\right\}\mathbb{E}\left\{\left(\sum_{x=1}^{M}|\rho_{\ell,x}|\right)^2\right\}}{\sigma^2} \tag{4.10}$$

Here we have assumed that g_{PA} is statistically independent of the data and the channel's covariance coefficients. While (4.7) contradicts this assumption, this is an affordable and common simplification to attain a closed form approximation of the average output SNR [58]. Moreover, for large M this becomes more justifiable, as derived by the law of large numbers. To further proceed, we should derive the statistical properties of random variable $|\rho_{\ell,x}|$ which is presented in Table 4.1 (for a proof consult Appendix A). In the sequel and for the perfect CSI case, we assume that $\mathrm{vec}\,(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0},\mathbf{I})$ and consequently the statistical properties of $|\rho_{\ell,x}|$ can be obtained by setting σ_h^2 equal to one. Therefore after some straightforward manipulations, (4.10) can be expressed as

$$\eta'_{\mathrm{PA}} = \frac{P\,\mathbb{E}\left\{g_{\mathrm{PA}}^2\right\}}{\sigma^2} N\left[(N+1)+(M-1)\left(1+\sqrt{N\pi}+(M-2)\,\frac{\pi}{4}\right)\right] \tag{4.11}$$

To calculate η'_{PA} in (4.11), we first consider the following theorem:

Theorem 4.1: $\mathbb{E}\left\{g_{\mathrm{PA}}^2\right\}$ can be represented as

$$\mathbb{E}\left\{g_{\mathrm{PA}}^2\right\} = \frac{1}{N\,(M+N)\,\mathrm{Tr}\left[\mathbf{R}^{-1}\right]} \tag{4.12}$$

Proof: Based on (4.7), we have

$$g_{\mathrm{PA}}^2 = \frac{1}{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]} \tag{4.13}$$

Therefore to calculate $\mathbb{E}\left\{g_{\mathrm{PA}}^2\right\}$, we should find the value of $\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]\right\}^{-1}$. Since both $\mathbf{R}_\theta^2$ and $\mathbf{R}^{-1}$ are Hermitian matrices, by using eigen-decomposition we have

$$\mathbf{R}_\theta^2 = \mathbf{U}_\theta\mathbf{\Lambda}_\theta\mathbf{U}_\theta^{\mathsf{H}} \tag{4.14}$$

$$\mathbf{R}^{-1} = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^{\mathsf{H}} \tag{4.15}$$

where $\mathbf{U}_\theta$ and $\mathbf{U}_r$ are unitary matrices containing the eigenvectors of $\mathbf{R}_\theta^2$ and $\mathbf{R}^{-1}$, respectively, and $\mathbf{\Lambda}_\theta$ and $\mathbf{\Lambda}_r$ are diagonal matrices consisting of eigenvalues of $\mathbf{R}_\theta^2$ and $\mathbf{R}^{-1}$,

respectively. Note that since we assumed i.i.d. input signaling, $\mathbf{R}_\theta^2$ is a random matrix, so is $\mathbf{U}_\theta$. Therefore $\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]\right\}$ can be written as

$$\begin{aligned}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]\right\}&=\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{U}_\theta\mathbf{\Lambda}_\theta\mathbf{U}_\theta^{\mathsf{H}}\mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^{\mathsf{H}}\right]\right\}\\&=\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{U}_r^{\mathsf{H}}\mathbf{U}_\theta\mathbf{\Lambda}_\theta\mathbf{U}_\theta^{\mathsf{H}}\mathbf{U}_r\mathbf{\Lambda}_r\right]\right\}\end{aligned} \tag{4.16}$$

Now if we define $\mathbf{U}\triangleq\mathbf{U}_r^{\mathsf{H}}\mathbf{U}_\theta$, (4.16) can be shown by

$$\begin{aligned}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]\right\}&=\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{U}\mathbf{\Lambda}_\theta\mathbf{U}^{\mathsf{H}}\mathbf{\Lambda}_r\right]\right\}\\&=\mathbb{E}\left\{\sum_{i=1}^{M}\sum_{k=1}^{M}\lambda_{r,i}\lambda_{\theta,k}\left|u_{i,k}\right|^2\right\}\end{aligned} \tag{4.17}$$

where $\lambda_{r,i}$ and $\lambda_{\theta,k}$ are the ith and the kth diagonal elements of matrices $\mathbf{\Lambda}_r$ and $\mathbf{\Lambda}_\theta$, respectively, and $u_{i,k}$ denotes the (i,k)th element of unitary matrix $\mathbf{U}$.
To continue, we consider the following lemma [3]:

Lemma 4.1: If the Hermitian unitary invariant[1] random matrix $\mathbf{W}$ can be eigen-decomposed as $\mathbf{W}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{H}}$, then the unitary matrix $\mathbf{U}$, which is a Haar[2] matrix, is independent of the diagonal matrix $\mathbf{\Lambda}$.

Since both $\mathbf{U}_\theta$ and $\mathbf{U}_r$ are random matrices, so is $\mathbf{U}$, hence $\mathbb{E}\left\{\left|u_{i,k}\right|^2\right\}=\dfrac{1}{M}$ [3], and based on lemma 4.1, we can write (4.17) as

$$\begin{aligned}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\mathbf{R}^{-1}\right]\right\}&=\sum_{i=1}^{M}\sum_{k=1}^{M}\lambda_{r,i}\,\mathbb{E}\left\{\lambda_{\theta,k}\right\}\mathbb{E}\left\{\left|u_{i,k}\right|^2\right\}\\&=\frac{1}{M}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}_\theta^2\right]\right\}\mathrm{Tr}\left[\mathbf{R}^{-1}\right]\end{aligned} \tag{4.18}$$

Note that since $\mathbf{R}^{-1}$ is a square matrix, $\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{R}^{-1}\right]\right\}$ does not exist [3], therefore we continue with $\mathrm{Tr}\left[\mathbf{R}^{-1}\right]$. To further proceed, we consider the following lemma:

Lemma 4.2: $\mathrm{Tr}\left[\mathbf{R}_\theta^2\right]=\mathrm{Tr}\left[\mathbf{R}^2\right]$

Proof: To continue with proof we just need to show that the ℓth diagonal elements of $\mathbf{R}_\theta^2$ and $\mathbf{R}^2$ are the same. Since matrix $\mathbf{R}_\theta$ is Hermitian, we have $\mathbf{R}_\theta^2=\mathbf{R}_\theta\mathbf{R}_\theta^{\mathsf{H}}$. Hence we

[1] A Hermitian random matrix $\mathbf{W}$ is called unitary invariant if the joint distribution of its entries equals that of $\mathbf{V}\mathbf{W}\mathbf{V}^{\mathsf{H}}$ for any unitary matrix $\mathbf{V}$ independent of $\mathbf{W}$.

[2] A $M\times M$ random matrix $\mathbf{U}$ is a *Haar* matrix (also called isotropic in the multiantenna literature) if it is uniformly distributed on the set of $M\times M$ unitary matrices.

can write

$$\dot{r}_{\ell,\ell} = \sum_{i=1}^{M} r_{\ell,i} r_{\ell,i}^{\mathsf{H}} \tag{4.19}$$

where $\dot{r}_{\ell,\ell}$ is the ℓth diagonal element of $\mathbf{R}_\theta^2$, and $r_{\ell,i}$ is the (ℓ, i)th element of $\mathbf{R}_\theta$. Based on (4.3), $r_{\ell,i} = |\rho_{\ell,i}|\, c_\ell c_i^{\mathsf{H}}$ where c_ℓ is the ℓth element of the data vector $\mathbf{c}$. Therefore we can rewrite (4.19) as

$$\begin{aligned}\dot{r}_{\ell,\ell} &= \sum_{i=1}^{M} |\rho_{\ell,i}|\, c_\ell c_i^{\mathsf{H}} \left(|\rho_{\ell,i}|\, c_\ell c_i^{\mathsf{H}}\right)^{\mathsf{H}} = \sum_{i=1}^{M} |\rho_{\ell,i}|\, c_\ell c_i^{\mathsf{H}} c_i c_\ell^{\mathsf{H}} \, |\rho_{\ell,i}|^{\mathsf{H}} \\ &= \sum_{i=1}^{M} |\rho_{\ell,i}|^2 = \sum_{i=1}^{M} \rho_{\ell,i} \rho_{\ell,i}^{\mathsf{H}}\end{aligned} \tag{4.20}$$

So from (4.20), we can deduce that the ℓth diagonal element of $\mathbf{R}_\theta^2$, i.e., $\dot{r}_{\ell,\ell}$, is equal to that of $\mathbf{R}^2 = \mathbf{R}\mathbf{R}^{\mathsf{H}}$ which is denoted by $\sum_{i=1}^{M} \rho_{\ell,i} \rho_{\ell,i}^{\mathsf{H}}$ ■

Therefore we have

$$\mathbb{E}\left\{g_{\text{PA}}^2\right\} = \frac{M}{\mathbb{E}\left\{\text{Tr}\left[\mathbf{R}^2\right]\right\} \text{Tr}\left[\mathbf{R}^{-1}\right]} \tag{4.21}$$

and since $\mathbb{E}\left\{\text{Tr}\left[\mathbf{R}^2\right]\right\} = MN\left(M+N\right)$ [3], $\mathbb{E}\left\{g_{\text{PA}}^2\right\}$ can be shown as

$$\mathbb{E}\left\{g_{\text{PA}}^2\right\} = \frac{1}{N\left(M+N\right) \text{Tr}\left[\mathbf{R}^{-1}\right]} \tag{4.22}$$

■

Therefore by considering theorem 4.1, (4.11) can now be written as

$$\eta'_{\text{PA}} = \frac{P\left[(N+1) + (M-1)\left(1 + \sqrt{N\pi} + (M-2)\,\frac{\pi}{4}\right)\right]}{(M+N)\,\sigma^2 \,\text{Tr}\left[\mathbf{R}^{-1}\right]} \tag{4.23}$$

Analogous to the same procedures of CI precoding in equations (3.8)–(3.9), and with respect to equation (4.23) and by considering $\text{Tr}\left[\mathbf{R}^{-1}\right] = \sum_{\ell=1}^{M} \left[\mathbf{R}^{-1}\right]_{\ell,\ell}$, the average output SNR of PA precoding for the ℓth user can be represented by

$$\eta'_{\text{PA}} = \frac{\text{snr}_{\text{PA}}}{\left[\mathbf{R}^{-1}\right]_{\ell,\ell}} \quad 1 \le \ell \le M \tag{4.24}$$

where

$$\mathrm{snr}_{\mathrm{PA}} = \frac{P\left[(N+1)+(M-1)\left(1+\sqrt{N\pi}+(M-2)\,\frac{\pi}{4}\right)\right]}{M\,(M+N)\,\sigma^2} \tag{4.25}$$

Now, the output SNR of PA precoding in (4.24) is of similar form to that of CI precoding in (3.9) in the sense that both of these equations have the same denominator. As mentioned before, this treatment of output SNR of PA precoding will help us compare the power efficiency of PA precoding to that of CI precoding and also facilitates the selection of a regularization parameter for the proposed RPA precoding in the following section.

4.3 Regularized Phase Alignment

In the previous section, we showed that PA precoding aims to rotate the phases of the transmitted symbols such that for each MT, the interferences of remaining $M-1$ streams add up coherently and consequently we can glean higher output SNRs for all MTs; however, since PA inherently uses channel inversion (see (4.6)), it is still problematic when the channel is ill-conditioned. To overcome this deficiency, we propose to use the concept of RCI precoding by adding a multiple of the identity matrix (i.e., $\varepsilon\mathbf{I}$) to $\mathbf{R}$ before inverting. Since ε controls the amount of interference introduced to each user, the most important point is how to choose ε to get the optimum performance since ε can take on any positive value. In section 3.4, we mentioned that under perfect CSI and for RCI precoding, this amount of ε is equal to $\varepsilon = M\gamma^{-1}$ which is optimal when M is large and works well even with small M, as also discussed in [13]. Moreover, under imperfect CSI, we derived an appropriate regularization parameter described in (3.45). Now in this section, we similarly derive optimum regularization parameters for the proposed RPA precoding under both perfect and imperfect CSI.

4.3.1 Precoder Design under Perfect CSI

In this subsection, we derive a regularization parameter for RPA precoding under perfect CSI, and the corresponding output SINR at each user is derived. Since there is a one-to-one mapping from PA to RPA precoding which is similar to that of CI to RCI precoding and since all these four precoders are linear, comparison can be used to seek a regularization parameter.

Since the output SNRs of CI and PA precoding resemble each other (see (3.9) and (4.24)); analogous to RCI precoding, and by comparing (4.24) with (3.9), it turns out that

one good choice of ε for the proposed RPA precoding can now be obtained via the inverse of (4.25), i.e., $\frac{1}{\mathrm{snr}_{\mathtt{PA}}}$. In subsection 4.3.2, we will show that this regularization parameter is optimum under perfect CSI. In this case, the transmitted signal is given by

$$\mathbf{s}_{\mathtt{RPA}} = g_{\mathtt{RPA}} \boldsymbol{\Psi}_{\mathtt{RPA}} \mathbf{c} \tag{4.26}$$

where

$$\boldsymbol{\Psi}_{\mathtt{RPA}} = \mathbf{H}^{\mathsf{H}} \left(\mathbf{R} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{R}_{\theta} \tag{4.27}$$

is the precoding matrix and the scaling factor can be shown as

$$g_{\mathtt{RPA}} = \frac{1}{\sqrt{\mathrm{Tr}\left[\mathbf{R} \left(\mathbf{R} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{R}_{\theta}^{2} \left(\mathbf{R} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \right]}} \tag{4.28}$$

Therefore the received signal can now be represented as

$$\begin{aligned} \mathbf{y}_{\mathtt{RPA}} &= \sqrt{P}\, g_{\mathtt{RPA}} \mathbf{H} \mathbf{H}^{\mathsf{H}} \left(\mathbf{R} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{R}_{\theta} \mathbf{c} + \mathbf{z} \\ &= \sqrt{P}\, g_{\mathtt{RPA}} \mathbf{H} \left(\mathbf{H}^{\mathsf{H}} \mathbf{H} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{H}^{\mathsf{H}} \mathbf{R}_{\theta} \mathbf{c} + \mathbf{z} \end{aligned} \tag{4.29}$$

Since $\mathbf{R}_{\theta} = |\mathbf{R}| \odot \mathbf{C}$, we define

$$\overline{\mathbf{c}} \triangleq \mathbf{R}_{\theta} \mathbf{c} = \left(\begin{array}{c} c_1 \sum_{x=1}^{M} |\rho_{1,x}|, \ldots, c_M \sum_{x=1}^{M} |\rho_{M,x}| \end{array} \right)^{\top} \tag{4.30}$$

Let $\mathbf{h}_{\ell}^{\mathsf{H}} \in \mathbb{C}^{1 \times N}$ denote the ℓth row of $\mathbf{H}$ and $\mathbf{H}_{\ell} \in \mathbb{C}^{(M-1) \times N}$ designate the submatrix obtained by striking $\mathbf{h}_{\ell}^{\mathsf{H}}$ out of $\mathbf{H}$. We also assume that $\overline{c}_{\ell}$ denotes the ℓth element of $\overline{\mathbf{c}}$ and $\overline{\mathbf{c}}_{\ell}$ stands for the subvector obtained by removing $\overline{c}_{\ell}$ from $\overline{\mathbf{c}}$. Then, the received signal at the ℓth user can be shown as

$$\begin{aligned} y_{\mathtt{RPA}}^{[\ell]} &= \sqrt{P} g_{\mathtt{RPA}} \mathbf{h}_{\ell}^{\mathsf{H}} \left(\mathbf{H}^{\mathsf{H}} \mathbf{H} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{H}^{\mathsf{H}} \overline{\mathbf{c}} + z_{\ell} \\ &= \sqrt{P} g_{\mathtt{RPA}} \mathbf{h}_{\ell}^{\mathsf{H}} \left(\mathbf{H}^{\mathsf{H}} \mathbf{H} + \frac{1}{\mathrm{snr}_{\mathtt{PA}}} \mathbf{I} \right)^{-1} \mathbf{h}_{\ell} \overline{c}_{\ell} \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{x\neq\ell}^{M} \sqrt{P} g_{\text{RPA}} \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}^{\mathsf{H}}\mathbf{H} + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_x \overline{c}_x + z_\ell \\
&= \underbrace{\sqrt{P} g_{\text{RPA}} \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}^{\mathsf{H}}\mathbf{H} + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell \overline{c}_\ell}_{\texttt{desired signal}} \\
&+ \underbrace{\sqrt{P} g_{\text{RPA}} \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}^{\mathsf{H}}\mathbf{H} + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{H}_\ell^{\mathsf{H}} \overline{\mathbf{c}}_\ell}_{\texttt{interference}} + z_\ell
\end{aligned} \tag{4.31}
$$

Based on the matrix inverse lemma 3.4 on page 38, we have

$$
\left(\mathbf{H}^{\mathsf{H}}\mathbf{H} + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell = \frac{\left(\mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell}{1 + \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell} \tag{4.32}
$$

by considering i.i.d. input signaling and with respect to lemma 3.3 on page 30, the output SINR of the ℓth user based on RPA precoding is equal to

$$
\eta_{\text{RPA}} = \frac{P\, g_{\text{RPA}}^2 G_\ell^2 F_\ell}{P\, g_{\text{RPA}}^2 D_\ell + (1+G_\ell)^2 \sigma^2} \tag{4.33}
$$

where $F_\ell = \left(\sum_{x=1}^{M} |\rho_{\ell,x}| \right)^2$, $G_\ell = \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell$,

$$
D_\ell = \mathbf{h}_\ell^{\mathsf{H}} \left(\mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{H}_\ell^{\mathsf{H}} \boldsymbol{\Upsilon}_\ell \mathbf{H}_\ell \left(\mathbf{H}_\ell^{\mathsf{H}}\mathbf{H}_\ell + \frac{1}{\text{snr}_{\text{PA}}}\mathbf{I} \right)^{-1} \mathbf{h}_\ell
$$

such that $\boldsymbol{\Upsilon}_\ell = \text{diag}\left\{F_1, \ldots, F_{\ell-1}, F_{\ell+1}, \ldots, F_M\right\}$ and $\text{diag}\left\{\cdot\right\}$ is the diagonal operator.

4.3.2 Precoder Design under Imperfect CSI

Since in practice, acquiring perfect CSI is not pragmatic and only partial CSI may be accessible, in this subsection, we propose an adaptive RPA precoding which outperforms the original RPA precoding (defined in subsection 4.3.1) in the case of CSI mismatch. In this case, we derive an appropriate regularization parameter for the proposed adaptive RPA precoding. In this case, we assume that instead of perfect CSI, i.e., $\mathbf{H}$, only imperfect CSI $\widehat{\mathbf{H}}$ is available at BS, where $\widehat{\mathbf{H}}$ and $\mathbf{H}$ are related to each other through (3.2) and (3.4).

Moreover, we assume that the BS is in possession of the variance of the channel estimation error, i.e., τ, which is possible to be known in advance, as discussed in subsection 3.2.2.

With respect to (4.27) and by considering the availability of imperfect CSI at BS, we define

$$\widehat{\mathbf{\Psi}}_{\text{RPA}} = \ddot{\mathbf{\Psi}}\widehat{\mathbf{R}}_{\theta} \tag{4.34}$$

such that

$$\widehat{\mathbf{R}}_{\theta} = \left|\widehat{\mathbf{R}}\right| \odot \mathbf{C} \quad \text{where} \quad \widehat{\mathbf{R}} = \widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} \tag{4.35}$$

and the scaling factor is equal to

$$\widehat{g}_{\text{RPA}} = \frac{1}{\sqrt{\text{Tr}\left[\widehat{\mathbf{\Psi}}_{\text{RPA}}^{\mathsf{H}}\widehat{\mathbf{\Psi}}_{\text{RPA}}\right]}} \tag{4.36}$$

The precoder under imperfect CSI can be found by using the following optimization criterion:

$$\arg\min_{\ddot{\mathbf{\Psi}}} \mathbb{E}\left\{\left\|\sqrt{P}\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}} + \ddot{f}\mathbf{z} - \sqrt{P}\widehat{\mathbf{c}}\right\|_2^2\right\} \tag{4.37}$$

where

$$\widehat{\mathbf{c}} = \widehat{\mathbf{R}}_{\theta}\mathbf{c} \tag{4.38}$$

and

$$\ddot{f} = \frac{1+\tau}{\widehat{g}_{\text{RPA}}} = (1+\tau)\sqrt{\text{Tr}\left[\widehat{\mathbf{\Psi}}_{\text{RPA}}^{\mathsf{H}}\widehat{\mathbf{\Psi}}_{\text{RPA}}\right]} = (1+\tau)\sqrt{\text{Tr}\left[\widehat{\mathbf{R}}_{\theta}^{\mathsf{H}}\ddot{\mathbf{\Psi}}^{\mathsf{H}}\ddot{\mathbf{\Psi}}\widehat{\mathbf{R}}_{\theta}\right]} \tag{4.39}$$

Note that the inclusion of $\ddot{f}$ in (4.37) is due to the fact that in all precoding schemes, the power of noise is affected by the precoding matrix, and consequently, this effect can be reflected through a multiplicative factor like $\ddot{f}$. This can be perceived with respect to the fact that at transmit side, the transmitted signals get scaled by g to meet the power constraints; consequently at receive side, to have an unbiased detection, the received signals should be scaled back by $(1+\tau)/g$ which further appears as a multiplicative factor for the noise vector.

To further proceed, we consider the following two lemma:

Lemma 4.3: $\mathbb{E}\left\{\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\right\} = \varrho\mathbf{I}$ where

$$\varrho = N(1+\tau)^2\left[(N+1) + (M-1)\left(1 + \sqrt{N\pi} + (M-2)\frac{\pi}{4}\right)\right] \tag{4.40}$$

Proof: Note that (4.38) can be rewritten as

$$\widehat{\mathbf{c}} = \left(c_1 \sum_{x=1}^{M} |\widehat{\rho}_{1,x}| \, , \ldots , c_M \sum_{x=1}^{M} |\widehat{\rho}_{M,x}| \right)^{\top} \tag{4.41}$$

Also by considering (3.2), we have $\mathrm{vec}\left(\widehat{\mathbf{H}}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, (1+\tau)\,\mathbf{I}\right)$. Therefore, the statistical properties of $|\widehat{\rho}_{\ell,x}|$, which is the (ℓ, x)th element of $\widehat{\mathbf{R}}$ in (4.35) can be obtained by setting $\sigma_h^2 = 1+\tau$ in the values of Table 4.1. Therefore, after some straightforward manipulations and by considering the fact that $\mathbb{E}\left\{\mathbf{c}\mathbf{c}^{\mathsf{H}}\right\} = \mathbf{I}$, the claim follows. ■

Lemma 4.4: $\mathbb{E}\left\{\widehat{\mathbf{R}}_{\theta}^{2}\right\} = \kappa\,\mathbf{I}$ where

$$\kappa = N\,(1+\tau)^2\,(M+N) \tag{4.42}$$

Proof: Based on (4.35), we have

$$\mathbb{E}\left\{\widehat{\mathbf{R}}_{\theta}^{2}\right\} = \mathrm{diag}\left(\mathbb{E}\left\{ \left[\widehat{\mathbf{R}}^2\right]_{1,1} \right\}, \ldots, \mathbb{E}\left\{ \left[\widehat{\mathbf{R}}^2\right]_{M,M} \right\} \right) \tag{4.43}$$

With respect to the fact that $\widehat{\mathbf{R}}$ is a Hermitian matrix, we have

$$\left[\widehat{\mathbf{R}}^2\right]_{\ell,\ell} = \sum_{i=1}^{M} |\widehat{\rho}_{\ell,i}|^2 \tag{4.44}$$

Thus, based on Table 4.1 and by considering the fact that in this case we set $\sigma_h^2 = 1+\tau$, $\mathbb{E}\left\{ \left[\widehat{\mathbf{R}}^2\right]_{\ell,\ell} \right\} = N\,(1+\tau)^2\,(M+N)$. Thus, the claim follows. ■

Due to lemma 4.3 and lemma 4.4, the objective function in (4.37) can be shown as

$$\begin{aligned}
\ddot{F} &= \mathbb{E}\left\{ \mathrm{Tr}\left[\left(\sqrt{P}\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}} + \ddot{f}\mathbf{z} - \sqrt{P}\widehat{\mathbf{c}}\right)\left(\sqrt{P}\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}} + \ddot{f}\mathbf{z} - \sqrt{P}\widehat{\mathbf{c}}\right)^{\mathsf{H}} \right] \right\} \\
&= \mathbb{E}\Big\{ \mathrm{Tr}\Big[P\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + \ddot{f}^2\mathbf{z}\mathbf{z}^{\mathsf{H}} - P\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}} - P\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + P\widehat{\mathbf{c}}\widehat{\mathbf{c}}^{\mathsf{H}} \\
&+ \sqrt{P}\,\ddot{f}\mathbf{H}\ddot{\mathbf{\Psi}}\widehat{\mathbf{c}}\mathbf{z}^{\mathsf{H}} + \sqrt{P}\ddot{f}\mathbf{z}\widehat{\mathbf{c}}^{\mathsf{H}}\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} - \sqrt{P}\ddot{f}\widehat{\mathbf{c}}\mathbf{z}^{\mathsf{H}} - \sqrt{P}\ddot{f}\mathbf{z}\widehat{\mathbf{c}}^{\mathsf{H}} \Big] \Big\} \\
&\overset{①}{=} P\varrho\,\mathrm{Tr}\left[\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\ddot{\mathbf{\Psi}}\right] + M\sigma^2\,(1+\tau)^2\,\kappa\,\mathrm{Tr}\left[\ddot{\mathbf{\Psi}}^{\mathsf{H}}\ddot{\mathbf{\Psi}}\right] - P\varrho\,\mathrm{Tr}\left[\mathbf{H}\ddot{\mathbf{\Psi}}\right] \\
&- P\varrho\,\mathrm{Tr}\left[\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\right] + PM\varrho
\end{aligned} \tag{4.45}$$

where ① follows lemma 3.3 on page 30. To obtain the sought precoder, we can differenti-

ate $\ddot{F}$ with respect to $\ddot{\mathbf{\Psi}}$ by first considering the following assumptions [50, 51]:

1. $\ddot{\mathbf{\Psi}}$ and $\ddot{\mathbf{\Psi}}^{\mathsf{H}}$ are treated as independent variables.
2. $\dfrac{\partial \operatorname{Tr}\left[\mathbf{A}\ddot{\mathbf{\Psi}}\right]}{\partial \ddot{\mathbf{\Psi}}} = \dfrac{\partial \operatorname{Tr}\left[\ddot{\mathbf{\Psi}}\mathbf{A}\right]}{\partial \ddot{\mathbf{\Psi}}} = \mathbf{A}.$

Following the preceding assumptions, the differentiation of $\ddot{F}$ with respect to $\ddot{\mathbf{\Psi}}$ gives

$$\begin{aligned}\frac{\partial \ddot{F}}{\partial \ddot{\mathbf{\Psi}}} &= P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{H} + M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} - P\varrho\mathbf{H} \\ &\overset{②}{=} P\varrho\ddot{\mathbf{\Psi}}^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau} + \breve{\mathbf{H}}\right)^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}}{1+\tau} + \breve{\mathbf{H}}\right) + M\sigma^2(1+\tau)^2\kappa\ddot{\mathbf{\Psi}}^{\mathsf{H}} \\ &- P\varrho\left(\frac{\widehat{\mathbf{H}}}{1+\tau} + \breve{\mathbf{H}}\right) \end{aligned} \tag{4.46}$$

where ② follows from (3.4). The adaptive precoder can then be found by setting $\dfrac{\partial \ddot{F}}{\partial \ddot{\mathbf{\Psi}}}$ equal to zero and taking the expectation over $\breve{\mathbf{H}}$. Therefore with respect to the fact that due to lemma 3.1 on page 28, $\widehat{\mathbf{H}}$ and $\breve{\mathbf{H}}$ are independent, we can represent the precoding matrix as

$$\mathbb{E}_{\breve{\mathbf{H}}}\left\{\frac{\partial \ddot{F}}{\partial \ddot{\mathbf{\Psi}}}\right\} = \mathbf{0} \implies \ddot{\mathbf{\Psi}} = \widehat{\mathbf{H}}^{\mathsf{H}}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{\mathsf{H}} + \ddot{\varepsilon}\,\mathbf{I}\right)^{-1} \tag{4.47}$$

where the regularization parameter $\ddot{\varepsilon}$ can now be expressed as

$$\begin{aligned}\ddot{\varepsilon} &= M(1+\tau)\left(\tau + \frac{\gamma^{-1}(1+\tau)^3\kappa}{\varrho}\right) \\ &= M(1+\tau)\left(\tau + \frac{\gamma^{-1}(1+\tau)^3(M+N)}{(N+1) + (M-1)\left(1+\sqrt{N\pi} + (M-2)\dfrac{\pi}{4}\right)}\right)\end{aligned} \tag{4.48}$$

Remark 4.3: Note that by setting $\tau = 0$, $\ddot{\varepsilon}$ boils down to $\dfrac{1}{\mathrm{snr}_{\mathrm{PA}}}$ which is the appropriate regularization parameter for RPA precoding under perfect CSI, where $\mathrm{snr}_{\mathrm{PA}}$ is defined

in (4.25). This verifies that the derived regularization parameter $\frac{1}{\mathrm{snr}_{\mathtt{PA}}}$ in subsection 4.3.1 is optimal under perfect CSI. Moreover, this implies that the proposed adaptive design of RPA precoding is a generalized version of the original RPA such that it accommodates the perfect CSI as a special scenario.

Therefore, for the proposed adaptive RPA, the transmitted signal from BS can be shown as

$$\widehat{\mathbf{s}}_{\texttt{adaptive RPA}} = \ddot{g}\ddot{\mathbf{\Psi}}\widehat{\mathbf{R}}_{\theta}\mathbf{c} \tag{4.49}$$

where $\ddot{\mathbf{\Psi}}$ is defined in (4.47), and $\ddot{g}$ is the scaling factor which can be represented as

$$\ddot{g} = \frac{1}{\sqrt{\mathrm{Tr}\left[\widehat{\mathbf{R}}\left(\widehat{\mathbf{R}}+\ddot{\varepsilon}\mathbf{I}\right)^{-1}\widehat{\mathbf{R}}_{\theta}^{2}\left(\widehat{\mathbf{R}}+\ddot{\varepsilon}\mathbf{I}\right)^{-1}\right]}} \tag{4.50}$$

4.4 Power Efficiency

In this section, we investigate the ability of PA and RPA precoding techniques to save the transmit power, which is more appropriate in the sense of green communications [36]. Moreover, we assume that only perfect CSI is available at BS, since evaluating the power efficiency of linear precoders under imperfect CSI is generally inconclusive. We want to investigate that for an averagely constant output SINR at each user, how much power saving RPA (PA) precoding achieves in comparison with RCI (CI) precoding. By considering (4.24) and (3.9), if $P_{\mathtt{PA}}$ and $P_{\mathtt{CI}}$, respectively, represent the deployed power for each user by PA and CI precoding (via replacing P), then for the same output SNR of PA and CI precoding we have

$$\begin{aligned}\dot{\eta}_{\mathtt{CI}} = \eta'_{\mathtt{PA}} \Longrightarrow \frac{P_{\mathtt{CI}}}{M\sigma^2\left[\mathbf{R}^{-1}\right]_{\ell,\ell}} &= \frac{P_{\mathtt{PA}}\left[(N+1)+(M-1)\left(1+\sqrt{N\pi}+(M-2)\frac{\pi}{4}\right)\right]}{M\left(M+N\right)\sigma^2\left[\mathbf{R}^{-1}\right]_{\ell,\ell}} \\ \Longrightarrow \xi \triangleq \frac{P_{\mathtt{PA}}}{P_{\mathtt{CI}}} &= \frac{M+N}{(N+1)+(M-1)\left(1+\sqrt{N\pi}+(M-2)\frac{\pi}{4}\right)}\end{aligned} \tag{4.51}$$

which means that with PA precoding, we can reduce the deployed power by a factor of ξ, to preserve the same average output SNR as CI precoding, and this results in power efficiency of

$$\text{Power Efficiency} = 10\log_{10}\left(\xi^{-1}\right) \quad \text{dB} \tag{4.52}$$

for each user. We will show that this analytical result closely matches the simulations.

If we define $P_{\texttt{RPA}}$ and $P_{\texttt{RCI}}$ as the deployed power by RPA and RCI precoding, respectively, by using the numerical simulations in the following section, we show that still $P_{\texttt{RPA}}/P_{\texttt{RCI}} \approx \xi$. Unfortunately due to the complexity of MMSE expressions, it is not possible to prove it mathematically; however, conceptually we can say that since there is a one-to-one mapping from PA to RPA precoding which is similar to that of CI to RCI precoding and since all these four precoders are linear, therefore we can expect that $P_{\texttt{RPA}}/P_{\texttt{RCI}} \approx \xi$.

One interesting observation from (4.51) is that, the larger both M and N, the more power we can save at transmit side. For example, in the following section we show that when $M = N = 16$, we can decrease the transmit power of PA precoding by 9.8dB (a nearly 10-fold reduction in transmit power) to deliver the same average output SNR to each user compared to CI precoding. This 10-fold reduction is also valid for RPA precoding compared to RCI precoding which makes the proposed RPA precoding very vital at low input SNR ranges.

Also based on (4.51), one can conclude that for a fixed number of transmit antennas at BS, i.e., N, the smaller the M, the less power efficiency can be gleaned. The most power-efficient case is related to $M = N$. This implies that for a fixed number of transmit antennas, i.e., N, the less number of users results in less power saving at BS. This highlights the importance of multiuser diversity.

4.5 Numerical Results

In this section, we provide numerical results to show the superior performance of the proposed RPA precoding compared to the other three precoders.

In simulations and without loss of generality, we assume that each user has one receive antenna. Also we consider the same fading model as the one discussed in subsection 3.2.1. Moreover, the output SINRs of CI, RCI, PA, and RPA precoding are related to equations (3.9), (3.39), (4.9), and (4.33), respectively.

To verify the accuracy of the derived SINR in (4.33), we evaluated the SER based on analytical and simulated results in Fig. 4.2 where it can be seen that the SER curves of these two methods closely match. This confirms the accuracy of the derived formula of the output SINR of RPA precoding.

Fig. 4.3 shows the comparison of the average SER based on CI, RCI, PA and RPA

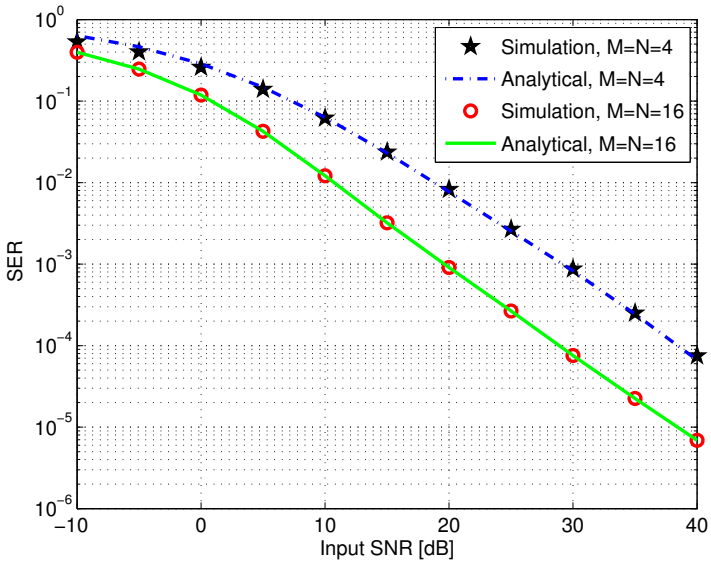

Figure 4.2: *Comparison of the average SER of RPA precoding, based on analytical and simulated results for $M = N = 4$ and $M = N = 16$ and QPSK constellation.*

precoding, where by increasing the number of users (and increasing the number of transmit antennas accordingly) from $M = N = 4$ to $M = N = 16$, the following behaviors are observed:

1. *CI Precoding:* each user experiences inferior SER performance.
2. *RCI Precoding:* the SER performance of each user remains constant at low input SNRs, and improves slightly at high input SNRs, as also shown in [13].
3. *PA Precoding:* the SER performance of each user remains almost constant for all input SNR ranges.
4. *RPA Precoding:* each user experiences significantly better SER performance for all input SNR ranges.

As seen, for $M = N = 4$ and to achieve a fixed SER, RPA yields 2.5dB gain compared to the PA, and for $M = N = 16$ this gain is about 10dB at low input SNRs and 15dB at high input SNRs.

Fig. 4.4 depicts the performance of CI, RCI, PA, and RPA precoding for $M = 10$, $N = 16$, and under QPSK signaling. As revealed, the proposed RPA precoding is able to achieve better performance even when the number of receive antennas is less than the number of transmit antennas at BS. However, when $N > M$, the performance of the non-regularized precoders becomes very close to that of the regularized ones.

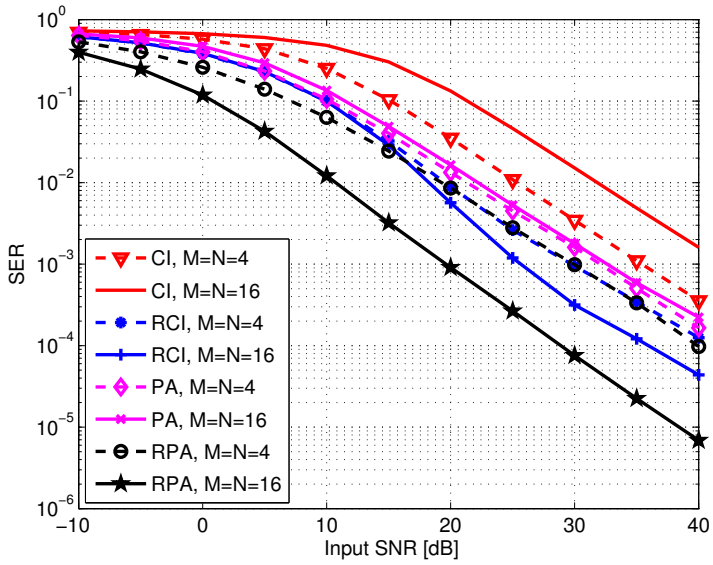

Figure 4.3: *Comparison of the average SER of CI, RCI, PA and RPA precoding for* $M = N = 4$ *and* $M = N = 16$ *and QPSK constellation.*

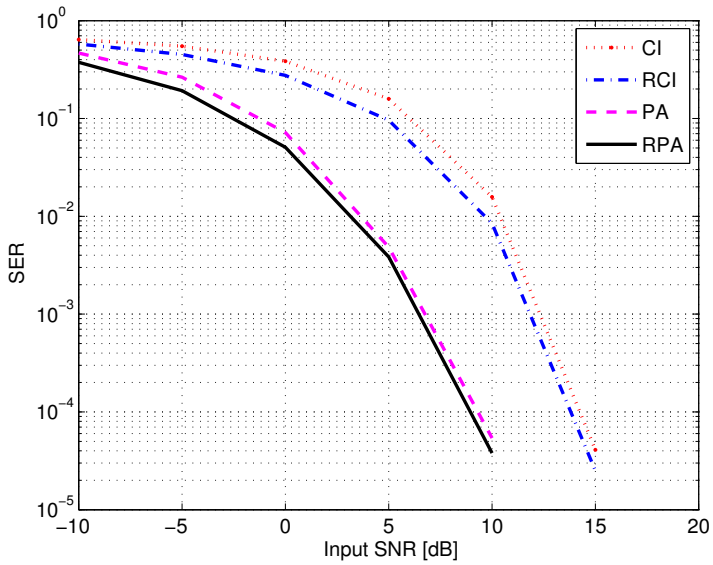

Figure 4.4: *Comparison of the average SER of CI, RCI, PA and RPA precoding for* $M = 10$ *and* $N = 16$ *and QPSK constellation.*

In Figs. 4.5–4.10, we assess the power-efficiency of PA and RPA precoding. Fig. 4.5 depicts the probability density of the output SINR of each user based on different precoding schemes for the case $M = N = 16$. Based on our discussions in section 4.4 and with respect to equation (4.51), we expect that for $M = N = 16$ and for a fixed

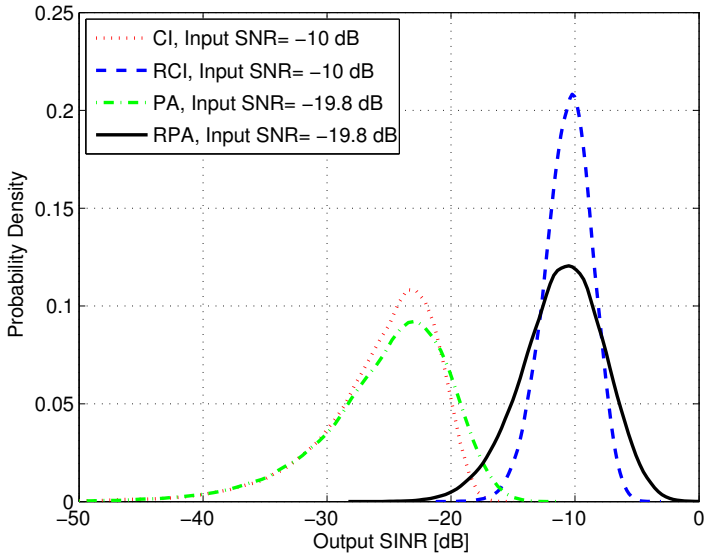

Figure 4.5: *Probability density of output SINR of each user based on different precoding techniques and for $M = N = 16$. For CI and RCI precoding, the input SNR is equal to -10dB while for PA and RPA precoding the input SNR is equal to -19.8 dB.*

average output SINR, PA and RPA precoding, respectively, achieve 9.8dB decrease in the transmit power compared to CI and RCI precoding for each user. Fig. 4.5 verifies this behavior. For example, as seen, the mean of the output SINR of each user based on RPA precoding with input SNR -19.8dB is almost the same as that of RCI precoding with input SNR -10dB.

In Fig. 4.6, we compare the complementary cumulative distributions (CCDs) of the output SINR of each user for the case $M = N = 16$. As observed, the CCD of the output SNR of PA precoding with input SNR -19.8dB is almost the same as that of CI precoding with input SNR -10dB. Also for 40% of channel realizations, the minimum output SINR of each user based on RPA precoding with input SNR -19.8dB is the same as that of RCI precoding with input SNR -10dB.

Fig. 4.7 depicts the probability density of the output SINR of each user based on different precoding schemes for the case $M = 10,\ N = 16$. Based on our discussions in section 4.4 and with respect to equation (4.51), we expect that for $M = 10,\ N = 16$ and for a fixed average output SINR, PA and RPA precoding, respectively, achieve 7.5dB decrease in the transmit power compared to CI and RCI precoding for each user. Fig. 4.7 verifies this behavior. For example, as seen, the mean of the output SINR of each user based on RPA precoding with input SNR -17.5dB is almost the same as that of RCI

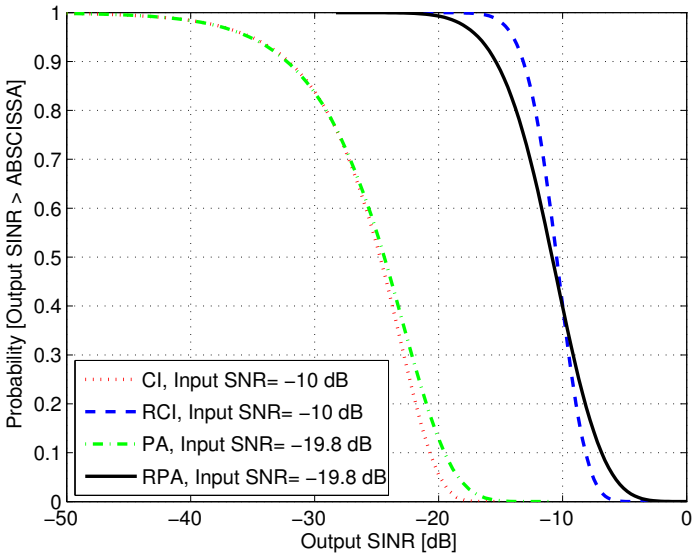

Figure 4.6: *Complementary cumulative distributions of output SINR of each user based on different precoding techniques and for $M = N = 16$. For CI and RCI precoding, the input SNR is equal to -10dB while for PA and RPA precoding the input SNR is equal to -19.8 dB.*

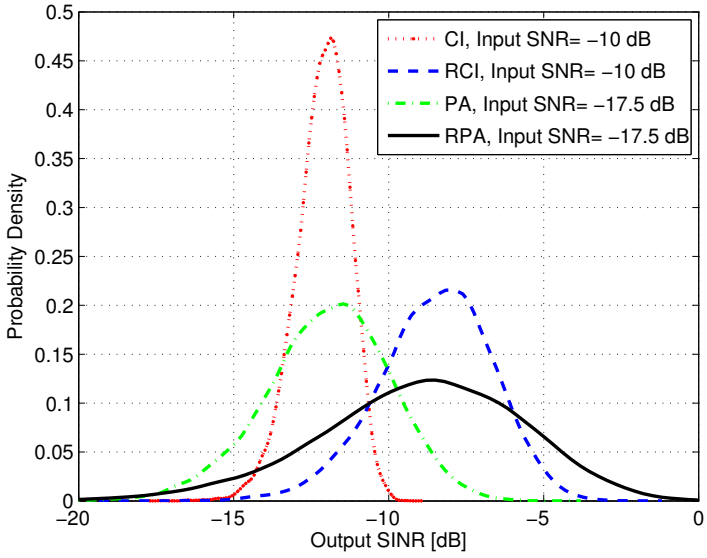

Figure 4.7: *Probability density of output SINR of each user based on different precoding techniques and for $M = 10,\ N = 16$. For CI and RCI precoding, the input SNR is equal to -10dB while for PA and RPA precoding the input SNR is equal to -17.5 dB.*

precoding with input SNR -10dB.

In Fig. 4.8, we compare the CCDs of the output SINR of each user for the case

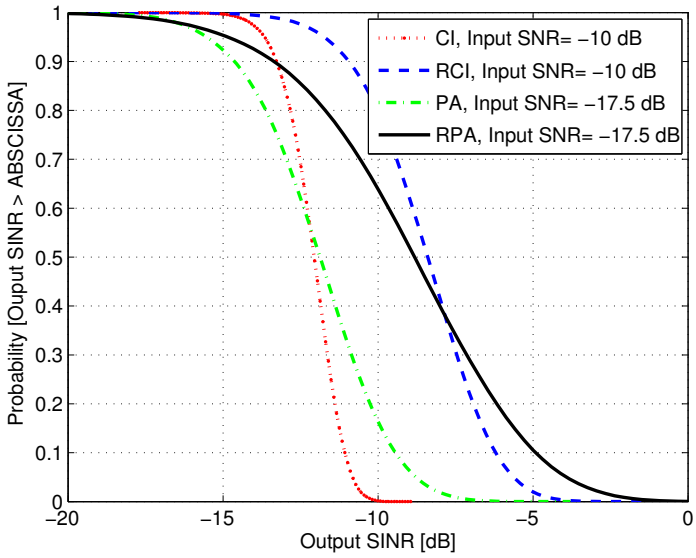

Figure 4.8: *Complementary cumulative distributions of output SINR of each user based on different precoding techniques and for $M = 10,\ N = 16$. For CI and RCI precoding, the input SNR is equal to -10dB while for PA and RPA precoding the input SNR is equal to -17.5 dB.*

$M = 10,\ N = 16$. As observed, for 60% of channel realizations, the minimum output SINR of each user based on PA precoding with input SNR -17.5dB is the same as that of CI precoding with input SNR -10dB. Also for 40% of channel realizations, the minimum output SINR of each user based on RPA precoding with input SNR -17.5dB is the same as that of RCI precoding with input SNR -10dB.

Fig. 4.9 shows the power efficiency of RPA (PA) to RCI (CI) for the case $M = N$. As depicted, the larger the M and N, the more power we can save at transmit side. This demonstrates the importance of PA and RPA precoding in the context of massive MIMO. For example with $M = N = 100$, the proposed RPA precoding enables us to save nearly 17dB (a 50-fold reduction) transmit power compared to RCI precoding for each user, which is significant at low input SNRs.

Fig. 4.10 shows the power efficiency of RPA (PA) to RCI (CI) as a function of N for a fixed M, e.g., $M = 4$. As revealed, the larger the N, the less power we can save at transmit side. This demonstrates the importance of PA and RPA precoding when M and N increase proportionally.

The throughput of different linear precoding techniques are illustrated in Fig. 4.11 for the case $M = N = 16$ and for both QPSK and 8-PSK constellations under perfect CSI. In

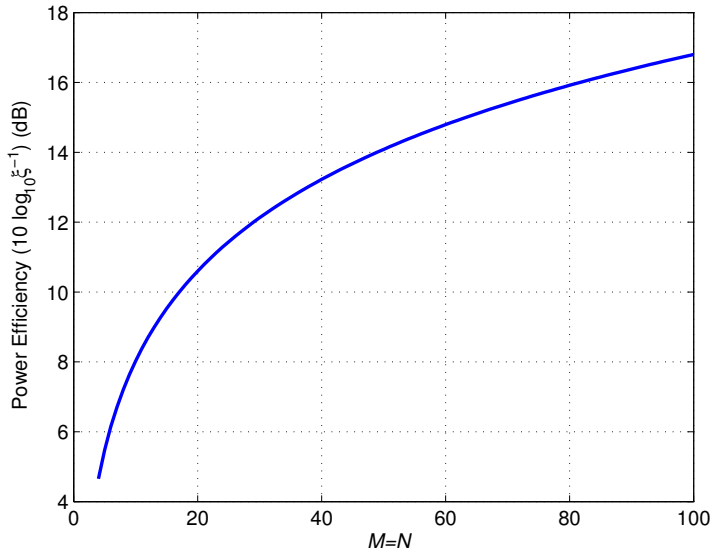

Figure 4.9: *Average power efficiency of RPA (PA) to RCI (CI) precoding for* $M = N$.

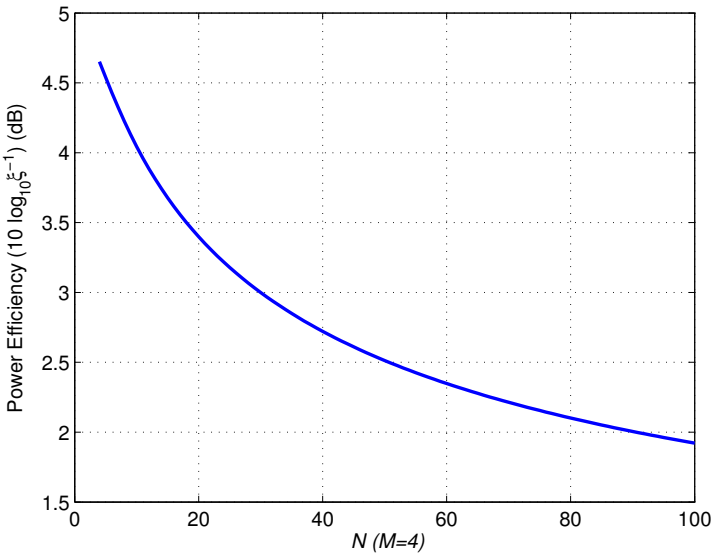

Figure 4.10: *Average power efficiency of RPA (PA) to RCI (CI) precoding as a function of* N *when* $M = 4$.

the results depicted, the throughput is expressed as

$$(1 - \mathrm{blkerr})\, M \log_2 \mathcal{M} \quad \text{bits per channel use} \tag{4.53}$$

where blkerr is the block error rate (here we considered each block consists of 128 sym-

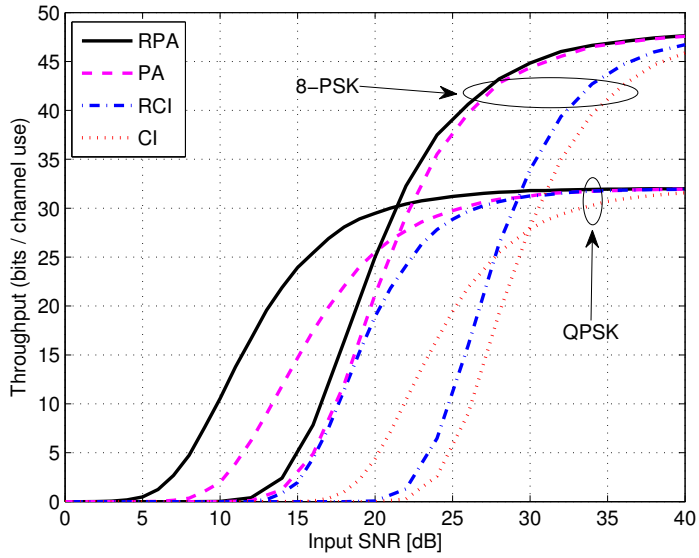

Figure 4.11: *Average throughput for CI, RCI, PA and RPA precoding, for $M = N = 16$ and for QPSK and 8-PSK constellations.*

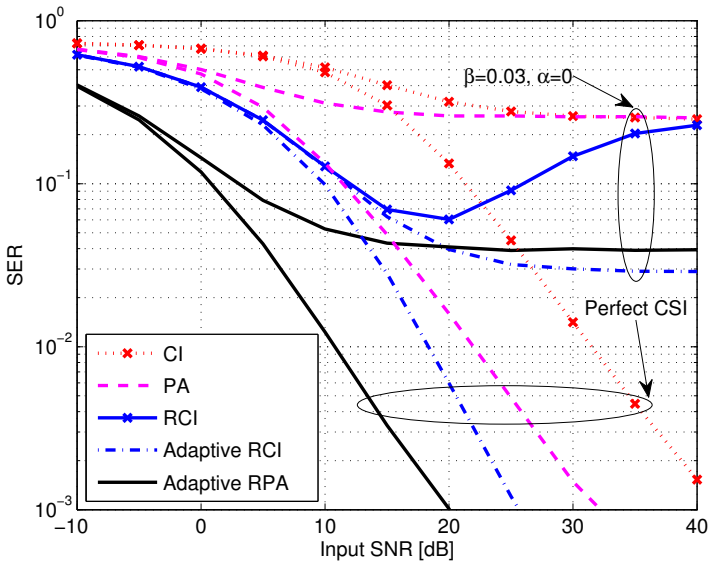

Figure 4.12: *Average SER performance of CI, RCI, PA and RPA precoding, for $M = N = 16$ under QPSK signaling and for the input SNR-independent error model, i.e., $\beta = 0.03,\ \alpha = 0$.*

bols), $\mathcal{M} = 4$ for QPSK and $\mathcal{M} = 8$ for 8-PSK constellations. As seen, the proposed RPA precoding achieves better throughput compared to the other three precoders. For example, at input SNR 7.5dB and for QPSK modulation, while CI, RCI and PA precoders give no throughput, that is attained by RPA precoding is equal to 5 bits per channel use.

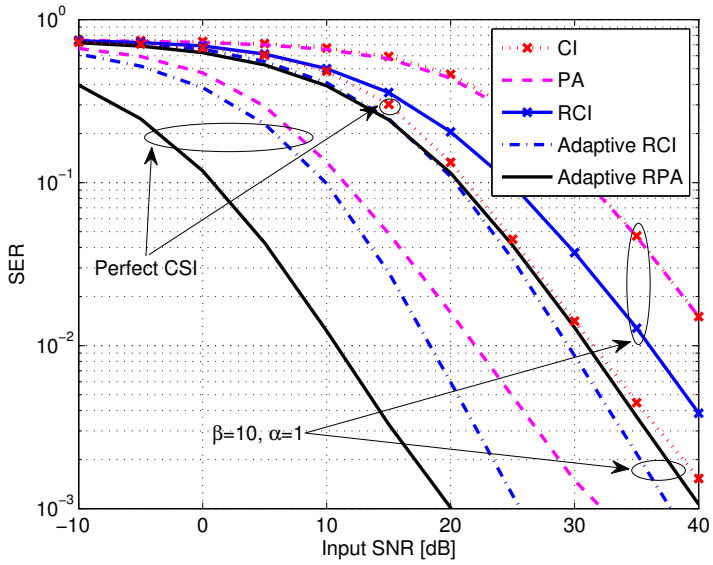

Figure 4.13: *Average SER performance of CI, RCI, PA and RPA precoding, for* $M = N = 16$ *under QPSK signaling and for the input SNR-dependent error model, i.e.,* $\beta = 10,\ \alpha = 1$.

So far, we have assumed perfect CSIT is available at BS. However since practically it is not easy to obtain perfect CSIT, we examine the performance (or sensitivity) of PA and RPA under channel estimation errors in comparison with CI and RCI. Without loss of generality we place our focus on merely SER performance analysis. Figs. 4.12–4.13 show the average SER of different precoding schemes under QPSK signaling and for the case $M = N = 16$, when only imperfect CSI is available at BS. The SER results of perfect CSIT are also repeated for comparison. Note that under perfect CSI, adaptive RCI and adaptive RPA boils down to standard RCI and RPA, respectively. We modeled the imperfect CSI based on the error variance $\tau = \beta\gamma^{-\alpha}$ as demonstrated in (3.3) within subsection 3.2.2. More specifically, the results of Fig. 4.12 and Fig. 4.13 are respectively related to $\beta = 0.03,\ \alpha = 0$ which represents the CSI feedback scenario (which is independent of input SNR), and $\beta = 10,\ \alpha = 1$ which reflects the reciprocal channels (which is dependent on input SNR).

As revealed, even under imperfect CSI, the SER performance of PA is better than (for $\alpha = 0$) or equal to (for $\alpha = 1$) that of CI. Also the proposed adaptive RPA described in subsection 4.3.2 notably outperforms standard RCI in [13] subject to CSI mismatch. However, the proposed adaptive RCI described in section 3.5 achieves slightly better performance than adaptive RPA for the case $\alpha = 1$. Nevertheless, the proposed adaptive RPA achieves better performance than adaptive RCI at low input SNRs for the case $\alpha = 0$.

This nonmonotonic behavior of standard RCI in Fig. 4.12 and for the case of $\alpha = 0$ is due to the fact that at low-to-intermediate SNRs (i.e., the noise-limited regime), standard RCI outperforms CI; however at high SNRs (i.e., the interference-limited regime), standard RCI and CI become almost equivalent [49].

4.6 Summary

We considered linear precoders in multiantenna downlink communications. We reformulated PA precoding which aims to rotate the phases of transmit symbols such that they cause constructive interference. Unlike CI precoding where we null out the interference completely, there is no need to remove the interference by using PA precoding. Consequently and by considering a fixed transmit power, PA precoding delivers more output SNR to each user compared to CI precoding. However, PA precoding is still problematic when channel is ill-conditioned. Therefore, we proposed an enhanced version of PA precoding (named RPA) by deriving an optimum regularization parameter, and showed that it achieves better SER and throughput than CI, RCI and PA precoding specially when the number of transmit and receive antennas becomes larger. It was also shown that PA and RPA precoding enable us to decrease the deployed power at transmit side to achieve the same average output SINR for each user, compared to CI and RCI precoding, respectively. This transmit-power reduction is more significant when there is a large number of transmit and receive antennas. We also illustrated that even with imperfect CSIT, the performance trend of PA and RPA follows that of standard precoders. Additionally, we proposed an adaptive RPA precoding, which with the knowledge of error variance in advance, is able to notably outperform CI, standard RCI and PA precoding under CSI mismatch. We did so by deriving an appropriate regularization parameter which is a function of error variance.

Chapter 5

Interference Alignment under CSI Mismatch

5.1 Introduction

In chapters 3 and 4, we analyzed the performance of linear precoders in the downlink cellular communications restricted to merely one cell. In this chapter and also the subsequent ones, we consider a more complex scenario where several transmitters communicate with their corresponding receivers at the same frequency and time slots. This scenario is commonly referred as IC or wireless interference networks.

Capacity characterization of wireless interference networks has been a hot topic for years since it provides a useful tool to analyze their performance limits. In spite of intensive research on this subject, the capacity region of such networks is still unknown even for small number of users. This prompted the researchers to derive various approximations of the capacity region, for example, the achievable DoF.

One of the promising techniques to achieve more than one DoF in wireless interference networks is IA. Unlike orthogonal medium access techniques, e.g., TDMA and FDMA, IA is able to achieve significant throughput such that the achievable DoF can be linearly scaled up with the number of users. In other words, it has been shown that in a K-user IC with a single antenna at each node, and with time-varying or frequency-selective channel coefficients, it is possible to achieve $\frac{K}{2}$ DoF by coding across sufficiently large symbol extension of the channel [17]. This implies that the length of the symbol extension must tend to infinity which is not pragmatic. Therefore in this chapter and also the subsequent ones, we quantify the performance of IA based on signal space alignment instead of

aligning interfering signals in time. This can be done by deploying multiple antennas at transmit/receive nodes which dissolve the need of symbol extension [23–34].

Nonetheless, for all IA techniques, the availability of perfect CSI is necessary to achieve full DoF. Unfortunately due to the realistic communication scenarios and also deployment challenges, only partial CSI may be accessible, which can adversely affect the achievable throughput and the total DoF in the network. During the latest few years, many researches have been done regarding the performance analysis of IA under CSI uncertainties. Most of the relevant efforts placed their focus on quantized feedback strategies with channel-aware receivers (see e.g., [59–64]). Interestingly, it has been shown that for IA utilizing quantized feedback and for multi-tap single-input single-output (SISO) [59] or MIMO IC [60], full DoF is achievable only if the number of fedback bits scales fast enough with SNR. Aside from quantized feedback, performance analysis of IA under a generalized imperfect CSI model is of particular interest. However, due to its relative intractability, some simpler forms of CSI uncertainties have been investigated, e.g., the performance of IA under transmit-side correlation with imperfect CSI has been investigated in [65] followed by [66] wherein the performance of IA under analog feedback has been evaluated. Also [67] derived upper and lower bounds on the sum mutual information where the variance of CSI error has been considered as a constant.

In this chapter, we consider performance analysis of constant MIMO IA subject to CSI mismatch. First, we analyze the performance of IA under a rather generalized imperfect CSI model thereby we quantify new bounds. To do so, we treat the CSI error variance as a function of SNR. We then show that when this error variance scales with the inverse of SNR, full DoF can be attained, and an upper bound on asymptotic mean loss in sum rate in comparison with the perfect CSI case is established. We likewise derive a bound on the achievable DoF when the error variance depends on the inverse of SNR to a power of a constant. Using numerical simulations, we substantiate the analytically derived bounds.

We also consider the performance improvement of Max-SINR algorithm described in [23] under CSI mismatch. Max-SINR is an interesting algorithm since it tries to maximize the SINR on a stream-by-stream basis instead of explicitly minimizing the leaked interference as being done by Min-WLI [23] and Alt-Min [24, 34] algorithms, and it thus achieves better performance. Because of its importance, some literature particularly focused on performance analysis of Max-SINR. For example, it has been established that Max-SINR is optimal within the class of linear beamformers at high SNRs [68], and it has been further shown that Max-SINR averagely achieves better throughput than sum-rate gradient algorithms at low-to-intermediate SNRs [30]. Its convergence has been also

addressed in [69]. However, performance analysis and improvement of Max-SINR under CSI mismatch has not been seriously considered so far. Therefore in this chapter, we additionally address this issue. First, it is shown that subject to CSI mismatch, the comparative improvement of Max-SINR over interference leakage minimization algorithms becomes negligible. We then propose an adaptive Max-SINR algorithm, which with the knowledge of error variance in advance, can tangibly improve the performance of original Max-SINR subject to CSI uncertainties.

Section 5.2 contains the basic concepts including the system model for standard IA, imperfect CSI model, and signal detection. In Section 5.3, we derive novel bounds regarding the asymptotic mean loss in sum rate and achievable DoF. In Section 5.4, we propose an adaptive Max-SINR algorithm to improve the performance of original Max-SINR under CSI mismatch. Section 5.5 contains numerical results wherein we substantiate the undergone analyses in Sections 5.3–5.4. Finally, Section 5.6 contains a summary of the analyses presented in this chapter.

5.2 Preliminaries

5.2.1 System Model and Standard IA

Consider a symmetric K-user MIMO interference channel consisting of $2K$ nodes, K of which are denoted as transmitters while the other K are receivers. Each transmitter is paired with a single receiver in a one-to-one mapping as denoted in Fig. 5.1. Specifically, each N-antenna transmitter communicates with its corresponding M-antenna receiver by sending d independent data streams. The channel output at receiver k is given by

$$\mathbf{y}_k = \mathbf{H}_{k,k}\mathbf{x}_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{z}_k \tag{5.1}$$

where $\mathbf{y}_k \in \mathbb{C}^{M\times 1}$ is the received signal, $\mathbf{x}_k \in \mathbb{C}^{N\times 1}$ is the transmitted signal from transmitter k and $\mathbf{x}_j \in \mathbb{C}^{N\times 1}$ is the interference received from transmitter j. $\mathbf{H}_{k,j} \in \mathbb{C}^{M\times N}$ describes the channel from transmitter j to receiver k. The magnitude of fading coefficients is assumed to be bounded away from zero and infinity. We also consider block fading model where all links are static for the duration of a transmission but may change between successive transmission, i.e., constant MIMO scenario. More specifically, the elements of channel matrices between each transmitter and receiver can be modeled

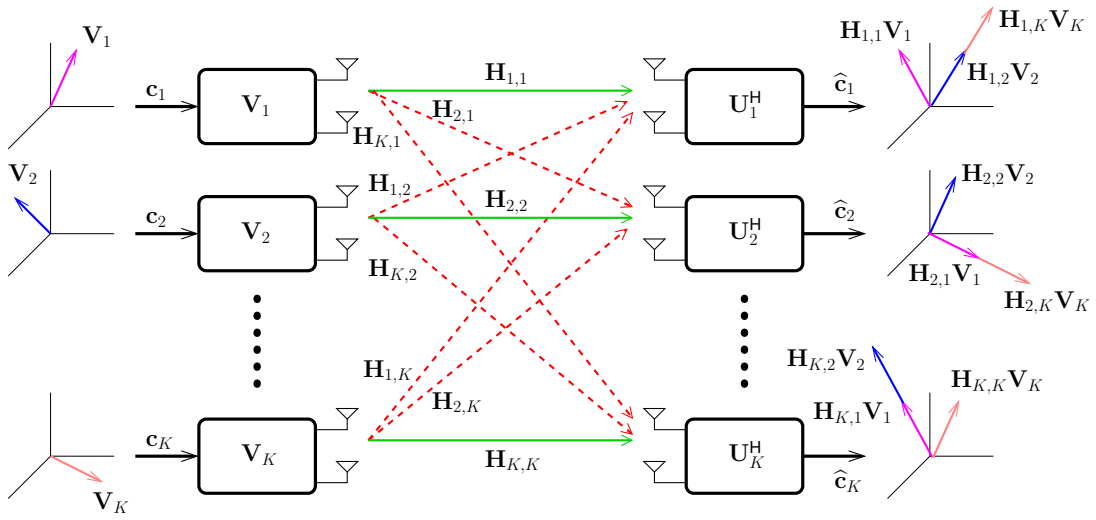

Figure 5.1: *K-user interference channel where green solid arrows denote direct links and red dash arrows represent interfering (cross) links. $\mathbf{c}_k$ designates the input data of transmit node k whereas $\widehat{\mathbf{c}}_k$ indicates the recovered data at receive node k.*

by i.i.d. Gaussian random variables with zero mean and unit variance, i.e., $\mathrm{vec}\left(\mathbf{H}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \mathbf{I}\right)$. $\mathbf{z}_k \in \mathbb{C}^{M\times 1}$ is the circularly symmetric additive white Gaussian noise with zero mean and variance σ^2 per entry, i.e., $\mathbf{z}_k \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$. We also consider that each transmitted signal $\mathbf{x}_k$ is equal to $\mathbf{V}_k\mathbf{c}_k$, where $\{\mathbf{V}_k\}_{k=1}^K \in \mathbb{C}^{N\times d}$ are truncated unitary transmit beamforming matrices (precoders), and $\mathbf{c}_k \in \mathbb{C}^{d\times 1}$ is the data stream intended for receiver k such that $\mathbb{E}\left\{\mathbf{c}_k\mathbf{c}_k^{\mathsf{H}}\right\} = P\mathbf{I}$. Without loss of generality, we assume uniform power allocation across all users and DoF which is asymptotically optimal. In this case, $\gamma = P/\sigma^2$ is defined as the nominal SNR.

Considering a ZF receiver, the conditions for perfect interference alignment can be described as [20, 23]

$$\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{V}_j = \mathbf{0} \quad , \quad \forall j \neq k \tag{5.2}$$

$$\mathrm{rank}\left(\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k\right) = d \tag{5.3}$$

where $\{\mathbf{U}_j\}_{j=1}^K$ are truncated unitary interference suppression matrices (combiners).

In other words, IA aims to design precoders $\mathbf{V}_k$ such that at each receive node, the unwanted interferences are aligned within a reduced subspace of the received signal space which is supposed to be independent of the desired signal subspace. This is shown in Fig. 5.1 wherein, for example, at the first receive node the interference subspaces $\mathbf{H}_{1,2}\mathbf{V}_2$ and $\mathbf{H}_{1,K}\mathbf{V}_K$ are aligned and are independent of the desired signal subspace $\mathbf{H}_{1,1}\mathbf{V}_1$. Therefore, with respect to (5.2), the first receive nodes premultiplies its received signal with $\mathbf{U}_1^{\mathsf{H}}$

which nulls out the aligned interference subspaces without suppressing the desired signal subspace.

5.2.2 Imperfect CSI Model

Similar to the same assumption as in [65–67] and regardless of distributed or centralized processing, we assume that all precoders and combiners are constructed with the knowledge of unified CSI mismatch. Analogous to the same imperfect CSI model for the broadcast channel discussed in subsection 3.2.2, we further model the CSI mismatch as

$$\widehat{\mathbf{H}}_{k,j} = \mathbf{H}_{k,j} + \mathbf{E}_{k,j} \tag{5.4}$$

where the channel measurement error $\mathbf{E}_{k,j}$ is thought to be independent of the actual channel matrix $\mathbf{H}_{k,j}$, and we consider $\mathbf{E}_{k,j}$ as a Gaussian matrix consisting of i.i.d. elements with mean zero and variance τ, i.e.,

$$\mathrm{vec}\left(\mathbf{E}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \tau\mathbf{I}\right) \text{ with } \tau \triangleq \beta\gamma^{-\alpha}, \quad \beta > 0, \quad \alpha \geq 0 \tag{5.5}$$

With respect to lemma 3.1 on page 28, we can further write

$$\mathbf{H}_{k,j} = \frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j} + \breve{\mathbf{H}}_{k,j} \tag{5.6}$$

where $\mathrm{vec}\left(\breve{\mathbf{H}}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{\tau}{1+\tau}\mathbf{I}\right)$ is statistically independent of $\widehat{\mathbf{H}}_{k,j}$.

5.2.3 Signal Postprocessing at Receive Nodes

In this subsection, we briefly address the data recovery at receive nodes. Without loss of generality, we assume that each receive node uses a linear ZF equalizer. It is also worth mentioning that the results of this subsection can be readily generalized to the case of channel inversion, i.e., preprocessing the signals at transmit nodes instead of postprocessing the signals at receive nodes.

We first assume that perfect CSI is available. In this case, if we define

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_K \end{pmatrix} \tag{5.7}$$

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{U}_K \end{pmatrix} \tag{5.8}$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,K} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{K,1} & \mathbf{H}_{K,2} & \cdots & \mathbf{H}_{K,K} \end{pmatrix} \tag{5.9}$$

with respect to (5.2), we have

$$\mathbf{U}^{\mathsf{H}}\mathbf{H}\mathbf{V} = \begin{pmatrix} \overline{\mathbf{H}}_{1,1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \overline{\mathbf{H}}_{K,K} \end{pmatrix} \tag{5.10}$$

where $\overline{\mathbf{H}}_{k,k} = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$. More specifically, by premultiplying the received signal at receiver k by $\mathbf{U}_k^{\mathsf{H}}$ we have

$$\begin{aligned} \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k &= \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{x}_k + \underbrace{\mathbf{U}_k^{\mathsf{H}}\sum_{\substack{j=1 \\ j\neq k}}^{K}\mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k}_{\overline{\mathbf{z}}_k} \\ \Rightarrow\ \overline{\mathbf{y}}_k &= \overline{\mathbf{H}}_{k,k}\mathbf{c}_k + \overline{\mathbf{z}}_k \end{aligned} \tag{5.11}$$

where $\overline{\mathbf{y}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k$, and following (5.2), at high enough SNRs we have $\overline{\mathbf{z}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k$. Therefore the transmitted symbol vector $\mathbf{c}_k$ can be easily recovered through premultiplying $\overline{\mathbf{y}}_k$ by (pseudo-) inverse of $\overline{\mathbf{H}}_{k,k}$, i.e., $\mathrm{pinv}\left(\overline{\mathbf{H}}_{k,k}\right)$.

Now we assume that all precoders and combiners are constructed based on imperfect CSI. Consequently, (5.2) can be written as

$$\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{V}}_j = \mathbf{0}\ ,\ \ \forall j \neq k \tag{5.12}$$

where all $\widehat{\mathbf{U}}_k$ and $\widehat{\mathbf{V}}_j$ are calculated based on the fact that only imperfect channel estimations $\widehat{\mathbf{H}}_{k,j}$ are available. In this case the received signal at node k in (5.1) can be written

as

$$\widehat{\mathbf{y}}_k = \mathbf{H}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k + \sum_{\substack{j=1\\j\neq k}}^{K}\mathbf{H}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \mathbf{z}_k \tag{5.13}$$

If the perfect direct link, i.e., $\mathbf{H}_{k,k}$, is available at receive node k, recovering $\mathbf{c}_k$ at the corresponding node is rather straightforward; however, if we assume that the receive node k has also access to imperfect direct link, i.e., $\widehat{\mathbf{H}}_{k,k}$, data recovery becomes a bit tricky. Therefore in the remainder of this subsection, we address this issue by considering the fact that only imperfect direct link $\widehat{\mathbf{H}}_{k,k}$ is available at receive node k. Now at the kth receive node, the received signal gets premultiplied by $\widehat{\mathbf{U}}_k^{\mathsf{H}}$, and therefore we have

$$\begin{aligned}
\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{y}}_k &= \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k + \sum_{\substack{j=1\\j\neq k}}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{H}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k \\
&\overset{①}{=} \widehat{\mathbf{U}}_k^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)\widehat{\mathbf{V}}_k\mathbf{c}_k+\sum_{\substack{j=1\\j\neq k}}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j}+\breve{\mathbf{H}}_{k,j}\right)\widehat{\mathbf{V}}_j\mathbf{c}_j+\widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k \\
&\overset{②}{=} \underbrace{\frac{1}{1+\tau}\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{V}}_k\mathbf{c}_k}_{\texttt{desired term}} + \underbrace{\sum_{j=1}^{K}\widehat{\mathbf{U}}_k^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{V}}_j\mathbf{c}_j + \widehat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{z}_k}_{\texttt{interference plus noise term}}
\end{aligned} \tag{5.14}$$

where ① follows from (5.6), and ② is due to (5.12). Thus to recover $\mathbf{c}_k$, $\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{y}}_k$ should be premultiplied by $(1+\tau)\times\operatorname{pinv}\left(\widehat{\mathbf{U}}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{V}}_k\right)$. In other words, when the receive node is in possession of imperfect direct link $\widehat{\mathbf{H}}_{k,k}$, to have an unbiased detection, the received signal should be scaled up by $(1+\tau)$.

5.3 Achievable Sum Rates and DoF

In this section, we quantify the achievable sum rates and DoF of constant MIMO IA under imperfect CSI model described in subsection 5.2.2. To do so, first sum rate and total DoF achieved by perfect CSI are going to be considered since this will be helpful to go through the imperfect CSI case in the forthcoming subsection.

5.3.1 Achievable Sum Rate and DoF under Perfect CSI

By assuming diagonalized subchannels, the IA conditions in (5.2)–(5.3) can be respectively restated as

$$\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{v}_{jm} = 0 \ , \ \forall\,(j,m) \neq (k,\ell) \tag{5.15}$$

$$\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right| > 0 \ , \ \forall k,\ell \tag{5.16}$$

where $\mathbf{u}_{k\ell}$ denotes the ℓth column of $\mathbf{U}_k$ and $\mathbf{v}_{jm}$ refers to the mth column of $\mathbf{V}_j$.

Now, with considering i.i.d. Gaussian input signaling, the sum rate achieved can be expressed as

$$R = \sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2}{J_{k\ell}+\sigma^2}\right) \tag{5.17}$$

where the leakage interference is treated as noise and

$$J_{k\ell} = \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{v}_{jm}\right|^2 + \sum_{\substack{m=1\\ m\neq \ell}}^{d} P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{km}\right|^2 \tag{5.18}$$

With the presence of perfect CSI, the IA condition in (5.15) is satisfied, and consequently $J_{k\ell} = 0$. In this case the sum rate in (5.17) reads

$$R_{\texttt{Perfect CSI}} = \sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2}{\sigma^2}\right) \tag{5.19}$$

To further proceed, we consider the following lemma:

Lemma 5.1: $\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}$ is a Gaussian random variable with mean zero and variance one.

Proof: First, it should be noted that $\mathbf{v}_{k\ell}$ are independent of $\mathbf{H}_{k,k}$. This can be trivially deduced since in IA solutions (e.g., interference leakage minimization algorithms in [23, 34]), each $\mathbf{V}_k$ is only a function of interfering links $\mathbf{H}_{k,j}, j \neq k$ and not the direct link $\mathbf{H}_{k,k}$. Since all channel coefficients are i.i.d., $\mathbf{v}_{k\ell}, \ell = 1,\ldots,d$ are independent of $\mathbf{H}_{k,k}$. Due to the truncated unitary precoders, it can be further assumed that $\mathbf{H}_{k,k}\mathbf{v}_{k,\ell}$ is independent of $\mathbf{H}_{k,k}\mathbf{v}_{k,m},\ \forall m \neq \ell$. The interference subspace observed by the ℓth stream of user k can be defined as

$$\boldsymbol{\mathcal{V}}_{k\ell} = \left(\mathbf{H}_{k,1}\mathbf{V}_1,\ldots,\mathbf{H}_{k,k-1}\mathbf{V}_{k-1},\mathbf{H}_{k,k}\mathbf{V}_k^{[\ell]},\mathbf{H}_{k,k+1}\mathbf{V}_{k+1},\ldots,\mathbf{H}_{k,K}\mathbf{V}_K\right) \tag{5.20}$$

where $\mathbf{V}_k^{[\ell]}$ is obtained by striking out the ℓth column of $\mathbf{V}_k$, i.e., $\mathbf{v}_{k\ell}$. Having met the feasibility conditions of IA [21] and with respect to (5.15), we can consider $\mathbf{u}_{k\ell}$ as the null space of $\boldsymbol{\mathcal{V}}_{k\ell}^{\mathsf{H}}$. Alternatively, we can calculate $\mathbf{u}_{k\ell}$ as the least dominant left singular vector of $\boldsymbol{\mathcal{V}}_{k\ell}$. Consequently $\mathbf{u}_{k\ell}$ is independent of $\mathbf{H}_{k,k}\mathbf{v}_{k\ell}$. Since $\mathbf{H}_{k,k}$ is a standard Gaussian matrix, it is bi-unitarily invariant [1] [3]. Thus $\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. ■

Corollary 5.1: Since $\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell} \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2$ is exponentially distributed with both mean and variance equal to one. Having met the feasibility conditions of IA, we can represent the total achievable DoF as

$$D_{\texttt{Perfect CSI}} = \lim_{P\to\infty} \frac{R_{\texttt{Perfect CSI}}}{\log_2 P} = Kd \tag{5.21}$$

where $R_{\texttt{Perfect CSI}}$ is defined in (5.19).

5.3.2 Achievable Sum Rate and DoF under Imperfect CSI

In this subsection, first we derive bounds on the asymptotic mean loss in sum rate under imperfect CSI compared to the perfect CSI, and then we establish a bound on achievable DoF under CSI mismatch. Also as mentioned earlier, for the imperfect CSI case, we assume that all precoders and combiners are obtained based on imperfect channel measurements $\widehat{\mathbf{H}}_{k,j},\ \forall k,j$ in (5.4). In this case the leaked interference is defined as

$$\widehat{J}_{k\ell} = \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,j}\widehat{\mathbf{v}}_{jm}\right|^2 + \sum_{\substack{m=1\\ m\neq \ell}}^{d} P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{km}\right|^2 \tag{5.22}$$

and with respect to (5.12), we can write

$$\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm} = 0 \ \ , \ \ \forall\,(j,m) \neq (k,\ell) \tag{5.23}$$

To further proceed, the following lemma should be noticed:

Lemma 5.2: $\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}$ is a Gaussian random variable with mean zero and variance one.

Proof: Notice the unit-norm vectors $\widehat{\mathbf{v}}_{k\ell}$ are calculated based on $\widehat{\mathbf{H}}_{k,j}, \forall j \neq k$ and are thus independent of $\widehat{\mathbf{H}}_{k,k}$. Since due to equation (5.4), $\widehat{\mathbf{H}}_{k,k}$ is related to $\mathbf{H}_{k,k}$, consequently $\widehat{\mathbf{v}}_{k\ell}$ are independent of $\mathbf{H}_{k,k}$ as well. Therefore, analogous to the same approach

[1] A rectangular random matrix $\mathbf{H}$ is called bi-unitarily invariant if the joint distribution of its entries equals that of $\mathbf{U}^{\mathsf{H}}\mathbf{H}\mathbf{V}$ for any unitary matrices $\mathbf{U}$ and $\mathbf{V}$ independent of $\mathbf{H}$.

as in the proof of lemma 5.1, it is easy to discuss that $\widehat{\mathbf{u}}_{k\ell}$ is independent of $\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}$. Thus $\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. ■

Corollary 5.2: Since $\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell} \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2$ is exponentially distributed with both mean and variance equal to one.

Now that we established both $\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}$ and $\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}$ are standard Gaussian random variables, we continue with the following theorem:

Theorem 5.1: Let ΔR denote the mean loss in sum rate. At asymptotically high SNRs and for $\alpha > 1$ and $\alpha < 1$, ΔR tends to zero and infinity, respectively. However, for $\alpha = 1$, ΔR is upper bounded as

$$\Delta R \leq Kd \log_2\left(1 + (Kd-1)\,\beta\right) \tag{5.24}$$

Proof: Considering $R_{\texttt{Perfect CSI}}$ as the achievable sum rate with perfect CSI described in (5.19), we define the mean loss in sum rate as

$$\Delta R = \mathbb{E}_{\mathbf{H}}\left\{R_{\texttt{Perfect CSI}}\right\} - \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{R_{\texttt{Imperfect CSI}}\right\} \tag{5.25}$$

which further yields the following upper bound

$$\begin{aligned}
\Delta R &= \mathbb{E}_{\mathbf{H}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2}{\sigma^2}\right)\right\} \\
&- \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2}{\widehat{J}_{k\ell}+\sigma^2}\right)\right\} \\
&= \mathbb{E}_{\mathbf{H}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2}{\sigma^2}\right)\right\} \\
&- \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{\widehat{J}_{k\ell}+P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2}{\sigma^2}\right)\right\} \\
&+ \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{\widehat{J}_{k\ell}}{\sigma^2}\right)\right\} \\
&\overset{③}{\leq} \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{\widehat{J}_{k\ell}}{\sigma^2}\right)\right\} \\
&\overset{④}{\leq} \sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\widehat{J}_{k\ell}\right\}}{\sigma^2}\right)
\end{aligned} \tag{5.26}$$

where ③ is due to integrating lemma 5.1 and lemma 5.2 to the following inequality

$$\begin{aligned}&\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{\widehat{J}_{k\ell}+P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2}{\sigma^2}\right)\right\}\geq\\&\mathbb{E}_{\mathbf{H}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right|^2}{\sigma^2}\right)\right\}\end{aligned}\tag{5.27}$$

and ④ is a result of Jensen's inequality. $\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\widehat{J}_{k\ell}\right\}$ can be further written as

$$\begin{aligned}\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\widehat{J}_{k\ell}\right\} &= \sum_{\substack{j=1\\j\neq k}}^{K}\sum_{m=1}^{d}P\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,j}\widehat{\mathbf{v}}_{jm}\right|^2\right\}+\sum_{\substack{m=1\\m\neq\ell}}^{d}P\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{km}\right|^2\right\}\\&\overset{⑤}{=}\sum_{\substack{j=1\\j\neq k}}^{K}\sum_{m=1}^{d}P\mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau}+\breve{\mathbf{H}}_{k,j}\right)\widehat{\mathbf{v}}_{jm}\right|^2\right\}\\&+\sum_{\substack{m=1\\m\neq\ell}}^{d}P\mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau}+\breve{\mathbf{H}}_{k,k}\right)\widehat{\mathbf{v}}_{km}\right|^2\right\}\\&\overset{⑥}{=}\sum_{\substack{j=1\\j\neq k}}^{K}\sum_{m=1}^{d}P\mathbb{E}_{\breve{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\right|^2\right\}+\sum_{\substack{m=1\\m\neq\ell}}^{d}P\mathbb{E}_{\breve{\mathbf{H}}}\left\{\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\breve{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\right|^2\right\}\\&\overset{⑦}{=}P\left(Kd-1\right)\frac{\tau}{1+\tau}\end{aligned}\tag{5.28}$$

where ⑤ follows from (5.6), and ⑥ is due to (5.23). ⑦ follows the fact that all $\widehat{\mathbf{u}}_{k\ell}$ and $\widehat{\mathbf{v}}_{jm},\ \forall k,j,\ell,m$ are calculated upon $\widehat{\mathbf{H}}_{k,j}$ which due to lemma 3.1 on page 28 are independent of $\breve{\mathbf{H}}_{k,j}$. Due to the fact that $\breve{\mathbf{H}}_{k,j}$ are bi-unitarily invariant, $\widehat{\mathbf{u}}_{k\ell}\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm},\ \forall k,j,\ell,m$ are Gaussian random variables with mean zero and variance $\tau/\left(1+\tau\right)$. Thus, following (5.26) we have

$$\Delta R\leq\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\gamma\left(Kd-1\right)\frac{\tau}{1+\tau}\right)=Kd\log_2\left(1+\left(Kd-1\right)\frac{\beta\gamma^{1-\alpha}}{1+\beta\gamma^{-\alpha}}\right)\tag{5.29}$$

which for asymptotically high SNRs yields

$$\lim_{\gamma\to\infty}\Delta R\begin{cases}=\infty & 0\leq\alpha<1\\ \leq Kd\log_2\left(1+\left(Kd-1\right)\beta\right) & \alpha=1\\ =0 & 1<\alpha\end{cases}\tag{5.30}$$

■

Thus, theorem 5.1 states that if the error variance is proportional to the inverse of SNR, the asymptotic mean loss in sum rate is bounded above by the right hand side (RHS) of (5.24), and by using numerical results in section 5.5, we will further show that this upper bound is not excessively loose.

In the following theorem, a novel bound on achievable DoF is derived:

Theorem 5.2: For a constant MIMO IA, when the variance of the CSI error is equal to $\tau = \beta\gamma^{-\alpha},\ 0 \leq \alpha < 1$, the achievability of an α-fraction of the total DoF, i.e., $\alpha D_{\texttt{Perfect CSI}}$ DoF, is guaranteed.

Proof:

$$\begin{aligned}
D_{\texttt{Imperfect CSI}} &= \lim_{P\to\infty}\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{R_{\texttt{Imperfect CSI}}\right\}}{\log_2 P}\\
&= \lim_{P\to\infty}\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(1+\frac{P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2}{\widehat{J}_{k\ell}+\sigma^2}\right)\right\}}{\log_2 P}\\
&= \lim_{P\to\infty}\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2+\widehat{J}_{k\ell}+\sigma^2\right)\right\}}{\log_2 P}\\
&- \lim_{P\to\infty}\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(\widehat{J}_{k\ell}+\sigma^2\right)\right\}}{\log_2 P}\\
&\overset{⑧}{\geq} \lim_{P\to\infty}\frac{\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(P\left|\widehat{\mathbf{u}}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right|^2\right)\right\}}{\log_2 P}\\
&- \lim_{P\to\infty}\frac{\sum_{k=1}^{K}\sum_{\ell=1}^{d}\log_2\left(\mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\left\{\widehat{J}_{k\ell}\right\}+\sigma^2\right)}{\log_2 P}
\end{aligned}$$

$$\overset{⑨}{=} Kd - \lim_{P\to\infty} \frac{\sum_{k=1}^{K}\sum_{\ell=1}^{d} \log_2\left(\frac{\beta P^{1-\alpha}\sigma^{2\alpha}}{1+\beta P^{-\alpha}\sigma^{2\alpha}}(Kd-1)+\sigma^2\right)}{\log_2 P}$$

$$= \begin{cases} Kd & 1 \le \alpha \\ \alpha Kd & 0 \le \alpha < 1 \end{cases} \tag{5.31}$$

where ⑧ is due to discarding interference plus noise in the first term and applying Jensen's inequality to the second one. ⑨ is also due to integrating lemma 5.2 to the first term and inserting the last term of equation (5.28) to the second one. ■

Remark 5.1: Note that the results of (5.31) are inherently related to those in (5.30). For example, for the case of $\alpha < 1$, while (5.31) implies that the achievable DoF is αKd, (5.30) indicates that for this case and by increasing SNR, the asymptotic mean loss in sum rate is unboundedly increasing. Also (5.31) implies that when $\alpha \geq 1$, full DoF is achievable, that is, the asymptotic mean loss in sum rate is a bounded value. In this case, (5.30) implies that when $\alpha = 1$, this asymptotic mean loss converges to a non-zero constant which is upper bounded by (5.24) whereas for $\alpha > 1$, it tends to zero.

5.4 Adaptive Max-SINR

In section 5.3, we derived general bounds regarding the mean loss in sum rate and achievable DoF of IA subject to CSI mismatch. In this section, we specifically place our focus on performance analysis and improvement of one particular IA technique, namely Max-SINR defined in [23]. It is shown that while with the presence of perfect CSI, Max-SINR outperforms interference leakage minimization algorithms, this promised improvement becomes negligible especially at high SNRs subject to imperfect CSI. Accordingly, we propose an adaptive Max-SINR, which with the knowledge of error variance in advance, can improve the performance of original Max-SINR under CSI mismatch.

Notice due to the coupled nature of the problem and regardless of what algorithm is being used, there are no closed form solutions for IA except for a very few particular cases, see e.g., [17, 70]. Consequently, finding precoders and combiners requests an iterative procedure in general. Therefore, first we fix the precoders and seek the combiners, and then we fix the combiners and seek the precoders. Given randomly initialized precoders and with respect to the fact that only imperfect channel estimations $\widehat{\mathbf{H}}_{k,j}$ are available, the interference plus noise covariance matrix associated with the ℓth stream of user k can

$$
\begin{aligned}
\mathbf{Q}_k^{\ell} &= \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} P\mathbf{H}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \sum_{\substack{m=1\\ m\neq \ell}}^{d} P\mathbf{H}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} + \sigma^2\mathbf{I} \\
&\overset{⑩}{=} \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j} + \breve{\mathbf{H}}_{k,j}\right)\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j} + \breve{\mathbf{H}}_{k,j}\right)^{\mathsf{H}} \\
&+ \sum_{\substack{m=1\\ m\neq \ell}}^{d} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k} + \breve{\mathbf{H}}_{k,k}\right)\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k} + \breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}} + \sigma^2\mathbf{I} \\
&= \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} \frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \sum_{\substack{m=1\\ m\neq \ell}}^{d} \frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \\
&+ \underbrace{\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} \frac{P}{1+\tau}\left(\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}^{\mathsf{H}} + \breve{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}\right)}_{\mathbf{J}_1} \\
&+ \underbrace{\sum_{\substack{m=1\\ m\neq \ell}}^{d} \frac{P}{1+\tau}\left(\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\breve{\mathbf{H}}_{k,k}^{\mathsf{H}} + \breve{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\right)}_{\mathbf{J}_2} \\
&+ \underbrace{\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m=1}^{d} P\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}^{\mathsf{H}} + \sum_{\substack{m=1\\ m\neq \ell}}^{d} P\breve{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\breve{\mathbf{H}}_{k,k}^{\mathsf{H}} + \sigma^2\mathbf{I}}_{\mathbf{J}_3}
\end{aligned}
\tag{5.32}
$$

be shown as (5.32) wherein ⑩ follows from (5.6). To further proceed, we consider the following lemmas:

Lemma 5.3: $\mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}^{\mathsf{H}}\right\} = \mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\breve{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}\right\} = \mathbf{0} \quad \forall k,j,m$

Proof: All precoders and combiners are constructed upon channel estimations $\widehat{\mathbf{H}}_{k,j}$ which based on lemma 3.1 on page 28 are independent of $\breve{\mathbf{H}}_{k,j}$. ■

Lemma 5.4: If $\mathbf{A} \in \mathbb{C}^{M\times N}$ represents a Gaussian matrix with i.i.d. elements of mean zero and variance a, and $\mathbf{b} \in \mathbb{C}^{N\times 1}$ refers to a unit-norm vector independent of $\mathbf{A}$, then $\mathbb{E}_{\mathbf{A}}\left\{\mathbf{A}\mathbf{b}\mathbf{b}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\right\} = a\mathbf{I}$.

Proof: Since $\mathbf{A}$ is a Gaussian matrix, it is bi-unitarily invariant, and consequently the

joint distribution of its entries equals that of $\mathbf{Ab}$ for any unit-norm vector $\mathbf{b}$ independent of $\mathbf{A}$. Therefore $\mathbf{Ab}$ is a zero-mean Gaussian vector with covariance matrix $a\mathbf{I}$. ■

Following lemma 5.3 and lemma 5.4, we substitute those parts of (5.32) including $\breve{\mathbf{H}}_{j,k}\ \forall j,k$ with their expected values, i.e,

$$\mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\mathbf{J}_1\right\} = \mathbb{E}_{\widehat{\mathbf{H}},\breve{\mathbf{H}}}\left\{\mathbf{J}_2\right\} = \mathbf{0}$$

and

$$\mathbb{E}_{\breve{\mathbf{H}}}\left\{\mathbf{J}_3\right\} = P\left(Kd-1\right)\frac{\tau}{1+\tau}\mathbf{I}$$

This way, we can approximate $\mathbf{Q}_k^\ell$ in (5.32) with a simpler form, i.e., $\widehat{\mathbf{Q}}_k^\ell$, as follows:

$$\begin{aligned}\widehat{\mathbf{Q}}_k^\ell &= \sum_{\substack{j=1\\j\neq k}}^{K}\sum_{m=1}^{d}\frac{P}{\left(1+\tau\right)^2}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \sum_{\substack{m=1\\m\neq\ell}}^{d}\frac{P}{\left(1+\tau\right)^2}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\\ &+ \left(P\left(Kd-1\right)\frac{\tau}{1+\tau}+\sigma^2\right)\mathbf{I}\end{aligned} \tag{5.33}$$

With respect to $\widehat{\mathbf{Q}}_k^\ell$ in (5.33), the SINR maximizing receive filter $\widehat{\mathbf{u}}_{k\ell}$ is thus given by

$$\widehat{\mathbf{u}}_{k\ell} = \frac{\left(\widehat{\mathbf{Q}}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}}{\left\|\left(\widehat{\mathbf{Q}}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right\|_2} \tag{5.34}$$

which yields the unit-norm combiner of the ℓth stream of user k.

As mentioned earlier, due to the coupled nature of the problem, finding precoders and combiners requests an iterative algorithm in general. Plus, with respect to the fact that only imperfect channel estimations $\widehat{\mathbf{H}}_{k,j}$ are available, and with the knowledge of error variance τ in advance, the proposed algorithm is concisely presented in the following.

Adaptive Max-SINR

1. Set $\varepsilon \coloneqq \gamma^{-1}\left(1+\tau\right)^2 + \tau\left(\tau+1\right)\left(Kd-1\right)$
2. Initialize random unit-norm vectors $\widehat{\mathbf{v}}_{k\ell},\ \forall\, k,\ \ell$
3. $\mathbf{Q}_k^{\ell} = \sum\limits_{\substack{j=1\\ j\neq k}}^{K}\sum\limits_{m=1}^{d} \widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jm}\widehat{\mathbf{v}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \sum\limits_{\substack{m=1\\ m\neq \ell}}^{d} \widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} + \varepsilon\,\mathbf{I}$
4. $\widehat{\mathbf{u}}_{k\ell} = \dfrac{\left(\mathbf{Q}_k^{\ell}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}}{\left\|\left(\mathbf{Q}_k^{\ell}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right\|_2}$
5. $\widehat{\mathbf{U}}_k \Longleftarrow \mathrm{orth}\left(\widehat{\mathbf{U}}_k\right)$
6. $\mathbf{R}_k^{\ell} = \sum\limits_{\substack{j=1\\ j\neq k}}^{K}\sum\limits_{m=1}^{d} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{jm}\widehat{\mathbf{u}}_{jm}^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k} + \sum\limits_{\substack{m=1\\ m\neq \ell}}^{d} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{km}\widehat{\mathbf{u}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k} + \varepsilon\,\mathbf{I}$
7. $\widehat{\mathbf{v}}_{k\ell} = \dfrac{\left(\mathbf{R}_k^{\ell}\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{k\ell}}{\left\|\left(\mathbf{R}_k^{\ell}\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{k\ell}\right\|_2}$
8. $\widehat{\mathbf{V}}_k \Longleftarrow \mathrm{orth}\left(\widehat{\mathbf{V}}_k\right)$
9. Go to Step 3 and repeat

In algorithm above, "$\Longleftarrow$" denotes assignment through an in-place manner, and $\mathrm{orth}\left(\mathbf{A}\right)$ denotes an orthonormal basis for the range of $\mathbf{A}$, e.g., the unitary part of the orthogonal-triangular (QR) decomposition.

Remark 5.2: Note that similar to the original Max-SINR in [23], although the precoders and combiners calculated by adaptive Max-SINR have unit-norm columns, they are not unitary which may result in rank-deficient precoders and/or combiners for at least one user and hence at high SNRs, they may not be able to achieve full multiplexing gain and thus degrade the performance. Therefore if the goal is to attain full DoF, one trick to make precoders and combiners unitary is to insert orthogonalization steps after obtaining $\mathbf{U}_k$ and $\mathbf{V}_k$, which ensure no performance degradation at high SNR ranges. This additional orthogonalization steps have been indicated in Steps 5 and 8 of algorithm above. However, if the intention is performance evaluation at low-to-intermediate SNRs (which does not include the achievable DoF), omitting the aforementioned orthogonalization steps does not compromise the performance.

Remark 5.3: By setting $\tau = 0$ in the first step of algorithm above and irrespective of the orthogonalization steps, adaptive Max-SINR boils down to the original Max-SINR in [23] subject to imperfect CSI. If we further replace $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$, the aforementioned

algorithm becomes original Max-SINR under perfect CSI. In other words, while for the original Max-SINR, the scaling factor of the identity matrix is γ^{-1}, for the proposed adaptive Max-SINR, this scaling factor is equal to ε which is defined in Step 1 of adaptive Max-SINR algorithm. Therefore, the proposed adaptive Max-SINR adds no extra computational complexity compared to the original Max-SINR while achieving notably better performance under CSI mismatch.

5.5 Numerical Results

In this section and by using numerical results, we corroborate the undergone analyses in this chapter. More specifically, we substantiate the analytically derived bounds in equations (5.30) and (5.31) and manifest the improved performance achieved by adaptive Max-SINR (discussed in section 5.4) compared to the original Max-SINR under CSI mismatch. For the case of imperfect CSI, we assume that the channel estimation error variance obeys (3.3), i.e., $\tau = \beta\gamma^{-\alpha}$.

By considering i.i.d. Gaussian input signaling and uniform power allocation, we evaluate the achievable sum rates as [67]

$$R = \sum_{k=1}^{K} \log_2 \det \left(\mathbf{I} + \left(\gamma^{-1}\mathbf{I} + \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{\Phi}_{k,j} \right)^{-1} \mathbf{\Phi}_{k,k} \right) \tag{5.35}$$

where $\mathbf{\Phi}_{k,j} = \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} \mathbf{U}_k$ and in the case of imperfect CSI, all precoders and combiners are constructed based on erroneous channel estimations in (5.4).

To evaluate the performance of IA in this chapter and also the subsequent chapters, it is more appropriate to recall some of the standard constant MIMO IA algorithms. Although there are many iterative IA algorithms in the literature, see e.g., [23–34], the most representative ones are Max-SINR [23], RCRM [26], Alt-Min [34], and Min-WLI [23]. Since Max-SINR in [23] is a spacial case of the proposed adaptive Max-SINR (by setting $\tau = 0$), the latter three are respectively represented in the following.

For RCRM the optimization criterion is based on maximizing the sum of interference free dimensions as follows [26]:

$$\min_{\substack{\{\mathbf{V}_\ell\}_{\ell=1}^{K} \\ \{\mathbf{U}_\ell\}_{\ell=1}^{K}}} \sum_{k=1}^{K} \operatorname{rank}\left(\boldsymbol{\mathcal{J}}_k\right)$$

$$\text{s.t.}: \qquad \text{rank}\left(\boldsymbol{\mathcal{S}}_k\right) = d, \quad \forall k \tag{5.36}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{S}}_k &\triangleq \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k \\ \boldsymbol{\mathcal{J}}_k &\triangleq \mathbf{U}_k^{\mathsf{H}} \left[\mathbf{H}_{k,1}\mathbf{V}_1 \ \ldots \ \mathbf{H}_{k,k-1}\mathbf{V}_{k-1} \ \mathbf{H}_{k,k+1}\mathbf{V}_{k+1} \ \ldots \ \mathbf{H}_{k,K}\mathbf{V}_K\right] \end{aligned} \tag{5.37}$$

Note that in the above RCRM criterion, the orthogonality constraints on precoders and combiners are omitted. However, since the aforementioned optimization is not readily solvable, the following approximation is used which is based on relaxing the rank cost function to a nuclear norm one:

IA via RCRM

1. Initialize $\mathbf{U}_k, \forall\, k$
2. $$\begin{cases} \min\limits_{\{\mathbf{V}_\ell\}_{\ell=1}^K} \sum\limits_{k=1}^K \|\boldsymbol{\mathcal{J}}_k\|_* \\ \quad \text{s.t.}: \quad \boldsymbol{\mathcal{S}}_k \succeq \mathbf{0} \\ \quad\quad\ \ \lambda_{\min}\left(\boldsymbol{\mathcal{S}}_k\right) \geq \epsilon, \quad \forall k \end{cases}$$
3. $$\begin{cases} \min\limits_{\{\mathbf{U}_\ell\}_{\ell=1}^K} \sum\limits_{k=1}^K \|\boldsymbol{\mathcal{J}}_k\|_* \\ \quad \text{s.t.}: \quad \boldsymbol{\mathcal{S}}_k \succeq \mathbf{0} \\ \quad\quad\ \ \lambda_{\min}\left(\boldsymbol{\mathcal{S}}_k\right) \geq \epsilon, \quad \forall k \end{cases}$$
4. Go to Step 2 and repeat
5. $\mathbf{V}_k = \text{orth}\left(\mathbf{V}_k\right) \quad \forall k$
6. $\mathbf{U}_k = \text{orth}\left(\mathbf{V}_k\right) \quad \forall k$

where $\succeq$ denotes positive semi-definiteness, $\lambda_{\min}$ represents the smallest eigenvalue, ϵ is an arbitrarily small positive real number, and $\|\cdot\|_*$ designates the nuclear norm, i.e., the sum of the singular values.

For interference leakage minimization algorithms like Alt-Min [34] and Min-WLI [23], the optimization criterion is based on minimizing the sum of the interference leakage

which is defined as follows:

$$\sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{U}_k^{\mathsf{H}} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{H}_{k,\ell} \mathbf{V}_\ell \mathbf{V}_\ell^{\mathsf{H}} \mathbf{H}_{k,\ell}^{\mathsf{H}} \mathbf{U}_k\right] \tag{5.38}$$

At each step of the above optimization, either the set of precoders or the set of combiners is fixed, and minimization of (5.38) is performed over the free set of variables. This interference leakage minimization can therefore be carried out by using the following two algorithms through an iterative manner:

Alt-Min

1. Initialize random unitary matrices $\mathbf{V}_k, \ \forall\, k$
2. $\mathbf{Q}_k = \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}}$
3. $\mathbf{G}_k = \mathrm{eig}\left(\mathbf{Q}_k\right)_{d+1:M}$
4. $\mathbf{U}_k = \mathbf{I} - \mathbf{G}_k \mathbf{G}_k^{\mathsf{H}}$
5. $\mathbf{R}_k = \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \left(\mathbf{I} - \mathbf{G}_j \mathbf{G}_j^{\mathsf{H}}\right) \mathbf{H}_{j,k}$
6. $\mathbf{V}_k = \mathrm{eig}\left(\mathbf{R}_k\right)_{1:d}$
7. Go to Step 2 and repeat

Min-WLI

1. Initialize random unitary matrices $\mathbf{V}_k, \forall\, k$
2. $\mathbf{Q}_k = \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}}$
3. $\mathbf{U}_k = \mathrm{eig}\left(\mathbf{Q}_k\right)_{1:d}$
4. $\mathbf{R}_k = \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k}$
5. $\mathbf{V}_k = \mathrm{eig}\left(\mathbf{R}_k\right)_{1:d}$
6. Go to Step 2 and repeat

wherein algorithms above, $\mathrm{eig}\left(\cdot\right)_{1:d}$ represents the eigenvectors corresponding to the d smallest eigenvalues, and $\mathrm{eig}\left(\cdot\right)_{d+1:M}$ represents the eigenvectors corresponding to the

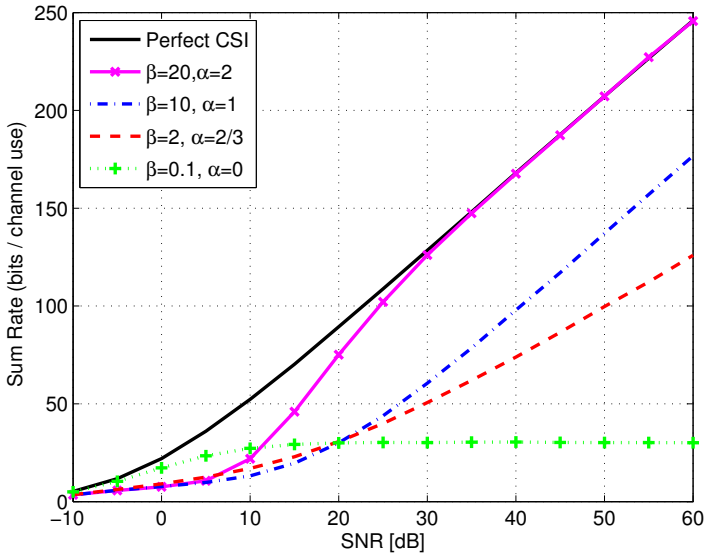

Figure 5.2: *Average sum rates achieved by Min-WLI algorithm for* $K = 3,\ d = 4$ *and* $M = N = 8$ *under various CSI qualities.*

$M - d$ largest eigenvalues.

Remark 5.4: Note that although Alt-Min and Min-WLI have been set up based on different design criteria, they achieve the same performance across all SNR ranges. Furthermore, to evaluate the performance of Alt-Min and Min-WLI under imperfect CSI, all perfect CSI links, i.e., $\mathbf{H}_{k,j}$, are replaced by channel estimations $\widehat{\mathbf{H}}_{k,j}$.

In the following, we verify the validity of the derived bounds in (5.30) and (5.31). Although there are various IA algorithms, without loss of generality we use Min-WLI algorithm to depict the results in Fig. 5.2.

Since the error variance depends on the inverse of SNR, i.e., $\tau = \beta\gamma^{-\alpha}$, to have the effect of CSI mismatch for a wider range of SNRs, it is assumed that the larger the α, the much larger the β.

Fig. 5.2 depicts the average sum rates of Min-WLI algorithm under various CSI qualities. We consider a symmetric constant MIMO IA with $K = 3$ and $d = 4$. To meet the sufficient conditions of feasibility for IA, we set $M = N = 8$ [21]. The solid black curve is related to the perfect CSI case where it can be seen that $Kd = 12$ DoF have been achieved. The blue dash-dot curve refers to the case $\beta = 10,\ \alpha = 1$. Based on (5.30), we expect that in this case, the asymptotic mean loss in sum rate should be bounded above by

$$Kd \log_2 \left(1 + (Kd - 1)\,\beta\right) \approx 81.5 \quad \text{bits per channel use} \tag{5.39}$$

As revealed in Fig. 5.2, this rate loss at $\gamma = 50$ dB is equal to 70 bits per channel use which is close to the upper bound 81.5 obtained through theorem 5.1. This verifies that the derived upper bound in (5.30) is not excessively loose.

The red dash curve denotes the case of $\beta = 2,\ \alpha = \frac{2}{3}$. In this case and based on (5.31), we expect that $\frac{2}{3}Kd = 8$ DoF should be achievable, which is verified in Fig. 5.2.

The solid curve with marker "$\times$" denotes the case $\beta = 20,\ \alpha = 2$. As expected by the last term of equation (5.30), for this value of α, the asymptotic mean loss in sum rate should be equal to zero. This is confirmed in Fig. 5.2 where the respective curve overlaps with the one corresponding to the perfect CSI at SNRs of larger than 35 dB.

The dot curve with marker "$+$" refers to the case $\beta = 0.1,\ \alpha = 0$. Based on the first term of equation (5.30), we expect that for the case of $\alpha = 0$, the asymptotic mean loss in sum rate should be unboundedly increasing with SNR. Also based on equation (5.31), for this value of α, the achievable DoF should be equal to zero. All these bounds match with the depicted results in Fig. 5.2.

To evaluate the achievable sum rates of original Max-SINR, if the goal is to evaluate the achievable DoF, with respect to remark 5.2, we consider Max-SINR with orthogonalization steps. In this case and with respect to remark 5.3, we need to set $\tau = 0$ in the first step of the proposed adaptive Max-SINR algorithm. In Fig. 5.3, we illustrate the achievable sum rate of Max-SINR with and without orthogonalization under perfect CSI and for a symmetric constant MIMO IA with $K = 4$ and $d = 2$. To meet the sufficient conditions of feasibility for IA, we set $M = 4,\ N = 6$ [21]. As shown, Max-SINR without orthogonalization (as primarily proposed in [23]) is not able to achieve full DoF. However, Max-SINR with orthogonalization rectify this degraded performance such that full DoF is now achievable. Plus, at low-to-intermediate SNRs, Max-SINR with orthogonalization achieves the same sum rate as the sans orthogonalization. This highlights the impact of inserted orthogonalization steps for the proposed adaptive Max-SINR as discussed in remark 5.2.

Fig. 5.4 depicts the achievable sum rates of Max-SINR with orthogonalization for the case $K = 4,\ d = 2,\ M = 4,$ and $N = 6$ under different CSI qualities. As revealed, for the case $\alpha = 1.7$, full DoF has been achieved, and the asymptotic mean loss in sum rate is zero such that the corresponding curve overlaps with the one representing the perfect CSI at SNR of 40 dB.

For the case $\alpha = 1$, the corresponding curve has the same slope as the one denoting the perfect CSI which implies that in this case full DoF have been achieved. In the case of $\alpha = 1,\ \beta = 15$ and according to (5.30), we expect that the asymptotic mean loss in

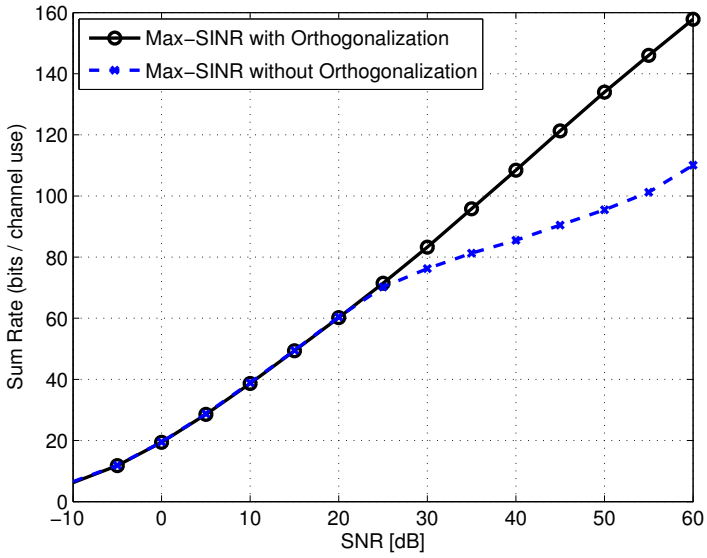

Figure 5.3: *Average sum rates achieved by Max-SINR algorithm with and without orthogonalization for* $K = 4,\ d = 2$ *and* $M = 4,\ N = 6$ *under perfect CSI.*

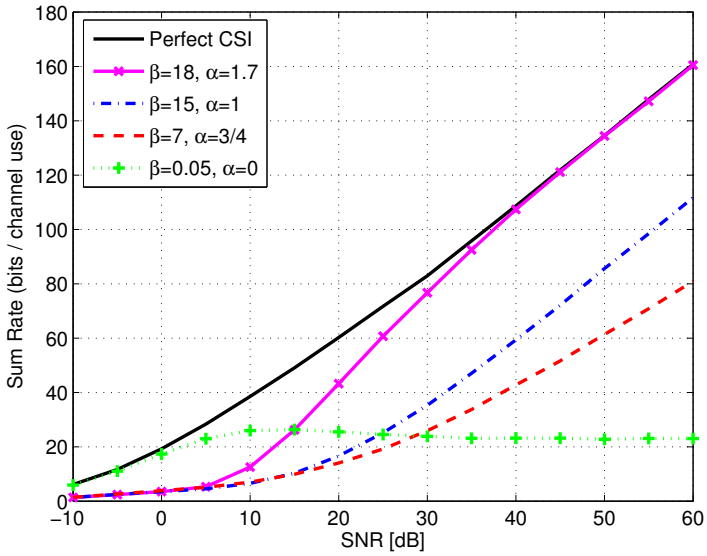

Figure 5.4: *Average sum rates achieved by Max-SINR algorithm with orthogonalization for* $K = 4,\ d = 2$ *and* $M = 4,\ N = 6$ *under various CSI qualities.*

sum rate should be no more than 53.8 bits per channel use. As seen in simulated results, at SNR of 50 dB, this gap is equal to 50 bits per channel use which is close enough to the analytically derived upper bound.

For $\alpha < 1$ and based on (5.31), it is expected that an α-fraction of the total DoF should

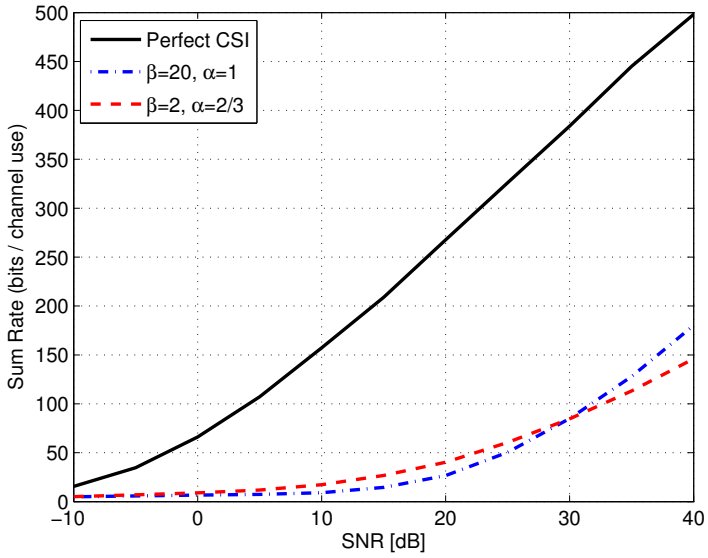

Figure 5.5: *Average sum rates achieved by Min-WLI algorithm for* $K = 9,\ d = 4$ *and* $M = N = 20$ *and related to* $\alpha = 1$ *and* $\alpha = \frac{2}{3}$.

be achievable. For example, when $\alpha = 3/4$, the achievable DoF should be equal to 6, which is again certified by considering the slope of the corresponding curve compared to the one denoting the perfect CSI. Also for both $\alpha = 0$ and $\alpha = 3/4$ and based on (5.30), it is expected that as SNR tends to infinity, the asymptotic mean loss in sum rate goes to infinity as well. This is again confirmed in Fig. 5.4, wherein the larger the SNR, the wider the gap between the curve representing the perfect CSI and the one denoting $\alpha < 1$.

Although Figs. 5.2–5.4 verify the derived bounds in theorem 5.1 and theorem 5.2, the analysis introduced might suggest that the bounds may not be rigorous enough for large K and/or d especially the upper bound of case $\alpha = 1$ in theorem 5.1. Therefore in Fig. 5.5, we depict the average sum rates under different CSI conditions for the case of $K = 9,\ d = 4$ using Min-WLI algorithm. To meet the feasibility conditions of IA, we set $M = N = 20$ [21]. The blue dash-dot curve refers to the case $\beta = 20, \alpha = 1$. Based on theorem 5.1, we expect that in this case, the asymptotic mean loss in sum rate should be bounded above by

$$Kd \log_2 \left(1 + (Kd - 1)\, \beta\right) \approx 340 \quad \text{bits per channel use} \tag{5.40}$$

As revealed in Fig. 5.5, this rate loss at $\rho = 40$ dB is equal to 320 which is very close to the upper bound 340 obtained in (5.30). This verifies that for larger values of K, d and

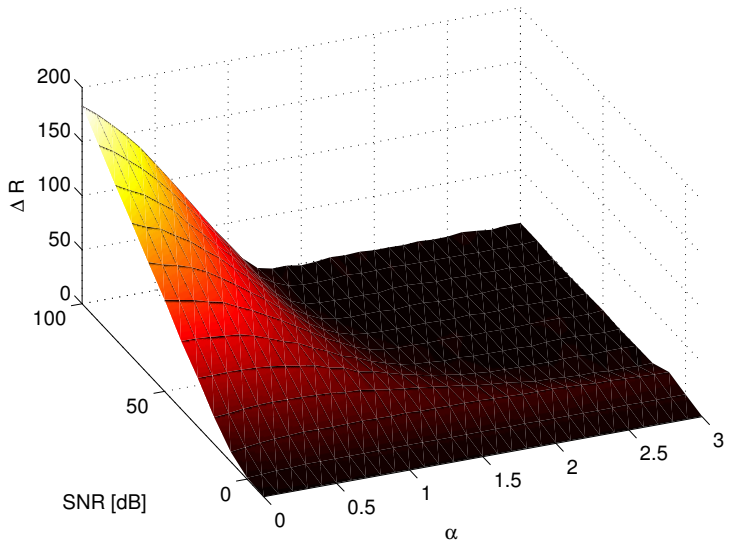

Figure 5.6: *Mean loss in sum rate of Min-WLI algorithm as a function of* α *for* $K = 3,\ d = 2$ *and* $M = N = 4$.

even β, the derived upper bound of theorem 5.1 is not excessively loose. Also the red dash curve denotes the case of $\beta = 2,\ \alpha = \frac{2}{3}$. Again in this case and based on theorem 5.2, we expect that $\frac{2}{3}Kd = 24$ DoF should be achievable, which is verified by Fig. 5.5.

Fig. 5.6 depicts the mean loss in sum rate of Min-WLI algorithm as a function of α for $K = 3,\ d = 2$ and $M = N = 4$. We increased α from 0 to 3 with steps of length 0.1. Accordingly, we changed β as $\beta = 10\alpha^2 + 0.1$. ΔR is also defined in (5.25). As shown, the major mean loss in sum rate occurs at high SNRs and for $\alpha < 1$.

Although the promised improvement of adaptive Max-SINR can be gleaned for various values of α, we focus on two representative cases: $\alpha = 0$ (which mimics the CSI feedback scenario), and $\alpha = 1$ (which mimics the reciprocal channels). We also consider a symmetric constant MIMO IA with $K = 3$ and $d = 4$ with $M = N = 8$.

Figs. 5.7 and 5.8, respectively, depict the average sum rate and SER for $\beta = 0.1,\ \alpha = 0$ and $\beta = 10,\ \alpha = 1$. The results due to perfect CSI are also depicted for comparison. To assess the SER performance, we considered communications under QPSK signaling where each transmitted block consists of 100 QPSK symbols. As observed, with the presence of perfect CSI, Max-SINR outperforms Alt-Min algorithm. However, subject to CSI mismatch, while the achieved improvement by Max-SINR is negligible compared to Alt-Min algorithm especially at high SNRs, adaptive Max-SINR achieves notably better performance. For example, adaptive Max-SINR achieves at least 18 dB gain compared to

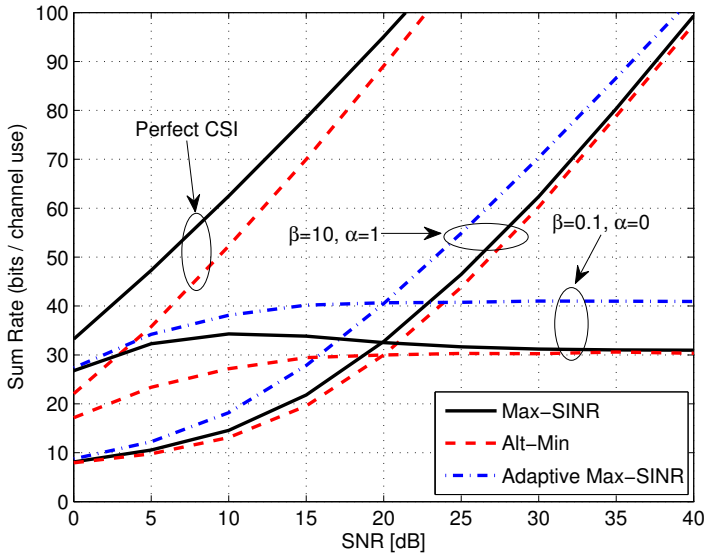

Figure 5.7: *Average sum rate for* $K = 3,\ d = 4,\ M = N = 8$ *and for the cases* $\beta = 0.1,\ \alpha = 0$, *and* $\beta = 10,\ \alpha = 1$.

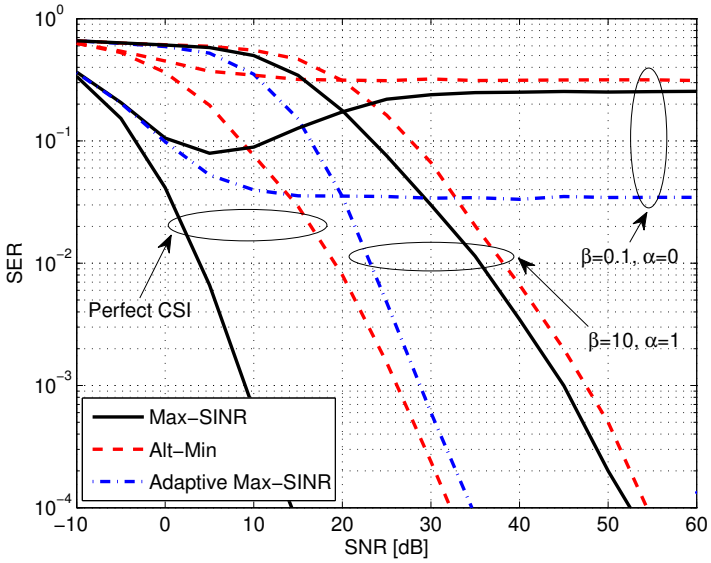

Figure 5.8: *SER for* $K = 3,\ d = 4,\ M = N = 8$ *and for the cases* $\beta = 0.1,\ \alpha = 0$, *and* $\beta = 10,\ \alpha = 1$. *Each transmitted block consists of* 100 *QPSK symbols.*

Max-SINR to achieve the same SER of 10^{-4} for the case $\alpha = 1$. Also at high SNRs, adaptive Max-SINR achieves 10 bits per channel use gain in sum rate compared to Max-SINR for the case $\alpha = 0$. Another interesting observation is that for the case of $\alpha = 0$, both sum rate and SER curves of Max-SINR show a nonmonotonic behavior whereas those

of the adaptive Max-SINR manifest a monotonic trend. This nonmonotonic behavior of Max-SINR is due to the fact that at low-to-intermediate SNRs, Max-SINR outperforms Alt-Min, but at high SNRs since the system becomes interference limited, Max-SINR and Alt-Min almost achieve the same performance.

5.6 Summary

Interference alignment will be one of the most dominating interference management techniques in future wireless networks since it enables us to increase the achievable sum rates proportional to the number of served users. However, similar to the other major communication techniques, full benefits of IA can be gleaned with perfect CSI, which its availability is not possible in general. In this chapter, we quantified the performance of IA under a generalized imperfect CSI model thereby we derived novel bounds regarding the asymptotic mean loss in sum rate and achievable DoF, where it has been shown that if the intention is to keep the asymptotic mean loss in sum rate bounded, the error variance must at least scales with the inverse of SNR. We then considered performance analysis of Max-SINR algorithm which maximizes the SINR of each stream of each user and therefore achieves better performance than interference leakage minimization algorithms under perfect CSI. Subject to imperfect CSI, however, Max-SINR achieves negligible improvement compared to interference leakage minimization algorithms, especially at high SNRs. We thus proposed an adaptive Max-SINR, which with the knowledge of error variance in advance, can notably improve the performance of original Max-SINR under CSI mismatch, without incurring extra computational complexity.

Chapter 6

Interference Alignment in Coordinated Networks

6.1 Introduction

In chapter 5, we represented the idea of interference alignment which aims to improve the achievable throughput in wireless interference networks such that the achievable DoF can be linearly scaled up with the number of users.

However, for a symmetric IC where each transmitter has N antennas and each receiver has M antennas and each transceiver pair requests d DoF, to achieve IA using multiple antennas instead of time extension, the necessary condition of feasibility is $M + N \geq d(K + 1)$ [20]. Moreover, it has been shown that if M and N divide d, the sufficient conditions of feasibility are met, i.e., the lower bound is tight and therefore the minimum required aggregate number of transmit and receive antennas per transceiver pair is exactly equal to $d(K + 1)$ [21].

Although for this symmetric system model, Kd DoF are achievable for K-user IC, all standard IA methods are based on the fact that at least $d(K + 1)$ antennas per transceiver pair are required to achieve this number of DoF. In this chapter, we propose IA algorithms such that Kd DoF are achievable even with less number of antennas per transceiver pair than $d(K + 1)$. The proposed approach relies on partially coordinated reception where averagely half of the total decoded data (which can be erroneous) are needed to be shared by receive nodes. It is worthwhile to point out that, from the practical point of view, coordinated reception and/or transmission are supported scenarios in future wireless networks [71–73], which facilitate the implementation of the partially coordinated receivers

being considered in this chapter. More importantly, most of these infrastructures are designed to support full coordination among transmit/receive nodes whereas here, we just consider half receive coordination which can introduce less burden to practical deployment.

We propose IA algorithms such that on average, half of the total cross CSI is needed to be available at receive nodes, which leads the proposed schemes to be a compromise between standard IA techniques. We show that for the proposed IA algorithms, the feasibility condition can now be expressed as $M+N \geq \frac{d(K+4)}{2}$. This implies a reduced number of antennas to achieve the same number of DoF compared to standard IA techniques like the ones in [23,24]. More specifically, for asymptotically large K and to achieve the same number of DoF, the required number of antennas of the proposed IA techniques can be half of that of standard IA methods. Apart from achieving the same number of DoF, we show that even with this reduced number of antennas, the achievable throughput of the proposed IA schemes is still comparable to that of standard IA methods. We also consider the effect of both error propagations and channel estimation errors on the performance of the proposed IA algorithms by proposing an adaptive algorithm, which with the knowledge of error variance in advance, is able to achieve notably better performance. Simulation results show that in this case, the proposed adaptive design enables us to glean better performance when only imperfect CSI is available.

This chapter is organized as follows: Section 6.2 deals with the idea of IA in partially coordinated receivers. We address the issue of the required amount of distributed CSI and the convergence of the proposed algorithms in Section 6.3. In Section 6.4, we derive the feasibility conditions of the proposed IA algorithms. We support the analysis of Section 6.4 by using numerical simulations in Section 6.5, and finally Section 6.6 contains a summary of the presented materials within this chapter.

6.2 IA in Partially Coordinated Receivers

6.2.1 Beamformer Design under Perfect CSI

By relying on the fact that coordinated transmission and/or reception are supported scenarios in future wireless networks and the corresponding infrastructures are provisioned [71–73], we propose IA algorithms adjusted for this configuration such that first, they enable us to achieve the same number of DoF as standard IA techniques but with less number of antennas; and second, even with this reduced number of antennas, the throughput of the

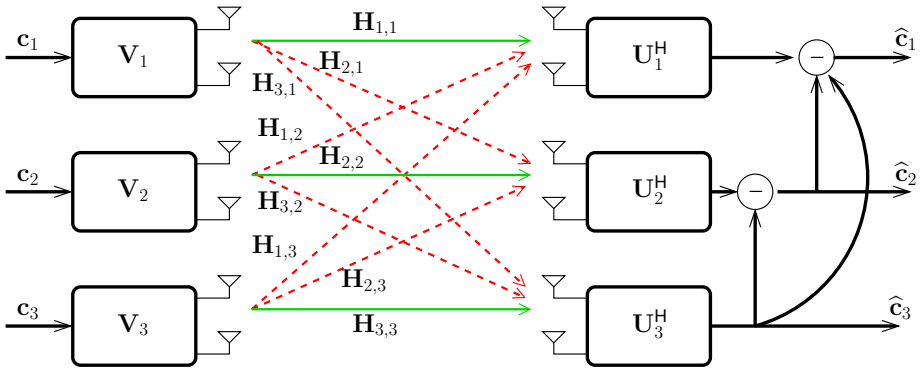

Figure 6.1: *3-user MIMO interference channel with partially coordinated receivers where solid green lines denote direct links and dashed red lines represent interfering (cross) links.* $\mathbf{c}_k$ *designates the input data of the* k*th transmit node whereas* $\widehat{\mathbf{c}}_k$ *indicates the decoded data at receive node* k*.*

proposed schemes is still comparable to that of standard IA methods. However, instead of assuming full information exchange among nodes, we just consider partially coordinated reception; that is, we arbitrarily index the receive nodes such that the $K-k$ receive nodes forward their decoded data to the kth receive node via high-bandwidth low-latency backhaul links. This implies that on average, half of the total decoded data is needed to be shared by receive nodes and consequently enables us to deploy successive interference cancelation (SIC). This imitates the case of Vertical-Bell Laboratories Layered Space-Time (V-BLAST) architectures with SIC which result in significant throughput for standard MIMO systems [74–76]. A simple description of the proposed scenario for the special case of $K = 3$ is illustrated in Fig. 6.1 where the third receive node forwards its decoded data to both the first and the second ones whereas the second receive node forwards its decoded data only to the first one. In this case, to achieve the same number of DoF per transceiver pair, the ZF constraint in (5.2) (meant for standard IA techniques) can now be expressed as

$$\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{V}_j = \mathbf{0} \;\;,\;\; \forall j \in \{1,\ldots,k-1\} \tag{6.1}$$

Similar to the standard IA algorithms, there are no closed form solutions for IA based on coordinated reception, except for a very few particular cases. Therefore, beamformer design requires an iterative procedure in general. With respect to this, we propose two different IA algorithms adjusted for partially coordinated reception. The first algorithm does not consider the effect of noise to design beamformers whereas the second one does

so. By considering (6.1) and with respect to the fact that perfect CSI is available, the proposed algorithms can be concisely described as follows:

Algorithm 1

1. Initialize random unitary matrices $\mathbf{V}_k,\ \forall\ k$
2. if $k = 1$
3. $\quad \mathbf{U}_k = \text{orth}\left(\mathbf{H}_{k,k}\mathbf{V}_k\right)$
4. else
5. $\quad \mathbf{Q}_k = \sum_{j=1}^{k-1} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}$
6. $\quad \mathbf{U}_k = \text{eig}\left(\mathbf{Q}_k\right)_{1:d}$
7. end
8. if $k = K$
9. $\quad \mathbf{V}_k = \text{orth}\left(\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k\right)$
10. else
11. $\quad \mathbf{R}_k = \sum_{j=k+1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\mathbf{H}_{j,k}$
12. $\quad \mathbf{V}_k = \text{eig}\left(\mathbf{R}_k\right)_{1:d}$
13. end
14. Go to Step 2 and repeat

Algorithm 2

1. Initialize random unit-norm vectors $\mathbf{v}_{k\ell},\ \forall\ k,\ \ell$
2. $\mathbf{Q}_k^{\ell} = \sum_{j=1}^{k-1}\sum_{n=1}^{d} P\mathbf{H}_{k,j}\mathbf{v}_{jn}\mathbf{v}_{jn}^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \sum_{m=1}^{\ell-1} P\mathbf{H}_{k,k}\mathbf{v}_{km}\mathbf{v}_{km}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} + \sigma^2\mathbf{I}$
3. $\mathbf{u}_{k\ell} = \dfrac{\left(\mathbf{Q}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{\left\|\left(\mathbf{Q}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right\|_2}$
4. $\mathbf{R}_k^{\ell} = \sum_{j=k+1}^{K}\sum_{n=1}^{d} P\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{u}_{jn}\mathbf{u}_{jn}^{\mathsf{H}}\mathbf{H}_{j,k} + \sum_{m=\ell+1}^{d} P\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{km}\mathbf{u}_{km}^{\mathsf{H}}\mathbf{H}_{k,k} + \sigma^2\mathbf{I}$
5. $\mathbf{v}_{k\ell} = \dfrac{\left(\mathbf{R}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}}{\left\|\left(\mathbf{R}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}\right\|_2}$
6. Go to Step 2 and repeat

A variational of Algorithm 1 based on alternating minimization can be described as follows:

Algorithm 3

1. Initialize random unitary matrices $\mathbf{V}_k,\ \forall\ k$
2. if $k = 1$
3. $\quad \mathbf{A}_k = \text{orth}\left(\mathbf{H}_{k,k}\mathbf{V}_k\right)$
4. $\quad \mathbf{U}_k = \mathbf{A}_k\mathbf{A}_k^{\mathsf{H}}$
5. $\quad \mathbf{G}_k = \mathbf{I} - \mathbf{U}_k$
6. else
7. $\quad \mathbf{Q}_k = \sum_{j=1}^{k-1} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}$
8. $\quad \mathbf{G}_k = \text{eig}\left(\mathbf{Q}_k\right)_{d+1:M}$
9. $\quad \mathbf{U}_k = \mathbf{I} - \mathbf{G}_k\mathbf{G}_k^{\mathsf{H}}$
10. end
11. if $k = K$
12. $\quad \mathbf{V}_k = \text{orth}\left(\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k\right)$
13. else
14. $\quad \mathbf{R}_k = \sum_{j=k+1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}}\left(\mathbf{I} - \mathbf{G}_j\mathbf{G}_j^{\mathsf{H}}\right)\mathbf{H}_{j,k}$
15. $\quad \mathbf{V}_k = \text{eig}\left(\mathbf{R}_k\right)_{1:d}$
16. end
17. Go to Step 2 and repeat

Note that Algorithm 3 is analogous to Algorithm 1 in the sense that it does not consider the effect of noise to design beamformers. Since Algorithm 3 achieves the same performance as Algorithm 1, hereafter, we place our focus on Algorithm 1.

Remark 6.1: Algorithm 1 only considers the interfering subchannels for aligning interference. In other words, for the kth user, Algorithm 1 does not consider neither the effect of direct link, i.e., $\mathbf{H}_{k,k}\mathbf{V}_k$, nor the noise power to design the appropriate combiners. However, Algorithm 2 utilizes the knowledge of both direct link and the noise power for beamformer design. Therefore, it is expected that Algorithm 2 should achieve better performance than Algorithm 1.

Remark 6.2: While for Algorithm 1, $\mathbf{V}_j$ and $\mathbf{U}_j$ are unitary matrices, this does not hold for Algorithm 2, which may result in rank-deficient precoders (receive combiners)

for at least one user and hence at high SNRs, it may not be able to achieve full DoF and thus reduces the throughput. However, one simple solution to make precoding and receive combining matrices unitary is using orthogonalization steps after obtaining $\mathbf{U}_k$ and $\mathbf{V}_k$, which ensure unitary precoding and receive combining matrices without incurring any performance degradation in total DoF and achievable throughput at high SNR ranges. However, if the intention is performance analysis at low-to-intermediate SNRs, there is no need for additional orthogonalization steps.

6.2.2 Beamformer Design under Imperfect CSI

In subsection 6.2.1, we proposed three algorithms under the assumption of the availability of perfect CSI. However, since from the practical point of view, perfect CSI is not readily accessible, it is desirable to consider an adaptive design to achieve better performance under CSI mismatch. In this subsection, we propose an adaptive version of Algorithm 2, which enables us to glean better performance subject to imperfect CSI. We consider the imperfect CSI model as the one discussed in subsection 5.2.2. Similar to the standard IA algorithms, we assume that all precoders and combiners are constructed with the knowledge of unified CSI mismatch. In this case, by replacing $\mathbf{H}_{k,j}$ with the RHS of (5.6), the covariance matrix $\mathbf{Q}_k^{\ell}$ in Step 2 of Algorithm 2 can be rewritten as

$$\begin{aligned}\mathbf{Q}_k^{\ell} &= \sum_{j=1}^{k-1}\sum_{n=1}^{d} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j}+\breve{\mathbf{H}}_{k,j}\right)\widehat{\mathbf{v}}_{jn}\widehat{\mathbf{v}}_{jn}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,j}+\breve{\mathbf{H}}_{k,j}\right)^{\mathsf{H}} \\ &+ \sum_{m=1}^{\ell-1} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}}+\sigma^2\mathbf{I} \quad (6.2)\end{aligned}$$

Now by taking the expectation over the redundant channel measurement error $\breve{\mathbf{H}}_{k,j}$ and by considering lemma 5.3 and lemma 5.4 on page 90, we can approximate $\mathbf{Q}_k^{\ell}$ in (6.2) with a simpler form, i.e., $\widehat{\mathbf{Q}}_k^{\ell}$ as follows:

$$\begin{aligned}\widehat{\mathbf{Q}}_k^{\ell} &= \sum_{j=1}^{k-1}\sum_{n=1}^{d}\frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jn}\widehat{\mathbf{v}}_{jn}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\sum_{m=1}^{\ell-1}\frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \\ &+ \left(P\left(kd+\ell-d-1\right)\frac{\tau}{1+\tau}+\sigma^2\right)\mathbf{I} \quad (6.3)\end{aligned}$$

Similarly, by replacing $\mathbf{H}_{k,j}$ with the RHS of (5.6), the covariance matrix $\mathbf{R}_k^\ell$ in Step 4 of Algorithm 2 can be rewritten as

$$\begin{aligned}\mathbf{R}_k^\ell &= \sum_{j=k+1}^{K}\sum_{n=1}^{d} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{j,k}+\breve{\mathbf{H}}_{j,k}\right)^{\mathsf{H}}\widehat{\mathbf{u}}_{jn}\widehat{\mathbf{u}}_{jn}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{j,k}+\breve{\mathbf{H}}_{j,k}\right)\\ &+ \sum_{m=\ell+1}^{d} P\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}}\widehat{\mathbf{u}}_{km}\widehat{\mathbf{u}}_{km}^{\mathsf{H}}\left(\frac{1}{1+\tau}\widehat{\mathbf{H}}_{k,k}+\breve{\mathbf{H}}_{k,k}\right)+\sigma^2\mathbf{I} \quad (6.4)\end{aligned}$$

Now by taking the expectation over the redundant channel measurement error $\breve{\mathbf{H}}_{k,j}$ and by considering lemma 5.3 and lemma 5.4 on page 90, we can approximate $\mathbf{R}_k^\ell$ in (6.4) with a simpler form, i.e., $\widehat{\mathbf{R}}_k^\ell$ as follows:

$$\begin{aligned}\widehat{\mathbf{R}}_k^\ell &= \sum_{j=k+1}^{K}\sum_{n=1}^{d}\frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{jn}\widehat{\mathbf{u}}_{jn}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}+\sum_{m=\ell+1}^{d}\frac{P}{(1+\tau)^2}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{km}\widehat{\mathbf{u}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}\\ &+ \left(P\Big((K-k+1)\,d-\ell\Big)\frac{\tau}{1+\tau}+\sigma^2\right)\mathbf{I} \quad (6.5)\end{aligned}$$

Therefore, by considering (6.3) and (6.5), the adaptive modification of Algorithm 2 (which hereafter we call it Algorithm 4) is illustrated as follows:

Algorithm 4

1. Initialize random unit-norm vectors $\mathbf{v}_{k\ell},\ \forall\, k,\ \ell$
2. Set $\varepsilon := \gamma^{-1}\left(1+\tau\right)^2+\tau\left(\tau+1\right)\left(kd+\ell-d-1\right)$
3. $\mathbf{Q}_k^\ell = \sum_{j=1}^{k-1}\sum_{n=1}^{d}\widehat{\mathbf{H}}_{k,j}\widehat{\mathbf{v}}_{jn}\widehat{\mathbf{v}}_{jn}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\sum_{m=1}^{\ell-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{km}\widehat{\mathbf{v}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}+\varepsilon\,\mathbf{I}$
4. $\mathbf{u}_{k\ell} = \dfrac{\left(\mathbf{Q}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}}{\left\|\left(\mathbf{Q}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}\widehat{\mathbf{v}}_{k\ell}\right\|_2}$
5. Set $\mu := \gamma^{-1}\left(1+\tau\right)^2+\tau\left(\tau+1\right)\Big(\left(K-k+1\right)d-\ell\Big)$
6. $\mathbf{R}_k^\ell = \sum_{j=k+1}^{K}\sum_{n=1}^{d}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{jn}\widehat{\mathbf{u}}_{jn}^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k}+\sum_{m=\ell+1}^{d}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{km}\widehat{\mathbf{u}}_{km}^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}+\mu\,\mathbf{I}$
7. $\mathbf{v}_{k\ell} = \dfrac{\left(\mathbf{R}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{k\ell}}{\left\|\left(\mathbf{R}_k^\ell\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\widehat{\mathbf{u}}_{k\ell}\right\|_2}$
8. Go to Step 2 and repeat

Remark 6.3: Analogous to the discussions in remark 6.2 regarding the orthonormality of the beamformers of Algorithm 2, Algorithm 4 does not yield unitary beamformers. Therefore, to make precoding and receive combining matrices unitary, we can insert orthogonalization steps after obtaining $\mathbf{U}_k$ and $\mathbf{V}_k$, which ensure unitary precoding and receive combining matrices without incurring any performance degradation in total DoF and achievable throughput at high SNR ranges. However, if the intention is performance analysis at low-to-intermediate SNRs, there is no need for additional orthogonalization steps.

Remark 6.4: Unlike Algorithm 2, Algorithm 4 is able to adaptively design beamformers based on the knowledge of the error variance in advance. As observed, by setting $\tau = 0$, Algorithm 4 boils down to Algorithm 2 under imperfect CSI. If we further replace $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$, Algorithm 4 becomes Algorithm 2 under perfect CSI. In other words, Algorithm 4 is a generalized version of Algorithm 2.

6.2.3 Signal Postprocessing at Receive Nodes

If we define $\mathbf{V}$, $\mathbf{U}$ and $\mathbf{H}$ as (5.7)–(5.9), for standard IA, at high enough SNRs and with the assumption of perfect CSI, $\mathbf{U}^{\mathsf{H}}\mathbf{H}\mathbf{V}$ yields a matrix in similar from as in (5.10). However, if $\{\mathbf{V}_j\}_{j=1}^{K}$ and $\{\mathbf{U}_j\}_{j=1}^{K}$ are obtained by one of the proposed Algorithms 1 and 2, then by considering (6.1), the effective channel can be described by

$$\mathbf{U}^{\mathsf{H}}\mathbf{H}\mathbf{V} = \begin{pmatrix} \overline{\mathbf{H}}_{1,1} & \overline{\mathbf{H}}_{1,2} & \cdots & \overline{\mathbf{H}}_{1,K} \\ \mathbf{0} & \overline{\mathbf{H}}_{2,2} & \cdots & \overline{\mathbf{H}}_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \overline{\mathbf{H}}_{K,K} \end{pmatrix} \tag{6.6}$$

such that $\overline{\mathbf{H}}_{k,j} = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{V}_j$. Therefore, unlike standard IA algorithms wherein by premultiplying the received signal at each receiver by the related receive combining matrix, all the interferences are suppressed (as shown in (5.11)), for the proposed algorithms and at the kth receive node, the interferences of just transmitters 1 to $k-1$ are subject to elimination, as can be deduced from (6.1). For example, for the kth node the received signal can be expressed as (5.1). Then after premultiplying $\mathbf{y}_k$ by $\mathbf{U}_k^{\mathsf{H}}$ obtained by one of

Algorithms 1 and 2, we have

$$\begin{aligned}\mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k=&\underbrace{\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{k-1}\mathbf{H}_{k,j}\mathbf{x}_j}_{\mathbf{0}}+\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{x}_k+\mathbf{U}_k^{\mathsf{H}}\sum_{j=k+1}^{K}\mathbf{H}_{k,j}\mathbf{x}_j+\mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k\\ =&\overline{\mathbf{H}}_{k,k}\mathbf{c}_k+\sum_{j=k+1}^{K}\overline{\mathbf{H}}_{k,j}\mathbf{c}_j+\overline{\mathbf{z}}_k\end{aligned} \tag{6.7}$$

In this case, first the Kth receive node tries to recover the transmitted symbol vector $\mathbf{c}_K$, then the $(K-1)$th node tries to recover $\mathbf{c}_{K-1}$ by first subtracting $\widehat{\mathbf{c}}_K$ from its received signal, and so on.

More specifically, first we arbitrarily index the nodes and obtain the precoding and receive combining matrices accordingly. Then by this assumption that each receive node uses LZF filtering to recover the transmitted data, the detection process can be summarized as follows

$$\begin{aligned}\widehat{\mathbf{c}}_K &= \mathcal{F}\left(\text{pinv}\left(\overline{\mathbf{H}}_{K,K}\right)\overline{\mathbf{y}}_K\right)\\ \widehat{\mathbf{c}}_{K-1} &= \mathcal{F}\left(\text{pinv}\left(\overline{\mathbf{H}}_{K-1,K-1}\right)\left[\overline{\mathbf{y}}_{K-1}-\overline{\mathbf{H}}_{K-1,K}\widehat{\mathbf{c}}_K\right]\right)\\ &\vdots\\ \widehat{\mathbf{c}}_k &= \mathcal{F}\left(\text{pinv}\left(\overline{\mathbf{H}}_{k,k}\right)\left[\overline{\mathbf{y}}_k-\sum_{j=k+1}^{K}\overline{\mathbf{H}}_{k,j}\widehat{\mathbf{c}}_j\right]\right)\end{aligned} \tag{6.8}$$

where $\overline{\mathbf{y}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k$ such that $\mathbf{y}_k$ is defined in (5.1), $\text{pinv}\left(\overline{\mathbf{H}}_{k,k}\right) = \left(\overline{\mathbf{H}}_{k,k}^{\mathsf{H}}\overline{\mathbf{H}}_{k,k}\right)^{-1}\overline{\mathbf{H}}_{k,k}^{\mathsf{H}}$, $\mathcal{F}(\cdot)$ denotes a demodulating operator like QAM demodulator, and $\widehat{\mathbf{c}}_k$ represents the decoded data at receive node k, which can be erroneous.

6.3 Discussions

6.3.1 CSI Availability

Depending on the algorithm, all IA schemes need certain amount of CSI to be available. In a centralized processing, the total CSI is available at an access point (AP) whereas in a distributed processing, each node must have access to certain amount of CSI. Without loss of generality, we consider distributed processing. While some standard IA techniques like

the ones in [17, 25] are based on the global channel knowledge at each node, some other IA techniques like the ones in [23] require only local CSI at each node (each receiver only needs to know the direct channel to its desired transmitter, i.e., $\mathbf{H}_{k,k}$) and the covariance matrix of the effective noise (consisting of the AWGN and the interferences from all remaining $K-1$ transmitters). However, for the proposed IA techniques in this work, the kth node requires the knowledge of direct link $\mathbf{H}_{k,k}$, the cross CSI of $K-k$ nodes, i.e., $\mathbf{H}_{k,j}, j = k+1, \ldots, K$, and the covariance matrix of the effective noise (consisting of the AWGN and the interferences of just $k-1$ transmit nodes). This implies that on average, half of the total cross CSI is needed to be shared by receive nodes. Consequently, from the CSI availability perspective, the proposed IA algorithms give a compromise between the ones in [17, 25] and those of [23, 24].

6.3.2 Convergence of Algorithms

The methodology introduced above might suggest that the convergence of the proposed algorithms is not guaranteed. Therefore, in this subsection and with respect to the fact that the feasibility conditions are met, we address this issue.

The proof of convergence of Algorithm 1 is trivial since transmitters and receivers take turns to adjust the beamforming vectors to reduce interference leakage under the assumption of channel reciprocity. Therefore Algorithm 1 monotonically reduces the total interference leakage and consequently converges. This causes Algorithm 1 to be similar to Min-WLI. However for Algorithm 2, the situation is different. Note that Algorithm 2 and Max-SINR are similar in the sense that both of these algorithms consider the effect of noise for aligning interference. Although in simulation results of Algorithm 2 (and also Max-SINR), the sum rate appears to converge, unfortunately due to the complexity of the sum rate and MMSE expressions, it is not possible to prove the convergence mathematically. However, it is easy to show that Algorithm 2 (Algorithm 1) converges faster than Max-SINR (Min-WLI) since in Max-SINR (Min-WLI), both the upper and the lower triangular parts of the equivalent channel matrix must be transformed into blocks of zeros, as shown in (5.10), whereas for Algorithms 2 (Algorithm 1) just the lower triangular part is needed to be made zero, as revealed in (6.6). Consequently Algorithm 2 (Algorithm 1) needs fewer constraints to meet in comparison with Max-SINR (Min-WLI). Therefore if Max-SINR is supposed to converge numerically, the situation for Algorithm 2 is more relaxed. In fact, in numerical simulations, it has been observed that for the same number of iterations, the percentage of the leaked interference of Algorithm 2 is less than that of Max-SINR, which

validates the above discussion.

It is also worthwhile to mention that regardless of the convergence speed of IA algorithms which can be affected by the initial point of precoders/combiners, they are able to converge to a solution.

6.4 Feasibility Conditions

The feasibility condition of standard IA techniques has been first investigated in [20] based on counting the number of variables (i.e., the number of nonsuperfluous elements of subspaces spanned by precoders and combiners) and number of equations which satisfy the ZF condition in (5.2). It has been shown that a symmetric system is proper if and only if the number of equations is not larger than the number of variables which leads to $M + N \geq d(K + 1)$. Later [21] solidified the condition in [20] where it has been shown that if M and N divide d, the lower bound on the aggregate number of antennas is tight, and the proper system becomes feasible, i.e., $M + N = d(K + 1)$ implies a feasible system. However, since the configuration considered in this work is different from that of the standard IA methods, therefore following the same approach of [21] to derive the feasibility conditions of the proposed algorithms is not straightforward and is generally inconclusive. Consequently, we use a similar approach to [20] to evaluate whether the system is proper and obtain the necessary conditions of feasibility; that is, the total number of variables should not be less than the total number of equations. It is worthwhile to note that this way of deriving the necessary conditions of feasibility based on counting the number of variables and equations is popular in an IA-based context (e.g., [20, 77, 78]). However, a proper system does not always guarantee IA feasibility. Since acquiring the sufficient conditions of feasibility is not as trivial as the necessary conditions, analogous to the same approach in [20, 23], we further measure the percentage of the leaked interference, i.e., the fraction of the interference power that is existent in the dimensions reserved for the desired signal. If this percentage is equal to zero, the considered system becomes plainly feasible and sufficient conditions of feasibility are met. Note that the interference percentage at the kth receive node is directly related to

$$p_k = \frac{\sum_{i=1}^{d} \lambda_i \left[\mathbf{Q}_k\right]}{\mathrm{Tr}\left[\mathbf{Q}_k\right]} \tag{6.9}$$

where λ_i is the ith smallest eigenvalue of $\mathbf{Q}_k$, and $\mathbf{Q}_k$ represents the scaled interference covariance matrix at receive node k which can be shown as

$$\mathbf{Q}_k = \sum_{j=1}^{k-1} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} \tag{6.10}$$

Hence, before meeting the sufficient conditions of feasibility by measuring the percentage of the leaked interference, we obtain the necessary conditions of feasibility by counting the number of variables and equations. Let N_v and N_e, respectively, denote the number of variables and the number of equations. The necessary condition for this system to be underdetermined (and consequently proper) is that the number of equations does not exceed the number of variables. This underlines the procedure being discussed in this section.

We reexpress the condition of (6.1) as

$$\begin{aligned} \mathbf{u}_{km}^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{v}_{jn} &= 0 \\ \text{s.t.}: \quad & k = 1, \ldots, K \\ & j \in \{1, \cdots, k-1\} \\ & m, n = 1, \ldots, d \end{aligned} \tag{6.11}$$

Therefore, we can obtain N_e from (6.11) as

$$N_e = \sum_{k=1}^{K}\sum_{j=1}^{k-1} d^2 = \frac{K\,(K-1)}{2}d^2 \tag{6.12}$$

Nevertheless, counting the number of variables is less straightforward, since there are some superfluous variables that should not be taken into account. We consider the following two criteria to calculate N_v:

1. As evident from Algorithm 1, in each iteration, to obtain receive combining matrices (or precoding matrices in the reverse iteration), the number of variables is confined to $K-1$ receive combining (or precoding) matrices. For example, as seen in Steps 2–7 of Algorithm 1, we pick up $\mathbf{U}_1$ as an arbitrarily unitary matrix and the rest of $\mathbf{U}_k,\ k = 2, \ldots, K$ are computed based on the covariance of the interference in Step 5. Note that although we choose $\mathbf{U}_1 = \mathrm{orth}\,(\mathbf{H}_{1,1}\mathbf{V}_1)$, but this is just a desired selection for $\mathbf{U}_1$ since as we mentioned, $\mathbf{U}_1$ can arbitrarily be selected and there is

no constraint on $\mathbf{U}_1$ except being unitary. Therefore the number of variables related to $\mathbf{U}_1$ (and also similarly $\mathbf{V}_K$ in the reverse iteration) should not be considered.

2. Now we consider the number of variables confined to each precoding or receive combining matrix. Let us first consider the precoding matrix $\mathbf{V}_k \in \mathbb{C}^{N\times d}$ which forms a basis for the transmitted signal space of transmitter k. If we postmultiply $\mathbf{V}_k$ by an invertible matrix, this does not change the space spanned by the columns of $\mathbf{V}_k$. We partition matrix $\mathbf{V}_k$ as

$$\mathbf{V}_k = \begin{pmatrix} \dot{\mathbf{V}}_k \\ \ddot{\mathbf{V}}_k \end{pmatrix} \tag{6.13}$$

where $\dot{\mathbf{V}}_k \in \mathbb{C}^{d\times d}$ and $\ddot{\mathbf{V}}_k \in \mathbb{C}^{(N-d)\times d}$. In this case, we have

$$\mathrm{span}\left(\mathbf{V}_k\right) = \mathrm{span}\left(\mathbf{V}_k \dot{\mathbf{V}}_k^{-1}\right) \tag{6.14}$$

where $\mathbf{A} = \mathrm{span}\left(\mathbf{V}_k\right)$ means that $\mathbf{A}$ spans the same space as the columns of $\mathbf{V}_k$. If we further define

$$\overline{\mathbf{V}}_k = \mathbf{V}_k \dot{\mathbf{V}}_k^{-1} = \begin{pmatrix} \mathbf{I}_{d\times d} \\ \ddot{\mathbf{V}}_k \dot{\mathbf{V}}_k^{-1} \end{pmatrix} \tag{6.15}$$

where $\mathbf{I}_{d\times d}$ represents the identity matrix of size $d \times d$, the transmit signal space is now spanned by $\overline{\mathbf{V}}_k$. It is easy to argue that there is no other basis representation for the kth transmitted signal space with fewer variables.

Consequently, by removing superfluous variables, the number of remaining variables related to the kth precoding matrix $\overline{\mathbf{V}}_k$ is equal to $d\left(N-d\right)$. Similarly, the number of nonsuperfluous variables that should be taken into account for receive combining matrix $\overline{\mathbf{U}}_k$ is $d\left(M-d\right)$.

As a result of the aforementioned criteria, the total number of variables in the network is equal to

$$N_v = \sum_{k=1}^{K-1} d\left(M+N-2d\right) = d\left(K-1\right)\left(M+N-2d\right) \tag{6.16}$$

Now we can discuss that the necessary condition of feasibility is $N_v \geq N_e$, which further results in

$$\begin{aligned} N_v \geq N_e \quad &\Rightarrow \quad d\left(K-1\right)\left(M+N-2d\right) \geq \frac{K\left(K-1\right)}{2} d^2 \\ &\Rightarrow \quad M+N \geq \frac{d(K+4)}{2} \end{aligned} \tag{6.17}$$

Corollary 6.1: For standard IA techniques, $\min(M+N)$ is equal to $d(K+1)$, while for the proposed algorithms, this amount is equal to $\frac{d(K+4)}{2}$. This indicates that to achieve the same number of DoF, the proposed IA algorithms require less number of antennas than standard IA techniques. In other words, for asymptotically large K, the required number of antennas for the proposed IA schemes can be half of that of standard IA techniques to secure the same number of DoF.

6.5 Numerical Results

In this section and by using numerical results, we support the analysis related to the feasibility conditions of the proposed algorithms in the previous section. Although there are several well-known IA algorithms is the literature (see e.g., [23, 24]), we compare the performance of the proposed Algorithms 1, 2, and 4 with Min-WLI and Max-SINR. This is because both Algorithm 1 and Min-WLI are based on just interference leakage minimization without considering the effect of direct links and noise whereas Algorithms 2, 4, and Max-SINR consider the effect of direct links and noise for beamformer design. It is also worth mentioning that under perfect CSI, Algorithm 4 boils down to Algorithm 2. Moreover, we do not consider other standard IA techniques since most of them are closely related to Min-WLI or Max-SINR. For the case of imperfect CSI, we assume that the channel estimation error variance obeys (3.3), i.e., $\tau = \beta\gamma^{-\alpha}$.

In Figs. 6.2–6.3, we consider beamformer design under perfect CSI with no error propagation at receive side whereas in Figs. 6.4–6.5, we assume beamformer design under imperfect CSI and with error propagation at receive side.

In Fig. 6.2, we depict the average sum rate for the case $K = 3$ and $d = 4$ such that each receive node uses optimum filtering to recover the transmitted data. For the standard IA schemes and based on discussions in [20], the aggregate number of transmit and receive antennas must satisfy $M + N \geq 16$. Based on discussions in [21] indicating that if M and N divide d, the lower bound is tight and the proper system becomes plainly feasible, we choose $M = N = 8$. For the proposed Algorithms 1 and 2 and based on (6.17), the aggregate number of transmit and receive antennas must satisfy $M + N \geq 14$. Consequently, we choose $M = 6, N = 8$ (or $M = 8, N = 6$). We also measured the percentage of the leaked interference in the desired signal subspaces based on (6.9), where with this choice of number of transmit and receive antennas, it turns out to be zero, which further implies the feasibility of the proposed algorithms.

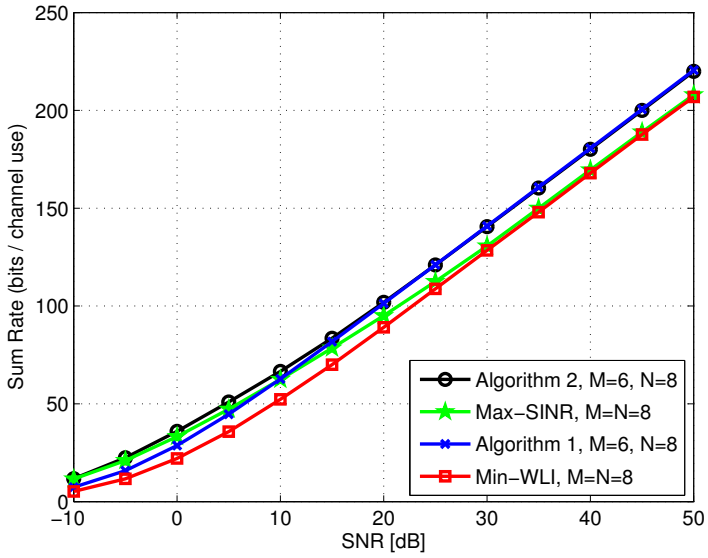

Figure 6.2: *Average sum rate for $K = 3$ and $d = 4$ with the assumption of perfect CSI and no error propagation. Each receive node uses optimum filtering to recover the transmitted symbols.*

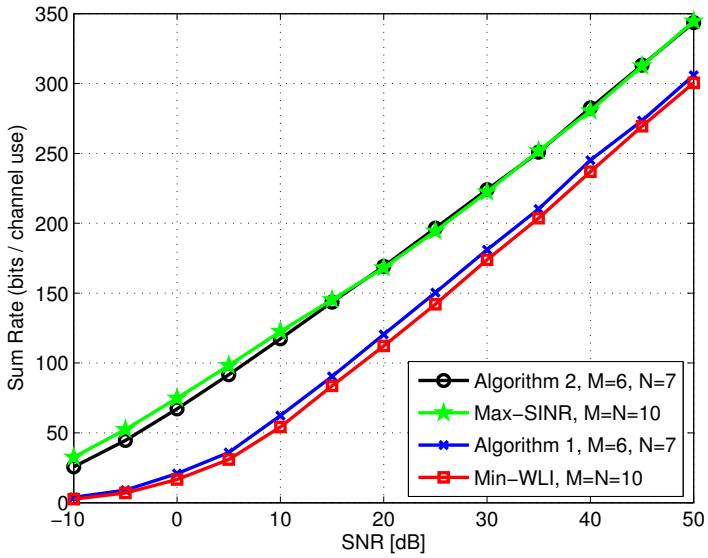

Figure 6.3: *Average sum rate for $K = 19$ and $d = 1$ with the assumption of perfect CSI and no error propagation. Each receive node uses optimum filtering to recover the transmitted symbols.*

Fig. 6.2 reflects the two important features of the proposed IA schemes: first, by using the proposed algorithms, we can decrease the number of transmit or receive antennas associated with each transceiver pair by two, to achieve the same number of DoF as standard IA schemes; second, even with this reduced number of antennas, the proposed Algorithms

1 and 2 achieve better throughput than Min-WLI and Max-SINR, respectively.

In Fig. 6.3, we illustrate the average sum rate for the case $K = 19$ and $d = 1$, and with considering optimum filtering for each decoupled subchannel. For the standard IA schemes and based on discussions in [21], the aggregate number of transmit and receive antennas must satisfy $M + N \geq 20$. Therefore, we choose $M = N = 10$. For the proposed algorithms and based on (6.17), the aggregate number of transmit and receive antennas must satisfy $M + N \geq 12$. However, as we measured the percentage of leaked interference in each desired signal subspace based on (6.9), it turns out that the percentage of the leaked interference is not zero which further implies that the proposed algorithms are not feasible when $M + N = 12$. However, with $M = 6, N = 7$ (or $M = 7, N = 6$) implying that $M + N = 13$, this percentage is equal to zero and the proposed algorithms become feasible.

Fig. 6.3 indicates that by using the proposed schemes for the case $K = 19$ and $d = 1$, we can decrease the number of transmit and receive antennas associated with each transceiver pair by $20 - 13 = 7$ to achieve the same number of DoF as standard IA schemes. This means that since $K = 19$, we have a total decrease of 133 antennas in the entire network compared to standard IA schemes. Moreover as shown, even with this reduced number of antennas, the achievable throughput of the proposed Algorithms 1 and 2 is still comparable to that of Min-WLI and Max-SINR, respectively.

So far, in Figs 6.2–6.3, we evaluated the performance of Algorithms 1–2 under perfect CSI and with no error propagation. It is however desirable to assess their performance under imperfect CSI and with error propagation. In Fig. 6.4, we evaluate the performance of Algorithms 1–2 and also the adaptive design in Algorithm 4 under the CSI mismatch model discussed in section 5.2.2. Note that based on remark 6.4, under perfect CSI, Algorithm 4 boils down to Algorithm 2. We consider two different cases for the imperfect CSI: $\beta = 10,\ \alpha = 1$ and $\beta = 0.05,\ \alpha = 0$. We further assume that we have $K = 4$ pairs of users each communicating with its corresponding receiver with 2 DoF such that each transmitted block consists of 100 QPSK symbols. For this case, and based on (6.17), the minimum aggregate number of transmit and receive antennas is equal to $M + N \geq 8$ whereas for the standard IA techniques the minimum aggregate number of transmit and receive antennas is equal to 10. By setting $M = N = 4$, the interference leakage percentage in (6.9) becomes zero and the feasibility conditions are met. We further assume that each receive node has access to perfect direct link, and the effect of error propagation has been considered as well. As revealed, the proposed adaptive design, i.e., Algorithm 4, enables us to glean more gains under imperfect CSI compared to Algorithms 1 and 2, e.g.,

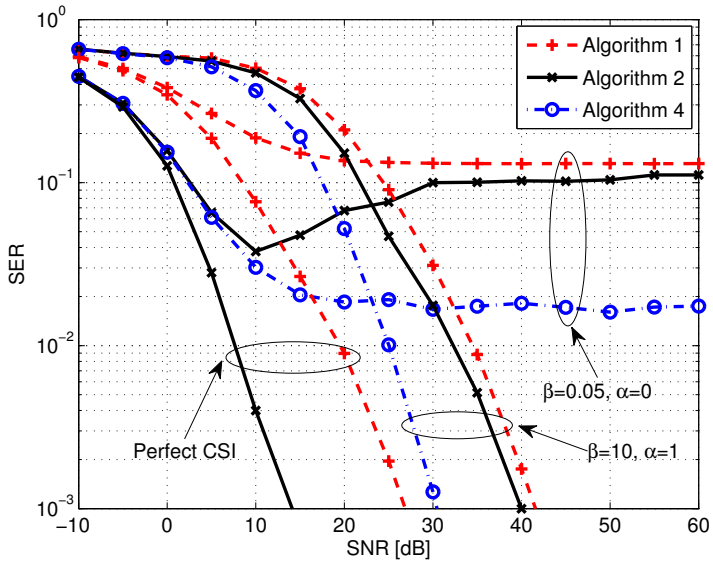

Figure 6.4: *Average SER for $K = 4,\ d = 2$ and $M = N = 4$ under perfect and imperfect CSI with error propagation. Each receive node uses LZF filtering to recover the transmitted symbols and each transmitted block consists of 100 QPSK symbols*

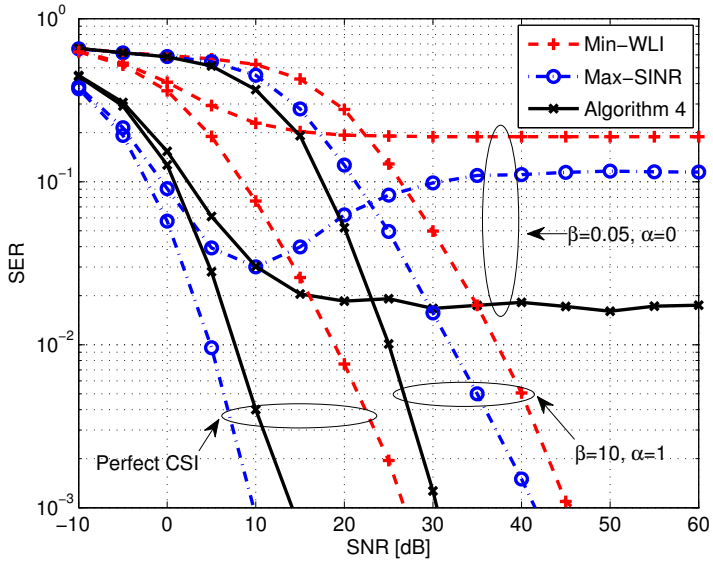

Figure 6.5: *Average SER for $K = 4,\ d = 2$. For Algorithm 4, $M = N = 4$ whereas for Max-SINR and Min-WLI algorithms, $M = 4,\ N = 6$. The plots are depicted under perfect and imperfect CSI while Algorithm 4 experiences error propagation. Each receive node uses LZF filtering to recover the transmitted symbols and each transmitted block consists of 100 QPSK symbols.*

to achieve the SER of 10^{-3}, Algorithm 4 enables us to save 10 and 12 dB gains compared to Algorithms 2 and 1, respectively.

In Fig. 6.4, we showed that Algorithm 4 outperforms Algorithm 1 and Algorithm 2 under imperfect CSI. Fig. 6.5 illustrates the average SER of Algorithm 4 in comparison with standard IA techniques, i.e., Min-WLI and Max-SINR, under imperfect CSI. With respect to the fact that Algorithm 4 boils down to Algorithm 2 under perfect CSI, the plots corresponding to this case are also depicted for comparison. Note that while for Algorithm 4 we have $M = N = 4$, for Max-SINR and Min-WLI algorithms, we have $M = 4,\ N = 6$. As revealed, although Algorithm 4 experiences error propagation, it can still outperform Max-SINR and Min-WLI under imperfect CSI. For example, when $\beta = 10,\ \alpha = 1$, Algorithm 4 respectively achieves 11 dB and 15 dB gain compared to Min-WLI and Max-SINR to reach the SER of 10^{-3}. Also for the case $\beta = 0.05,\ \alpha = 0$, Algorithm 4 decreases the SER by a factor of nearly $\frac{1}{10}$ compared to Min-WLI and Max-SINR at SNRs of larger than 20 dB.

6.6 Summary

In wireless interference networks, to attain full DoF using IA, the aggregate number of transmit and receive antennas at each transceiver pair must satisfy $M + N \geq d\,(K + 1)$. In this chapter, by relying on partially coordinated reception where averagely half of the total decoded data (which can be erroneous) are shared by receive nodes, we proposed IA algorithms such that their feasibility condition can now be expressed as $M+N \geq d\left(\frac{K+4}{2}\right)$. This implies less required number of antennas to secure the same number of DoF compared to standard IA techniques. Also by using numerical simulations, we showed that even with this reduced number of antennas, the achievable throughput of the proposed algorithms is still comparable to that of standard IA schemes. Moreover, we also proposed an adaptive IA design under the availability of imperfect CSI. By relying on the knowledge of error variance in advance, the proposed approach is able to notably improve the achievable performance compared to standard IA techniques subject to CSI mismatch. In a nutshell, even by considering imperfect CSI and error propagation, the proposed algorithms are able to achieve comparable performance to standard IA techniques while enabling us to decrease the aggregate number of transmit and receive antennas.

Chapter 7

LS and MMSE based Beamformer Design for IC

7.1 Introduction

In chapter 5, we analyzed the performance of constant MIMO IA under CSI mismatch wherein we derived novel bounds regarding the asymptotic mean loss in sum rate and achievable DoF which are generally applicable to any IA scheme. Also in chapter 6, we employed the idea of partially coordinated reception to design IA algorithms which enable us to decrease the number of deployed antennas at transmit and/or receive nodes in order to secure the same number of DoF compared to standard IA techniques.

Although there are various IA algorithms in the literature [23–34], in this chapter, we place our focus on designing two novel IA algorithms based on LS and MMSE criteria which hereafter are referred as LS and MMSE based designs for multiuser MIMO IC. Unlike standard IA methods which are primarily designed based on the availability of perfect CSI, the optimization criteria of the proposed algorithms are set up based on the availability of only imperfect channel estimation. This makes the proposed algorithms be adaptive in a sense that we are allowed to design beamformers based on the knowledge of channel estimation error variance. In this case, the beamformer design under perfect CSI becomes a special scenario when the error variance is set to zero. This makes the proposed schemes more general than standard IA techniques.

Most of the previously proposed IA algorithms are closely related to either Min-WLI or Max-SINR algorithms. For example, the Alt-Min algorithm defined in [34] is algorithmically identical to Min-WLI, or the joint signal and interference alignment algorithm

in [28] is a modified version of Alt-Min algorithm in [34]. Also the MMSE IA in [29] is a special case of Max-SINR and the weighted MMSE IA defined in [24]. Therefore, without loss of generality, in this chapter we compare the performance of the proposed algorithms with that of Min-WLI and Max-SINR algorithms.

It is shown that the proposed LS based design is able to outperform interference leakage minimization algorithms under both perfect and imperfect CSI. This is due to the fact that unlike Min-WLI and Alt-Min algorithms which do not consider the direct links to design the beamformers, the proposed LS based scheme does so.

MMSE IA was first introduced in [29] for single-stream-per-user transmission, and then generalized to the case of multi-stream-per-user communication (see e.g., [24]). In this chapter, however, we propose a novel MMSE based IA which yields unitary precoders and combiners. Since the proposed approach is implicitly built on MMSE to design beamformers, it may be slightly suboptimal compared to the weighted MMSE IA techniques like the ones in [24,29]. However, note that the precoders obtained by the weighted MMSE IA are not unitary, and they further need an extra optimization step to meet the power constraint. Moreover, this power-constraint optimization step has no closed-form solution and has to be done numerically within each iteration. Nevertheless, the beamformers obtained by the proposed MMSE based scheme are unitary, and they do not require such an extra power-constraint optimization step, which consequently results in much simpler implementation.

Compared to Max-SINR, while the proposed MMSE based design is a user-by-user approach, the former is a stream-by-stream approach. Consequently, the proposed MMSE based IA needs less CSI to be available compared to Max-SINR in order to calculate the beamformers. Plus, the former possesses less computational complexity compared to the latter. We also prove that the proposed MMSE based design achieves the same performance as Max-SINR under perfect CSI. Subject to imperfect CSI, however, the former outperforms the latter. Using simulation results, we demonstrate the improved performance achieved by the proposed algorithms compared to standard IA techniques.

Moreover, it is also shown that the proposed LS based design results in diagonalized subchannels for all SNR ranges, under perfect CSI. In other words, after premultiplying the received signal by the corresponding combiner, the interferences are first suppressed and the desired subchannels are then diagonalized. This is in contrast to the previously proposed IA schemes like interference leakage minimization algorithms wherein the resulted subchannels are full matrices. In this case and in order to use waterfilling to achieve higher sum rates, while standard IA schemes like Min-WLI and Alt-Min need to use sin-

gular value decomposition (SVD), the proposed LS based design dissolve the need of such decomposition since the decoupled subchannels have been already diagonalized. It is also worthwhile to point out that since at sufficiently high SNRs, MMSE based IA boils down to LS based IA, the resulted subchannels of MMSE based IA are also diagonalized at high enough SNRs.

This chapter is organized as follows: in Sections 7.2 and 7.3, respectively, we propose LS and MMSE based IA algorithms by setting up the optimization problem based on the knowledge of imperfect CSI and channel estimation error variance. Section 7.4 contains discussions regarding the proposed algorithms. In Section 7.5, we use numerical results to show that the proposed schemes outperform standard IA techniques under both perfect and imperfect CSI. Finally the chapter ends up with a summary in Section 7.6.

7.2 LS based IA

In this section, we propose an optimized beamformer design for MIMO IC based on LS criterion and the knowledge of imperfect CSI. The system model under both perfect and imperfect CSI is similar to those in subsections 5.2.1 and 5.2.2, respectively. Without loss of generality, we consider a symmetric interference channel where there are K pairs of transceivers such that each N-antenna transmitter communicates with its corresponding M-antenna receiver by sending d independent data streams as illustrated in Fig. 5.1. In this case, the received signal at receive node k can be expressed as

$$\mathbf{y}_k = \mathbf{H}_{k,k}\mathbf{x}_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{z}_k \tag{7.1}$$

where the notations $\mathbf{y}_k$, $\mathbf{x}_k = \mathbf{V}_k\mathbf{c}_k$, $\mathbb{E}\left\{\mathbf{c}_k\mathbf{c}_k^{\mathsf{H}}\right\} = P\mathbf{I}$, $\mathrm{vec}\left(\mathbf{H}_{k,j}\right) \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\mathbf{I}\right)$, and $\mathbf{z}_k \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0},\sigma^2\mathbf{I}\right)$ are consistent with those defined in subsection 5.2.1. $\gamma = P/\sigma^2$ is also considered as the nominal SNR. Therefore, the kth receive node premultiplies the received signal by $\mathbf{U}_k^{\mathsf{H}}$ which yields

$$\begin{aligned} \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k &= \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{x}_k + \underbrace{\mathbf{U}_k^{\mathsf{H}}\sum_{\substack{j=1 \\ j\neq k}}^{K}\mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k}_{\overline{\mathbf{z}}_k} \\ \Longrightarrow\ \overline{\mathbf{y}}_k &= \overline{\mathbf{H}}_{k,k}\mathbf{c}_k + \overline{\mathbf{z}}_k \end{aligned} \tag{7.2}$$

where $\overline{\mathbf{y}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k$, $\overline{\mathbf{H}}_{k,k} = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$, and $\mathbf{V}_k$ and $\mathbf{U}_k$ are unitary precoders and combiners, respectively. Due to (5.2), at high enough SNRs, we have $\overline{\mathbf{z}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k$. Thus, the data vector $\mathbf{c}_k$ can be readily recovered via premultiplying $\overline{\mathbf{y}}_k$ by (pseudo-) inverse of $\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$.

As discussed in earlier chapters, due to the coupled nature of the beamformer design for multiuser MIMO IC, there are no closed form solutions for IA, except for a few particular cases. Consequently, finding precoders and combiners requests an iterative procedure in general. Therefore, first the precoders get fixed and the combiners are sought, and then the combiners get fixed and the precoders are sought through an iterative manner.

For the proposed LS based design and without loss of generality, we consider the adaptive design under imperfect CSI as the major optimization problem, and as it will be shown later, the standard design under perfect CSI is a special case of the adaptive design by setting error variance equal to zero. In order to calculate the beamformers adaptively, we assume that the variance of channel estimation error is known in advance. Given randomly initialized precoders and with respect to (7.2), the optimization criterion based on LS design can be considered as

$$\arg\min_{\mathbf{U}_k} \mathbb{E}\left\{\left\|\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j - \mathbf{c}_k\right\|_2^2\right\} \quad \forall k = 1,\ldots,K \tag{7.3}$$

Now the objective function of the optimization problem in (7.3) can be defined as

$$\dot{F}_{\text{LS}} = \mathbb{E}\left\{\text{Tr}\left[\left(\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j - \mathbf{c}_k\right)\left(\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j - \mathbf{c}_k\right)^{\mathsf{H}}\right]\right\} \tag{7.4}$$

To further continue, we consider the following lemma:

Lemma 7.1: We assume that the transmitted data vector $\mathbf{c}_k$ consists of i.i.d. symbols, i.e., $\mathbb{E}\left\{\mathbf{c}_k\mathbf{c}_j^{\mathsf{H}}\right\} = \mathbf{0},\ j \neq k$, and also the noise vector $\mathbf{z}_k$ is independent of the data vector $\mathbf{c}_k$ as well as the channel matrices $\mathbf{H}_{k,j}$.

With respect to lemma 7.1, $\dot{F}_{\text{LS}}$ in (7.4) can be rewritten as

$$\dot{F}_{\text{LS}} = \text{Tr}\left[P\,\mathbf{U}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}\mathbf{U}_k - P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k - P\,\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k\right] + Pd \tag{7.5}$$

To obtain combiners based on minimizing the objective function in (7.5), we differentiate $\dot{F}_{\text{LS}}$ with respect to $\mathbf{U}_k$ by first considering the following assumptions [50, 51]:

1. $\mathbf{U}_k$ and $\mathbf{U}_k^{\mathsf{H}}$ are treated as independent variables.

2. $\dfrac{\partial \operatorname{Tr}[\mathbf{A}\mathbf{U}_k]}{\partial \mathbf{U}_k} = \dfrac{\partial \operatorname{Tr}[\mathbf{U}_k\mathbf{A}]}{\partial \mathbf{U}_k} = \mathbf{A}.$

Therefore, with respect to the two preceding assumptions, we have

$$\frac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k} = P\,\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} \tag{7.6}$$

The imperfect CSI model, that is going to be considered throughout this chapter, is similar to the one that has been discussed in subsection 5.2.2. Therefore, by considering (5.6), equation (7.6) can be further rewritten as

$$\frac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k} = P\,\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j}\right)\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j}\right)^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau} + \breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}} \tag{7.7}$$

Note that $\dfrac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k}$ is now dependent on both $\widehat{\mathbf{H}}_{k,j}$ and $\breve{\mathbf{H}}_{k,j}$. To make $\dfrac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k}$ dependent only on $\widehat{\mathbf{H}}_{k,j}$, we can take the expectation with respect to $\breve{\mathbf{H}}_{k,j}$ by noticing the following two lemmas:

Lemma 7.2:

$$\mathbb{E}_{\breve{\mathbf{H}}|\widehat{\mathbf{H}}}\left\{\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\breve{\mathbf{H}}_{k,j}^{\mathsf{H}}\right\} = \mathbb{E}_{\breve{\mathbf{H}}|\widehat{\mathbf{H}}}\left\{\breve{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}\right\} = \mathbf{0} \tag{7.8}$$

Proof: All precoders and combiners are constructed upon channel estimates $\widehat{\mathbf{H}}_{k,j}$ which based on lemma 3.1 on page 28 are independent of $\breve{\mathbf{H}}_{k,j}$. ■

Lemma 7.3: If $\mathbf{A} \in \mathbb{C}^{M\times N}$ represents a Gaussian matrix with i.i.d. elements of mean zero and variance a, and $\mathbf{B} \in \mathbb{C}^{N\times d}$ refers to a truncated unitary matrix independent of $\mathbf{A}$, then $\mathbb{E}_{\mathbf{A}}\left\{\mathbf{A}\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\right\} = ad\,\mathbf{I}$.

Proof: Since $\mathbf{A}$ is a Gaussian matrix, it is bi-unitarily invariant, and consequently the joint distribution of its entries equals that of $\mathbf{A}\mathbf{B}$ for any truncated unitary matrix $\mathbf{B}$ independent of $\mathbf{A}$ [3]. Therefore $\mathbf{A}\mathbf{B}$ is equivalent to a Gaussian matrix with i.i.d. elements of mean zero and variance a. Since $\mathbf{A}\mathbf{B}$ has d independent columns, $\mathbb{E}_{\mathbf{A}}\left\{\mathbf{A}\mathbf{B}\mathbf{B}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\right\} = ad\,\mathbf{I}$. ■

Following lemmas 7.2 and 7.3 and by taking the expectation of $\dfrac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k}$ over $\breve{\mathbf{H}}_{k,j}$, we

have

$$\mathbb{E}_{\check{\mathbf{H}}_{k,j}}\left\{\frac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k}\right\}=P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\frac{Kd\tau}{1+\tau}\mathbf{I}\right)-\frac{P}{1+\tau}\,\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \tag{7.9}$$

Now the sought combiner $\mathbf{U}_k$ can be obtained by setting $\mathbb{E}_{\check{\mathbf{H}}_{k,j}}\left\{\frac{\partial \dot{F}_{\text{LS}}}{\partial \mathbf{U}_k}\right\}$ equal to zero, which yields

$$P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\frac{Kd\tau}{1+\tau}\mathbf{I}\right)-\frac{P}{1+\tau}\,\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}=\mathbf{0}$$

$$\implies \mathbf{U}_k=(1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\varepsilon_{\text{LS}}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\mathbf{V}_k \tag{7.10}$$

such that $\varepsilon_{\text{LS}}=\tau\,(1+\tau)\,Kd$.

Now we turn our focus to obtain the precoders. Given randomly initialized combiners, the optimization criterion based on LS design can be considered as

$$\arg\min_{\mathbf{V}_k}\ \mathbb{E}\left\{\left\|\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j-\mathbf{c}_k\right\|_2^2\right\}\quad\forall k=1,\ldots,K \tag{7.11}$$

Now the objective function of the optimization problem in (7.11) can be defined as

$$\ddot{F}_{\text{LS}}=\mathbb{E}\left\{\text{Tr}\left[\left(\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j-\mathbf{c}_k\right)\left(\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j-\mathbf{c}_k\right)^{\mathsf{H}}\right]\right\} \tag{7.12}$$

With respect to lemma 7.1, $\ddot{F}_{\text{LS}}$ in (7.12) can be rewritten as

$$\ddot{F}_{\text{LS}}=\text{Tr}\left[P\,\mathbf{V}_k^{\mathsf{H}}\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\mathbf{H}_{j,k}\mathbf{V}_k-P\,\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k-P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k\right]+Pd \tag{7.13}$$

To obtain precoders based on minimizing the objective function in (7.13), we differentiate $\ddot{F}_{\text{LS}}$ with respect to $\mathbf{V}_k$ by first considering the following assumptions [50, 51]:

1. $\mathbf{V}_k$ and $\mathbf{V}_k^{\mathsf{H}}$ are treated as independent variables.

2. $\dfrac{\partial \operatorname{Tr}\left[\mathbf{A}\mathbf{V}_k\right]}{\partial \mathbf{V}_k} = \dfrac{\partial \operatorname{Tr}\left[\mathbf{V}_k\mathbf{A}\right]}{\partial \mathbf{V}_k} = \mathbf{A}.$

Therefore, with respect to the two preceding assumptions, we have

$$\frac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k} = P\,\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k} - P\,\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \tag{7.14}$$

Thus, by considering (5.6), equation (7.14) can be further rewritten as

$$\frac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k} = P\,\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \left(\frac{\widehat{\mathbf{H}}_{j,k}}{1+\tau} + \breve{\mathbf{H}}_{j,k} \right)^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{j,k}}{1+\tau} + \breve{\mathbf{H}}_{j,k} \right) - P\,\mathbf{U}_k^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau} + \breve{\mathbf{H}}_{k,k} \right) \tag{7.15}$$

Note that $\dfrac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k}$ is now dependent on both $\widehat{\mathbf{H}}_{j,k}$ and $\breve{\mathbf{H}}_{j,k}$. To make $\dfrac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k}$ dependent only on $\widehat{\mathbf{H}}_{j,k}$, we can take the expectation with respect to $\breve{\mathbf{H}}_{j,k}$. Consequently, following lemma 7.2 and lemma 7.3 and by taking the expectation of $\dfrac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k}$ over $\breve{\mathbf{H}}_{j,k}$, we have

$$\mathbb{E}_{\breve{\mathbf{H}}_{j,k}} \left\{ \frac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k} \right\} = P\,\mathbf{V}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) - \frac{P}{1+\tau}\,\mathbf{U}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k} \tag{7.16}$$

Now the sought combiner $\mathbf{V}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{j,k}} \left\{ \dfrac{\partial \ddot{F}_{\text{LS}}}{\partial \mathbf{V}_k} \right\}$ equal to zero, which yields

$$\begin{aligned} P\,\mathbf{V}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) - \frac{P}{1+\tau}\,\mathbf{U}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k} &= \mathbf{0} \\ \Longrightarrow \mathbf{V}_k = (1+\tau) \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \varepsilon_{\text{LS}}\,\mathbf{I} \right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}} \mathbf{U}_k & \end{aligned} \tag{7.17}$$

where $\varepsilon_{\text{LS}} = \tau\,(1+\tau)\,Kd$.

Analogous to the standard IA techniques, due to the coupled nature of the problem, finding precoders and combiners requires an iterative algorithm in general. Therefore, with respect to the fact that the unitary precoders and combiners are more desirable, the proposed algorithm can be concisely presented as follows:

LS based IA

1. Set $\varepsilon_{\text{LS}} \coloneqq \tau\left(1+\tau\right) Kd$
2. Initialize random unitary matrices $\mathbf{V}_k, \forall\ k$
3. $\mathbf{U}_k = \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}} + \varepsilon_{\text{LS}}\,\mathbf{I}\right)^{-1} \widehat{\mathbf{H}}_{k,k}\mathbf{V}_k$
4. $\mathbf{U}_k \Longleftarrow \text{orth}\left(\mathbf{U}_k\right)$
5. $\mathbf{V}_k = \left(\sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k} + \varepsilon_{\text{LS}}\,\mathbf{I}\right)^{-1} \widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\mathbf{U}_k$
6. $\mathbf{V}_k \Longleftarrow \text{orth}\left(\mathbf{V}_k\right)$
7. Go to Step 3 and repeat

Remark 7.1: Although the above mentioned algorithm has been designed based on the channel estimations $\widehat{\mathbf{H}}_{k,j}$, it can be readily used for the perfect CSI. In this case, we need to set $\tau = 0$ in Step 1, and replace $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$.

Remark 7.2: As revealed in LS based IA, to design beamformers for user k, the direct link, i.e., $\mathbf{H}_{k,k}$, is taken into account. This is in contrast to interference leakage minimization algorithms where $\mathbf{H}_{k,k}$ is not considered to design beamformers for user k, as denoted in Min-WLI and Alt-Min algorithms. Therefore, it is expected that the proposed approach should outperform interference leakage minimization methods under both perfect and imperfect CSI.

7.3 MMSE based IA

In section 7.2, we proposed an LS based beamformer design for multiuser MIMO IC. Nonetheless, as it is well-known, MMSE based designs outperform LS based schemes since they further consider the effect of noise to design beamformers. Note that similar to the approach in [79] for MIMO precoders, it may be possible to design weighted MMSE based IA schemes which may achieve better performance than non-weighted MMSE based techniques; however, since we are supposed to use orthogonalization steps for the proposed MMSE based IA, we do not consider the case of weighted MMSE based IA. This is due to the fact that weighted MMSE based IA schemes are more beneficial without orthogonalization steps (see e.g., [24]). Therefore, in this section, we propose an MMSE based IA which similar to Max-SINR algorithm, considers the noise power to calculate

beamformers. Since we will largely compare the performance of the proposed MMSE based IA with Max-SINR, it is appropriate to recall this algorithm. However, as noted in remark 5.2, Max-SINR without orthogonalization can not achieve full DoF, since at least one of its precoders and/or combiners may become rank-deficient, and this can further result in reduced multiplexing gain. Therefore, when using Max-SINR, if the goal is to preserve full DoF, the inclusion of orthogonalization steps is mandatory. Consequently, in the sequel, we use Max-SINR with orthogonalization. Under perfect CSI and with respect to remark 5.3, Max-SINR algorithm with orthogonalization steps is a spacial case of the proposed adaptive Max-SINR algorithm in section 5.4 when τ is set to zero. Consequently, we represent the Max-SINR algorithm in a different but more tractable form as follows:

Max-SINR with Orthogonalization

1. Set $\varepsilon_{\texttt{Max-SINR}} \coloneqq \gamma^{-1}$
2. Initialize random unit-norm vectors $\mathbf{v}_{k\ell},\ \forall\ k,\ \ell$
3. $\mathbf{T}_k^{\ell} = \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} - \mathbf{H}_{k,k}\mathbf{v}_{k\ell}\mathbf{v}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} + \varepsilon_{\texttt{Max-SINR}}\,\mathbf{I}$
4. $\mathbf{u}_{k\ell} = \dfrac{\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{\left\|\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right\|} \qquad \ell = 1,\ldots,d$
5. $\mathbf{U}_k \Longleftarrow \mathrm{orth}\left(\mathbf{U}_k\right)$
6. $\mathbf{R}_k^{\ell} = \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\mathbf{H}_{j,k} - \mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}\mathbf{u}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k} + \varepsilon_{\texttt{Max-SINR}}\,\mathbf{I}$
7. $\mathbf{v}_{k\ell} = \dfrac{\left(\mathbf{R}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}}{\left\|\left(\mathbf{R}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{u}_{k\ell}\right\|} \qquad \ell = 1,\ldots,d$
8. $\mathbf{V}_k \Longleftarrow \mathrm{orth}\left(\mathbf{V}_k\right)$
9. Go to Step 3 and repeat

Hereafter and for the sake of simplicity, we call Max-SINR in lieu of Max-SINR with orthogonalization.

To derive the desired beamformers based on the MMSE criterion and similar to the approach in section 7.2, we consider an adaptive design based on the knowledge of imperfect CSI as the major optimization problem, and as it will be shown later, the standard design under perfect CSI is a special case of the adaptive design under imperfect CSI by setting error variance equal to zero. Given randomly initialized precoders and with respect

to (7.2), the optimization criterion based on MMSE design can be considered as

$$\arg \min_{\mathbf{U}_k} \mathbb{E}\left\{\left\|\mathbf{U}_k^{\mathsf{H}}\mathbf{y}_k - \mathbf{c}_k\right\|_2^2\right\} \quad \forall k = 1, \ldots, K \tag{7.18}$$

where $\mathbf{y}_k$ is defined in (7.1). In this case, the objective function in (7.18) can de shown as

$$\begin{aligned}\dot{F}_{\text{MMSE}} = \mathbb{E}\Bigg\{ \mathrm{Tr}\Bigg[&\left(\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k - \mathbf{c}_k\right) \\ &\left(\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{c}_j + \mathbf{U}_k^{\mathsf{H}}\mathbf{z}_k - \mathbf{c}_k\right)^{\mathsf{H}}\Bigg]\Bigg\}\end{aligned} \tag{7.19}$$

With respect to lemma 7.1, (7.19) reads

$$\begin{aligned}\dot{F}_{\text{MMSE}} \;=\; &\mathrm{Tr}\Bigg[P\,\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}\mathbf{U}_k + \sigma^2\mathbf{U}_k^{\mathsf{H}}\mathbf{U}_k \\ &- P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\mathbf{U}_k - P\,\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k\Bigg] + Pd\end{aligned} \tag{7.20}$$

Since minimizing the optimization problem in (7.20) requires differentiation, we have

$$\frac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k} = P\,\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \sigma^2\mathbf{U}_k^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}} \tag{7.21}$$

By considering (5.6), equation (7.21) can be further rewritten as

$$\begin{aligned}\frac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k} = &P\,\mathbf{U}_k^{\mathsf{H}} \sum_{j=1}^{K} \left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j}\right)\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,j}}{1+\tau} + \breve{\mathbf{H}}_{k,j}\right)^{\mathsf{H}} \\ &+\sigma^2\mathbf{U}_k^{\mathsf{H}} - P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau} + \breve{\mathbf{H}}_{k,k}\right)^{\mathsf{H}}\end{aligned} \tag{7.22}$$

Note that $\dfrac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k}$ is now dependent on both $\widehat{\mathbf{H}}_{k,j}$ and $\breve{\mathbf{H}}_{k,j}$. To make $\dfrac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k}$ dependent only on $\widehat{\mathbf{H}}_{k,j}$, and with respect to lemma 7.2 and lemma 7.3, we take the expectation of

$\frac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k}$ over $\breve{\mathbf{H}}_{k,j}$ which yields

$$\begin{aligned}\mathbb{E}_{\breve{\mathbf{H}}_{k,j}}\left\{\frac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k}\right\} &= P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\frac{Kd\tau}{1+\tau}\mathbf{I}\right)\\ &+ \sigma^2\mathbf{U}_k^{\mathsf{H}}-\frac{P}{1+\tau}\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\end{aligned} \tag{7.23}$$

Now the sought combiner $\mathbf{U}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{k,j}}\left\{\frac{\partial \dot{F}_{\text{MMSE}}}{\partial \mathbf{U}_k}\right\}$ equal to zero, which yields

$$\begin{aligned}&P\,\mathbf{U}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\frac{Kd\tau}{1+\tau}\mathbf{I}\right)+\sigma^2\mathbf{U}_k^{\mathsf{H}}-\frac{P}{1+\tau}\mathbf{V}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}=\mathbf{0}\\ &\qquad\Longrightarrow \mathbf{U}_k=(1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\varepsilon_{\text{MMSE}}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\mathbf{V}_k\end{aligned} \tag{7.24}$$

where

$$\varepsilon_{\text{MMSE}}=\gamma^{-1}(1+\tau)^2+\tau(1+\tau)Kd \tag{7.25}$$

Now we turn our focus to obtain the precoders. Given randomly initialized combiners, the optimization criterion based on MMSE design can be considered as

$$\arg\min_{\mathbf{V}_k}\ \mathbb{E}\left\{\left\|\mathbf{V}_k^{\mathsf{H}}\ddot{\mathbf{y}}_k-\mathbf{c}_k\right\|_2^2\right\}\quad\forall k=1,\ldots,K \tag{7.26}$$

where $\ddot{\mathbf{y}}_k$ is defined as

$$\ddot{\mathbf{y}}_k=\sum_{j=1}^{K}\mathbf{H}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{c}_j+\ddot{\mathbf{z}}_k \tag{7.27}$$

where $\ddot{\mathbf{z}}_k\sim\mathcal{N}_{\mathbb{C}}(\mathbf{0},\sigma^2\mathbf{I})$.

In this case, the objective function in (7.26) can de shown as

$$\ddot{F}_{\text{MMSE}} = \mathbb{E}\left\{ \text{Tr}\left[\left(\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{c}_j + \mathbf{V}_k^{\mathsf{H}} \ddot{\mathbf{z}}_k - \mathbf{c}_k \right) \left(\mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{c}_j + \mathbf{V}_k^{\mathsf{H}} \ddot{\mathbf{z}}_k - \mathbf{c}_k \right)^{\mathsf{H}} \right] \right\} \tag{7.28}$$

With respect to lemma 7.1, (7.28) reads

$$\begin{aligned} \ddot{F}_{\text{MMSE}} &= \text{Tr}\left[P\, \mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k} \mathbf{V}_k + \sigma^2 \mathbf{V}_k^{\mathsf{H}} \mathbf{V}_k \right. \\ &- \left. P\, \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k - P\, \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \mathbf{U}_k \right] + Pd \end{aligned} \tag{7.29}$$

Since minimizing the optimization problem in (7.29) requires differentiation, we have

$$\frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k} = P\, \mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \mathbf{H}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \mathbf{H}_{j,k} + \sigma^2 \mathbf{V}_k^{\mathsf{H}} - P\, \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \tag{7.30}$$

By considering (5.6), equation (7.30) can be further rewritten as

$$\begin{aligned} \frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k} = & P\, \mathbf{V}_k^{\mathsf{H}} \sum_{j=1}^{K} \left(\frac{\widehat{\mathbf{H}}_{j,k}}{1+\tau} + \breve{\mathbf{H}}_{j,k} \right)^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{j,k}}{1+\tau} + \breve{\mathbf{H}}_{j,k} \right) \\ & + \sigma^2 \mathbf{V}_k^{\mathsf{H}} - P\, \mathbf{U}_k^{\mathsf{H}} \left(\frac{\widehat{\mathbf{H}}_{k,k}}{1+\tau} + \breve{\mathbf{H}}_{k,k} \right) \end{aligned} \tag{7.31}$$

Note that $\frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k}$ is now dependent on both $\widehat{\mathbf{H}}_{j,k}$ and $\breve{\mathbf{H}}_{j,k}$. To make $\frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k}$ dependent only on $\widehat{\mathbf{H}}_{j,k}$, and with respect to lemma 7.2 and lemma 7.3, we take the expectation of $\frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k}$ over $\breve{\mathbf{H}}_{j,k}$ which yields

$$\begin{aligned} \mathbb{E}_{\breve{\mathbf{H}}_{j,k}} \left\{ \frac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k} \right\} &= P\, \mathbf{V}_k^{\mathsf{H}} \left(\frac{1}{(1+\tau)^2} \sum_{j=1}^{K} \widehat{\mathbf{H}}_{j,k}^{\mathsf{H}} \mathbf{U}_j \mathbf{U}_j^{\mathsf{H}} \widehat{\mathbf{H}}_{j,k} + \frac{Kd\tau}{1+\tau} \mathbf{I} \right) \\ &+ \sigma^2 \mathbf{V}_k^{\mathsf{H}} - \frac{P}{1+\tau} \mathbf{U}_k^{\mathsf{H}} \widehat{\mathbf{H}}_{k,k} \end{aligned} \tag{7.32}$$

Now the sought combiner $\mathbf{V}_k$ can be obtained by setting $\mathbb{E}_{\breve{\mathbf{H}}_{j,k}}\left\{\dfrac{\partial \ddot{F}_{\text{MMSE}}}{\partial \mathbf{V}_k}\right\}$ equal to zero, which yields

$$P\,\mathbf{V}_k^{\mathsf{H}}\left(\frac{1}{(1+\tau)^2}\sum_{j=1}^{K}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}+\frac{Kd\tau}{1+\tau}\mathbf{I}\right)+\sigma^2\mathbf{V}_k^{\mathsf{H}}-\frac{P}{1+\tau}\,\mathbf{U}_k^{\mathsf{H}}\widehat{\mathbf{H}}_{k,k}=\mathbf{0}$$
$$\Longrightarrow \mathbf{V}_k=(1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k}+\varepsilon_{\text{MMSE}}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\mathbf{U}_k \tag{7.33}$$

where

$$\varepsilon_{\text{MMSE}}=\gamma^{-1}\,(1+\tau)^2+\tau\,(1+\tau)\,Kd \tag{7.34}$$

With respect to the fact that the unitary precoders and combiners are more desirable, and since only imperfect channel estimates $\widehat{\mathbf{H}}_{k,j}$ are available, the proposed algorithm, which iteratively optimizes the precoders and combiners, can be concisely presented as follows:

MMSE based IA

1. Set $\varepsilon_{\text{MMSE}} \coloneqq \gamma^{-1}\,(1+\tau)^2+\tau\,(1+\tau)\,Kd$
2. Initialize random unitary matrices $\mathbf{V}_k, \forall\, k$
3. $\mathbf{U}_k=(1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{k,j}^{\mathsf{H}}+\varepsilon_{\text{MMSE}}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}\mathbf{V}_k$
4. $\mathbf{U}_k \Longleftarrow \text{orth}\,(\mathbf{U}_k)$
5. $\mathbf{V}_k=(1+\tau)\left(\sum_{j=1}^{K}\widehat{\mathbf{H}}_{j,k}^{\mathsf{H}}\mathbf{U}_j\mathbf{U}_j^{\mathsf{H}}\widehat{\mathbf{H}}_{j,k}+\varepsilon_{\text{MMSE}}\,\mathbf{I}\right)^{-1}\widehat{\mathbf{H}}_{k,k}^{\mathsf{H}}\mathbf{U}_k$
6. $\mathbf{V}_k \Longleftarrow \text{orth}\,(\mathbf{V}_k)$
7. Go to Step 3 and repeat

Remark 7.3: Similar to the discussions in remark 7.1, the proposed MMSE based design can be applied to the case of perfect CSI by setting $\tau=0$ and replacing $\widehat{\mathbf{H}}_{k,j}$ with $\mathbf{H}_{k,j}$.

7.4 Discussions

7.4.1 CSI Availability

By comparing Min-WLI with LS based IA, one can conclude that to calculate one specific combiner, Min-WLI needs $K-1$ covariance matrices whereas LS based IA requires K covariance matrices. Therefore, the proposed LS based design needs slightly more CSI to be available than Min-WLI.

Both Max-SINR and the proposed MMSE based IA consider the effect of noise to derive precoders and combiners. However, while the former is based on a stream-by-stream approach, the latter is based on a user-by-user approach. As seen in Step 3 of Max-SINR and MMSE based IA algorithms, to calculate the kth combiner, Max-SINR needs $K+d-1$ covariance matrices whereas MMSE based IA requires K covariance matrices. Therefore, the proposed MMSE based design needs less CSI to be available than Max-SINR. This also makes sense since Max-SINR is a stream-by-stream approach whereas MMSE based IA is a user-by-user approach, and as it is well-known the user-by-user schemes need less CSI to be available.

7.4.2 Computational Complexity

In this subsection, we compare the computational complexity of the proposed IA algorithms with that of standard IA schemes. Without loss of generality, we consider beamformer design under perfect CSI.

Note that for both Min-WLI and LS based IA, the beamformers are designed independent of the nominal SNR, i.e., the precoders and combiners are once calculated and then can be used for any SNR. Therefore, the computational complexity of these two schemes are comparable and none of them gives considerable advantage over the other in terms of decreasing the computational complexity. However, this does not hold for the proposed MMSE based IA and Max-SINR, since the beamformers obtained by either of these algorithms are dependent on SNR. In other words, the beamformers are to be calculated each time the nominal SNR changes. Therefore if any of these algorithms has a slight advantage over the other for one specific nominal SNR, this can lead to a huge reduction in computational complexity for a wide range of SNRs. To demonstrate this superior performance of the proposed MMSE based IA over Max-SINR, we consider the calculations involved in just one iteration of either of these two algorithms and for one specific combiner, i.e., the kth combiner. As demonstrated in Steps 3–4 of Max-SINR, to compute

$\mathbf{U}_k$, we need to calculate d matrix inverses whereas for the proposed MMSE based IA, we have to calculate only one matrix inverse of the same size. The same is true to calculate the kth precoder $\mathbf{V}_k$. Therefore, within one iteration, the proposed MMSE based IA calculates 2 matrix inverses whereas Max-SINR calculates $2d$ matrix inverses. By considering the fact that these algorithms need at least hundreds of iterations to efficiently calculate the precoders and combiners, the proposed MMSE based IA results in much less computational complexity compared to Max-SINR, and this automatically translates to the reduced running time for the proposed MMSE based IA.

7.4.3 The Equivalence of MMSE based IA and Max-SINR

In this subsection, it is shown that Max-SINR is equivalent to the proposed MMSE based IA under perfect CSI, i.e., these two algorithms achieve exactly the same beamformers under perfect CSI. To do so, we should demonstrate that in each iteration, the precoders and combiners obtained by one of these algorithms are the same as the other's. Therefore and without loss of generality, we just show that the kth combiner in Step 5 of Max-SINR is the same as the one in Step 4 of the proposed MMSE based IA. To do so, we assume that τ in Step 1 of the proposed MMSE based IA has been set to zero and $\widehat{\mathbf{H}}_{j,k}$ has been further replaced with $\mathbf{H}_{j,k}$. To further proceed, we consider the following lemma [48]:

Lemma 7.4: If $\mathbf{a} \in \mathbb{C}^{M\times 1}$ and $\mathbf{A} \in \mathbb{C}^{M\times M}$ then

$$\left(\mathbf{A} - \mathbf{a}\mathbf{a}^{\mathsf{H}}\right)^{-1}\mathbf{a} = \frac{\mathbf{A}^{-1}\mathbf{a}}{1 - \mathbf{a}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{a}} \tag{7.35}$$

With respect to lemma 7.4, $\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}$ in Step 4 of Max-SINR algorithm can be rewritten as

$$\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell} = \frac{\left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{1 - \mathbf{v}_{k\ell}^{\mathsf{H}}\mathbf{H}_{k,k}^{\mathsf{H}}\left(\sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}} \tag{7.36}$$

By considering the fact that the ℓth column of $\mathbf{U}_k$ is equal to

$$\mathbf{u}_{k\ell} = \frac{\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}}{\left\|\left(\mathbf{T}_k^{\ell}\right)^{-1}\mathbf{H}_{k,k}\mathbf{v}_{k\ell}\right\|_2} \tag{7.37}$$

and by horizontally concatenating the d columns of $\mathbf{U}_k$, we have

$$\mathbf{U}_k = \left(\sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1} \mathbf{I} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k \begin{pmatrix} t_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_d \end{pmatrix} \tag{7.38}$$

such that

$$t_\ell = \frac{1}{\left\| \left(\sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \gamma^{-1} \mathbf{I} \right)^{-1} \mathbf{H}_{k,k} \mathbf{v}_{k\ell} \right\|_2} \qquad \ell = 1, \ldots, d \tag{7.39}$$

Note that although $\mathbf{U}_k$ in (7.38) has unit-norm columns, it is not unitary, i.e., $\mathbf{U}_k^{\mathsf{H}} \mathbf{U}_k \neq \mathbf{I}$. However, the orthogonalization in Step 5 of Max-SINR substitutes $\mathbf{U}_k$ with the unitary part of its QR decomposition which is exactly the same as the one obtained in Step 4 of the proposed MMSE based IA. Similarly, it can be readily proved that in each iteration, the kth precoder of Max-SINR is the same as the one obtained through MMSE based IA. Consequently, the two algorithms achieve the same precoders and combiners and thus achieve the same performance under perfect CSI.

7.4.4 On Diagonalized Subchannels

In this part, it is shown that the proposed LS based design leads to diagonalized subchannels for all SNR ranges. In other words, using LS based IA, $\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k$ becomes a diagonal matrix. This is in contrast to the previously proposed IA schemes wherein the resulted subchannels are full matrices. This removes the need of SVD for the LS based IA in order to employ waterfilling since the decoupled subchannels have been already diagonalized. Therefore, in the sequel, it is shown that $\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_{k,k} \mathbf{V}_k$ is a diagonal matrix, provided that $\mathbf{U}_k$ and $\mathbf{V}_k$ are obtained by LS based IA. Since, this condition is met under the assumption of perfect CSI, we assume that in the proposed LS based IA, τ has been set to zero and all imperfect CSI $\widehat{\mathbf{H}}_{k,j}$ have been replaced by the perfect CSI, i.e., $\mathbf{H}_{k,j}$.

With respect to Step 3 of LS based IA algorithm, we have

$$\left(\sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k$$

$$= \left(\sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \mathbf{H}_{k,k} \mathbf{V}_k \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k$$
$$= \left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{H} \right)^{-1} \boldsymbol{\mathcal{V}}_k \tag{7.40}$$

where $\boldsymbol{\mathcal{V}}_k = \mathbf{H}_{k,k} \mathbf{V}_k \in \mathbb{C}^{M \times d}$ and

$$\boldsymbol{\mathcal{V}}_{-k} = [\mathbf{H}_{k,1} \mathbf{V}_1, \ldots, \mathbf{H}_{k,k-1} \mathbf{V}_{k-1}, \mathbf{H}_{k,k+1} \mathbf{V}_{k+1}, \ldots, \mathbf{H}_{k,K} \mathbf{V}_K] \in \mathbb{C}^{M \times (K-1)d}$$

By considering the fact that the feasibility conditions of IA have been met [20, 21] and the interferences have been aligned within the reduced subspace of the received signal space, the components of the desired space $\boldsymbol{\mathcal{V}}_k$ become linearly independent of the components of the interference space $\boldsymbol{\mathcal{V}}_{-k}$ [17] such that

$$\operatorname{rank}\left(\boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \in \mathbb{C}^{M \times M} \right) = d \qquad \text{and} \qquad \operatorname{rank}\left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} \in \mathbb{C}^{M \times M} \right) = M - d$$

To further continue, we consider the following lemma:

Lemma 7.5: Let $\mathbf{A}$ be a horizontal concatenation of two submatrices, i.e., $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2] \in \mathbb{C}^{M \times M}$ such that $\mathbf{A}_1 \in \mathbb{C}^{M \times a}$ and $\mathbf{A}_2 \in \mathbb{C}^{M \times (M-a)}$ have independent columns. In this case, the ZF condition implies that

$$\mathbf{A}_1^{\mathsf{H}} \left(\mathbf{A} \mathbf{A}^{\mathsf{H}} \right)^{-1} \mathbf{A}_1 = \mathbf{A}_1^{\mathsf{H}} \left(\mathbf{A}_1 \mathbf{A}_1^{\mathsf{H}} + \mathbf{A}_2 \mathbf{A}_2^{\mathsf{H}} \right)^{-1} \mathbf{A}_1 = \mathbf{I}_{a \times a} \tag{7.41}$$

Hence, due to lemma 7.5 and with respect to the fact that the interferences have been aligned within a $(M-d)$-dimensional subspace of the M-dimensional received signal space, we have

$$\boldsymbol{\mathcal{V}}_k^{\mathsf{H}} \left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{H} \right)^{-1} \boldsymbol{\mathcal{V}}_k$$
$$= \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \left(\sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{V}_j^{\mathsf{H}} \mathbf{H}_{k,j}^{\mathsf{H}} + \mathbf{H}_{k,k} \mathbf{V}_k \mathbf{V}_k^{\mathsf{H}} \mathbf{H}_{k,k}^{\mathsf{H}} \right)^{-1} \mathbf{H}_{k,k} \mathbf{V}_k = \mathbf{I}_{d \times d} \tag{7.42}$$

We further consider the QR decomposition of $\left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{H} \right)^{-1} \boldsymbol{\mathcal{V}}_k$ as

$$\left(\boldsymbol{\mathcal{V}}_{-k} \boldsymbol{\mathcal{V}}_{-k}^{\mathsf{H}} + \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{V}}_k^{H} \right)^{-1} \boldsymbol{\mathcal{V}}_k = \mathbf{U}_k \boldsymbol{\mathcal{R}}_k \tag{7.43}$$

where $\mathbf{U}_k$ is the unitary part. Therefore with respect to (7.42), we have

$$\mathbf{U}_k^{\mathsf{H}}\boldsymbol{\mathcal{V}}_k = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k = \left(\boldsymbol{\mathcal{R}}_k^{\mathsf{H}}\right)^{-1} \tag{7.44}$$

However, note that due to the properties of IA and LS criteria, $\boldsymbol{\mathcal{R}}_k$ is now a diagonal matrix instead of an upper triangular. Furthermore, since the diagonal elements of the triangular part of the QR decomposition of any matrix are real numbers, this implies that $\mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,k}\mathbf{V}_k$ is a diagonal matrix consists of real numbers.

It is also worthwhile to point out that since at high enough SNRs, MMSE based IA boils down to LS based IA, the resulted subchannels of MMSE based IA become diagonalized at sufficiently high SNRs.

7.5 Numerical Results

In this section and by using numerical results, we corroborate the improved performance achieved by the proposed LS and MMSE based IA compared to standard IA techniques. To have a fair comparison, we also consider the performance of Max-SINR and Min-WLI algorithms.

Without loss of generality, we just consider a symmetric constant MIMO IA with $K = 4$ and $d = 2$. To meet the sufficient conditions of feasibility for IA, we set $M = 4,\ N = 6$ [21].

For the case of imperfect CSI, we assume that the channel estimation error variance obeys (3.3), i.e., $\tau = \beta\gamma^{-\alpha}$. Regarding the performance analysis under CSI mismatch, although the promised improvement of the proposed LS and MMSE based IA can be gleaned for various values of α, we focus on two representative cases: $\alpha = 0$ (which imitates the CSI feedback scenario), and $\alpha = 1$ (which imitates the reciprocal channels).

The performance trend of the proposed LS and MMSE based designs under perfect CSI can be obtained with respect to remark 7.1 and remark 7.3, respectively.

Plus, by considering i.i.d. Gaussian input signaling and uniform power allocation, we evaluate the achievable sum rates as [67]

$$R = \sum_{k=1}^{K} \log_2 \det \left(\mathbf{I} + \left(\gamma^{-1}\mathbf{I} + \sum_{j=1, j\neq k}^{K} \boldsymbol{\Phi}_{k,j} \right)^{-1} \boldsymbol{\Phi}_{k,k} \right) \tag{7.45}$$

where $\boldsymbol{\Phi}_{k,j} = \mathbf{U}_k^{\mathsf{H}}\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\mathsf{H}}\mathbf{H}_{k,j}^{\mathsf{H}}\mathbf{U}_k$, such that in the case of imperfect CSI, all precoders

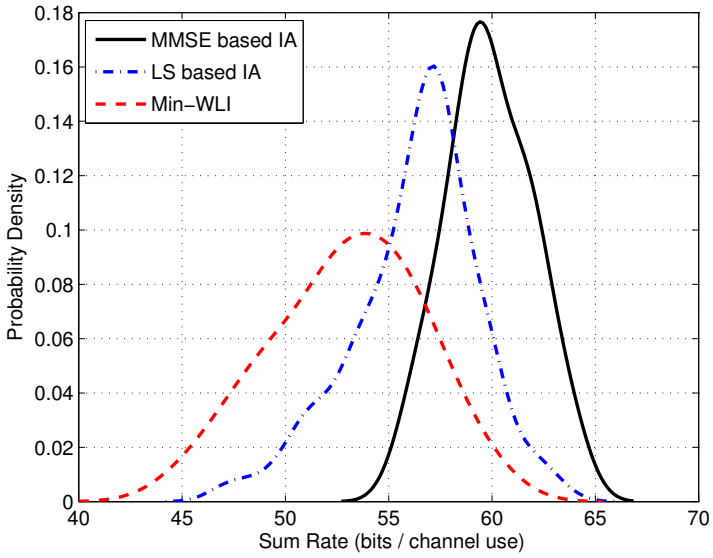

Figure 7.1: *Probability density of the sum rate for the solutions obtained from different random initialization for Min-WLI, LS and MMSE based IA for the case* $K = 4,\ d = 2,\ M = 4,\ N = 6$ *at SNR of* 20 *dB and under perfect CSI.*

and combiners are constructed based on erroneous channel estimations in (5.4), i.e., all $\mathbf{H}_{k,j}$ are replaced by $\widehat{\mathbf{H}}_{k,j}$

In Figs. 7.1–7.2, we consider a fixed random channel initialization for the case $K = 4,\ d = 2$, and a fixed SNR of 20 dB. For this scenario, we ran Min-WLI, LS and MMSE based IA algorithms 100 times, each one starting from different random unitary precoders. The number of iterations for each algorithm was 3000, which assures that the interferences are almost perfectly aligned within a reduced subspace of each received signal space.

Fig. 7.1 depicts the probability density of the sum rate for three different IA algorithm. As shown, although all algorithms have been initialized from the same unitary precoders, the final sum rate of the proposed LS and MMSE based algorithms is more concentrated around higher values, which implies that with the same initializations and under the same channel realization, the mean sum rate achieved by LS and MMSE based IA is higher than that of Min-WLI.

Fig. 7.2 illustrates the CCD of the sum rate for three different IA algorithms. As revealed, although all algorithms have been initialized from the same unitary precoders, the proposed algorithms achieve better performance than Min-WLI.

Fig. 7.3 shows the convergence of the sum rate of Min-WLI, LS and MMSE based IA algorithms (averaged over 200 channel realizations) for the case $K = 4,\ d = 2$ at SNR of

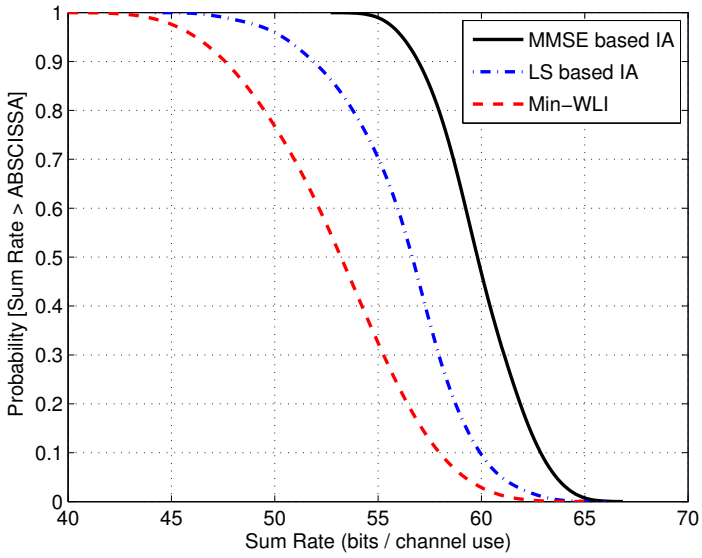

Figure 7.2: *Complementary cumulative distribution of the sum rate for the solutions obtained from different random initialization for Min-WLI, LS and MMSE based IA for the case* $K = 4,\ d = 2,$ $M = 4,\ N = 6$ *at SNR of* 20 *dB and under perfect CSI.*

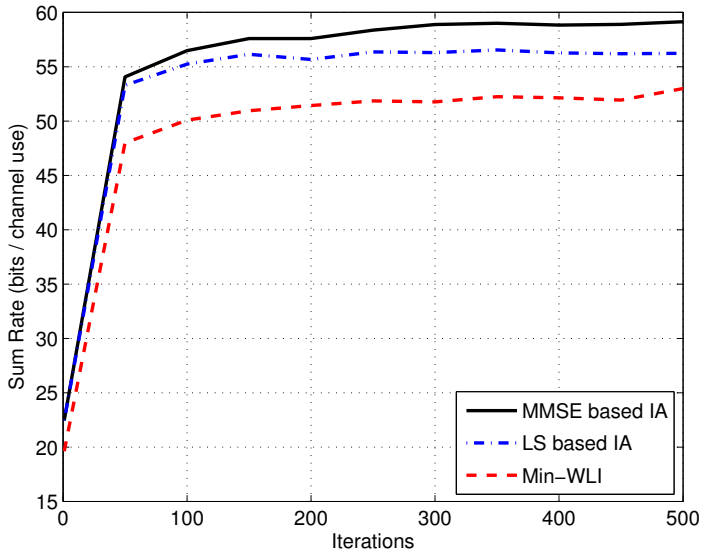

Figure 7.3: *Convergence of sum rate for Min-WLI, LS and MMSE based IA for the case* $K = 4,\ d = 2,\ M = 4,\ N = 6$ *at SNR of* 20 *dB and under perfect CSI.*

20 dB. As revealed the proposed LS and MMSE based IA have the same convergence rate as Min-WLI. Moreover, for any number of iterations, they achieve higher sum rates than Min-WLI.

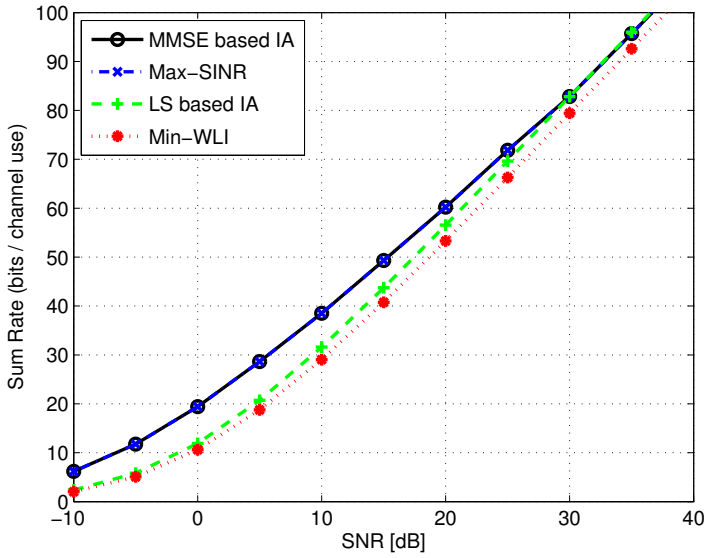

Figure 7.4: *Average sum rates for $K = 4,\ d = 2,\ M = 4,\ N = 6$ and under perfect CSI.*

Fig. 7.4 illustrates the sum rate under perfect CSI. As revealed, with the presence of perfect CSI, MMSE based IA achieves the same sum rate as Max-SINR while outperforming both LS based IA and Min-WLI. However, the proposed LS based IA outperforms Min-WLI such that the achieved gain in sum rate is 3 bits per channel use.

Figs. 7.5 and 7.6, respectively, depict the average sum rate for $\beta = 10,\ \alpha = 1$ and $\beta = 0.1,\ \alpha = 0$, respectively. As shown, while under perfect CSI, MMSE based design achieves the same sum rate as Max-SINR, under imperfect CSI, MMSE based IA outperforms Max-SINR. Also the proposed LS based IA is able to achieve better performance than Min-WLI. For example, for the case of $\beta = 10,\ \alpha = 1$ and at SNR of 30 dB, LS and MMSE based IA achieve 10 and 7 bits per channel use gain in sum rate compared to Min-WLI and Max-SINR, respectively. Similarly, for the case of $\beta = 0.1,\ \alpha = 0$, and at SNR of 30 dB, LS and MMSE based IA achieve 9 and 7 bits per channel use gain in sum rate compared to Min-WLI and Max-SINR algorithms, respectively.

As revealed in Figs. 7.5 and 7.6, the proposed LS based design is able to achieve almost the same performance as the MMSE based design under imperfect CSI. Also, it achieves better performance than Max-SINR for $\alpha = 1$. However, for $\alpha = 0$, while at low SNRs, Max-SINR outperforms LS based design, the latter achieves better performance than the former at high SNRs.

Fig. 7.7 illustrates the average SER of Min-WLI, Max-SINR and the proposed MMSE based IA for $K = 4,\ d = 2$ under the imperfect CSI cases $\beta = 10,\ \alpha = 1$ and $\beta =$

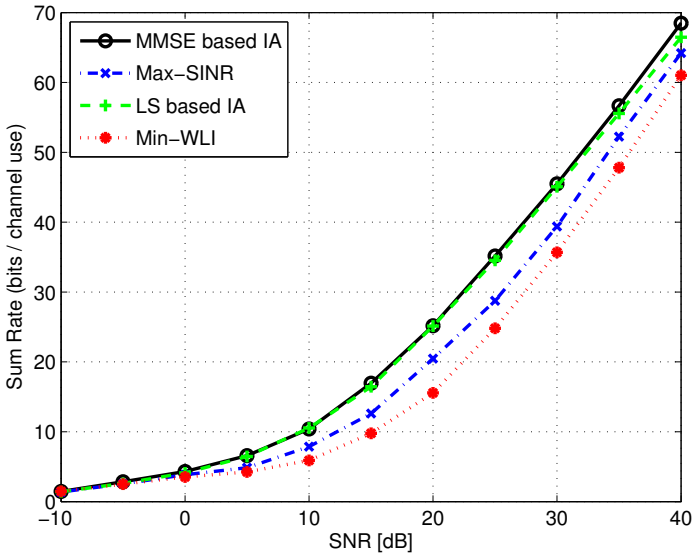

Figure 7.5: *Average sum rates for* $K = 4,\ d = 2,\ M = 4,\ N = 6$ *when* $\beta = 10,\ \alpha = 1$.

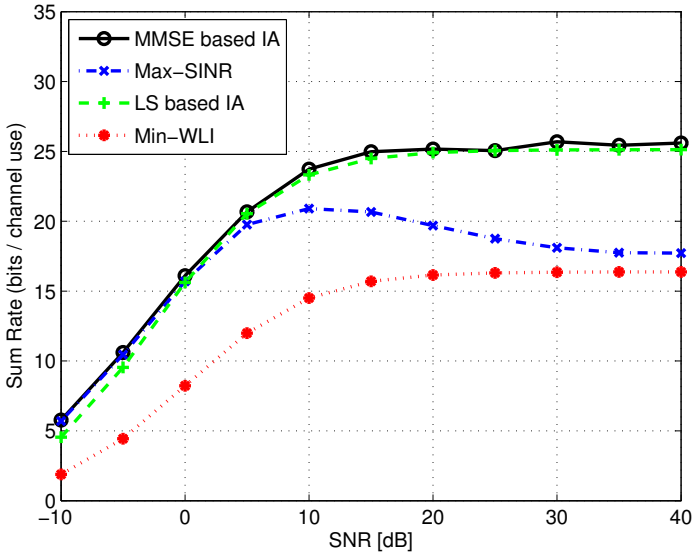

Figure 7.6: *Average sum rates for* $K = 4,\ d = 2,\ M = 4,\ N = 6$ *when* $\beta = 0.1,\ \alpha = 0$.

$0.05,\ \alpha = 0$. We assumed that each transmitted block consists of 100 QPSK symbols. As seen MMSE based IA outperforms both Min-WLI and Max-SINR. For example, when $\beta = 10,\ \alpha = 1$, MMSE based IA respectively achieves 18 dB and 14 dB gain compared to Min-WLI and Max-SINR to reach the SER of 10^{-3}. Also for the case $\beta = 0.05,\ \alpha = 0$, the MMSE based IA decreases the SER by a factor of at least $\frac{1}{10}$ compared to Min-WLI

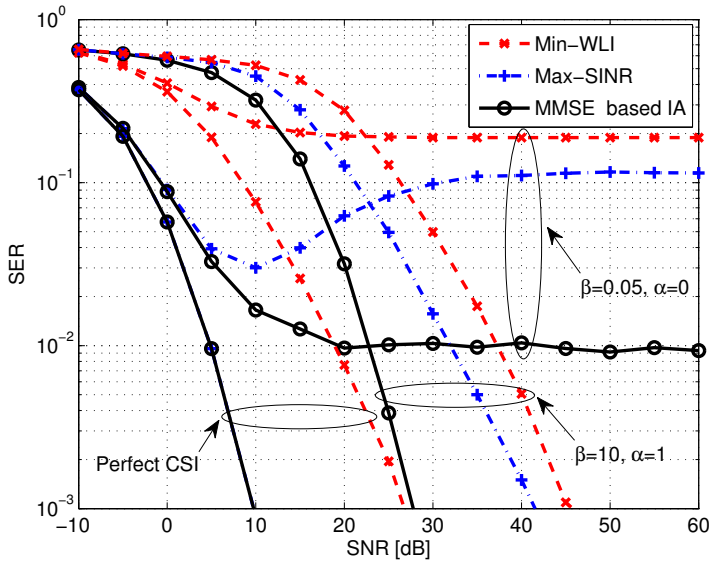

Figure 7.7: *Average SER for* $K = 4,\ d = 2,\ M = 4,\ N = 6$ *under the imperfect CSI* $\beta = 10,\ \alpha = 1$ *and* $\beta = 0.05,\ \alpha = 0$. *Each transmitted block consists of 100 QPSK symbols.*

and Max-SINR at SNRs of larger than 20 dB.

From Figs. 7.6–7.7, one interesting observation is that, for $\alpha = 0$, the performance trend of Max-SINR is nonmonotonic whereas that of MMSE based IA is monotonic. Also by considering the slope of the curves in both Figs. 7.5 and 7.7, another interesting point is that when $\alpha = 1$, i.e., the error variance scales with the inverse of SNR, full multiplexing gain can be achieved. This is consistent with analytically derived bounds in chapter 5 and implies that the derived bounds in (5.30)–(5.31) are generally applicable to any IA scheme.

7.6 Summary

With the presence of perfect CSI, interference alignment enables us to achieve full DoF. However, when subject to imperfect CSI, the full benefits of IA may not be readily achievable. In this chapter, we proposed two novel IA algorithms such that the optimization criteria were set up based on the knowledge of imperfect CSI. While the LS based scheme does not consider the effect of noise to design beamformers, the MMSE based method does so. This causes the latter to outperform the former under both perfect and imperfect CSI. We also compared the proposed algorithms with standard IA methods. It was shown that the LS based algorithm outperforms interference leakage minimization algorithms un-

der both perfect and imperfect CSI. However, while MMSE based IA achieves the same performance as Max-SINR under perfect CSI, the former outperforms the latter subject to imperfect CSI. We showed that even with this superior performance, the proposed MMSE based IA needs less CSI to be available and has less computational complexity compared to Max-SINR.

Chapter 8

Conclusions

8.1 Summary and Conclusions

Future wireless networks will be dominated by the application of smart phones and tablets which demand high data rate wireless communications. Due to the scarcity of radio resources, spectrum and energy efficient design of such networks become very important. Consequently and with respect to the fact that in almost all communication scenarios only imperfect CSI may be accessible, advanced and optimized communication techniques are inevitable means to satisfy such demands in future cellular networks. In this thesis, we addressed advanced interference management techniques which enable us to meet the aforementioned design criteria by relying on a generalized imperfect CSI model.

Under the considered imperfect CSI model introduced in subsection 3.2.2, we derived novel bounds on asymptotic mean loss in sum rate and achievable DoF for both multiuser MIMO downlink and wireless interference networks with IA. For example, it has been shown that if the intention is to keep the asymptotic mean loss in sum rate bounded (and consequently maintain the full DoF), the error variance must at least scales with the inverse of SNR.

While standard RCI outperforms CI under perfect CSI, the former is not able to significantly outperform the latter under CSI mismatch. Therefore, we proposed an adaptive RCI technique, which with the knowledge of error variance in advance, is able to outperform CI. We also proposed an enhanced linear precoding technique, namely RPA. We showed that the proposed RPA precoding is able to outperform standard linear precoders (i.e., CI, RCI, and PA) under both perfect and imperfect CSI. Also it was shown that PA and RPA precoding techniques enable us to decrease the deployed power to secure the same average

output SINR for each user compared to CI and RCI precoding, respectively.

As it is well-known, Max-SINR outperforms interference leakage minimization algorithms under the availability of perfect CSI. We showed that under imperfect CSI, however, the former fails to maintain the same comparable improvement. Therefore, we proposed an adaptive Max-SINR algorithm, which with the knowledge of error variance in advance, is able to significantly outperform interference leakage minimization algorithms. Moreover, we proposed adaptive LS and MMSE based IA techniques which are able to outperform standard IA schemes under perfect and imperfect CSI.

When the number of users or the number of DoF becomes increased, standard IA techniques require large number of antennas. By relying on the concept of partially coordinated reception in wireless interference networks, we proposed IA algorithms which enable us to decrease the number of deployed antennas compared to standard IA schemes to maintain the same achievable DoF. This requires that on average, half of the total decoded data is needed to be shared by receive nodes which consequently leads to SIC. It was shown that even with reduced number of antennas, the proposed IA techniques are able to outperform standard IA methods under perfect and imperfect CSI.

8.2 Limitations and Future Work

As earlier mentioned in chapter 4, the proposed PA and RPA precoding can achieve better throughput and SER than standard linear precoders, i.e., CI and RCI. On the other hand, they also enable us to save transmit power at BS in order to meet the same average output SINR at each user compared to CI and RCI precoding, respectively. Nevertheless, while both CI and RCI are based on Gaussian input signaling, PA and RPA precoding are merely applicable when the utilized constellation is $\mathcal{M}$-ary PSK. However, since the future cellular networks are provisioned to support the high order constellations like QAM, one appealing direction for future research is to extend the benefits of PA and RPA precoding to rectangular constellations.

Also in chapter 4, although the SER performance of PA and RPA under both perfect and imperfect CSI was evaluated, the transmit-power efficiency of PA and RPA precoding was merely considered under the assumption of perfect CSIT. In other words, the derived formula in (4.51) is just meant for the case when perfect CSIT is available at BS. Since the availability of imperfect CSIT is more realistic, it is also desirable to evaluate the capability of PA and RPA precoding to save transmit power subject to the availability of

CSIT mismatch. In other words, it is more appealing to determine the amount of transmit-power reduction to secure the same average output SINR for each user, when PA (RPA) is used in lieu of CI (RCI), when only imperfect CSIT is available at BS.

In chapter 6, we proposed IA algorithms which enable us to decrease the number of deployed antennas at transmit and/or receive nodes to secure the same number of DoF compared to standard IA techniques. However, the proposed schemes require partial co-ordination at receive side which implies that averagely half of the total decoded data are needed to be shared by receive nodes. With respect to the fact that coordinated down-link transmission is highly supported in future cellular networks, one interesting research direction is to seek IA algorithms designed for transmit-side coordination instead of the receive-side coordination, i.e., in the case of BS cooperation.

In chapter 5, we evaluated the asymptotic performance of IA under a generalized CSI mismatch model. It is however more desirable to specifically assess the performance of IA under digital feedback. Quantized feedback strategies are not only appealing in single-cell broadcast channels but also has been received significant attention in interference channels. Although there are some literature addressed the feedback topology design for IA, see e.g., [61–63], the optimum feedback strategy in IC is still unknown. Therefore, finding optimized feedback scenarios is of particular interest. This includes designing codebooks optimized with respect to the quantization errors and feedback delays.

Regardless of digital feedback, analog feedback is also another potential approach to provide access to some sort of CSI in interference channels. This type of feedback is more relevant to the reciprocal channels where the downlink channel can be estimated through pilots sent over the uplink channel and consequently the channel measurement error depends on the noise level at BS as well as the pilot power. Therefore, in reciprocal channels, the CSI quality depends on the training sequence. With respect to this, one approach to provide higher CSI quality for IA in reciprocal channels is to seek an optimized channel training scheme.

Irrespective of IA which has been proposed to achieve the maximum DoF in IC, another interesting interference scenario is interference broadcast channel (IBC) wherein each transmitter communicates with more than one receive node. This is denoted in Fig. 8.1 where each BS communicates with two MTs. This is a more generalized scenario of IC wherein each transmitter communicates with merely one receiver as depicted in Fig. 2.5. More specifically, while standard IA is able to provide interference free communications for the proposed scenario in Fig. 2.5, it requires more sophisticated designs to handle the case of Fig. 8.1, provided that all BSs transmitted at the same time and frequency slots.

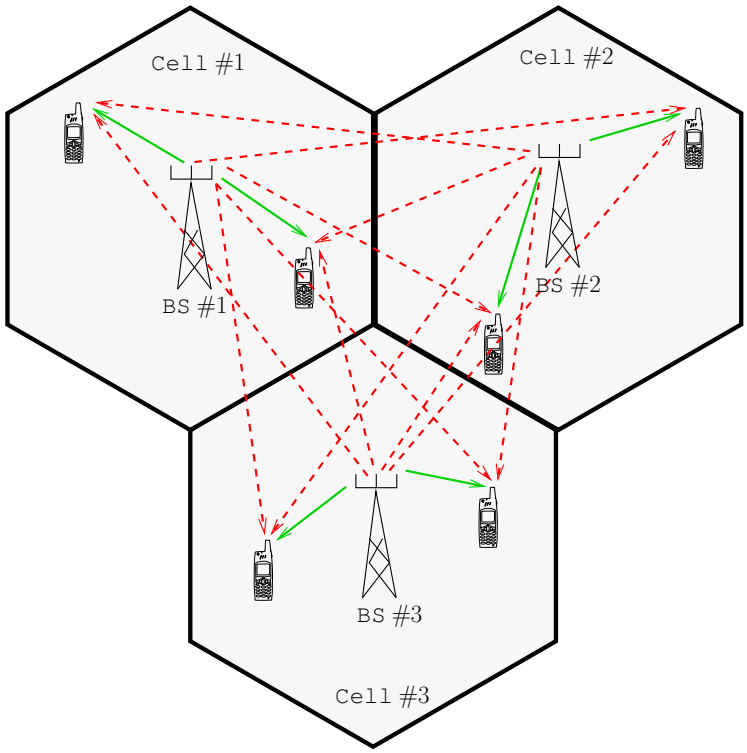

Figure 8.1: *3-cell interference broadcast channel where all BSs transmit at the same time and frequency slots. Each BS communicates to two corresponding MTs while causing interference to the other MTs in adjacent cells. In this case, dash red arrows represent inter-cell interference while solid green arrows denote desired links.*

Therefore, IBC requests different approaches for beamformer design. Although there are few schemes to communicate over IBC, see e.g., [80, 81], they were merely proposed under the assumption of the availability of perfect CSI. However, as already mentioned, since from the practical point of view, having access to perfect CSI is very optimistic, new adaptive algorithms are needed to be designed for the case when only imperfect CSI is available. This may include optimized digital and analog feedback topology designs for IBC. Similar to the case of IC, optimization can be sought over codebooks regarding digital feedback or channel training schemes in the case of analog feedback.

The imperfect CSI model that has been considered so far in this thesis is only related to the instantaneous CSI, i.e., it has been assume that the current CSI has some errors. Apart from current CSI mismatch, performance analysis of communication systems under perfect delayed CSI is also of particular interest. This case represents the scenario wherein the perfect CSI is available but is related to the previous time slots. In other words, in delayed CSI it is assume that at time t, the only knowledge of the channel state is related to the perfect CSI at time slot $t-1$ and/or the more outdated ones. However, so far and due

to its intractability, the impact of delayed CSI on the performance of broadcast channels has been mainly evaluated in the literature, see e.g., [82]. Therefore, it is likewise desirable to evaluate the impact of perfect delayed CSI in IC and IBC.

Although the current CSI mismatch and the perfect delayed CSI models are able to simulate many practical scenarios, performance evaluation of broadcast channels, IC and IBC under both imperfect and delayed CSI is of particular interest. This case represents the mixed CSI, where not only the available CSI is outdated but also it is not perfect. In this case, adaptive and optimized beamformer design (even for the simple case of single-cell broadcast channel) under mixed CSI is a promising future work.

Apart from availability of perfect or imperfect CSI, performance analysis of blind IA has received lots of attention recently [83]. The idea is that the transmitters' knowledge of channel coherence intervals alone (without any knowledge of the values of channel coefficients) can be surprisingly useful in a multiuser setting. Beyond the concept of IA, we expect that there may be a variety of settings which give rise to opportunities for blind schemes to be applicable in IBC. The search for such settings is an interesting direction for future work.

Appendix A

Statistical Properties of $|\rho_{\ell,x}|$

Consider $\mathbf{H} \in \mathbb{C}^{M\times N}$ and $\mathrm{vec}\,(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}\,(\mathbf{0}, \sigma_h^2\mathbf{I})$. By expanding the complex multiplications of matrix $\mathbf{R} = \mathbf{H}\mathbf{H}^{\mathsf{H}}$ for the case $\ell \neq x$, we have

$$|\rho_{\ell,x}| = \left(\left[\sum_{n=1}^{N} \left(h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i \right) \right]^2 + \left[\sum_{n=1}^{N} \left(h_{\ell,n}^i h_{x,n}^r - h_{\ell,n}^r h_{x,n}^i \right) \right]^2 \right)^{\frac{1}{2}} \tag{A.1}$$

where the notations $h_{\ell,n}^r = \Re\left(h_{\ell,n}\right), h_{\ell,n}^i = \Im\left(h_{\ell,n}\right)$, are used for convenience, and $h_{\ell,n}$ is used to denote the generic channel coefficient of the nth transmit antenna to the ℓth receive antenna. Since we assumed that $h_{\ell,n}^r, h_{\ell,n}^i, h_{x,n}^r, h_{x,n}^i \in \mathcal{N}_{\mathbb{C}}\left(0, \frac{\sigma_h^2}{2}\right)$, we have

$$\mathbb{E}\left\{h_{\ell,n}^r h_{x,n}^r\right\} = 0 \quad \text{and} \quad \mathrm{var}\left\{h_{\ell,n}^r h_{x,n}^r\right\} = \frac{\sigma_h^4}{4} \tag{A.2}$$

The same applies to all combinations of real and imaginary coefficient that appear in (A.1). Therefore

$$\begin{aligned} \mathbb{E}\left\{h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right\} &= 0 \\ \mathrm{var}\left\{h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i\right\} &= \frac{\sigma_h^4}{2} \end{aligned} \tag{A.3}$$

and

$$\mathbb{E}\left\{ \sum_{n=1}^{N} \left(h_{\ell,n}^r h_{x,n}^r + h_{\ell,n}^i h_{x,n}^i \right) \right\} = 0$$

$$\text{var}\left\{\sum_{n=1}^{N}\left(h_{\ell,n}^{r}h_{x,n}^{r}+h_{\ell,n}^{i}h_{x,n}^{i}\right)\right\}=\frac{N\sigma_{h}^{4}}{2}\triangleq\vartheta_{1} \tag{A.4}$$

Due to the symmetry of the real and imaginary parts of the channel taps, the values of (A.4) also apply to the second term on the right side of (A.1). Consequently $|\rho_{\ell,x}|$ is a Rayleigh variable with $\mathbb{E}\left\{|\rho_{\ell,x}|\right\}=\sqrt{2\vartheta_{1}}\Gamma\left(\frac{3}{2}\right)=\frac{\sigma_{h}^{2}\sqrt{N\pi}}{2}$ and $\mathbb{E}\left\{|\rho_{\ell,x}|^{2}\right\}=2\vartheta_{1}\Gamma\left(2\right)=N\sigma_{h}^{4}$ [53], where $\Gamma\left(\cdot\right)$ is the gamma function such that $\Gamma\left(1\right)=1$, $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$, and $\Gamma\left(1+t\right)=t\Gamma\left(t\right)$.

For the case of $\ell=x$ we have

$$|\rho_{\ell,\ell}|=\sum_{n=1}^{N}|h_{\ell,n}|^{2}=\sum_{n=1}^{N}\left[\left(h_{\ell,n}^{r}\right)^{2}+\left(h_{\ell,n}^{i}\right)^{2}\right] \tag{A.5}$$

Since $h_{\ell,n}^{r},h_{\ell,n}^{i}\in\mathcal{N}_{\mathbb{C}}\left(0,\frac{\sigma_{h}^{2}}{2}\right)$, $|\rho_{\ell,\ell}|$ is a χ-square random variable with $2N$ degrees of freedom, i.e., $|\rho_{\ell,\ell}|\sim\chi_{2N}^{2}$, and $\mathbb{E}\left\{|\rho_{\ell,\ell}|\right\}=2N\vartheta_{2}=N\sigma_{h}^{2}$ and $\text{var}\left\{|\rho_{\ell,\ell}|\right\}=4N\vartheta_{2}^{2}=N\sigma_{h}^{4}$ [53] where $\vartheta_{2}\triangleq\text{var}\left\{h_{\ell,n}^{r}\right\}=\frac{\sigma_{h}^{2}}{2}$; therefore $\mathbb{E}\left\{|\rho_{\ell,\ell}|^{2}\right\}=\left[\mathbb{E}\left\{|\rho_{\ell,\ell}|\right\}\right]^{2}+\text{var}\left\{|\rho_{\ell,\ell}|\right\}=\sigma_{h}^{4}N\left(N+1\right)$.

Appendix B

List of Publications

B.1 Accepted Papers

1. **S. M. Razavi** and T. Ratnarajah, "Performance analysis of interference alignment under CSI mismatch," *To appear in IEEE Transactions on Vehicular Technology*.
2. **S. M. Razavi**, T. Ratnarajah, and C. Masouros, "Transmit-power efficient linear precoding utilizing known interference for the multiantenna downlink," *To appear in IEEE Transactions on Vehicular Technology*.
3. **S. M. Razavi** and T. Ratnarajah, "Interference alignment in K-user MIMO interference channels with partially coordinated receivers," *IET Communications*, vol. 8, no. 1, pp. 50–57, Jan. 2014.
4. **S. M. Razavi** and T. Ratnarajah, "Performance analysis of Max-SINR algorithm under CSI mismatch," *To appear in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
5. **S. M. Razavi** and T. Ratnarajah, "Asymptotic performance analysis of interference alignment under imperfect CSI," *To appear in IEEE Wireless Communications and Networking Conference (WCNC)*, 2014.
6. **S. M. Razavi**, T. Ratnarajah, C. Masouros, and M. Sellathurai "Regularized phase alignment precoding for the MISO downlink," *To appear in IEEE International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2014.
7. **S. M. Razavi** and T. Ratnarajah, "Interference alignment in partially coordinated multipoint receivers," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2013, pp. 1114–1118.
8. H. Zhou, **S. M. Razavi**, T. Ratnarajah, and M. Sellathurai, "Interference align-

ment with doubly layered signalling for constant SISO interference channels," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 4197–4201.

9. **S. M. Razavi**, T. Ratnarajah, "Interference alignment in coordinated multipoint systems," *Asilomar Conference on Signals, Systems, and Computers*, 2012, pp. 1576–1580.
10. **S. M. Razavi**, T. Ratnarajah, C. Masouros, and M. Sellathurai,"Joint interference and phase alignment in multiuser MIMO interference channels," *Asilomar Conference on Signals, Systems, and Computers*, 2012, pp. 1137–1141.
11. **S. M. Razavi** and T. Ratnarajah, "Enhanced linear precoding in coordinated multicell broadcast systems," *IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, 2012, pp. 249-253.
12. **S. M. Razavi** and T. Ratnarajah, "Subspace beamforming via block SVD for MIMO systems," *IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, 2012, 95–99.

Bibliography

[1] "Cisco VNI mobile forecast," 2013. [Online]. Available: http://www.cisco.com/

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[3] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[4] "Global System for Mobile Communications." [Online]. Available: http://en.wikipedia.org/wiki/GSM

[5] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[6] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[7] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[8] J. Maurer, J. Jaldén, D. Seethaler, and G. Matz, "Vector perturbation precoding revisited," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 315–328, Jan. 2011.

[9] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Computationally efficient vector perturbation precoding using thresholded optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1880–1890, May 2013.

[10] L. Sanguinetti and M. Morelli, "Non-linear pre-coding for multiple-antenna multiuser downlink transmissions with different QoS requirements," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 852–856, Mar. 2007.

[11] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Interference optimization for trans-

mit power reduction in Tomlinson-Harashima precoded MIMO downlinks," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2470–2481, May 2012.

[12] T. Haustein, C. von Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, "Performance of MIMO systems with channel inversion," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 1, May 2002, pp. 35–39.

[13] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 192–202, Jan. 2005.

[14] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, , and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[15] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[16] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[17] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

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

[19] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling—settling the Høst-Madsen–Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4552–4565, Sep. 2010.

[20] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[21] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Feb. 2012.

[22] S. W. Choi, S. A. Jafar, and S.-Y. Chung, "On the beamforming design for efficient interference alignment," *IEEE Commun. Lett.*, vol. 13, no. 11, pp. 847–849, Nov. 2009.

[23] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[24] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[25] H. Yu and Y. Sung, "Least squares approach to joint beam design for interference alignment in multiuser multi-input multi-output interference channels," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4960–4966, Sep. 2010.

[26] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4278–4288, Aug. 2012.

[27] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias, "Reweighted nuclear norm approach for interference alignment," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3754–3765, Sep. 2013.

[28] K. R. Kumar and F. Xue, "An iterative algorithm for joint signal and interference alignment," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 2293–2297.

[29] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. Asilomar Conf. Signal Syst. Comput. (ACSSC)*, 2009, pp. 1106–1110.

[30] I. Santamaria, O. Gonzalez, R. W. Heath, and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2010.

[31] S.-H. Park, H. Park, Y.-D. Kim, and I. Lee, "Regularized interference alignment based on weighted sum-MSE criterion for MIMO interference channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2010, pp. 1–5.

[32] H. Gao, J. Leithon, C. Yuen, and H. A. Suraweera, "New uplink opportunistic interference alignment: An active alignment approach," in *Proc. IEEE Wireless Commun. Net. Conf. (WCNC)*, 2013, pp. 3099–3104.

[33] S.-H. Park, H. Park, H. Sung, and I. Lee, "Regularized transceiver designs for multiuser MIMO interference channels," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2571–2579, Sep. 2012.

[34] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 2445–2448.

[35] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling

using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[36] C. Masouros, T. Ratnarajah, M. Sellathurai, C. B. Papadias, and A. K. Shukla, "Known interference in the cellular downlink: A performance limiting factor or a source of green signal power?" *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 162–171, Oct. 2013.

[37] A. D. Dabbagh and D. J. Love, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, Nov. 2008.

[38] V. K. Nguyen and J. S. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2008, pp. 1–4.

[39] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[40] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[41] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall, 1993.

[43] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[44] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 805–810, Mar. 2007.

[45] A. Edelman, "Eigenvalues and condition numbers of random matrices," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, Oct. 1988.

[46] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 185–194, Aug. 1999.

[47] A. Neumaier, "Solving ill-conditioned and singular linear systems: a tutorial on regularization," *SIAM Review*, vol. 40, no. 3, pp. 636–666, Sep. 1998.

[48] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore and

London: The Johns Hopkins Univ. Press, 1996.

[49] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2008–2026, Apr. 2011.

[50] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[51] A. Hjørungnes, *Complex-Valued Matrix Derivatives*. Cambridge University Press, 2011.

[52] M. R. McKay, I. B. Collings, and A. M. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 396–410, Jan. 2010.

[53] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2007.

[54] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252–262, Jan. 2011.

[55] C. Masouros and T. Ratnarajah, "Interference as a source of green signal power in cognitive relay assisted co-existing MIMO wireless transmissions," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 525–536, Feb. 2012.

[56] F. A. Khan, C. Masouros, and T. Ratnarajah, "Interference-driven linear precoding in multiuser MISO downlink cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2531–2543, Jul. 2012.

[57] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.

[58] X. Shao, J. Yuan, and Y. Shao, "Error performance analysis of linear zero forcing and MMSE precoders for MIMO broadcast channels," *IET Commun.*, vol. 1, no. 5, pp. 1067–1074, Oct. 2007.

[59] J. Thukral and H. Bolcskei, "Interference alignment with limited feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2009, pp. 1759–1763.

[60] R. T. Krishnamachari and M. K. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3908–3917, Aug. 2013.

[61] J.-S. Kim, S.-H. Moon, S.-R. Lee, and I. Lee, "A new channel quantization strategy for MIMO interference alignment with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 358–366, Jan. 2012.

[62] X. Rao, L. Ruan, and V. K. N. Lau, "CSI feedback reduction for MIMO interference alignment," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4428–4437, Sep. 2013.

[63] S. Cho, K. Huang, D. K. Kim, V. K. N. Lau, H. Chae, H. Seo, and B.-H. Kim, "Feedback-topology designs for interference alignment in MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6561–6575, Dec. 2012.

[64] J. Leithon, C. Yuen, H. A. Suraweera, and H. Gao, "A new opportunistic interference alignment scheme and performance comparison of MIMO interference alignment with limited feedback," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2012, pp. 1123–1127.

[65] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, Jun. 2011.

[66] O. El Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.

[67] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2009.

[68] J. Park, Y. Sung, and H. V. Poor, "On beamformer design for multiuser MIMO interference channels." [Online]. Available: http://arxiv.org/abs/1011.6121v1

[69] C. Wilson and V. V. Veeravalli, "A convergent version of the Max SINR algorithm for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2952–2961, Jun. 2013.

[70] R. Tresch, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," in *Proc. IEEE/SP Workshop Stat. Signal Process.*, 2009, pp. 277–280.

[71] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [coordinated and distributed MIMO]," *IEEE Wireless Commun. Mag.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[72] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[73] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 44–50, Nov. 2012.

[74] P. W. Wolniansky, G. J. Foschini., G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, 1998, pp. 295–300.

[75] H. Zhu, Z. Lei, and F. Chin, "On interference cancellation ordering of V-BLAST detectors," in *Proc. ICICS-PCM*, 2003, pp. 902–906.

[76] W.-T. Hsu and S. W. Kim, "Iterative post-successive interference cancellation in V-BLAST systems," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, 2007, pp. 1–6.

[77] B. Zhuang, R. A. Berry, and M. L. Honig, "Interference alignment in MIMO cellular networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2011, pp. 3356–3359.

[78] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, "User arrival in MIMO interference alignment networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 842–851, Feb. 2012.

[79] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[80] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[81] M. Razaviyayn, M. Baligh, A. Callard, and Z.-Q. Luo, "Joint user grouping and transceiver design in a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 85–94, Jan. 2014.

[82] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[83] S. A. Jafar, "Blind interference alignment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 3, pp. 216–227, Jun. 2012.