Queen's University Belfast

To be completed by Student

Degree:	PhD
Full Name:	Milad Fozooni

Thesis title:

Low-Cost Architectures for Future MIMO Systems

Summary: (max. 300 words)

Massive multiple-input multiple-output is a promising technique for the next generation of wireless communication systems which addresses most of the critical challenges associated with concurrent relaying systems, such as digital signal processing complexity, long processing delay, and low latency wireless communications. However, the deployment of conventional fully digital beamforming methods, dedicates one radio frequency (RF) chain to each antenna, is not viable enough due to the high fabrication/implementation cost and power consumption. In this thesis, we envision to address this critical issue by reducing the number of RF chains in a viable analog/digital configuration paradigm which is usually referred to hybrid structure.

From another viewpoint, the development of fifth generation enabling technologies brings new challenges to the design of power amplifiers (PAs). In particular, there is a strong demand for low-cost, nonlinear PAs which, however, introduce nonlinear distortions. On the other hand, contemporary expensive PAs show great power efficiency in their nonlinear region. Inspired by this trade-off between nonlinearity distortions and efficiency, finding an optimal operating point is highly desirable, and this is the second key contribution of this thesis.

To be completed by Examiner

EXAMINER CERTIFICATION OF SUBMITTED WORK

I hereby certify that this is the final accepted copy of the submitted work and that all required amendments have been completed and submitted within the required deadline.

Name of Examiner: Youngwook Ko

Signature of Examiner: **Date:** 6/11/2017

Student Declaration:

I give permission for my thesis to be made available, under regulations determined by the University, for inclusion in the University Library, consultation by readers in the School, inter-library lending for use in another library and to be photocopied, electronically reproduced and to be stored and made available publicly in electronic format
Please tick as appropriate:

i)   Immediately ☐
             Or

ii)   After an embargo period of   1 year ☑   2 years ☐   3 years ☐   4 years ☐   5 years ☐

Reason for embargo: (applies to both print and e-thesis)

☑   The thesis is due for publication, either as a series of articles or as a monograph

☐   The thesis includes material that was obtained under a promise of confidentiality

☐   Would substantially prejudice the commercial interests of the author, the University or an external company

☐   Contains information which may endanger the physical/mental health or personal safety of an individual(s)

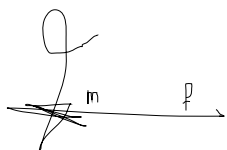**I wish to embargo the e-thesis copy permanently:** (applies to e-thesis only)

☐   The thesis contains material whose copyright belongs to a third party and the gaining of approval to publish the material electronically would be onerous or expensive; and the removal of the copyright material would compromise the thesis

---

CONFIRMATION OF DATA/HUMAN TISSUE SAMPLES HANDOVER

All research involving human participants, their tissue (e.g. blood, saliva, urine) or their data (interviews, consent forms, questionnaires) must be retained by the University for at least five years. These sources must be handed over to your supervisor. Please confirm, by ticking the appropriate box, that:

☐   I have provided my supervisor with all laboratory notebooks and/or primary source material pertaining to the study, including electronic data.

☐   I have identified for my supervisor the location of stored human tissue samples and provided an inventory of these.

☑   Due to the nature of the project I am not required to handover any data or samples relating to my thesis.

Signature of Student: **Date:** 1/11/2017

SUPERVISOR CONFIRMATION AND APPROVAL

I approve any embargo request above and confirm that the information given regarding Data/Human tissue samples is correct and that, where applicable*, the final copy of the thesis has also been made available in electronic format (e-thesis) via PURE.

Name of Supervisor: Dr Michail Matthaiou

Signature of Supervisor: **Date:**   6/11/2017

* RCUK funded students and <u>all</u> students who commence research from September 2016 must also submit an electronic copy of their thesis via Pure

Low-Cost Architectures for Future MIMO Systems

Milad Fozooni

School of Electronics, Electrical Engineering and Computer Science

Queen's University Belfast

A thesis submitted for the degree of

Doctor of Philosophy

20 August 2017

This page intentionally left blank.

Abstract

Massive multiple-input multiple-output is a promising technique for the next generation of wireless communication systems which addresses most of the critical challenges associated with concurrent relaying systems, such as digital signal processing complexity, long processing delay, and low-latency wireless communications. However, the deployment of conventional fully digital beamforming methods, dedicates one radio frequency (RF) chain to each antenna, is not viable enough due to the high fabrication/implementation cost and power consumption. In this thesis, we envision to address this critical issue by reducing the number of RF chains in a viable analog/digital configuration paradigm which is usually referred to hybrid structure.

From another viewpoint, the development of fifth generation enabling technologies brings new challenges to the design of power amplifiers (PAs). In particular, there is a strong demand for low-cost, nonlinear PAs which, however, introduce nonlinear distortions. On the other hand, contemporary expensive PAs show great power efficiency in their nonlinear region. Inspired by this trade-off between nonlinearity distortions and efficiency, finding an optimal operating point is highly desirable, and this is the second key contribution of this thesis.

Acknowledgements

I would like to thank my PhD supervisors, Dr. Michail Matthaiou and Dr. Trung Q. Duong for their precious time, patient encouragement, and invaluable advice. Indeed, the accomplishment of this study would not have been feasible without their kind support and cooperation. Also, I feel deeply honored in expressing my sincere thanks to Dr. Hien Q. Ngo for sharing his knowledge, worthwhile suggestions, and most importantly, his warm friendship during my studies. His unconditional support will always be remembered.

I would like to extend my heartfelt gratitude to my ever faithful parents, who have always been a source of moral and emotional support for me throughout my life and from start to the end of this dissertation. I would like to sincerely appreciate my sisters who have never left my side and are very special. Finally, my special feeling of gratitude goes out to my lovely wife, Faezeh, for all of her patience, devotion, and belief in me. All that I am or hope to be I owe to my family, and I feel deeply indebted to them. Indeed, without them, I would never have gotten to this point and this thesis would never have been written.

This thesis is dedicated to

my parents,

my sisters,

and my wife,

for being more than the sky to me.

Table of Contents

List of Tables

List of Figures

List of Notations

Non-bold letters denote scalars, while lower and upper case bold-face letters indicate vectors and matrices, respectively.

$\mathbb{C}^m$	set of complex $m \times 1$ vectors
$\mathbb{R}^m$	set of real $m \times 1$ vectors
$\mathbb{C}^{m \times n}$	set of complex $m \times n$ matrices
$\mathbb{R}^{m \times n}$	set of real $m \times n$ matrices
x, X	scalar
$|x|$	modulus of complex number x
$\measuredangle x$	phase of complex number x
$\mathbf{x}$	vector $\mathbf{x}$
$\mathbf{X}$	matrix $\mathbf{X}$
$[\mathbf{X}]_{i,j}$	(i, j)-th element of matrix $\mathbf{X}$
$\mathbf{x}_i$	i-th column of matrix $\mathbf{X}$
$(\cdot)^*$	complex conjugate
$(\cdot)^T$	transpose
$(\cdot)^H$	transpose conjugate
$\mathbf{X} \succeq \mathbf{0}$	positive semi-definite matrix
$\mathrm{Tr}(\mathbf{X})$	trace of the matrix $\mathbf{X}$
$\det(\mathbf{X})$	determinant of the matrix $\mathbf{X}$
$\lvert\lvert \cdot \rvert\rvert_2$	(vector argument:) Euclidian vector norm, i.e., $\lvert\lvert \mathbf{x} \rvert\rvert_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$

$\|\| \cdot \|\|_F$	(matrix argument:) Frobenius norm, i.e., $\|\|\mathbf{X}\|\|_F = \sqrt{\mathrm{Tr}\left(\mathbf{X}^H \mathbf{X}\right)}$
$\otimes$	Kronecker product
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{0}_N$	$N \times N$ matrix of all zeros
$\mathrm{diag}\left(a_1, \ldots, a_N\right)$	square diagonal matrix with $a_1, \ldots, a_N$ on the main diagonal
$\mathbb{E}_{\mathbf{X}}\left[\cdot\right]$	expectation of a random variable over $\mathbf{X}$ (subscript dropped when obvious)
$\mathrm{var}_{\mathbf{X}}\left(\cdot\right)$	variance of a random variable over $\mathbf{X}$ (subscript dropped when obvious)
$\mathcal{CN}\left(\mu, \boldsymbol{\Sigma}\right)$	complex Gaussian random vector with mean μ and covariance matrix $\boldsymbol{\Sigma}$
$\ln$	natural logarithm function function
$\lim$	limit operator
$\min\{\}, \max\{\}$	minimum and maximum function
$\lfloor \cdot \rfloor$	floor function
$\Gamma(\cdot)$	Gamma function
$\psi(\cdot)$	Euler digamma function
$\triangleq$	defined as
$\xrightarrow{\mathrm{a.s.}}$	almost sure convergence
$\xrightarrow{\mathrm{dist.}}$	convergence in distribution

List of Acronyms

3G third generation.

3GPP LTE Third Generation Partnership Project Long Term Evolution.

4G forth generation.

5G fifth generation.

ADC analog-to-digital converter.

AF amplify-and-forward.

AoA angle of arrival.

AoD angle of departure.

AWGN additive white Gaussian noise.

BS base station.

CAB constrained analog beamforming.

CF compress-and-forward.

CSI channel state information.

DAC digital-to-analog converter.

DF decode-and-forward.

DFT Discrete Fourier Transform.

DSP digital signal processing.

DTV digital television.

EAR ergodic achievable rate.

i.i.d. independent and identically distributed.

KKT Karush-Kuhn-Tucker.

LTE Long Term Evolution.

LTE-A LTE-Advanced.

MAN metropolitan area network.

MIMO multiple-input multiple-output.

MMSE minimum mean-squared error.

MRC maximum ratio combining.

MRC/MRT maximum ratio combining/maximum ratio transmission.

MU-MIMO multi-user MIMO.

PA power amplifier.

PDF probability density function.

SIMO single-input multiple-output.

SINR signal-to-interference-noise ratio.

SNR signal-to-noise ratio.

SU-MIMO single-user MIMO.

SVD singular value decomposition.

UAB unconstrained analog beamforming.

V-BLAST Vertical Bell Labs Space-Time.

WLAN wireless local area network.

ZF zero-forcing.

Chapter 1

Introduction

The widespread use of wireless communications has become an integral part of modern life. Wireless connectivity has become nearly ubiquitous due to the portable devices which are connected to the Internet all the time, almost anywhere. In the present days, the wireless communication technology offers a variety of new exciting applications such as voice, multimedia teleconferencing, online games, file transfer, and navigation. Moreover, the proliferation of such devices, of course, accelerates the process and consequently creates new traffic demands for the mobile operators and vendors. To handle the explosive demand for data transmission, a host of spectrally efficient transmission technologies should be developed for the next generations of cellular networks which at the same time have to be low-cost and power-efficient.

1.1   Cooperative Communications

The exponential growth of modern technologies and data traffic, along with the development of high speed vehicles and transportation, introduce new challenges for communication engineers to ensure ultra high-speed services with unprecedented reliability for the costumers. Indeed, the lack of sufficient bandwidth, power limitations, and high complexity of resource management should be treated in a cost-efficient way for the next generation of mobile networks.

In this sense, cooperative communication was first proposed for wireless appli-

cations in the seminal papers [2, 3]. After that, this technique has received a huge amount of research interest in Third Generation Partnership Project Long Term Evolution (3GPP LTE) and has been employed in LTE-Advanced (LTE-A) systems. Indeed, relay systems offer many benefits over conventional point-to-point networks [4]. This technique can combat the long-term path loss by effectively reducing the long distance between geographically separated user nodes. From this viewpoint, relaying is an energy efficient technique in wireless communication networks which can save a considerable amount of power by reducing the distance among the nodes. Moreover, relay nodes can circumvent intermediate obstacles, and consequently resolve the shadowing and blockage problem in an efficient way and ensure a reliable communication even in very high frequencies [5]. In addition, utilizing the wired backhaul for future ultra-dense networks has been recognized as a costly method which can be resolved by deploying relay nodes in future networks [6]. All in all, these properties render cooperative communication as a promising solution to overcome the data rate degradation and radio link outages in wireless communication networks. In the literature, there exist three prevalent relaying strategies, namely decode-and-forward (DF), compress-and-forward (CF), and amplify-and-forward (AF).

- DF relaying: In this wireless communication protocol, the relay receives the coded signals from sources, and decodes the message of each user all in one block. In the next block, the relay re-encodes the messages and forwards them toward the designated destinations. DF relaying is a hard decision protocol, as the forwarded signal does not include any additional information regarding the channel reliability of the source-relay link. Also, DF relaying is a resilient scheme against white noise, since it first cancels out the noise at the relay station, and then, sends the noise-free signal to the destinations [7].

- CF relaying: In this scheme, the relay station cannot decode the received signals from sources. However, it compresses and quantizes the observed signals in one block, and then forwards them toward the destination nodes in the following block. This type of relaying scheme can be further classified based on

2

compressing methods [8–10].

- AF relaying: This is the simplest relaying technique, where the relay station receives the signals from the sources in one time-slot, and then boosts whatever it receives including messages, noise and interference up to a certain level of power. Then, the relay station forwards the aggregated signal toward the destinations in the next time-slot. AF relaying requires less time, power, and computational complexity compared to the two other schemes as no decoding or compressing operation is applied. Since AF relaying works on a time-slot by time-slot basis, it exhibits less processing delay, and is most suitable to the high speed wireless communications. In this light, we consider the AF relaying protocol throughout of this thesis.

It is worth pointing out that by deploying more antennas at the relay station, the performance of relay networks can be further improved to enhance the area spectral efficiency and transmission reliability even more [11, 12]. This is due to the fact that multiple-input multiple-output (MIMO) techniques can offer a generous array and multiplexing gain by leveraging the spatial selectivity. However, we note that the great performance of MIMO relaying systems comes at the price of signaling and circuit complexity as multiple radio frequency (RF) chains are required at the relay stations.

1.2   MIMO Development

MIMO technology utilizes multiple antennas at both ends of a radio link to simultaneously transmit multiple data streams in wireless communication systems. This technology can support a high-speed and reliable link by exploiting the spatial degrees of freedom of the propagation channel. Today, MIMO is a relatively mature technique with a wide range of applications, such as digital television (DTV), wireless local area networks (WLANs), metropolitan area networks (MANs), and mobile communications. The great potential of MIMO systems has been recognized, for the first time, in the seminal papers [13–15]. The development of MIMO systems can be further clas-

sified in three symbiotic pillars: single-user MIMO (SU-MIMO), multi-user MIMO (MU-MIMO), and massive MIMO.

1.2.1 SU-MIMO

The development of MIMO appeared in the late 1990s [13–19]. SU-MIMO communication is the initial form of MIMO communications in which both the base station (BS) and a single user equipment are equipped with an array antenna. In this scheme, multiple antennas at the transmitter and receiver offer spatial selectivity, resulting in a great improvement in capacity, reliability, and resistance to interference. Although SU-MIMO can offer a generous multiplexing and beamforming power gain, however, its performance might be seriously limited by some practical considerations. First, dedicating many antennas to each user equipment significantly increases the burden of signal processing and implementation cost. Second, the advantage of multiplexing gain is deteriorated in the low power regime, hence, this kind of MIMO systems struggles to provide a descent throughput to cell-edge users which are more prone to interference from neighboring cells. Finally, when the propagation channel is ill-conditioned, the SU-MIMO performance is substantially degraded [20].

1.2.2 MU-MIMO

As a generalization of SU-MIMO, MU-MIMO techniques emerged and provided several key advantages over SU-MIMO communication systems. In MU-MIMO, the BS can simultaneously communicate with multiple single-antenna users in the same time-frequency blocks. From this viewpoint, the deployment of small and cheap user terminals is quite simpler than SU-MIMO communications in terms of signal processing and implementation cost. More fundamentally, MU-MIMO communication is more immune to the ill-conditioned propagation environments which are highly restrictive for the SU-MIMO ecosystem. Thanks to these great abilities, MU-MIMO has been employed into the 3GPP LTE, and evolving wireless broadband standards like forth generation (4G) Long Term Evolution (LTE) and LTE-A [21, 22]. Nevertheless, it is

notable that the advantages of MU-MIMO come at the cost of channel state information (CSI) acquisition, which requires substantial time or frequency resources to be set aside for pilot training and feedback signaling [23]. Hence, MU-MIMO is not a scalable technique with respect to the number of service antennas at the BS.

1.2.3 Massive MIMO

Massive MIMO is an enabling technique for the development of future wireless communications which can reap all the benefits of MU-MIMO, but on a much greater scale. Massive MIMO can handle orders of magnitude more data traffic by deploying the dozens or hundreds of antennas at the BS. Assuming that CSI is only required at the BS, and simple linear processing is used on both uplink and downlink, this technique is scalable with respect to the number of service antennas at the BS [24, 25]. Besides scalability, massive MIMO in cellular networks brings improvements in the following aspects [26–28]:

- Spectral efficiency: Deploying BSs with very large number of antennas can effectively amplify the both beamforming power gain and diversity gain, which consequently, can increase the capacity by 10 times or more.

- Signal processing: Massive MIMO avails of favorable propagation channel due to the channel hardening property [29]. Hence, very simple and linear signal processing like maximum ratio combining (MRC), zero-forcing (ZF), and minimum mean-squared error (MMSE) achieve the same performance of complicated non-linear signal processing methods, e.g, dirty paper coding.

- Robustness: Massive MIMO is a resilient technique against the small-scale fading and intentional jamming, by offering huge degrees of freedom and diversity gain.

- Energy efficiency: Due to the large number of service antennas in massive MIMO, the BSs can concentrate their radiated energy on particular destinations by forming very narrow beams. The more service antennas that are utilized,

the finer the spatial focusing can be. From this point, massive MIMO can enhance the energy-efficiency in the order of 100 times compared to conventional MU-MIMO systems.

- Latency: Latency on the air interface is one of the key challenges of concurrent wireless communication systems. It usually occurs when strong destructive interference happens in multipath environments, and the channel experiences deep fading. This phenomenon is more prohibitive in slow fading channels, when the channel may experience a deep fading and the receiver has to wait for a long-time for the channel to recover its good condition and provide a reasonable gain again. However, thanks to the law of large numbers, massive MIMO avails of the channel hardening with less gain fluctuation compared to conventional MIMO channels. Thus, it is unlikely that a destination terminal is trapped in a fading dip. Therefore, massive MIMO can naturally contribute to design of low-latency wireless links.

1.3 Massive MIMO Relaying

A considerable amount of attention has been paid to wireless network latency in recent years. The user plane latency achieved in LTE is approximately two times less than the corresponding latency in prevalent third generation (3G) technologies. This provides a direct service advantage for online and highly interactive application environments, such as sporting events, video conferencing, and multiplayer gaming. The effect of delay and latency is more pronounced in cooperative communications, where the relay node carries on the burden of complicated signal processing and power allocation algorithms. Therefore, cooperative relaying combined with massive MIMO, as a key enabling technology for future energy-efficient and low-latency communications, has received a great deal of research interest very recently.

In [30], the asymptotic performance of one-way massive relaying under perfect CSI was analyzed in three different power-scaling schemes. It was shown that the ra-

diated power of user and relay nodes can be scaled down inversely proportional to the number of relay antennas. Two-way massive relaying with the similar power strategy has been investigated in [31] evaluating the system performance in terms of both spectral efficiency and energy efficiency. In [32], the authors considered multipair two-way massive relaying, where perfect CSI is not available at the relay station. Then, they derived a closed-form expression for the achievable ergodic rate under MRC and ZF processing which provides an insightful scheme for simple power allocation. The energy efficiency of two-way relaying with unlimited relay nodes was also studied in [33]. Massive MIMO in the context of full-duplex relaying was investigated for the first time in [34], where the authors suppressed the interference loop in the spatial domain and optimized the energy efficiency. Also, some other efforts in this area can be found in [35–37].

Altogether, massive MIMO has the potential to address the challenges of relaying systems through channel hardening that averages out small-scale fading, and then reduce the latency considerably with a simple linear signal processing at the relay station. However, the extension of MIMO to massive MIMO relaying is not straightforward due to a plethora of practical constraints that are articulated in the following section.

1.4 Challenges and Motivations

1.4.1 Cost and Complexity

Although massive MIMO is a secure and robust technology which can address the high connectivity, growing capacity requirements, and low-latency communication in future fifth generation (5G) cellular networks, the practical implementation of massive MIMO induces some formidable challenges. In particular, having one RF chain per antenna element will boost to unprecedented levels the circuit complexity, fabrication/implementation cost and power consumption. In addition, wireless communication systems are continuously approaching higher carrier frequencies, and larger band-

widths. This technological shift involves much more complicated digital-to-analog converters (DACs), mixers, oscillators, and non-efficient power amplifiers (PAs) in each RF front-end. These critical issues were originally identified and addressed in [38, 39] by a cascade structure of an analog RF beamformer followed by digital baseband processor, which is normally referred to hybrid structure. The main idea stems from the fact that the multiplexing gain is restricted by the number of RF chains in MIMO systems, but, the array gain and diversity order can still be leveraged if suitable beamforming can be performed in the RF domain.

The emerging hybrid massive MIMO technology has recently attracted wide attention due to its potential benefits in cellular networks [40–43, 43–45]. Hybrid structures deployed in different fields of wireless communication systems can be found in Fig. 1.1. In these structures, the overall beamformer consists of a low dimensional digital baseband processor $\mathbf{F}_{\mathrm{BB}}$ and an analog RF beamformer $\mathbf{F}_{\mathrm{RF}}$ implemented by means of either switches or phase shifters. There are essentially three prevalent hybrid structures. A fully connected structure is illustrated in Fig. 1.1 (see subfigure *A*), where each RF chain is connected to all the antennas. In this model, the performance of the system may deteriorate due to the mutual coupling and interference among users, however this topology can provide the highest spectral efficiency and more accurate channel estimates [44]. In the partially connected structure, subfigure *B*, each RF chain is connected to a subset of array antennas. This topology is best suitable for the case that system suffers from high statistic and dynamic insertion loss. Different from fully and partially connected structure, topology *C* deploys many digital beamformers which is a generalization of the spatial sectorization in contemporary wireless communication systems. Note that we assume a fully connected structure throughout this thesis as we ignore the mutual coupling and insertion loss in our system model.

Although hybrid massive MIMO has already been introduced in different areas of wireless communications [39–45], to the best of our knowledge, there are few prior works that considered the hybrid massive MIMO structure in combination with cooperative communications [46, 47]. In [46], the authors developed an iterative sparse or-
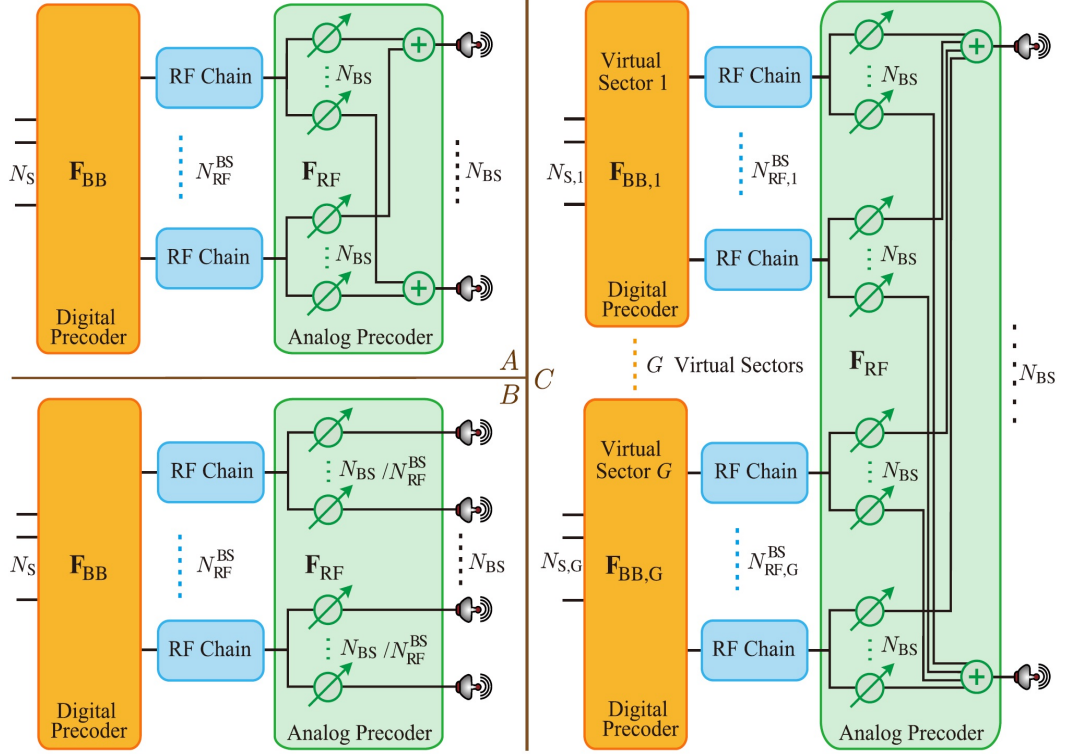
Figure 1.1: Typical block diagrams of hybrid beamforming structure at BS for a downlink transmission, where structure A, B, and C denote the fully connected, partially connected, and the virtual-sectorization structures, respectively [1].

thogonal matching pursuit algorithm to maximize the average rate for millimeter-wave MIMO systems. However, they did not consider a multiuser scenario, and also, they did not derive any closed-form expression for their achievable rate. In addition, the results are obtained under the assumptions of perfect CSI and accurate phase shifters which are too idealistic in reality. In [47], the authors evaluated the spectral and energy efficiency of a hybrid massive MIMO relaying for Rayleigh fading channels. They showed that the source and relay transmit power can respectively scale down by factors $\frac{1}{N}$ and $\frac{2K}{N}$ assuming that the pilot power is a constant parameter, where N and K denote the number of relay antennas and user pairs, respectively. Also, this paper reveals that the source and relay transmit power can scale down by $\frac{1}{\sqrt{N}}$ and $\frac{2K}{\sqrt{N}}$ if the pilot and payload data power are reduced in the same scale. We note that these results are under the assumptions that the propagation channels are uncorrelated, the number of RF chains (L) are exactly two times the number of user pairs, i.e. $L = 2K$, and also $N > \left\lfloor \frac{4L^2}{\pi} \right\rfloor$.

1.4.2 Power Amplifier Nonlinearities

There are increasing requirements for modern wireless communications systems to enhance the data rate. On the other hand, energy consumption becomes a crucial concern since wireless communication systems contribute significantly towards the global carbon footprint. To satisfy these ever conflicting requirements, emerging wireless communication systems are continuously using larger bandwidth, higher carrier frequencies, and energy efficient methods. This results in an ever increasing demand on the performance of RF chains, which at the same time should be power-efficient and low-cost. We recall that PAs consume about 57% of the operation energy in macro BSs [48]. Furthermore, it has been estimated that BSs dedicate 10% of their energy consumption just to cool the PAs. Hence, there is a great tendency to push the operating point of PAs closer to their nonlinear region, where the efficiency of PAs is much higher compared to the linear regime. From the other perspective, the number of PAs can be, and in many applications like massive MIMO must be, very high. Therefore, the design of extremely low-cost, low-complexity, and consequently, nonlinear PAs is of paramount importance.

Although there is a strong demand for low-cost PAs in future communication systems, the performance of MIMO systems can be severely degraded by the nonlinearities introduced by these inexpensive PAs. This issue has been recently studied in a few related papers. For instance, the impact of PAs nonlinearity was considered in [49], where the authors evaluated the channel estimation under nonlinearity distortions, and then proposed a novel quantized approach to minimize the bit-error-rate, and consequently, improve the spectral efficiency. A constellation-based compensation method is also proposed in [50] to reduce the in-band distortion for high-power amplifiers. There are also some other research efforts like [51–56], which have dealt with the PA nonlinearities in different wireless applications, but there is a dearth of literature considering the PAs nonlinearity in the context of MIMO systems.

1.5   Contributions

Inspired by the above discussion, in this thesis, we first consider a multipair AF relaying with hybrid structure at the relay station. Then, we investigate the performance assessment of the system under perfect CSI in three different power regimes. Moreover, we develop closed-form expressions for the asymptotic spectral efficiency, and provide some insightful tradeoffs between the number of service antennas and the transmitted power either in user nodes or at the relay station. In order to have a low-cost implementation of analog beamformers, we also relax the accuracy constraints of phase shifters to an arbitrary number of quantization bits. Then, we mathematically evaluate the impact of this relaxation which is a case of practical interest. In this scenario, analog beamformers should be able to adapt to the quick variations of instantaneous channels over time. This adaptation, not only requires perfect CSI at the relay station, but also is a challenging task due to less flexibility of analog beamformers compared to the digital beamformers. To have a better understanding of the performance of the proposed system with imperfect CSI, we assume a hybrid beamforming where the analog beamformer is a bank of phased shifters with the simplest possible phases, i.e. $\pm\pi$. Finally, we take the channel correlation into account and design our hybrid beamformer contains of a statistical-based analog beamformer followed by simple digital beamformers, i.e. either MRC or ZF processor. Our analytical results reveal that although multiplexing gain and power gain are restricted due to the limited number of RF chains and perfect CSI is not available at the relay station, the proposed scheme can still exploit the diversity gain and improve the spectral efficiency.

In addition, we consider a new system model that incorporates a general case by taking the PAs nonlinearities into account. We expand our analytical derivations to find how the desired signal and distortion interact together under a nonlinearity PAs. In order to optimize the spectral efficiency, we will go through the power allocation optimization, and then reduce it to a simple power control optimization problem. We finally derive simple lower and upper bounds on the original optimization problem which are quite useful for further research in this field. This thesis makes the following

specific contributions which are cited by our recent papers in these scopes:

1. We consider a multipair massive MIMO relaying, and propose a hybrid structure includes an online RF analog beamformer and a baseband digital processor. Then, we study the system performance in three different power regimes under the assumption that perfect CSI is available at the relay station [57].

2. We expand our analytical results for a practical case with an arbitrary quantization bit. Our numerical results reveal that our system is robust to coarse quantization and even with 2-bit resolution, the proposed configuration can capture more than 90% of the achievable rate offered by an unrealistic system with ideal phase shifters [57].

3. We assume a worst case scenario, where only imperfect CSI is available at the relay station, and analog beamformers are invariant over time with only two phases $\pm\pi$. Our analytical results show that with only half of the RF chains the system can achieve 75% of the achievable rate of fully digital structure [58].

4. Since analog beamforming restricts the performance of relay systems with channel correlation, we develop a correlation-based analog beamformer followed by linear digital processors. Then, we find a new lower bound and approximation bound on the achievable rate and evaluate the performance of the system for different correlation setups [59].

5. We assume a point-to-point MIMO communication link, and then evaluate the impact of nonlinear PAs on the ergodic achievable rate. We analyze how the desired signal and distortion affect the system performance and provide some insightful bounds and asymptotic results [60].

1.6   Organization of the Thesis

The thesis consists of six chapters which are outlined as follows. In Chapter 3, we consider a multipair AF relay system with a hybrid topology at the relay station. We

assume that perfect CSI is available only at the relay station while this information is not available at the user nodes. After that, we adjust the analog beamformers to cancel out the phase of channels at the relay station, and then we complete the processing at the digital signal processing (DSP) unit which utilizes the maximum ratio combining/maximum ratio transmission (MRC/MRT) scheme. To have a case of practical interest, we assume that the phase shifters have a limited resolution, and then figure out the achievable rate of the proposed system under this assumption.

Since the analog beamformer is implemented in the RF domain with less flexibility than the baseband processing, we assume the worst case in Chapter 4, where the analog beamformer is constant over time, and consists of phase shifters with phases $\pm\pi$. This assumption reduces the channel size to the number of RF chains and restricts the impact of array gain. However, it avails of a very simple structure which describes the basic relation between the spectral efficiency, number of RF chains, and active users, particularly in cases that only imperfect CSI is available at the relay station.

In Chapter 5, we consider a correlated channel, and adjust the analog beamformers to the variations of channel statistics rather than the instantaneous CSI. In this scenario, we estimate the propagation channels with uplink orthogonal pilots, and then design the analog beamformers to minimize the estimation error, and consequently improve the achievable rate. We develop new bounds on the achievable rate for both MRC/MRT and ZF schemes.

In Chapter 6, we evaluate the impact of low-cost PAs in the context of point-to-point MIMO communication links, where the modeling and analysis of PAs nonlinearities is extremely appealing. Indeed, PAs may operate within their nonlinear region, particularly when they are working in the high-power signal regimes to achieve higher efficiency. In this chapter, we simplify the nonlinear behavior of PAs by invoking the Bussgang's theorem. Next, we evaluate the achievable rate of the system, where we reduce our power allocation optimization to a simple power control problem, and then derive a closed-form expression on the achievable rate. As the closed-form expression is not very insightful, we derive lower and upper bounds on the achievable rate and

also analyze the performance of system in the low and high power regime.

Chapter 7 summarizes some important conclusions obtained from this thesis, and also gives some possible avenues for future work.

Chapter 2

Background

In this chapter, we provide a brief overview of the fundamental concepts of wireless communication systems which will be often used in the subsequent chapters. First, we briefly introduce the propagation manifestations which will be frequently encountered in our system model. Then, we explicitly discuss the main concepts of diversity order and multiplexing gain which are the key features of MIMO systems. Finally, we explain the linear MIMO receivers which are very popular due to their easy implementation.

2.1 Path Loss

The path loss is the main factor of signal power reduction, when the signal propagates through space. This natural phenomenon is commonly expressed by invoking the classical power law in dB

$$P_L(dB) = 10\log_{10}\left(\frac{P_t}{P_R}\right), \tag{2.1}$$

where P_t and P_R denote the signal power at transmitter and receiver, respectively. A simplified version of this formula can be expressed as

$$P_r = P_t K \left(\frac{d}{d_0}\right)^{-\nu} \qquad \text{for } d > d_0. \tag{2.2}$$

15

In this equation, K is a unitless constant which depends on the average channel attenuation and antenna properties, d is the distance between transmitter and receiver antennas while d_0 is a reference distance which normally takes a value in the range of $10-100\ m$ for outdoor communications and between $1-10\ m$ for indoor communications. The symbol ν represents the path loss exponent which is typically between $2-6$, where the 2 describes free space propagation. In general, a more cluttered environment leads to a higher path loss exponent [61].

2.2   Multipath Fading Channel

The wireless channel is highly unstable due to the scatterers, reflectors, and obstacles in the physical environment. Each scatterer has the potential to create a new copy of the transmitted signal and this phenomenon is well-known as multipath in wireless communications. Hence, multiple copies of a transmitted signal are received at the destination. Each signal travels through a different path, hence, they experience different attenuation, phase shifting, and delay. Hence, these signals may constructively or destructively add together and this intrinsically leads to a variation of the phase and amplitude of the aggregated signal. This property is referred as channel fading which can be divided into large-scale fading and small-scale fading.

Large-scale fading is a consequence of path loss and shadow fading which will be explained in following section. Small-scale fading can be classified from two different viewpoints: frequency and time domain. In frequency domain, the channel fading is flat if the signal bandwidth is less than the channel coherence bandwidth. Otherwise, the channel exhibits frequency selective fading. In the time domain, the channel is slow fading if the channel coherence time is larger relative to the delay requirement of the application. At the other extreme, a fast fading channel can appear if the transmitter, obstacles or receivers are moving very fast with respect to each other so that the Doppler phenomenon is more pronounced. However, we note that in a fast fading channel, the communication system avails of the variations in the channel conditions, i.e., time selectivity, to enhance the communication reliability using suitable equaliza-

tion schemes [62].

2.3   Shadow Fading Channel

Shadowing is often introduced by topographical elements and big objects, such as a hill or large buildings that interact with propagation paths over relatively large distances. In fact, an obstacle with depth of d can exponentially attenuate the transmitted signal so that

$$s(d) = \alpha \exp\left(-\beta d\right) \tag{2.3}$$

where α is an adjusting coefficient and β is an attenuation constant, which depends on the obstacle material. Now, let us assume that there exist M obstacles so the total attenuation can be expressed as

$$s(d_t) = \alpha \exp\left(-\sum_{i=1}^{M} \beta_i d_i\right), \tag{2.4}$$

where the index i refers to the i-th object, and d_t represents the sum of the obstacles depths through which the signal travels. In a propagation channel with many obstacles between transceivers we can utilize the law of large numbers and conclude that $\sum_{i=1}^{M} \beta_i d_i$ can be treated as a Gaussian random variable. From this viewpoint, we can claim that the logarithm function of shadow fading follows a Gaussian distribution, and consequently, shadow fading can be commonly modeled by a log-normal distribution as

$$f_\Psi(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{\mathrm{dB}}}\psi} \exp\left(-\frac{\left(\psi_{\mathrm{dB}} - \mu_{\psi_{\mathrm{dB}}}\right)^2}{2\sigma_{\psi_{\mathrm{dB}}}^2}\right), \qquad \text{for } \psi \geq 0, \tag{2.5}$$

where $\xi = \frac{10}{\ln(10)}$ and $\psi_{\mathrm{dB}} = 10\log_{10}(\psi)$. Also, $\mu_{\psi_{\mathrm{dB}}}$ and $\sigma_{\psi_{\mathrm{dB}}}$ represent the mean value and standard deviation of ψ_{dB}, respectively [63].

2.4   Rayleigh Fading Channel

In the case that there is no dominant line-of-sight between the transmitter and receiver, and the environment is rich enough to provide many propagation paths, the Rayleigh distribution can well describe the small-scale fading of transmitted signal. This model relies on the fact that there exist many reflectors and consequently paths in the channel. Hence, the real and imaginary part of the signal can be accurately modeled via a Gaussian distribution according to the law of large numbers. Hence, the wireless channel can be described by $h = h_R + jh_I$ where h_R and h_I are Gaussian random variables with distribution $\mathcal{N}(0, \sigma^2)$. In this model, the envelope of the aforementioned channel $R = \sqrt{h_R^2 + h_I^2}$ is a Rayleigh random variable with the following probability density function (PDF):

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r}{2\sigma^2}\right). \tag{2.6}$$

2.5   Spatial Channel Correlation

MIMO systems avail of spatial diversity to ensure a reliable connection among the transceivers by sending copies of the same data stream through independent channels. In addition, MIMO systems can greatly improve the capacity of wireless communication systems by proper beamforming methods that steer the radiated energy into the intended directions. However, both diversity gain and capacity achievement are significantly influenced by the channel spatial correlation [64, 65]. Spatial channel correlation in MIMO systems is mostly due to the short antenna separations and lack of reflectors and scatterers around the array antennas [65, 66]. Precisely speaking, space limitation either in mobile devices or BSs leads to the channel correlation as adjacent antennas receive approximately similar data streams. Furthermore, the BSs are usually located on rooftops to have a line-of-sight channel with the users. However, due to the lack of reflectors in such heights, MIMO systems may exhibit high spatial correlation. Considering a Rayleigh fading channel, one can decompose the propagation channel based

on the Kronecker model

$$\mathbf{G} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H} \left(\mathbf{R}_t^{\frac{1}{2}} \right)^H, \tag{2.7}$$

where $\mathbf{H}$ is a random matrix with i.i.d. $\mathcal{CN}(0, 1)$ elements. Also, $\mathbf{R}_r$ and $\mathbf{R}_t$ characterize the spatial correlation at the receiver and transmitter side, respectively. Under this channel model, the propagation channel can be rewritten as

$$\mathbf{H} \sim \mathcal{CN} \left(\mathbf{0}, \mathbf{R}_t \otimes \mathbf{R}_r \right). \tag{2.8}$$

It is worth pointing out that the Kronecker model is not the most accurate model, but it can provide a tractable analytical expression and insightful observation of the channel spatial correlation [67,68]. There are many different models to characterize the spatial correlation matrices, $\mathbf{R}_r$ and $\mathbf{R}_t$, depending on the application scenario. In general, the classical models focus on the distribution of signal power and Doppler shift effects. However, there are more elaborated correlation models, build on the principles of the classical channels, which take other factors, such as angle of arrival (AoA), angle of departure (AoD), and time delay into account. In Chapter 5, we utilize the Gaussian AoA and AoD models, which involve matrices with a spread inversely proportional to the product of the antenna spacing and angle spread. A detailed discussion of these models has already been provided in [64].

2.6 Diversity Gain

Wireless communication channels often experience a very large power penalty due to the large-scale (path-loss and shadowing) and small-scale (Rayleigh) fading. Hence, the error probability decays badly with the signal-to-noise ratio (SNR) due to the fact that the propagation channel can be in a deep fade in slow fading channels. One prominent solution is to increase the number of independent paths that information symbols pass through. Thus, the reliable communication can be ensured as long as there exists

at least one strong path between the transmitter and receiver. This technique is called diversity which can be achieved in many different domains such as time, frequency, macro-cell, and space. Regardless of the chosen diversity technique, the diversity gain (d) is defined as

$$d = \lim_{\text{SNR}\to\infty} -\frac{\log_2 P_e(\text{SNR})}{\log_2(\text{SNR})}, \tag{2.9}$$

where $P_e(\cdot)$ denotes the error probability function [69].

Diversity over time domain can be obtained by spreading the information symbols over different symbols or coherence blocks. In a similar spirit, frequency diversity is best suitable for frequency-selective channels where symbols can be sent through independent channels in frequency domain. Also, diversity can be achieved in cellular networks when users have this opportunity to choose a BS with the better channel conditions. This kind of diversity is called macro-diversity. Finally, spatial diversity can be achieved by deploying more than one antenna at transceivers. It has been shown that by having N_t service antennas at the transmitter and N_R service antennas at the receiver side, a MIMO system can extract up to $N_t N_r$ diversity gain in independent and identically distributed (i.i.d.) Rayleigh fading channels [70].

2.7   Spatial Multiplexing Gain

It has been recognized that MIMO systems under suitable channel fading conditions can simultaneously transfer multiple orthogonal data streams by utilizing separate spatial dimensions. This results in a substantial capacity gain at no additional power or bandwidth. Spatial multiplexing gain had a significant impact on introducing the MIMO systems in wireless technology. Based on the definition, a MIMO system achieves a spatial multiplexing gain r if

$$r = \lim_{\text{SNR}\to\infty} \frac{R(\text{SNR})}{\log_2(\text{SNR})}, \tag{2.10}$$

where $R(\cdot)$ is the data rate function [69]. It can be shown that, in i.i.d. Rayleigh fast fading channels, the capacity of MIMO system with N_t transmit antennas and N_r receive antennas linearly scales with minimum number of antennas in high SNR, i.e. $R(\text{SNR}) = n_{\min} \log_2(\text{SNR})$, where $n_{\min} = \min\{N_t, N_r\}$. This result implies that this system model can be decomposed to $n_{\min}$ parallel spatial channels which have the potential to simultaneously transfer $n_{\min}$ independent data streams.

2.8 Diversity-Multiplexing Tradeoff

The outage probability evaluates the performance of wireless communication system under slow fading channels. In this sense, diversity gain is the most important benchmark that explains how fast the error probability decays at high SNR. On the other hand, multiplexing gain demonstrates the performance of any MIMO system under fast fading channels. It is shown that there exists a trade off between diversity gain and multiplexing gain [69]. Fig. 2.1 showcases the optimal diversity gain d^* as a function of the multiplexing gain r in an i.i.d. Rayleigh fading channel. It has been proved that for a block code of length L, the function $d^*(r)$ is a pairwise linear function which connects the following points:

$$\Big(k, \big(N_t - k\big)\big(N_r - k\big)\Big), \qquad \text{for } k = 0, 1, \ldots, n_{\min}, \tag{2.11}$$

where it is assumed that $L \geq N_t + N_r - 1$ [69].

This theoretical result sheds light on how diversity and spatial multiplexing gain contribute to the performance of MIMO systems. On the one hand, MIMO system avails of maximum possible diversity gain $N_t N_r$ at the price of very limited multiplexing gain $(r \to 0)$. This strategy works well in slow fading channels where outage events, and consequently, diversity gain is much of a concern. However, it falls short to address the high data rate requirements in fast fading environments. On the other hand, MIMO system can capture full multiplexing gain $\big(r \to \min\{N_t, N_r\}\big)$ when the propagation channel varies very fast, but this system is not resilient against the deep
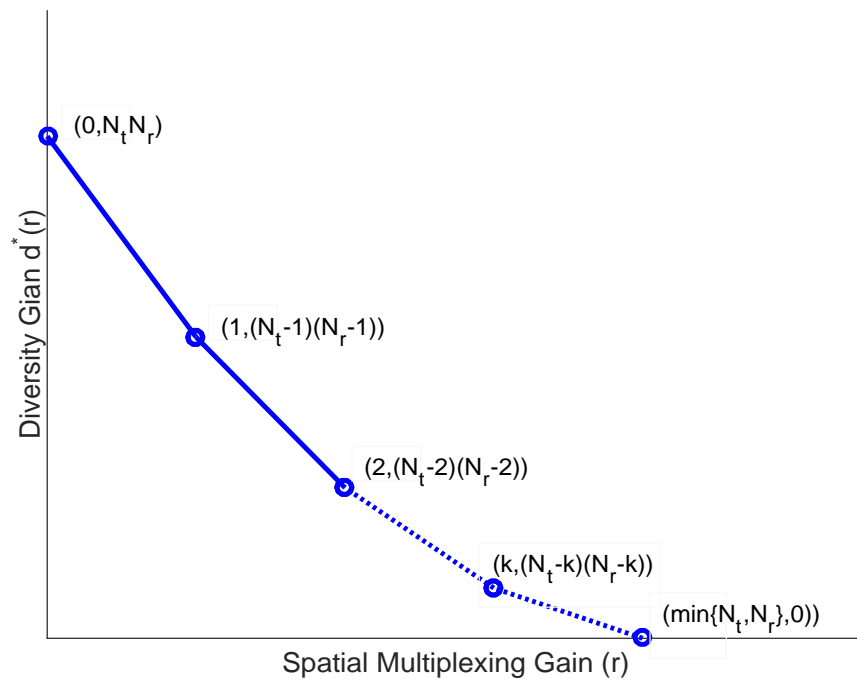
Figure 2.1: Diversity-multiplexing gain tradeoff for MIMO systems with N_t transmit antennas and N_r receive antennas.

fades of slow fading channels anymore. This analytical result provides other insightful results as well. For instance, adding one extra antennas at both the transmit and receive side enhances the multiplexing gain by 1, while the system still enjoys the same diversity gain as before.

2.9 MIMO Receiver Architectures

It is well recognized that Vertical Bell Labs Space-Time (V-BLAST) architecture can exploit the maximum diversity gain of a point-to-point MIMO system. By performing a singular value decomposition (SVD) of the MIMO channel, the transmitter pre-rotates the data streams and sends them trough the eigenmodes of the propagation channel toward the destination. Then, the receiver can recover the data streams without interference by utilizing the right singular vector matrix of the channel matrix. Although, V-BLAST is an optimal structure, perfect CSI is required at both transmitter and receiver side which is not a case of practical interest in terms of cost, complexity,

feedback and overhead signaling. Hence, various techniques have been already developed which only require the CSI at the receiver side. These so-called combiners can be divided in two general groups: linear and nonlinear receivers. Due to the low-cost nature of linear receivers, hereafter, we just focus on this group of receivers and briefly introduce the main linear receivers in the context of MIMO systems.

2.9.1 Maximum Ratio Combiner

Let us assume that $\mathbf{H} = \left[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_t}\right]$ denotes the flat fading MIMO channel and vector $\mathbf{x} = \left[x_1, x_2, \ldots, x_{N_t}\right]^T$ is the information symbol. Then, the received signal at the receiver of a point-to-point MIMO system is given by

$$\mathbf{y} = \mathbf{Hx} + \mathbf{w}, \tag{2.12}$$

or equivalently,

$$\mathbf{y} = \sum_{i=1}^{N_t} \mathbf{h}_i x_i + \mathbf{w}, \tag{2.13}$$

where $\mathbf{w}$ is a Gaussian random vector with i.i.d. $\mathcal{CN}\left(0, N_0\right)$ entries which represents the additive noise. It is clear that the k-th symbol can be extracted as follows

$$\mathbf{y} = \mathbf{h}_k x_k + \sum_{i \neq k}^{N_t} \mathbf{h}_i x_i + \mathbf{w}. \tag{2.14}$$

The MRC technique weighs the received signal in each branch proportional to the channel gain and coherently adds them together to maximize the desired signal power. It can be shown that the maximum SNR can be achieved if each antenna branch multiplies its received signal by the conjugate transpose of its channel vector. Then, the k-th output of MRC combiner is given by

$$\frac{\mathbf{h}_k^H}{\|\mathbf{h}_K\|}\mathbf{y} = \|\mathbf{h}_K\| x_k + \sum_{i \neq k}^{N_t} \frac{\mathbf{h}_k^H \mathbf{h}_i}{\|\mathbf{h}_K\|} x_i + \frac{\mathbf{h}_k^H}{\|\mathbf{h}_K\|}\mathbf{w}, \tag{2.15}$$

where we note that $\frac{\mathbf{h}_k^H}{\|\mathbf{h}_K\|}\mathbf{w}$ has the same distribution as the noise vector $\mathbf{w}$. This is due this fact that $\mathbf{w}$ has a symmetric distribution and is invariant to the rotational operator. It is noteworthy that the MRC scheme is an optimal receiver in terms of achievable data rate if the unwanted signal, including noise and interference, is a vector of uncorrelated random variables. Hence, in the low SNR regime, where the interference signal is negligible compared to the noise power, the MRC receiver approaches the optimal linear MMSE.

2.9.2 Zero-Forcing

The ZF or decorrelator is a linear algorithm which removes the interference from the received signal. To this end, the receiver projects all the received signal including desired signal, noise and interference, into a subspace which is orthogonal to the interference subspace. This method can be easily done by pre-multiplying the pseudo-inverse of propagation channel on equation (2.12)

$$\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{y} = \mathbf{x} + \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{w}. \tag{2.16}$$

This result implies that the ZF receiver is a powerful technique to remove the interference. Nevertheless, the ZF scheme will boost the noise power, and induce a high error probability into the system. Therefore, this scheme should be utilized in high SNR regime to ensure a reliable performance. Also, we note that, in a special case that the propagation channel matrix is square and full rank, i.e., number of transmit and receive antennas are equal, the ZF receiver simply reduces to the channel inversion receiver $\mathbf{H}^{-1}$.

2.9.3 Minimum Mean-Squared Error

It has been proved that the MRC receiver is the optimal linear receiver for single-input multiple-output (SIMO) systems with additive white noise. Motivated by this observation, MMSE receiver generalizes this idea to the MIMO channels with inter-

stream interference, where the aggregated noise and interference are not necessarily uncorrelated (white) with the desired signal. In fact, the MMSE receiver first utilizes a whitening strategy to transform the correlated noise and interference into a white additive vector. Then, by invoking the MRC scheme it follows the same optimal strategy of MISO receivers. Let us recall the original signal model of the k-th data stream

$$\mathbf{y} = \mathbf{h}_k x_k + \underbrace{\sum_{i \neq k}^{N_t} \mathbf{h}_i x_i + \mathbf{w}}_{\text{unwanted } \mathbf{z}_k}, \tag{2.17}$$

where the correlation matrix of the unwanted signal can be described as

$$\mathbf{R}_{Z_k} = N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H. \tag{2.18}$$

In this equation, we define P_i as the power of the i-th data stream. Then, we multiply the received signal vector in k-th branch by $\left(\mathbf{R}_{Z_k}\right)^{-\frac{1}{2}}$ to get the uncorrelated unwanted signal

$$\left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-\frac{1}{2}} \mathbf{y} = \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-\frac{1}{2}} \mathbf{h}_k x_k + \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-\frac{1}{2}} \mathbf{z}_k. \tag{2.19}$$

Now, we apply the MRC strategy and scale up the desired signal according to its weight

$$\mathbf{h}_k^H \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-1} \mathbf{y} = \mathbf{h}_k^H \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-1} \mathbf{h}_k x_k$$

$$+ \mathbf{h}_k^H \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-1} \mathbf{z}. \tag{2.20}$$

Altogether, the MMSE receiver applies the following linear transformation

$$\mathbf{h}_k^H \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H\right)^{-1}, \tag{2.21}$$

and offers the following corresponding signal-to-interference-noise ratio (SINR) in the k-th receive antenna

$$\mathrm{SINR}_k = P_k \mathbf{h}_k^H \left(N_0 \mathbf{I} + \sum_{i \neq k} P_i \mathbf{h}_i \mathbf{h}_i^H \right)^{-1} \mathbf{h}_k. \tag{2.22}$$

2.10 Linear MMSE estimator

Linear MMSE estimator has been recognized as a simple and accurate estimator which is widely used in wireless communication systems [71]. In this section, we introduce the fundamentals of this estimator in MIMO systems. Assume that $\mathbf{G} \in \mathbb{C}^{N \times K}$ is a propagation channel so that $\mathbf{G} = \mathbf{R}^{\frac{1}{2}} \mathbf{H}$, where $\mathbf{H} \in \mathbb{C}^{N \times K}$ denotes the small-scale fading, and $\mathbf{R} \in \mathbb{C}^{N \times N}$ represents a known correlation matrix. Then, we define the observation matrix $\mathbf{Y}$

$$\mathbf{Y} = \sqrt{p} \mathbf{F} \mathbf{G} + \mathbf{N}, \tag{2.23}$$

where $\mathbf{F} \in \mathbb{C}^{M \times N}$ is a given beamforming matrix, $\mathbf{N} \in \mathbb{C}^{M \times K}$ is an AWGN matrix including i.i.d. $\mathcal{CN}(0,1)$ entries, and p is a known constant which controls the SNR in this model. Hereafter, we estimate the l-th column of $\mathbf{G}$ by a linear transformation $\hat{\mathbf{g}}_l = \mathbf{X}_l \mathbf{y}_l$ to minimize the quadratic error function $\epsilon_l = \mathbb{E} \left[\| \mathbf{g}_l - \hat{\mathbf{g}}_l \|^2 \right]$

$$\min_{\mathbf{X}_l} \mathbb{E} \left[\| \mathbf{g}_l - \hat{\mathbf{g}}_l \|^2 \right], \quad \text{for } = 1, 2, \ldots, K. \tag{2.24}$$

The error function ϵ_l can be extended in the following manner:

$$\begin{aligned}
\mathbb{E} \left[\| \mathbf{g}_l - \hat{\mathbf{g}}_l \|^2 \right] &= \mathbb{E} \left[\| \mathbf{g}_l - \mathbf{X}_l \mathbf{y}_l \|^2 \right] \\
&= \mathbb{E} \left[\mathrm{Tr} \left(\mathbf{g}_l - \mathbf{X}_l \mathbf{y}_l \right) \left(\mathbf{g}_l - \mathbf{X}_l \mathbf{y}_l \right)^H \right] \\
&= \mathrm{Tr} \left(\mathbb{E} \left[\mathbf{g}_l \mathbf{g}_l^H \right] \right) - \mathrm{Tr} \left(\mathbb{E} \left[\mathbf{g}_l \mathbf{y}_l^H \right] \mathbf{X}_l^H \right) \\
&\quad - \mathrm{Tr} \left(\mathbf{X}_l \mathbb{E} \left[\mathbf{y}_l \mathbf{g}_l^H \right] \right) + \mathrm{Tr} \left(\mathbf{X}_l \mathbb{E} \left[\mathbf{y}_l \mathbf{y}_l^H \right] \mathbf{X}_l^H \right).
\end{aligned} \tag{2.25}$$

Next, we find the minimum point of the aforementioned error function by finding the roots of its first derivative, i.e., $\frac{\partial \epsilon_l}{\partial \mathbf{X}_l} = 0$ which yields

$$-2\mathbb{E}\Big[\mathbf{g}_l \mathbf{y}_l^H\Big] + 2\mathbf{X}_l \mathbb{E}\Big[\mathbf{y}_l \mathbf{y}_l^H\Big] = 0, \qquad (2.26)$$

and consequently, we find the linear MMSE estimate of the channel as follows

$$\mathbf{X}_l = \left(\mathbb{E}\Big[\mathbf{g}_l \mathbf{y}_l^H\Big] \right) \left(\mathbb{E}\Big[\mathbf{y}_l \mathbf{y}_l^H\Big] \right)^{-1}, \qquad (2.27)$$

which can be further simplified by leveraging the properties of Gaussian matrices

$$\mathbf{X}_l = \sqrt{p}\mathbf{R}\mathbf{F}^H \Big(p\mathbf{F}\mathbf{R}\mathbf{F}^H + \mathbf{I}_M \Big)^{-1}, \qquad (2.28)$$

which results in

$$\hat{\mathbf{g}}_l = \sqrt{p}\mathbf{R}\mathbf{F}^H \Big(p\mathbf{F}\mathbf{R}\mathbf{F}^H + \mathbf{I}_M \Big)^{-1} \mathbf{y}_l. \qquad (2.29)$$

It is notable that $\frac{\partial \epsilon_l}{\partial \mathbf{X}_l} = 2\mathbb{E}\Big[\mathbf{y}_l \mathbf{y}_l^H\Big]$ is a positive definite matrix. Therefore, the quadratic error function is a convex function which is minimized by the calculated $\mathbf{X}_l$. Finally, we can represent all the estimations into the following equation

$$\hat{\mathbf{G}} = \Big[\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \ldots, \hat{\mathbf{g}}_K\Big] = \sqrt{p}\mathbf{R}\mathbf{F}^H \Big(p\mathbf{F}\mathbf{R}\mathbf{F}^H + \mathbf{I}_M \Big)^{-1} \mathbf{Y}, \qquad (2.30)$$

which after some straightforward manipulations minimizes the total error as follows

$$\epsilon = \sum_{l=1}^{K} \epsilon_l = K\text{Tr}\Big(\big(\mathbf{R}^{-1} + p\mathbf{F}^H\mathbf{F}\big)^{-1} \Big). \qquad (2.31)$$

Chapter 3

Achievable Rate of Multipair Massive Relaying with Hybrid Processing and Perfect CSI

Massive MIMO is a fairly new research paradigm that offers orders of magnitude more spectral efficiency and the simplicity of transceiver design compared to today's examples. Moreover, this technique can tackle many of the drawbacks of multipair relaying topologies, namely complicated signal processing, delay, and noise/interference amplification. Yet, the practical implementation of massive MIMO induces some formidable challenges. In particular, having one RF chain dedicated to each antenna will boost to unprecedented levels the circuit complexity, fabrication/implementation cost and power consumption. In this chapter, we envision that these critical issues can be addressed by a cascade structure of an analog RF beamformer and digital baseband processor, namely hybrid structure. In particular, we consider a multipair AF relay system with MRC/MRT and we determine the asymptotic achievable rate for this hybrid analog/digital architecture. After that, we extend our analytical results to account for coarsely quantized analog phase shifters and show that the performance loss with 2 quantization bits is only 10%.

3.1   Introduction

Cooperative communications have been intensively investigated over the past decade due to their ability to extend the cellular coverage and enhance the communication reliability, especially at the cell edges [72–75]. However, the complexity of relaying has always been a major concern since even a small processing delay results in a loss of valuable physical resource, i.e. time, which should be available for other purposes like backhaul and access link operations [6]. Moreover, relays are typically supported by small power supply units, due to their physical limitations. So, energy efficiency plays an important role in this sense.

Recently, massive MIMO has come at the forefront of wireless communications' research [25]. Massive MIMO avails of *favorable propagation* conditions, where channels between different users and the BS become pairwise orthogonal. Most importantly, with massive MIMO, simple linear processing schemes such as MRC, ZF, and MMSE can achieve the same performance as other nonlinear methods, like dirty paper coding [27, 29, 76]. Moreover, massive MIMO with its array gain and also its great ability to suppress the noise and interference, delivers the required SINR to reduce the power consumption at the relay station [25, 77, 78]. The later property also offers a big advantage for AF relays, where the relay station amplifies whatever it receives including noise and interference.

Not surprisingly, massive MIMO relaying has very recently received a great deal of research interest from different viewpoints. In [30], the asymptotic performance of one-way massive relaying was analyzed in three different cases to scale down the power by the number of active antennas at the relay station. The energy efficiency of two-way relaying with unlimited relay nodes was derived in [33], while some other works in this context can be found in [31, 72, 79]. The potential of massive relaying to mitigate the self-interference in full-duplex relaying was originally developed in [34]. However, the practical deployment of massive MIMO relaying faces many critical challenges. More specifically, deploying one RF chain behind every single antenna, in all above cases, will scale badly the implementation/maintenance cost, DSP

complexity, power consumption and circuit complexity. This important issue has been addressed in other fields like [1,39,42–44] by a hybrid architecture including an analog RF precoder/combiner and a fully digital baseband processor, where all antennas are connected to a limited number of RF chains through a bank of phase shifters. However, there is a dearth of literature considering the hybrid solution for relaying schemes [46]. The authors in [46] proposed an algorithm based on SVD to maximize the average rate for millimeter-wave MIMO systems. However, they did not derive any closed-form expression for the achievable rate, and also these results are obtained under the assumption of accurate phase shifters which is too idealistic in practice. In [47], the authors evaluated the spectral and energy efficiency of a hybrid massive MIMO relaying for i.i.d. Rayleigh fading channels under the following assumptions: (a) $L = 2K$, and (b) $N > \left\lfloor \frac{4L^2}{\pi} \right\rfloor$. Where K, L, and N denote the number of users, RF chains, and antennas, respectively.

Motivated by the above discussion, this chapter investigates the performance of a multipair massive relaying where part of the DSP on the relay station is performed in the analog domain, using simple analog phase shifters. In particular, we analytically determine the asymptotic end-to-end achievable rate by considering MRC/MRT processing at the relay station, where the number of antennas grows up without bound. Then, we apply our analytical results on three power saving strategies and deduce their asymptotic power scaling laws. These laws reveal important physical insights and tradeoffs between the transmit power of user nodes and the relay station. Finally, we consider the case of quantized phase shifters and work out the performance degradation for small number of quantization bits. Our numerical results indicate that (a) hybrid processing can obtain the required data rate with lower number of RF chains compared to the conventional fully digital structures; and (b) 2 bits of quantization cause a minor performance degradation of about 10%.
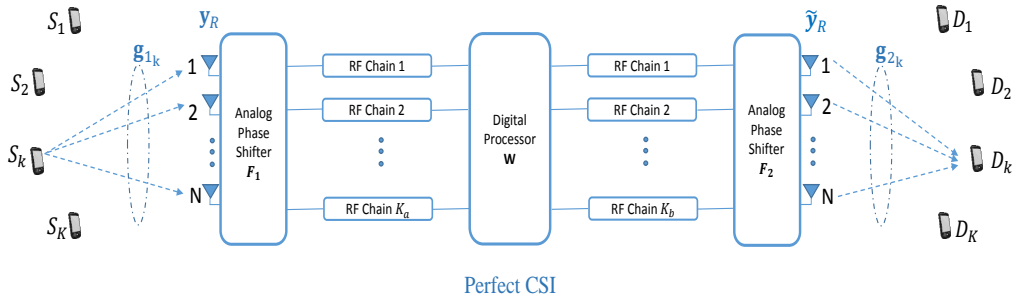
Figure 3.1: Simplified block diagram of a multipair relay system with a baseband digital processor combined with two analog RF beamformers implemented by means of quantized phase shifters. Channels are uncorrelated Rayleigh fading and perfect CSI is available at the relay station.

3.2 System Model

Consider a cooperative system model as illustrated in Fig. 3.1 where K pairs of single antenna users, one-to-one pairing S_k–D_k with $k = 1, \ldots, K$, intend to communicate with each other via the help of a relay with N service antennas on each side, where N is an unconventionally large number. Furthermore, we assume that the direct link among the K pairs does not exist due to large path loss and heavy shadowing; to keep our analysis simple, perfect CSI is available and we ignore hardware imperfections [80]. Users send their data stream through a narrow band flat-fading propagation channel in the same time-frequency block. To keep the implementation cost of this massive MIMO relaying topology at low levels, we consider K_a receive and K_b transmit RF chains at the relay station, with $K \leq K_a, K_b \ll N$. As mentioned earlier, by reducing the number of RF paths, we can avail of reduced power consumption (reduced numbers of mixers, PAs and analog-to-digital converters (ADCs) and reduced circuitry. Moreover, to reduce the power dissipation of DSP, we deploy an analog combiner $\mathbf{F}_1 \in \mathbb{C}^{K_a \times N}$ and precoder $\mathbf{F}_2 \in \mathbb{C}^{K_b \times N}$ at the relay station which perform phase matching at a much lower dimension compared to full DSP. Since analog processing alone is not flexible enough, the remaining portion of signal processing is performed in the digital domain through the matrix $\mathbf{W} \in \mathbb{C}^{K_b \times K_a}$. Under this model, the received signal at the relay and destinations can be mathematically expressed, respectively, as

$$\mathbf{y}_R = \sqrt{P_u}\mathbf{G}_1\mathbf{x} + \mathbf{n}_R, \tag{3.1}$$

$$\mathbf{y}_D = \sqrt{P_u}\mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{G}_1\mathbf{x} + \mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + \mathbf{n}_D, \tag{3.2}$$

where P_u represents the transmitted power of each source, and $\mathbf{x} = [x_1, x_2, \cdots, x_K]^T$ is the symbol vector so that $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_K$. Also, the received signal at the destinations is included in $\mathbf{y}_D \in \mathbb{C}^{K \times 1}$, while the N-dimensional vector $\mathbf{n}_R$ and K-dimensional vector $\mathbf{n}_D$ model the additive circularly symmetric complex Gaussian noise so that $\mathbf{n}_R \sim \mathcal{CN}(0, \sigma_{n_R}^2\mathbf{I}_N)$ and $\mathbf{n}_D \sim \mathcal{CN}(0, \sigma_{n_D}^2\mathbf{I}_K)$. Moreover, $\mathbf{G}_1, \mathbf{G}_2 \in \mathbb{C}^{N \times K}$ express the propagation channels between sources and relay, and between relay and destinations, respectively. More precisely, $\mathbf{G}_1 = \mathbf{H}_1\mathbf{D}_1^{\frac{1}{2}}$ and $\mathbf{G}_2 = \mathbf{H}_2\mathbf{D}_2^{\frac{1}{2}}$, where $\mathbf{H}_1$, $\mathbf{H}_2 \in \mathbb{C}^{N \times K}$ refer to small scale fading channels with i.i.d. entries, each of them following $\mathcal{CN}(0, 1)$. Besides, the diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2 \in \mathbb{C}^{K \times K}$ include the large-scale fading parameters, where we define $\beta_{1,k} \triangleq [\mathbf{D}_1]_{k,k}$ and $\beta_{2,k} \triangleq [\mathbf{D}_2]_{k,k}$. From (3.2) the received signal at the k-th destination with additive white Gaussian noise (AWGN) n_{D_k} is given by

$$y_{D_k} = \sqrt{P_u}\mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1_k}x_k + \sqrt{P_u}\sum_{i \neq k}^{K}\mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1_i}x_i + \mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + n_{D_k},$$

$$\tag{3.3}$$

where $\mathbf{g}_{1_k}$, and $\mathbf{g}_{2_k}$ denote the k-th column of the matrices $\mathbf{G}_1$ and $\mathbf{G}_2$, respectively. In (3.3), the first term points out to the desired signal, the second term refers to the interpair-interference, while the last two terms correspond to the amplified noise at the relay and noise at the destination, respectively. Thus, the instantaneous end-to-end SINR for the k-th pair is given by

$$\text{SINR}_k = \frac{P_u\left|\mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1_k}\right|^2}{P_u\sum_{i \neq k}^{K}\left|\mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1_i}\right|^2 + \|\mathbf{g}_{2_k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\|^2\sigma_{n_R}^2 + \sigma_{n_D}^2}. \tag{3.4}$$

Consequently, the ergodic achievable sum-rate, or simply achievable rate, of this multipair massive MIMO relaying system can be obtained by[1]

$$R = \sum_{k=1}^{K} \mathbb{E}\Big[\log_2 \left(1 + \text{SINR}_k \right) \Big] \qquad \text{bit/s.} \tag{3.5}$$

As mentioned before, the role of the analog combiners is to balance out the phase of the propagation matrices. It is noteworthy that the matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ can only perform analog phase shifting, hence, we normalize each entry of these matrices by $\frac{1}{\sqrt{N}}$ to avoid an unlimited gain in the analog domain. To this end, we have

$$\measuredangle\big[\mathbf{F}_1\big]_{i,j} = -\measuredangle\big[\mathbf{G}_1\big]_{j,i}, \tag{3.6}$$

$$\measuredangle\big[\mathbf{F}_2\big]_{i,j} = -\measuredangle\big[\mathbf{G}_2\big]_{j,i}, \tag{3.7}$$

$$\left|\big[\mathbf{F}_1\big]_{i,j}\right| = \left|\big[\mathbf{F}_2\big]_{i,j}\right| = \frac{1}{\sqrt{N}}. \tag{3.8}$$

On the other hand, the baseband precoder matrix $\mathbf{W}$ can modify both the amplitude and phase of the incoming vector. Moreover, we introduce the following long-term transmit power constraint for the output of the relay station:

$$\mathbb{E}\Big[\text{Tr}\big(\tilde{\mathbf{y}}_R \tilde{\mathbf{y}}_R^H\big)\Big] = P_r, \tag{3.9}$$

where, $\tilde{\mathbf{y}}_R = \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \mathbf{y}_R$ demonstrates the relay output signal.

In the rest of this chapter, we assume MRC to combine received signals at the relay, and also consider MRT to forward the received signals from the relay station to the destinations. We recall that MRC/MRT processing has been well integrated in the context of massive MIMO, since it offers near-optimal performance and can be implemented in a distributed manner [81]. Now, we define the following symbols that will be used in our subsequent analysis; $\mathbf{A}_1 \triangleq \mathbf{F}_1 \mathbf{G}_1$, $\mathbf{A}_2 \triangleq \mathbf{F}_2 \mathbf{G}_2$, $\mathbf{B}_1 \triangleq \mathbf{F}_1 \mathbf{F}_1^H$, and $\mathbf{B}_2 \triangleq \mathbf{F}_2 \mathbf{F}_2^H$. For the digital MRC/MRT transformation matrix, $\mathbf{W} \triangleq \alpha \mathbf{A}_2 \mathbf{A}_1^H$ where α is an amplification factor to fulfill the power constraint in (3.9). Therefore, we can

---

[1]In our subsequent analysis, we assume that the bandwidth is normalized to 1Hz such that equation (3.5) represents indeed the achievable rate.

obtain after some mathematical simplifications

$$\alpha = \sqrt{\frac{P_r}{P_u \|\mathbf{F}_2^H \mathbf{A}_2 \mathbf{A}_1^H \mathbf{A}_1\|_F^2 + \sigma_{n_R}^2 \|\mathbf{F}_2^H \mathbf{A}_2 \mathbf{A}_1^H \mathbf{F}_1\|_F^2}}. \tag{3.10}$$

It is notable that if we assume a short-term constraint, like $\mathrm{Tr}\big(\tilde{\mathbf{y}}_R \tilde{\mathbf{y}}_R^H\big) = P_r$, the fluctuation of amplification factor α will be very high particularly in fast fading channels. Therefore, it would be a challenging task for the relay station to establish a constant output power, i.e., P_r, based on the variations of α. Fortunately, the simulation result in Fig. 3.2 corroborates that these variations are not substantial around the average. Hence, assuming the long-term constraint in (3.9) can be considered as a reasonable assumption from a practical viewpoint.

3.3 Large N analysis

In this section, we asymptotically analyze the performance of the massive MIMO relay with hybrid processing in two dedicated subsections: Section 3.3.1 assuming ideal (continuous) phase shifters, and Section 3.3.2 assuming phase quantization.

3.3.1 Ideal (continuous) phase shifters

We now briefly review some asymptotic results that will be particularly useful in our analysis later on.

Lemma 3.1. *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be two* $n \times 1$ *mutually independent vectors whose elements are i.i.d random variables with variances* σ_p^2 *and* σ_q^2, *respectively. Then, based on the law of large numbers, we have*

$$\frac{1}{n}\mathbf{p}^H \mathbf{p} \xrightarrow{\text{a.s.}} \sigma_p^2 \quad \text{and} \quad \frac{1}{n}\mathbf{q}^H \mathbf{q} \xrightarrow{\text{a.s.}} \sigma_q^2 \quad, \text{as} \quad n \to \infty, \tag{3.11}$$

where $\xrightarrow{\text{a.s.}}$ *indicates almost sure convergence. Moreover, based on the Lindeberg–Lévy*
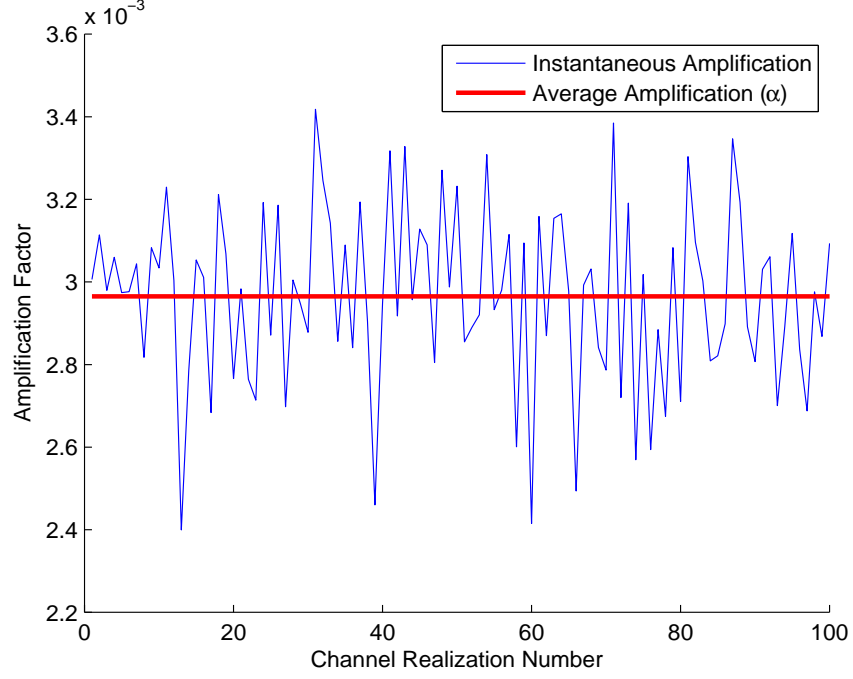
Figure 3.2: Fluctuation of relay amplification factor assuming phase cancelation at the analog beamformers and the MRC scheme at the DSP unit ($K = K_a = K_b = 10$, $N = 100$, $\frac{P_u}{N_0} = 0.2$, $\frac{P_r}{N_0} = 20$, $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}$).

central limit theorem we can write

$$\frac{1}{\sqrt{n}} \mathbf{p}^H \mathbf{q} \xrightarrow{\text{dist.}} \mathcal{CN}(0, \sigma_p^2 \sigma_q^2), \tag{3.12}$$

where $\xrightarrow{\text{dist.}}$ shows the convergence in distribution.

We can now turn our attention to the analog processing matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ which satisfy the following relationship.

Lemma 3.2. *As $N \to \infty$, the matrices $\mathbf{F}_1 \mathbf{F}_1^H$ and also $\mathbf{F}_2 \mathbf{F}_2^H$ converge pairwise to the identity matrix as follows*

$$\mathbf{F}_1 \mathbf{F}_1^H \xrightarrow{\text{a.s.}} \mathbf{I}_{K_a}, \tag{3.13}$$

$$\mathbf{F}_2 \mathbf{F}_2^H \xrightarrow{\text{a.s.}} \mathbf{I}_{K_b}. \tag{3.14}$$

Proof. See Appendix A.1. □

Corollary 3.3. *As $N \to \infty$, the analog phase shifter, $\mathbf{F}_1$, preserves the distribution of the AWGN due to its orthonormal rows.*

Lemma 3.4. *Let us define $\mathbf{I}_{a,b,r} \in \mathbb{R}^{a \times b}$ as an $a \times b$ rectangular diagonal matrix whose entries are entirely zero except the (i,i)-th elements of matrix which are one, where $i = \{1, 2, \ldots, r\}$. Then, we can conclude that*

$$\mathbf{F}_1 \mathbf{H}_1^H \xrightarrow{\text{a.s.}} \sqrt{\frac{N\pi}{4}} \mathbf{I}_{K_a, K, r_1}, \tag{3.15}$$

$$\mathbf{F}_2 \mathbf{H}_2^H \xrightarrow{\text{a.s.}} \sqrt{\frac{N\pi}{4}} \mathbf{I}_{K_b, K, r_2}, \tag{3.16}$$

where $r_1 = \min\{K_a, K\}$ and $r_2 = \min\{K_b, K\}$.

Proof. See Appendix A.2. $\qquad\square$

Turning now to (3.2) and using the aforementioned lemmas, when $N \to \infty$, it can be shown that

$$\mathbf{y}_D \to \sqrt{P_u} \alpha \left(\frac{N\pi}{4}\right)^2 \left(\mathbf{D}_2^{\frac{1}{2}}\right)^H \mathbf{D}_2^{\frac{1}{2}} \left(\mathbf{D}_1^{\frac{1}{2}}\right)^H \mathbf{D}_1^{\frac{1}{2}} \mathbf{x} + \alpha \left(\frac{N\pi}{4}\right)^{\frac{3}{2}} \left(\mathbf{D}_2^{\frac{1}{2}}\right)^H \mathbf{D}_2^{\frac{1}{2}} \left(\mathbf{D}_1^{\frac{1}{2}}\right)^H \mathbf{n}_R + \mathbf{n}_D, \tag{3.17}$$

which can be simplified for the k-th destination as

$$y_{D_k} \to \sqrt{P_u} \alpha \left(\frac{N\pi}{4}\right)^2 \beta_{2_k} \beta_{1_k} x_k + \alpha \left(\frac{N\pi}{4}\right)^{\frac{3}{2}} \beta_{2_k} \beta_{1_k}^{\frac{1}{2}} n_{R_k} + n_{D_k}, \tag{3.18}$$

where $k \in \{1, 2, \ldots, r\}$, and $r = \min\{K_a, K_b, K\}$. Thus, from (3.4) we can obtain the corresponding SINR for the k-th destination in the case that the number of antennas increases without bound

$$\text{SINR}_k \to \frac{\left(\frac{N\pi}{4}\right)^4 P_u \alpha^2 \beta_{1_k}^2 \beta_{2_k}^2}{\left(\frac{N\pi}{4}\right)^3 \sigma_{n_R}^2 \alpha^2 \beta_{1_k} \beta_{2_k}^2 + \sigma_{n_D}^2}, \qquad \text{as } N \to \infty. \tag{3.19}$$

In the following, we investigate three power scaling strategies and draw very interesting engineering insights. Our analysis can be divided into three main cases, namely, Case

1) fixed NP_u and NP_r while $N \to \infty$; Case 2) fixed NP_u while $N \to \infty$; Case 3) fixed NP_r while $N \to \infty$.

1. Let $\lim\limits_{N \to \infty} NP_u = E_u$ and $\lim\limits_{N \to \infty} NP_r = E_r$, where both E_u and E_r are finite constants. Therefore, we readily conclude that both P_u and P_r go to zero if we increase the number of antennas to infinity. Hence, from (3.10) we can get

$$N^3 \alpha^2 \to \frac{E_r}{\left(\frac{\pi}{4}\right)^3 E_t \sum\limits_{i=1}^{r} \beta_{1_i}^2 \beta_{2_i} + \left(\frac{\pi}{4}\right)^2 \sigma_{n_R}^2 \sum\limits_{i=1}^{r} \beta_{1_i} \beta_{2_i}}, \tag{3.20}$$

which finally yields for $k \in \{1, 2, \ldots, r\}$

$$\mathrm{SINR}_k \to \frac{\left(\frac{\pi}{4}\right)^2 E_u E_r \beta_{1_k}^2 \beta_{2_k}^2}{\left(\frac{\pi}{4}\right) E_r \sigma_{n_R}^2 \beta_{1_k} \beta_{2_k}^2 + \left(\frac{\pi}{4}\right) E_u \sigma_{n_D}^2 \sum\limits_{i=1}^{r} \beta_{1_i}^2 \beta_{2_i} + \sigma_{n_R}^2 \sigma_{n_D}^2 \sum\limits_{i=1}^{r} \beta_{1_i} \beta_{2_i}}. \tag{3.21}$$

As a consequence, under a perfect CSI assumption we can reduce the transmitted power and also relay power proportionally to $\frac{1}{N}$ if the number of relay antennas grows without bound. This result is consistent with [81].

2. Let $\lim\limits_{N \to \infty} NP_u = E_u$, where E_u is a finite constant. Therefore, we readily conclude that P_u goes to zero if we increase the number of antennas to infinity. Then, returning to (3.21) and after a few simplifications we obtain

$$\mathrm{SINR}_k \to \frac{\pi}{4} \frac{E_u \beta_{1_k}}{\sigma_{n_R}^2}, \tag{3.22}$$

which is associated with the following achievable rate

$$R_2 \to \frac{1}{2} \sum_{k=1}^{r} \log_2 \left(1 + \frac{\pi}{4} \frac{E_u \beta_{1_k}}{\sigma_{n_R}^2}\right). \tag{3.23}$$

The above result is quite intuitive. It shows that if the number of RF chains is, at least, equal to the number of users, i.e. $\min\{K_a, K_b\} \geq K$ or equiva-

lently $r = K$, we can enjoy full multiplexing gain and boost the achievable rate. Moreover, in comparison with a single-input single-output system without any intra-cell interference, our system model only suffers a $\frac{\pi}{4}$-fold reduction on the power gain due to the analog processing. All in all, this power gain penalty is quite acceptable as we have eliminated many relay RF chains, and consequently, we have substantially reduced the circuitry complexity and power consumption. Similar to Case 1, we can infer that we can scale down the transmit power analogously to the number of relay antennas and, still, maintain a non-zero achievable rate.

3. Let $\lim\limits_{N \to \infty} NP_r = E_r$, where E_r is a finite constant. Therefore, we readily conclude that P_r goes to zero if we increase the number of antennas to infinity. Then, we can find out the achievable rate in the same way as pointed out in Case 2 to get

$$R_3 \to \frac{1}{2} \sum_{k=1}^{r} \log_2 \left(1 + \frac{\pi}{4} \frac{E_r \beta_{1_k}^2 \beta_{2_k}^2}{\sigma_{n_D}^2 \sum\limits_{i=1}^{r} \beta_{1_i}^2 \beta_{2_i}} \right). \tag{3.24}$$

It is noteworthy that if we ignore large-scale fading effects, we get the same results in Case 2 and 3. However, Case 3 converges faster than Case 2 to its own asymptotic result. This can be observed from (3.21), where we can asymptotically derive both R_2 and R_3: In Case 3, we can ignore the constant term $E_r \sigma_{n_R}^2 \beta_{1_k} \beta_{2_k}^2$ in comparison with $E_u \sigma_{n_D}^2 \sum_{i=1}^{r} \beta_{1_i}^2 \beta_{2_i}$ even for moderate number of antennas. Unlike, in Case 2, a much higher number of antennas is required to ignore the constant term $E_u \sigma_{n_D}^2 \sum_{i=1}^{r} \beta_{1_i}^2 \beta_{2_i}$ vs. the scaled term $E_r \sigma_{n_R}^2 \beta_{1_k} \beta_{2_k}^2$ in (3.21).

3.3.2 Phase Quantization

Until now, we have assumed ideal analog phase shifters (beamformers) which generate any required phases. However, the implementation of such shifters with continuous phase is not feasible or, at least, is quite expensive due to the hardware limitations

[43, 82–84]. Most importantly, efficiently quantized analog beamformers are more attractive in limited feedback systems [85, 86]. In the rest of this chapter, the system performance will be assessed under quantized phases. Thus, the phase of each entry of $\mathbf{F}_1$ and $\mathbf{F}_2$ is chosen from a codebook Ψ based on the closest Euclidean distance.

$$\Psi = \left\{ 0, \pm\left(\frac{2\pi}{2^q}\right), \pm 2\left(\frac{2\pi}{2^q}\right), \cdots, \pm 2^{q-1}\left(\frac{2\pi}{2^q}\right) \right\}, \tag{3.25}$$

where, q denotes the number of quantization bits. As pointed out previously, the channel coefficients $\left[\mathbf{G}_1\right]_{m,n}$ and $\left[\mathbf{G}_2\right]_{m,n}$ all have uniform phase between 0 and 2π, such that $\angle\left[\mathbf{G}_i\right]_{m,n} = \phi_{m,n} \sim U(0, 2\pi)$, for $i = 1, 2$. Let us define $\epsilon_{m,n}$ as the error between the unquantized phase $\phi_{m,n}$ and quantized phase $\hat{\phi}_{m,n}$ chosen from the codebook

$$\epsilon_{m,n} \triangleq \phi_{m,n} - \hat{\phi}_{m,n}. \tag{3.26}$$

Due to the uniform distribution of phase, we can easily conclude that the error is an uniform random variable, i.e. $\epsilon_{m,n} \sim U[-\delta, +\delta)$, where we define $\delta \triangleq \frac{\pi}{2^q}$. This error affects Lemma 3.4, and in turn, the achievable rate. For this reason, we provide the following lemma to account for phase quantization. For the sake of presentation clarity, we use a hat sign for the variables that are associated with the quantized beamforming assumption in this chapter.

Lemma 3.5. *Let $\hat{\mathbf{F}}_1$ and $\hat{\mathbf{F}}_2$ denote that analog decoder and precoder, respectively. Then,*

$$\hat{\mathbf{F}}_1\mathbf{H}_1^H \xrightarrow{a.s.} \sqrt{\frac{N\pi}{4}} \operatorname{sinc}(\delta)\mathbf{I}_{K_a, K, r_1}, \tag{3.27}$$

$$\hat{\mathbf{F}}_2\mathbf{H}_2^H \xrightarrow{a.s.} \sqrt{\frac{N\pi}{4}} \operatorname{sinc}(\delta)\mathbf{I}_{K_b, K, r_2}, \tag{3.28}$$

where we define $\operatorname{sinc}(\delta) = \frac{\sin(\delta)}{\delta}$.

Proof. See Appendix A.3. □

Now, we incorporate Lemma 3.5 into the system model and signal description. The

modified normalization factor $\hat{\alpha}$ can be calculated as

$$\hat{\alpha} \rightarrow \sqrt{\frac{P_r}{P_u \operatorname{sinc}^6(\delta)\|\mathbf{F}_2^H \mathbf{A}_2 \mathbf{A}_1^H \mathbf{A}_1\|_F^2 + \sigma_{n_R}^2 \operatorname{sinc}^4(\delta)\|\mathbf{F}_2^H \mathbf{A}_2 \mathbf{A}_1^H \mathbf{F}_1\|_F^2}}. \qquad (3.29)$$

Furthermore, the received signal for the k-th destination can be obtained from the following formula

$$\hat{y}_{D_k} \rightarrow \sqrt{P_u} \operatorname{sinc}^4(\delta)\hat{\alpha}\left(\frac{N\pi}{4}\right)^2 \beta_{2_k}\beta_{1_k} x_k + \left(\frac{N\pi}{4}\right)^{\frac{3}{2}} \operatorname{sinc}^3(\delta)\hat{\alpha}\,\beta_{2_k}\beta_{1_k}^{\frac{1}{2}} n_{R_k} + n_{D_k}. \tag{3.30}$$

Phase quantization also affects the power scaling strategies considered in Cases 1–3 above. The corresponding results for these three cases under quantized analog processing can be respectively modified as follows

$$\hat{R}_1 \rightarrow \frac{1}{2}\sum_{k=1}^r \log_2\left(1 + \frac{(\frac{\pi}{4})^2 \operatorname{sinc}^4(\delta)E_u E_r \beta_{1_k}^2 \beta_{2_k}^2}{(\frac{\pi}{4})\operatorname{sinc}^2(\delta)E_r\sigma_{n_R}^2\beta_{1_k}\beta_{2_k}^2 + (\frac{\pi}{4})\operatorname{sinc}^2(\delta)E_u\sigma_{n_D}^2\sum\limits_{i=1}^r \beta_{1_i}^2\beta_{2_i} + \sigma_{n_R}^2\sigma_{n_D}^2\sum\limits_{i=1}^r \beta_{1_i}\beta_{2_i}}\right),$$

$$\hat{R}_2 \rightarrow \frac{1}{2}\sum_{k=1}^r \log_2\left(1 + \frac{\pi}{4}\frac{E_u\beta_{1_k}}{\sigma_{n_R}^2}\operatorname{sinc}^2(\delta)\right), \qquad (3.31)$$

$$\hat{R}_3 \rightarrow \frac{1}{2}\sum_{k=1}^r \log_2\left(1 + \frac{\pi}{4}\frac{E_r\beta_{1_k}^2\beta_{2_k}^2}{\sigma_{n_D}^2\sum\limits_{i=1}^r \beta_{1_i}^2\beta_{2_i}}\operatorname{sinc}^2(\delta)\right). \qquad (3.32)$$

Taken together, these results indicate a penalty function associated with quantized processing. Roughly speaking, $\operatorname{sinc}^2(\delta)$ is a good approximation of this power gain penalty. In a worst case, where we have only one quantization bit $q = 1$, the SINR will be reduced by a factor of $\operatorname{sinc}^2(\frac{\pi}{2}) = \frac{4}{\pi^2} \approx 40\%$. However, there are wide range of applications, each requiring a different quantization resolution [39, 43, 87]. In general, the beam width is inversely relative to the number of antennas, so if we double the number of antennas we need to increase the angle accuracy two times more. Hence, a reasonable rule-of-thumb is to add 1 bit resolution while the number of antennas doubles.
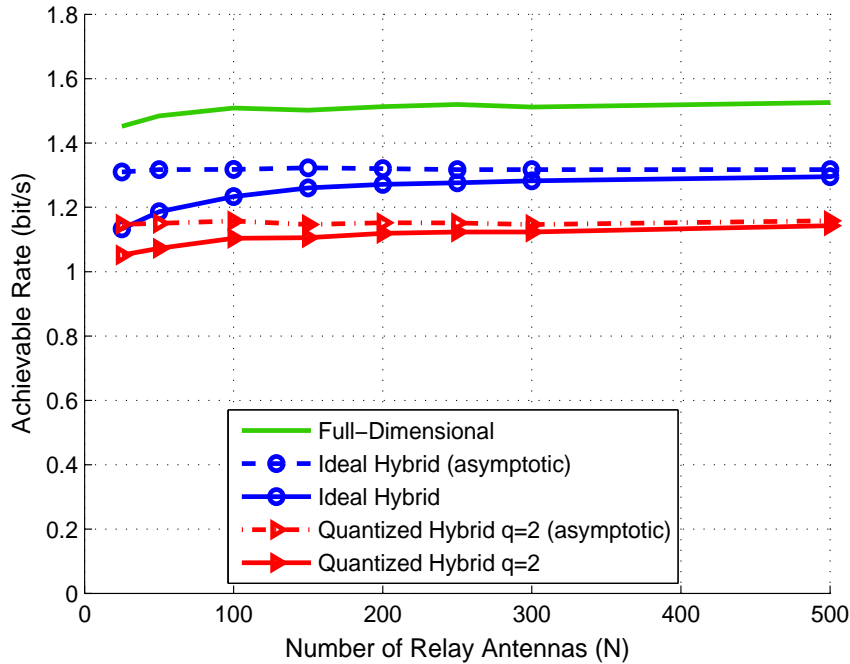
Figure 3.3: Achievable rate in Case 1 $\left(E_u = E_r = 13 \text{ dB}\right)$.

3.4 Simulation Results

In this section, Monte Carlo simulations are provided to assess the validity of the achievable rate of a multipair relay system. We assume that the relay station covers a circular area with a radius of 1000 meters. Users are located with a uniform random distribution around the relay with a guard zone of $r_g = 100$ meters. We consider a Rayleigh flat fading channel for small-scale fading effects. Also, the large-scale fading is modeled via a log-normal random variable, with standard deviation σ_{sh}, which is multiplied by $\left(\frac{r_g}{r_k}\right)^\nu$ to model the path-loss as well. Here, r_k is the distance between the k-th user and the relay, and also ν denotes the path loss exponent. Without loss of generality, we assume $\sigma_{n_R}^2 = \sigma_{n_D}^2 = 1$ to express the powers, i.e. P_u and P_r, in decibel. We also set $\nu = 3.8$, $K_a = K_b = K = 10$ and $\sigma_{sh} = 8$ dB for all simulations.

Figure 3.3 compares the performance of a full-dimensional topology, where all the amount of detection/precoding is performed in the digital domain, against that of a hybrid topology with continuous and quantized analog processing. A full-dimensional massive relay is equipped with N RF chains which seems to be infeasible in practice, while this number is reduced to only $K = 10$ in the hybrid structure. Moreover, a
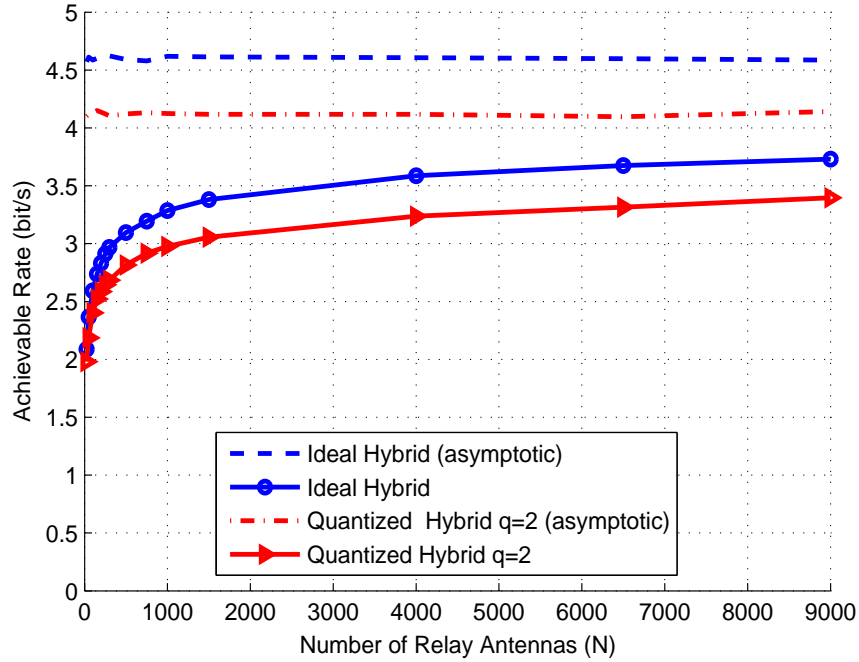
Figure 3.4: Achievable rate in Case 2 $\left(E_u = 13 \text{ dB}, P_r = 13 \text{ dB}\right)$.

hybrid relay deploys two inexpensive beamformers which can be actually implemented in the analog domain with the phase shifters. It can be also observed that the hybrid scheme performs very close to the conventional scheme, with about a 10% reduction in achievable rate but substantially reduced complexity. However, this reduction in achievable rate can be compensated by deploying more antennas at the relay station without any additional RF chains. Hence, this promising idea seems to be a viable alternative to the conventional relaying system. Moreover, this figure examines a more restricted case, where there is a severe phase control on beamformers with only 2 bit resolution. Our results confirm that the proposed method suffers a negligible reduction. Figures 3.4 demonstrates the similar results for Case 2. Clearly, as the number of relay antennas increases, the achievable rate approaches the saturation value which is expected by our analytical approximations. Also, we note that the curve scales slower in Case 2 in comparison with Case 1. In order to shed light on how quantization scheme affects the achievable rate, we simulate Case 3 with different quantization levels in Fig. 3.5. This figure confirms that, although the achievable rate is not outstanding with one quantization bit, 2-bit resolution can achieve about 90% of the achievable
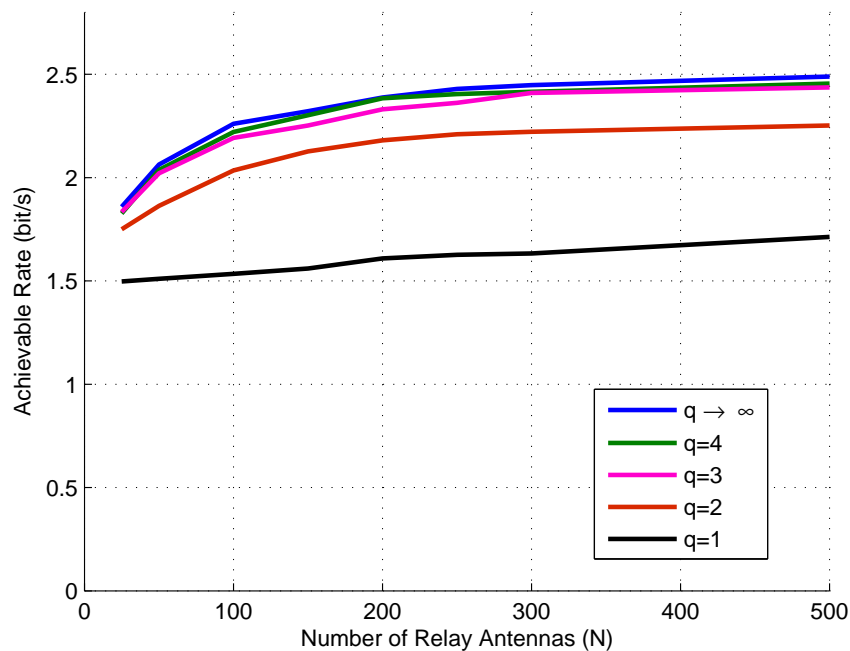
Figure 3.5: Achievable rate in Case 3 with different quantization levels $\left(P_u = 13 \text{ dB}, E_r = 13 \text{ dB}\right)$.

rate offered by the ideal phase shifters. Moreover, it can be seen that phase shifters with 3 quantization bits can perform very close to the ideal phase shifters.

3.5 Conclusion

Massive MIMO is a major candidate for the next generation of wireless systems. This technique combined with relays can enhance the cell coverage, provide uniform quality of service, with simple signal processing at the relay station. On the other hand, the high cost and power consumption of RF chains can be prohibitive due to the large number of mixers and PAs. For this reason, we proposed an analog/digital (hybrid) structure at the relay and also reduced the number of RF chains to the number of users while the system still enjoys a full multiplexing gain. Our results revealed that with only 10 RF chains at the relay station we can capture about 90% of asymptotic achievable rate of a fully digital structure with hundreds of RF chains. For this topology, we also analytically determined the system robustness even under coarse quantization. Our analysis revealed that the SINR under q quantization bit almost scales down as a function of $\text{sinc}^2\left(\frac{\pi}{2^q}\right)$.

Chapter 4

Achievable Rate of Multipair Massive Relaying with Hybrid Processing and Imperfect CSI

4.1 Introduction

In the previous chapter, we addressed the hardware constraints of massive MIMO relaying by proposing the hybrid analog/digital structure. However, in that scenario, phase shifters should be adopted to the quick variations of the instantaneous channels. This phase adaptation, not only requires the perfect CSI at the relay station, but also is a demanding task due to low flexibility of analog beamformers. Hence, in this chapter, we assume that perfect CSI is not available at the relay station, and the relay estimates the channels via uplink orthogonal pilots during the training phase. Also, we assume the worst case scenario, where we utilize Hadamard beamformers with one bit resolution and invariant over time. This assumption reduces the channel size to the number of RF chains and restricts the impact of array gain. However, it avails of a very simple structure which describes the basic relation between the achievable rate, number of RF chains, and active users. Besides, we employ an MRC/MRT scheme in the digital domain to keep the complexity of DSP unit at the relay station to affordable levels.

Overall, this chapter investigates the performance of multipair AF relaying with a large number of antennas but limited RF chains under the hybrid structure, and makes the following specific contributions:

- We investigate the performance of a hybrid analog and digital relaying topology with imperfect CSI. Our results show that by removing half of the RF chains in this scenario, the hybrid paradigm can still capture 75% the achievable rate of a fully digital relaying system. Moreover, to further simplify the described configuration, we assume that the analog beamformers are implemented using the simplest phase shifters whose phases $\pm\pi$ are constant over time.

- Since the analog processing is time invariant and independent of the propagation channel, we perform all the beamforming in the digital domain via MRC/MRT which avails of simple signal processing. For this scenario and for imperfect CSI, we derive a tractable lower bound on the achievable rate which involves only the large-scale fading coefficients by leveraging the properties of large Gaussian matrices.

4.2 System Model

In this chapter, we consider the same system model as we have already introduced in Section 3.2. However, we assume that CSI is not available not only at the users, but also at the relay side. This is a case of practical interest that we will address in this chapter. Considering this system model, illustrated in Fig. 4.1, we can recall our signaling models as follows

$$\mathbf{y}_R = \sqrt{P_u}\mathbf{G}_1\mathbf{x} + \mathbf{n}_R, \tag{4.1}$$

$$\tilde{\mathbf{y}}_R = \mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{y}_R, \tag{4.2}$$

$$\mathbf{y}_D = \sqrt{P_u}\mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{G}_1\mathbf{x} + \mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + \mathbf{n}_D, \tag{4.3}$$
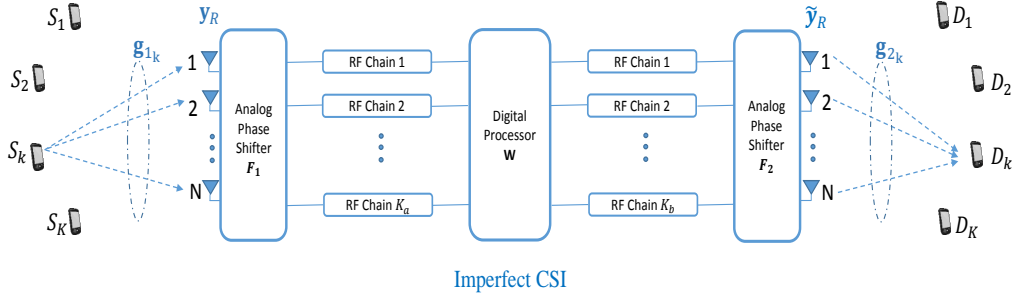
Figure 4.1: Simplified block diagram of a multipair massive relay system with a hybrid analog/digital beamformer. Channels are Rayleigh fading and perfect CSI is not available at the relay station.

where both small- and large-scale fading are incorporated into the channel matrices such that $\mathbf{G}_1 = \mathbf{H}_1\mathbf{D}_1^{\frac{1}{2}}$ and $\mathbf{G}_2 = \mathbf{H}_2\mathbf{D}_2^{\frac{1}{2}}$. In this model $\mathbf{H}_1$ and $\mathbf{H}_2 \in \mathbb{C}^{N \times K}$ denote the small-scale fading with i.i.d. $\mathcal{CN}(0,1)$ elements. Additionally, the diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2 \in \mathbb{C}^{K \times K}$ reflect the geometric attenuation and shadow fading, where their k-th diagonal elements are defined as $\beta_{1,k}$ and $\beta_{2,k}$, respectively. Then, from (4.3), the received signal at the k-destination user is given by

$$y_{D_k} = \sqrt{P_u}\mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1,k}x_k + \sqrt{P_u}\sum_{i \neq k}^{K}\mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1,i}x_i + \mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + n_{D_k},$$

(4.4)

where the first term refers to the desired signal and the second term corresponds to the inter-user interference, while the last two terms represent the compound noise.

4.2.1 Design Constraints

The system is subject to a set of practical constraints. First, all entries of $\mathbf{F}_1$, $\mathbf{F}_2$ should have equal modulus, because the hardware deployed to implement the RF precoder/combiner includes a bank of phase shifters. We fix this amplitude by $1/\sqrt{N}$ to avoid an unlimited gain in the analog domain. These analog schemes are usually implemented using phase shifters to do some part of beamforming as well. In general, we need high-resolution phase shifters to finely steer the beams and place the

nulls in predefined directions in each channel realization. This can increase the power consumption and implementation cost, and will undermine the realizable potential of the hybrid paradigm. To avoid this, phase shifters should be designed once for ever and preferably with the minimum resolution. Although Discrete Fourier Transform (DFT) matrices have been considered in some prior works [88–90], the implementation of such phase shifters is costly due to its high resolution, particularly in the massive MIMO ecosystem. Herein, we take a different approach and propose to use Hadamard matrices which can be constructed with phases $\pm\pi$. Sylvester's construction [91] is an easy way to generate such Hadamard matrices as follows:

$$
\mathbf{M}_1 = [1],
$$

$$
\mathbf{M}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},
$$

$$
\mathbf{M}_{2^n} = \begin{bmatrix} \mathbf{M}_{2^{n-1}} & \mathbf{M}_{2^{n-1}} \\ \mathbf{M}_{2^{n-1}} & -\mathbf{M}_{2^{n-1}} \end{bmatrix} \quad \text{for } n \geq 2, \tag{4.5}
$$

where $\mathbf{M}_N$ denotes an $N \times N$ Hadamard matrix which satisfies $\mathbf{M}_N \mathbf{M}_N^T = N\mathbf{I}_N$. Hence, any submatrix extracted from an $N \times N$ Hadamard matrix is suitable for our analog beamformers, i.e. $\mathbf{F}_1$ and $\mathbf{F}_2$. It is noteworthy that the rows of such matrices are orthogonal to each other which simplifies our analysis later on.[1] Hence, throughout this chapter we use the property of Hadamard matrices which states that $\mathbf{F}_1 \mathbf{F}_1^H = \mathbf{I}_{K_a}$ and $\mathbf{F}_2 \mathbf{F}_2^H = \mathbf{I}_{K_b}$, where we normalized each entry of these matrices by $\frac{1}{\sqrt{N}}$ to avoid an unlimited gain in the analog domain. The second constraint is related to the AF relaying where the received signal should be boosted up to a certain level of power P_r. Under this setting, the AF relay should satisfy the following long-term constraint as we discussed earlier in Chapter 3

$$
\mathbb{E}\left[\text{Tr}\left(\tilde{\mathbf{y}}_R \tilde{\mathbf{y}}_R^H\right)\right] = P_r. \tag{4.6}
$$

---

[1]It is notable that this property is also satisfied by the prevalent existing analog beamformer, i.e. DFT beamformer.

For the sake of presentation clarity, we define $\mathbf{A}_1 \overset{\triangle}{=} \mathbf{F}_1\mathbf{G}_1$ and $\mathbf{A}_2 \overset{\triangle}{=} \mathbf{F}_2\mathbf{G}_2$ as the effective channels in the remainder of this chapter. Thus, the received vector at the destinations in (4.3) can be mathematically re-expressed as

$$\mathbf{y}_D = \sqrt{P_u}\mathbf{A}_2^H\mathbf{W}\mathbf{A}_1\mathbf{x} + \mathbf{A}_2^H\mathbf{W}\tilde{\mathbf{n}}_R + \mathbf{n}_D, \tag{4.7}$$

where we define $\tilde{\mathbf{n}}_R = \mathbf{F}_1\mathbf{n}_R$, while its entries follow the same distribution of $\mathbf{n}_R$ elements due to the orthogonal rows of the Hadamard matrix $\mathbf{F}_1$.

4.2.2 Channel Estimation

As perfect CSI is not available at the relay node, hereafter, we assume that the relay estimates the channels based on τ_p uplink training symbols, while this information is not available at the sources and destination users. This assumption can relax the system model to a practical scenario with less overhead signaling and feedback transmission, and is widely used in massive MIMO works. Nevertheless, it is reasonable to assume that the users know the channel statistics.

By this preamble, let us stack all the pilots into the matrices $\sqrt{\tau_p P_p}\mathbf{\Phi}_1 \in \mathbb{C}^{\tau_p \times K}$ and $\sqrt{\tau_p P_p}\mathbf{\Phi}_2 \in \mathbb{C}^{\tau_p \times K}$, where P_p denotes the average pilot power. We assume that $\tau_p \geq 2K$ to guarantee that pilot sequences are mutually orthogonal such that $\mathbf{\Phi}_1^H\mathbf{\Phi}_1 = \mathbf{I}_K$, $\mathbf{\Phi}_2^H\mathbf{\Phi}_2 = \mathbf{I}_K$ and $\mathbf{\Phi}_1^H\mathbf{\Phi}_2 = \mathbf{0}_K$. Note that the process for the channel estimation should be done in the digital domain. Thus, we take into account the signals after $\mathbf{F}_1$ and $\mathbf{F}_2$ which exactly refer to the baseband signals before the DSP unit. Then, the baseband received signal before the DSP unit is given by

$$\mathbf{Y}_{pr} = \sqrt{\tau_p P_p}\mathbf{F}_1\mathbf{G}_1\mathbf{\Phi}_1^T + \mathbf{N}_{pr}, \tag{4.8}$$

where $\mathbf{N}_{pr}$ is an AWGN matrix including i.i.d. $\mathcal{CN}(0,1)$ entries. Then, after multiplying the received pilot sequences by $\mathbf{\Phi}_1^*$, we get the new observation matrix $\mathbf{Y}_1$ so

that

$$\mathbf{Y}_1 = \sqrt{\tau_p P_p} \mathbf{F}_1 \mathbf{G}_1 + \mathbf{N}_1 \tag{4.9}$$

where the $\mathbf{N}_1$ is an AWGN matrix whose entries follow the same distribution as $\mathbf{N}_{pr}$ entries due to the fact that columns of $\boldsymbol{\Phi}_1$ are orthonormal. Now, by employing a conventional MMSE estimator, the estimate of the effective channel is given by

$$\hat{\mathbf{A}}_1 = \sqrt{\tau_p P_p} \mathbf{Y}_1 \left(\tau_p P_p \mathbf{D}_1 + \mathbf{I}_K \right)^{-1} \mathbf{D}_1, \tag{4.10}$$

which implies that the k-th column of this estimation matrix is $\mathcal{CN}\left(\mathbf{0}, \hat{\beta}_{1,k}\mathbf{I}_K\right)$ where $\hat{\beta}_{1,k} = \frac{\tau_p P_p \beta_{1,k}^2}{1 + \tau_p P_p \beta_{1,k}}$. The estimation error can also be written from $\mathbf{A}_1 = \hat{\mathbf{A}}_1 + \mathbf{E}_1$ where $\hat{\mathbf{A}}_1$ and $\mathbf{E}_1$ are independent due to their jointly Gaussian distribution and the orthogonality principle of the MMSE estimator. Therefore, we can conclude that $\hat{\mathbf{A}}_1 = \hat{\mathbf{H}}_1 \hat{\mathbf{D}}_1^{1/2}$ and $\mathbf{E}_1 = \mathbf{H}_{e1} \mathbf{D}_{e1}^{1/2}$, where $\hat{\mathbf{H}}_1$ and $\mathbf{H}_{e1} \in \mathbb{C}^{K_a \times K}$ are random matrices with i.i.d. $\mathcal{CN}(0,1)$ elements. Also, $\hat{\mathbf{D}}_1 \triangleq \mathrm{diag}\{\hat{\beta}_{1,1}, \ldots, \hat{\beta}_{1,K}\}$ and $\mathbf{D}_{e1} \triangleq \mathrm{diag}\{\beta_{e1,1}, \ldots, \beta_{e1,K}\}$. The same results are analogous for $\mathbf{A}_2$ and $\mathbf{E}_2$.

4.3 Achievable Rate

In this section, we derive a lower bound on the achievable rate. By recalling (4.7), the received signal at the k-destination user can be treated as a desired signal plus an independent effective noise, n_{eff_k}, where we assume that the MRC/MRT transformation invoked in this scenario can be formulated as $\mathbf{W} = \alpha \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H$ with a normalization factor α to meet the total power constraint in (4.6)

$$y_{D_k} = \sqrt{P_u} \alpha \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{a}}_{1,k}\right] x_k + n_{\mathrm{eff}_k}. \tag{4.11}$$

In the above equation, the effective noise is given by

$$n_{\text{eff}_k} = \sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}x_k - \sqrt{P_u}\alpha\mathbb{E}\big[\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big]x_k$$

$$+ \sqrt{P_u}\alpha\sum_{i\neq k}\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,i}x_i + \sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x}$$

$$+ \sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{A}}_1\mathbf{x} + \sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x}$$

$$+ \alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + \alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + n_{D_k}, \tag{4.12}$$

where user nodes do not have access to the channel estimates at the relay station, and the entropy of the effective noise can be upper-bounded by a Gaussian model in the worst case. Hence, a lower bound on the ergodic sum achievable rate, or simply achievable rate, can be obtained by[1]

$$R = \frac{T - \tau_p}{2T}\sum_{k=1}^{K}\log_2\left(1 + \text{SINR}_k\right), \quad \text{for } K \leq K_a, K_b \leq N, \tag{4.13}$$

where the pre-log factor $\frac{T-\tau_p}{2T}$ accounts for the half-duplex relaying operation and pilot overhead, and the SINR_k is given by

$$\text{SINR}_k = \frac{P_u\alpha^2\left|\mathbb{E}\big[\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big]\right|^2}{P_u\alpha^2\left(t_1 + t_2 + t_3 + t_4 + t_5\right) + \alpha^2\left(t_6 + t_7\right) + \sigma_{n_D}^2}, \tag{4.14}$$

where the denominator terms are given as follows:

$$t_1 = \text{var}\big(\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big), \tag{4.15}$$

$$t_2 = \sum_{i\neq k}\mathbb{E}\Big[\big|\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,i}\big|^2\Big], \tag{4.16}$$

$$t_3 = \mathbb{E}\Big[\big\|\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\big\|_2^2\Big], \tag{4.17}$$

$$t_4 = \mathbb{E}\Big[\big\|\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{A}}_1\big\|_2^2\Big], \tag{4.18}$$

---

[1]In our subsequent analysis, we assume that the bandwidth is normalized to 1Hz such that equation (4.13) represents indeed the achievable rate.

$$t_5 = \mathbb{E}\left[\left\|\mathbf{e}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \mathbf{E}_1\right\|_2^2\right], \tag{4.19}$$

$$t_6 = \mathbb{E}\left[\left|\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \tilde{\mathbf{n}}_R\right|^2\right], \tag{4.20}$$

$$t_7 = \mathbb{E}\left[\left|\mathbf{e}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \tilde{\mathbf{n}}_R\right|^2\right]. \tag{4.21}$$

The following is a set of propositions to further simplify the SINR terms, and provide a closed-form expression.

Proposition 4.1. *Consider the estimates of the two effective channels $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$. Then, we can conclude that*

$$\left|\mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{a}}_{1,k}\right]\right|^2 = K_a^2 K_b^2 \hat{\beta}_{1,k}^2 \hat{\beta}_{2,k}^2. \tag{4.22}$$

Proof. We conclude the proposition by recalling $\hat{\mathbf{A}}_1 = \hat{\mathbf{H}}_1 \hat{\mathbf{D}}_1^{1/2}$, $\hat{\mathbf{A}}_2 = \hat{\mathbf{H}}_2 \hat{\mathbf{D}}_2^{1/2}$, and borrowing Lemma 1 from Appendix B.1 in $(L1)$ below, to calculate the (k,k)-th entry of $\mathbb{E}\left[\hat{\mathbf{A}}_2^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1\right]$, which is given by

$$\begin{aligned}
\mathbb{E}\left[\hat{\mathbf{A}}_2^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1\right]_{k,k} &= \mathbb{E}\left[\hat{\mathbf{D}}_2^{\frac{1}{2}} \hat{\mathbf{H}}_2^H \hat{\mathbf{H}}_2 \hat{\mathbf{D}}_2^{\frac{1}{2}} \hat{\mathbf{D}}_1^{\frac{1}{2}} \hat{\mathbf{H}}_1^H \hat{\mathbf{H}}_1 \hat{\mathbf{D}}_1^{\frac{1}{2}}\right]_{k,k} \\
&\stackrel{(L1)}{=\!=\!=} \left[K_a K_b \hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2\right]_{k,k} = K_a K_b \hat{\beta}_{1,k} \hat{\beta}_{2,k}. \tag{4.23}
\end{aligned}$$

$\square$

Proposition 4.2. *Given the estimates of two effective channels $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$, we have*

$$t_1 = K_a K_b (K_a + K_b) \hat{\beta}_{1,k}^2 \hat{\beta}_{2,k}^2 + K_a K_b \hat{\beta}_{1,k} \hat{\beta}_{2,k} \mathrm{Tr}\left(\hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2\right). \tag{4.24}$$

Proof. See Appendix B.2. $\square$

Proposition 4.3. *Consider the estimates of two effective channels $\hat{\mathbf{A}}_1$, $\hat{\mathbf{A}}_2$. Then, it can be shown that the multipair interference can be simplified as*

$$t_2 = K_b\big(K_a^2 + K_a\big)\hat{\beta}_{2,k} \sum_{i \neq k}^{K} \hat{\beta}_{1,i}^2 \hat{\beta}_{2,i} + K_a(K_b^2 + K_b)\hat{\beta}_{1,k}\hat{\beta}_{2,k}^2 \sum_{i \neq k}^{K} \hat{\beta}_{1,i}$$

$$+ K_a K_b \hat{\beta}_{2,k} \sum_{i \neq k}^{K} \sum_{m \neq i,k}^{K} \hat{\beta}_{1,i}\hat{\beta}_{1,m}\hat{\beta}_{2,m}. \tag{4.25}$$

Proof. In the first stage, we expand the matrix multiplication across the columns, and then we use the column orthogonality of matrices $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$ to get a simple expression. $\square$

Proposition 4.4. *Consider the estimates of two effective channels $\hat{\mathbf{A}}_1$, $\hat{\mathbf{A}}_2$ and the estimation error matrix $\mathbf{E}_1$. Then,*

$$t_3 = K_a K_b \hat{\beta}_{2,k} \text{Tr}\big(\mathbf{D}_{e1}\big)\Big(K_b \hat{\beta}_{1,k}\hat{\beta}_{2,k} + \text{Tr}\big(\hat{\mathbf{D}}_1\hat{\mathbf{D}}_2\big)\Big). \tag{4.26}$$

Proof. See Appendix B.4. $\square$

The following results can also be obtained in a similar fashion

$$t_5 = K_a K_b \beta_{e2,k} \text{Tr}\big(\mathbf{D}_{e1}\big)\text{Tr}\big(\hat{\mathbf{D}}_1\hat{\mathbf{D}}_2\big), \tag{4.27}$$

$$t_6 = K_a K_b \sigma_{n_R}^2 \hat{\beta}_{2,k}\Big(K_b \hat{\beta}_{1,k}\hat{\beta}_{2,k} + \text{Tr}\big(\hat{\mathbf{D}}_1\hat{\mathbf{D}}_2\big)\Big), \tag{4.28}$$

$$t_7 = K_a K_b \sigma_{n_R}^2 \beta_{e2,k} \text{Tr}\big(\hat{\mathbf{D}}_1\hat{\mathbf{D}}_2\big). \tag{4.29}$$

Proposition 4.5. *Let $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_2$ represent the effective channel estimates, then, we have*

$$t_4 = K_a^2 K_b \beta_{e2,k} \text{Tr}\big(\hat{\mathbf{D}}_1^2\hat{\mathbf{D}}_2\big) + K_a K_b \beta_{e2,k} \text{Tr}\big(\hat{\mathbf{D}}_1\big)\text{Tr}\big(\hat{\mathbf{D}}_1\hat{\mathbf{D}}_2\big). \tag{4.30}$$

Proof. See Appendix B.5. $\square$

As was pointed out earlier, the normalization factor α is chosen to meet the long-term power constraint at the AF relay station. Starting from (4.6), followed by Lemma

3 and then applying Lemma 1 from Appendix B.1, we obtain

$$\alpha = \sqrt{\frac{P_r}{K_a K_b \left(c_1 + c_2 + c_3 + c_4\right)}}, \tag{4.31}$$

where $c_1 = P_u K_a \text{Tr}\big(\hat{\mathbf{D}}_1^2 \hat{\mathbf{D}}_2\big)$, $c_2 = P_u \text{Tr}\big(\hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2\big) \text{Tr}\big(\hat{\mathbf{D}}_1\big)$, $c_3 = P_u \text{Tr}\big(\hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2\big) \text{Tr}\big(\mathbf{D}_{e1}\big)$, and $c_4 = \sigma_{n_R}^2 \text{Tr}\big(\hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2\big)$.

To get better insights into how different parameters affect the achievable rate, we consider a special case without the large-scale fading effects and with the same amount of RF chains at the both sides of the relay station, i.e. $K_a = K_b$. Also, we set $\tau_p = 2K$ to satisfy the minimum orthogonal pilot requirement. Next, we choose the dominant terms, which scale faster with K_a, K_b and K, in the numerator and denominator of SINR_k and ignore the other terms. Consequently, we get

$$R \approx \gamma \log_2 \left(1 + \frac{K_a^4}{2KK_a^3 + K^2K_a^2 + \frac{\sigma_{n_D}^2}{P_r}\left(KK_a^3 + K^2K_a^2\right)}\right), \quad \text{for } K \leq K_a, K_b \leq N, \tag{4.32}$$

where γ is defined as $\frac{K(T-2K)}{2T}$. Roughly speaking, this equation reveals that the ratio of number of RF chains to the scheduled users ($\frac{K_a}{K}$) boosts the power gain, while the number of scheduled users additionally contributes to the multiplexing gain (i.e. see the prelog factor). Moreover, the power of users does not appear in this approximation. Hence, this observation implies that the power of the relay station P_r plays a far more important role than the users' power P_u which has been already eliminated in this approximation.

4.4   Simulation Results

In this section, we provide Monte Carlo simulations to evaluate the performance of hybrid multipair massive relaying with the channel estimation and validate our proposed achievable bound. We assume that the small-scale fading encountered by the system model is i.i.d. Rayleigh. Unless otherwise specified, we assume that $K = 10$,

$N = 128$, $P_u = P_r = P_p = 13$ dB, $\sigma_{n_R}^2 = \sigma_{n_D}^2 = 0$ dB, $T = 196$, $\tau_p = 2K$, and $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}_K$.

Figure 4.2 represents the performance of multipair massive relaying for different number of RF chains. Since there are $N = 128$ relay antennas, deploying 128 RF chains represents a canonical fully digital massive relaying. The red curve is our proposed lower bound which evaluates the performance assessment of the signal model on (4.11):

$$
\begin{aligned}
y_{D_k} = & \underbrace{\sqrt{P_u}\alpha\mathbb{E}\big[\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big]x_k}_{\text{desired signal}} \\
& + \underbrace{\sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}x_k - \sqrt{P_u}\alpha\mathbb{E}\big[\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big]x_k}_{\text{fluctuation of desired signal}} \\
& + \underbrace{\sqrt{P_u}\alpha\sum_{i\neq k}\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,i}x_i}_{\text{inter}-\text{user interference}} + \underbrace{\sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x}}_{\text{estimation error}} \\
& + \underbrace{\sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{A}}_1\mathbf{x} + \sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x}}_{\text{estimation error}} \\
& + \underbrace{\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + \alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + n_{D_k}}_{\text{aggregated noise}}, \quad\quad\quad\quad (4.33)
\end{aligned}
$$

where we introduce the first term as a desirable signal due to the fact that the channel estimates are only available at the relay station. However, it is reasonable to assume that the long-term statistic of channels are available at the destinations. The second term represents a fluctuation over the desired signal which should be treated as a new source of ambiguity at the destination. The third term denotes the inter-user interference, while the rest describe the compound noise which includes AWGN and channel estimation errors. The red curve implies that by removing the half of RF chains, the hybrid system can still capture 75% of the spectral efficiency offered by the fully digital massive MIMO relaying. In order to understand how tight is our lower bound, Fig. 4.2 also illustrates an idealistic case, namely *upper bound*, which is generated by means of Monte Carlo simulation, see the blue curve. In this case, we assume that the
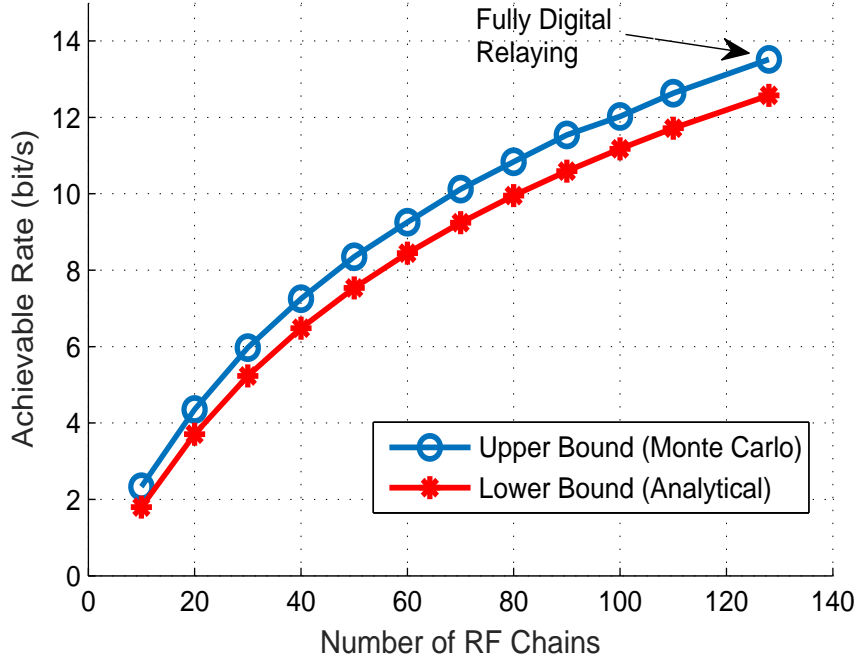
Figure 4.2: Achievable rate as a function of number of RF chains with imperfect CSI at the relay station.

destinations know the channel estimates $\hat{\mathbf{A}}_1$, $\hat{\mathbf{A}}_2$, and consequently this signal model provides a higher achievable rate. Precisely speaking, under this assumption, the received baseband signal in (4.33) can be substituted by the less strict model given as follows

$$
y_{D_k} = \underbrace{\sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}x_k}_{\text{desired signal}} + \underbrace{\sqrt{P_u}\alpha\sum_{i\neq k}\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,i}x_i}_{\text{inter}-\text{user interference}}
$$

$$
+ \underbrace{\sqrt{P_u}\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x} + \sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{A}}_1\mathbf{x}}_{\text{estimation error}}
$$

$$
+ \underbrace{\sqrt{P_u}\alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\mathbf{E}_1\mathbf{x}}_{\text{estimation error}} + \underbrace{\alpha\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + \alpha\mathbf{e}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\tilde{\mathbf{n}}_R + n_{D_k}}_{\text{aggregated noise}}, \quad (4.34)
$$

where the fluctuation of desired signal is not a matter anymore. Figure 4.3 showcases the difference of achievable rate between our analytical lower bound and Monte Carlo upper bound in percentage. It reveals that by increasing the number of RF chains, the gap between these bounds is diminished, and with half of the RF chain this gap is around 8%. Hence, the analytical lower bound is indeed tight even for moderate
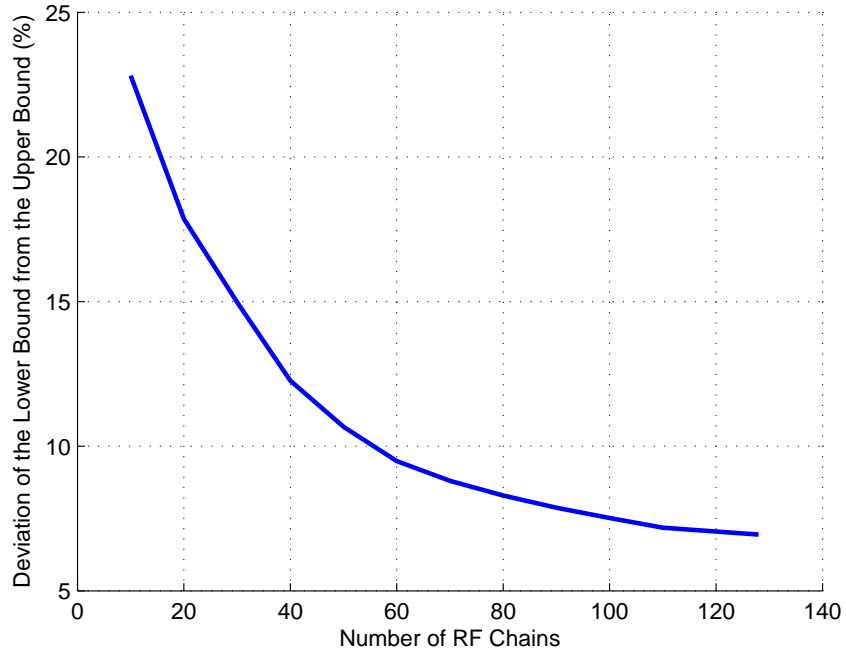
Figure 4.3: Gap between the proposed lower bound and Monte Carlo upper bound, assuming imperfect CSI at the relay station.

number of RF chains.

In general, there is a tradeoff between the data transmission rate and accuracy of the channel estimation. In other words, the more pilot signals are dedicated to estimate the channels, the less payload data we can pass through the system. In this light, the number of scheduled users plays an important role to maintain the balance between these two factors, and consequently, achieve the maximum system achievable rate, according to (4.13). Figure 4.4 depicts the achievable rate as a function of the number of scheduled users and illustrates the aforementioned trade-off. It can also be observed that the best operating point for this relaying system with $N = 128$ antennas and $K_a = K_b = 30$ RF chains can be achieved by $K = 11$ transmitter users which yields 6 bit/s achievable rate; interestingly, this achievable rate scales linearly with the number of RF chains.

We have also provided another simulation based on practical parameter values to evaluate the system performance under large-scale fading effects. In this scenario, we assume K users are uniformly distributed in a circular area with a radius of 1000 meters around the relay station, but no closer than $r_g = 100$ meters. The small-scale fading
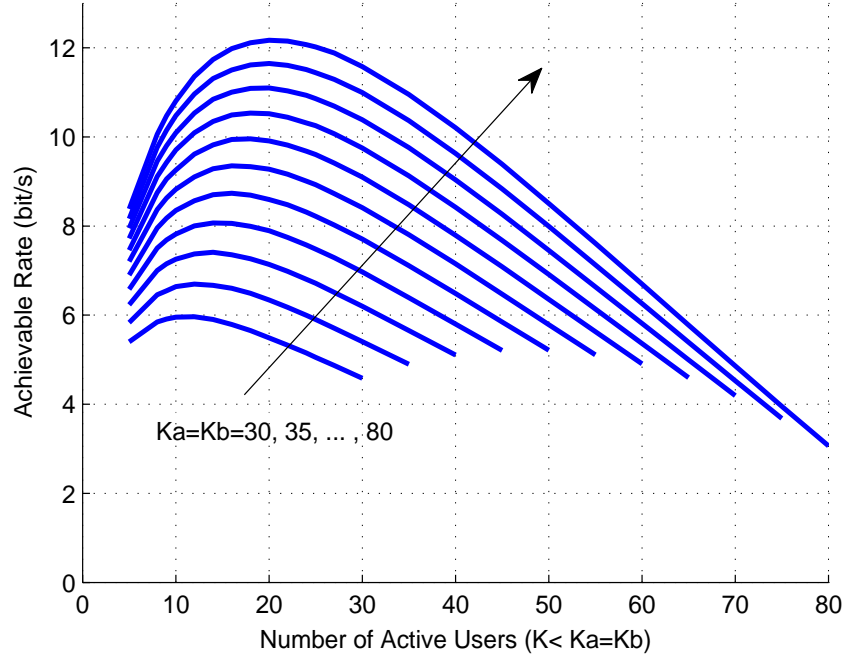
Figure 4.4: Achievable rate as a function of the scheduled transmitter users with imperfect CSI at the relay station.

encountered by the system model is Rayleigh which represents an isotropic scattering environment. Furthermore, we model the large-scale fading effects with a log-normal random variable, with standard deviation σ_{sh}, scaled by $(r_k/r_g)^{-\nu}$ to incorporate the path-loss effects. In this model, r_k denotes the distance of k-th user from the relay station, ν is the path-loss exponent, and the noise power is given by

$$\text{noise power} = \text{BW} \times k_B \times T_0 \times \text{noise figure}, \qquad (4.35)$$

where, BW denotes the spectral bandwidth in Hz, k_b is Boltzmann's constant in Joules

Parameter	Value	Parameter	Value
Number of RF chains	$K_a = K_b$	Coherence time	$T = 196$ symbols
Number of antennas	$N = 2K_a$	Pathloss-exponent	$\nu = 2$
Number of users	$K = 8$	Shadowing STD	$\sigma_{\text{sh}} = 4$ dB
Relay power	$P_r = 1$ W	Boltzmann	$k_b = 1.381 \times 10^{-23}$ J/K
User's power	$P_u = 0.1$ W	Temperature	$T_0 = 290$ K
Pilot power	$P_p = 0.1$ W	Noise figure	9 dB
Pilot length	$\tau_p = 2K$	Bandwidth	BW$= 20$ MHz

Table 4.1: Simulation parameters of system model with imperfect CSI at the relay station.
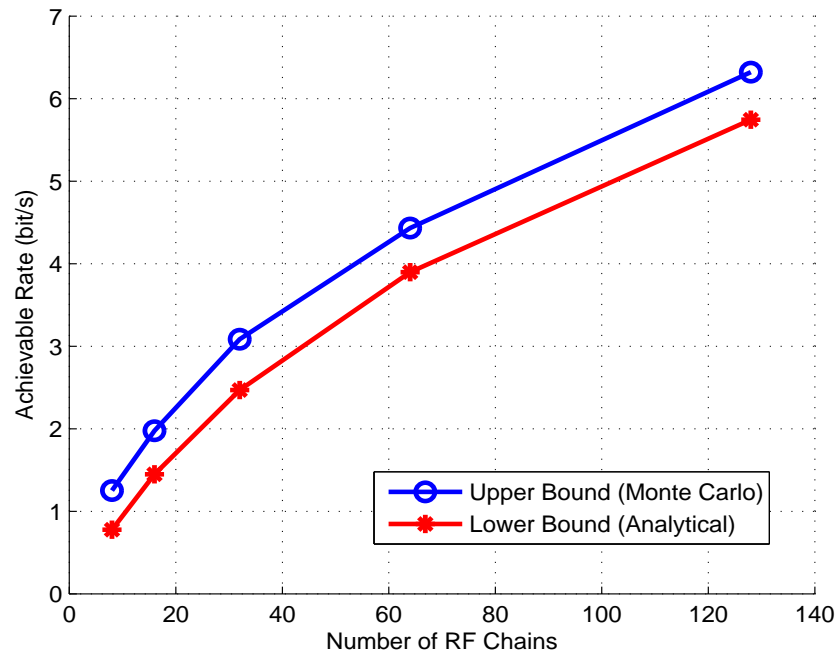
Figure 4.5: Achievable rate as a function of number of RF chains while $\frac{N}{K_a} = 2$. The parameter values are detailed in Table 4.1.

per Kelvin (J/K), T_0 is the outdoor absolute temperature in Kelvin (K), and then noise power is expressed in Watt.

Figure 4.5 represents our simulation result considering the parameter settings in Table 4.1. In this figure, the red curve showcases our analytical lower bound based on (4.33). On the other hand, the blue curve illustrates the Monte Carlo simulation result which validates our benchmark upper bound from the signal model in (4.34). This figure returns an acceptable gap between the proposed lower bound and the upper bound, and hence, our bound is tight even when the large-scale fading is also taken into account.

4.5 Conclusion

Massive relaying as a key technology of 5G induces a high implementation cost and power consumption due to the large number of RF chains. In this chapter, we considered a hybrid topology to address these challenges through a reduced number of RF chain. Then, we derived a tractable lower bound on the achievable rate, assuming that

perfect CSI is not available at the relay station. A high advantage of the proposed solution is that the analog beamforming was implemented using Hadamard matrices with the minimum resolution of phase shifters. This reduces the implementation complexity of our system even more. Most importantly, it was shown that the performance penalty compared to a fully digital configuration when deploying less than 50% RF chains was only 25%.

Chapter 5

Hybrid Processing Design for Multipair Massive MIMO Relaying with Channel Spatial Correlation

5.1 Introduction

As we discussed earlier, the practical implementation of massive MIMO induces some critical challenges. In particular, having one RF chain per antenna will boost to unprecedented levels the circuit complexity, fabrication/implementation cost and power consumption. In Chapter 3, we addressed part of these issues by a cascade structure of an analog RF beamformer and digital baseband processor referred as hybrid analog and digital (A/D) structure. Then, we relaxed the resolution of phase shifters in analog beamformers to arbitrary quantization bits, and also derived a closed-form expression for the achievable rate. However, in this scenario, the phase shifters should be able to adapt to the quick variations of the propagation channels over time. This phase adaptation, not only requires perfect CSI at the relay station, but also is a challenging task due to less flexibility of analog beamformers compared to the digital ones. To avoid this, in Chapter 4, we considered the fixed analog beamformers based on Hadamard matrices to alleviate the burden of signal processing at the relay station. However, in this

methodology, the propagation channel and analog beamformers are treated together as a single effective channel. Therefore, this assumption restricts the channel size to the number of RF chains, and consequently restricts the ability of the number of antennas to boost the power gain especially in the absence of perfect CSI.

At the other extreme, it has been long recognized that the achievable rate of point-to-point MIMO systems is deteriorated due to the spatial correlation [67,92,93]. Nevertheless, spatial correlation in multiuser MIMO systems can be exploited in the transceiver design offering performance improvements [94, 95]. To the best of our knowledge, there is no prior work in the context of hybrid MIMO relaying design under spatially correlated fading channels. In this chapter, motivated by the above discussion and due to the fact that analog beamformers are implemented in the RF domain with less flexibility than in baseband domain, we adjust the analog beamformers to the slow variation of channel statistics rather than the short-term fluctuations of channel. Thus, we will design our correlation-based analog beamformers to leverage the long-term channel spatial selectivity for the hybrid relaying structure. The main contributions of this chapter are summarized as follows:

- We consider a relay station with a hybrid architecture, and take the spatial correlation and imperfect CSI into account. In this scenario, we explicitly evaluate the role of digital beamformer by developing analytical bounds on the achievable rate for the prevalent digital schemes, i.e., MRC/MRT and ZF. Our closed-form expressions involve only the statistical parameters of the channels. Also, our numerical results reveal that the ZF scheme avails of channel correlation and offers higher achievable rate compared to the MRC/MRT scheme, even for low SNR values.

- We design a correlation-based analog beamformer which exploits the long-term eigenmodes of the propagation channel to maximize the achievable rate. Our analytical result showcases that, although, multiplexing and array gain are obviously restricted under this hybrid A/D architecture, the system can still avail of promising spatial diversity gain. Moreover, our simulation results illustrate that

with only 50 RF chains, in a typical massive MIMO system with 128 antennas elements, the hybrid A/D structure can nearly capture 90% of the achievable rate offered by the conventional fully digital structure with 128 RF chains.

- We numerically evaluate the achievable rate of hybrid structure under quantized phase shifter assumption. Our simulation results reveal that the hybrid configuration paradigm is robust to phase quantization, and this observation is more pronounced for the hybrid structure utilizing the MRC/MRT scheme.

5.2 System Model

5.2.1 Signal Model and Hybrid Architecture

We assume the same system model as in previous chapters, which is illustrated in Fig. 5.1. Considering that users are randomly located around the relay station, we assume that the propagation channels at the users' terminals are uncorrelated due to their large distance from each other, whereas there exists correlation among the relay antennas on both sides of the relay station. This correlation occurs if the relay antennas are insufficiently spaced from each other or there are a limited number of scatterers surrounding the relay station. We also assume that there are no direct links between K pairs due to the heavy shadowing and/or path loss attenuation. For the simplicity of analysis, and for obtaining a clear understanding of the impact of the antennas correlation, the large-scale fading is neglected. Nevertheless, it is notable that our results can be readily extended to propagation channels which include large-scale fading.[1] Having a case of practical interest, we assume that CSI is not available at the users' nodes, and the relay station has to estimate the channels via the uplink pilots during the training phase.

Under this configuration setup, we recall our signal models for the received and transmitted signals at the relay station, and also the model for the received signals at

---

[1]The impact of large-scale fading can be compensated by a power control scheme at the relay station which is proportional to the inverse of the channel long-term attenuation, including path-loss and shadow fading [96–99].
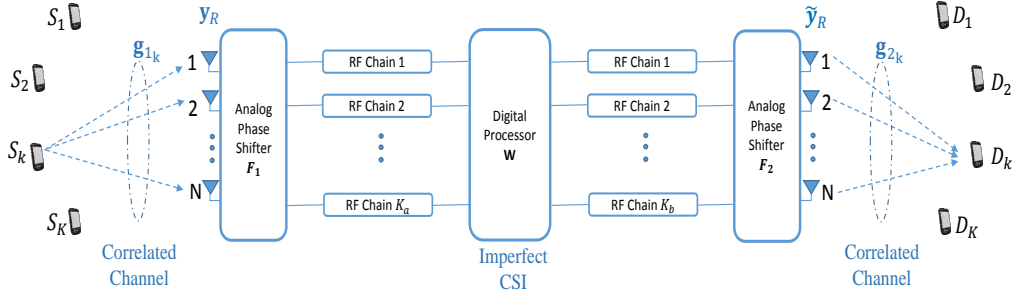
Figure 5.1: Simplified block diagram of a multipair relay system with a baseband digital processor combined with two analog RF beamformers. Channels are Rayleigh faded and correlated at the relay side. Only statistics and estimates of the propagation channels are available at the relay station.

K destinations in below

$$\mathbf{y}_R = \sqrt{P_u}\mathbf{G}_1\mathbf{x} + \mathbf{n}_R, \tag{5.1}$$

$$\tilde{\mathbf{y}}_R = \mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{y}_R, \tag{5.2}$$

$$\mathbf{y}_D = \sqrt{P_u}\mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{G}_1\mathbf{x} + \mathbf{G}_2^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + \mathbf{n}_D, \tag{5.3}$$

where the propagation channels from the K sources to the relay, and from the relay to the destinations are expressed as $\mathbf{G}_1$, $\mathbf{G}_2 \in \mathbb{C}^{N \times K}$, respectively. We represent the spatial correlation matrices between the relay antennas by $\mathbf{R}_1$, $\mathbf{R}_2 \in \mathbb{C}^{N \times N}$ which have been already incorporated into the channel matrices, i.e., $\mathbf{G}_1 = \mathbf{R}_1^{\frac{1}{2}}\mathbf{H}_1$ and $\mathbf{G}_2 = \mathbf{R}_2^{\frac{1}{2}}\mathbf{H}_2$. In this model, $\mathbf{H}_1$ and $\mathbf{H}_2 \in \mathbb{C}^{N \times K}$ denote the small-scale fading with i.i.d. $\mathcal{CN}(0,1)$ elements. Then, from (5.3), the received signal at the k-destination can be extracted as

$$y_{D_k} = \sqrt{P_u}\mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1,k}x_k + \sqrt{P_u}\sum_{i \neq k}^{K}\mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{g}_{1,i}x_i + \mathbf{g}_{2,k}^H\mathbf{F}_2^H\mathbf{W}\mathbf{F}_1\mathbf{n}_R + n_{D_k},$$

$$\tag{5.4}$$

where n_{D_k} is AWGN at the k-th destination. In this aggregated received signal, the first term refers to the desired signal and the second term signifies inter-user interference, while the last two terms represent the compound noise.

This system model is still under two major practical constraints. First, the analog combiner and precoder just contribute to the phase alignment as they are usually implemented using a bank of analog phase shifters. In this light, we have to assign an equal modulus to all the entries of matrices $\mathbf{F}_1$ and $\mathbf{F}_2$. We fix this modulus by $1/\sqrt{N}$ to avoid an unlimited gain in the analog domain. Second, the AF relay station receives the signals from all sources, and boosts them up to a certain level of power P_r before transmitting to the destinations. Thus, we consider the following long-term power constraint

$$\mathbb{E}\Big[\mathrm{Tr}\big(\tilde{\mathbf{y}}_R\tilde{\mathbf{y}}_R^H\big)\Big] = P_r. \tag{5.5}$$

5.2.2 Channel Estimation

To capture the advantages that massive MIMO relay can offer, CSI is required at the relay station. This CSI is used to design the digital processor W. The CSI acquisition at the relay station is done via the uplink pilots transmitted from the sources and the destinations. Let T be the length of each coherence interval, and τ_p be the duration for uplink training (in symbols). Then, the remaining part, $T - \tau_p$, is used for downlink data transmission. We assume that pilot sequences sent from all sources and destinations are mutually orthogonal. This requires $2K \leq \tau_p \leq T$. Let us stack all of these pilot sequences into the matrices $\sqrt{\tau_p P_p}\boldsymbol{\Phi}_1 \in \mathbb{C}^{\tau_p \times K}$ and $\sqrt{\tau_p P_p}\boldsymbol{\Phi}_2 \in \mathbb{C}^{\tau_p \times K}$, where P_p denotes the average pilot power. Then, we have $\boldsymbol{\Phi}_1^H\boldsymbol{\Phi}_1 = \mathbf{I}_K$, $\boldsymbol{\Phi}_2^H\boldsymbol{\Phi}_2 = \mathbf{I}_K$ and $\boldsymbol{\Phi}_1^H\boldsymbol{\Phi}_2 = \mathbf{0}_K$. The baseband received pilot signals at the receive and transmit sides of the relay station before the DSP unit can be expressed, respectively, as

$$\mathbf{Y}_{pr} = \sqrt{\tau_p P_p}\mathbf{F}_1\mathbf{G}_1\boldsymbol{\Phi}_1^T + \mathbf{N}_{pr}, \tag{5.6}$$

$$\mathbf{Y}_{pt} = \sqrt{\tau_p P_p}\mathbf{F}_2\mathbf{G}_2\boldsymbol{\Phi}_2^T + \mathbf{N}_{pt}, \tag{5.7}$$

where $\mathbf{N}_{pr} \in \mathbb{C}^{K_a \times \tau_p}$ and $\mathbf{N}_{pt} \in \mathbb{C}^{K_b \times \tau_p}$ are AWGN matrices including i.i.d. $\mathcal{CN}(0,1)$ entries. Hereafter, we just derive the results for the source side and the same result

can be similarly deduced for the destination side. We can obtain the matrix $\mathbf{Y}_1$ after projecting matrix $\mathbf{Y}_{pr}$ onto $\mathbf{\Phi}_1^*$

$$\mathbf{Y}_1 = \mathbf{Y}_{pr}\mathbf{\Phi}_1^* = \sqrt{\tau_p P_p}\mathbf{F}_1\mathbf{R}_1^{\frac{1}{2}}\mathbf{H}_1 + \mathbf{N}_1, \tag{5.8}$$

where the entries of $\mathbf{N}_1 \in \mathbb{C}^{K_a \times K}$ follow the same distribution as the entries of $\mathbf{N}_{pr}$ due to the fact that the columns of $\mathbf{\Phi}_1$ are orthonormal. Then, for a given channel correlation $\mathbf{R}_1$, the MMSE estimate of $\mathbf{G}_1$ can be obtained by [71]

$$\hat{\mathbf{G}}_1 = \sqrt{\tau_p P_p}\mathbf{R}_1\mathbf{F}_1^H\left(\tau_p P_p\mathbf{F}_1\mathbf{R}_1\mathbf{F}_1^H + \mathbf{I}_{K_a}\right)^{-1}\mathbf{Y}_1. \tag{5.9}$$

Since the MMSE estimator is a linear transformation, we can conclude that the channel estimation matrices $\hat{\mathbf{G}}_1$, $\hat{\mathbf{G}}_2$ and their errors $\mathbf{E}_1$, $\mathbf{E}_2$ are Gaussian random matrices, such that

$$\mathbf{G}_1 = \hat{\mathbf{G}}_1 + \mathbf{E}_1, \tag{5.10}$$

$$\mathbf{G}_2 = \hat{\mathbf{G}}_2 + \mathbf{E}_2. \tag{5.11}$$

After some standard manipulations, we can rewrite the channel estimate as

$$\begin{aligned}
\hat{\mathbf{G}}_1 &= \left(\frac{\mathbb{E}\left[\hat{\mathbf{G}}_1\hat{\mathbf{G}}_1^H\right]}{\mathrm{Tr}\left(\mathbb{E}\left[\hat{\mathbf{G}}_1\hat{\mathbf{G}}_1^H\right]\right)}\right)^{\frac{1}{2}}\hat{\mathbf{H}}_1\left(\mathbb{E}\left[\hat{\mathbf{G}}_1^H\hat{\mathbf{G}}_1\right]\right)^{\frac{1}{2}} \\
&= \left(\tau_p P_p\mathbf{R}_1\mathbf{F}_1^H\left(\tau_p P_p\mathbf{F}_1\mathbf{R}_1\mathbf{F}_1^H + \mathbf{I}_{K_a}\right)^{-1}\mathbf{F}_1\mathbf{R}_1\right)^{\frac{1}{2}}\hat{\mathbf{H}}_1 \\
&= \left(\mathbf{R}_1 - \left(\mathbf{R}_1^{-1} + \tau_p P_p\mathbf{F}_1^H\mathbf{F}_1\right)^{-1}\right)^{\frac{1}{2}}\hat{\mathbf{H}}_1, \tag{5.12}
\end{aligned}$$

where $\hat{\mathbf{H}}_1 \in \mathbb{C}^{N \times K}$ is a Gaussian random matrix with i.i.d. $\mathcal{CN}(0,1)$ entries, and the last equality is derived by invoking the matrix inversion lemma. We can derive the same result in a similar fashion for the error matrix. As a consequence, we gather both

results into the following simple expressions

$$\hat{\mathbf{G}}_1 = \mathbf{U}_1^{\frac{1}{2}} \hat{\mathbf{H}}_1, \tag{5.13}$$

$$\mathbf{E}_1 = \mathbf{U}_{e_1}^{\frac{1}{2}} \mathbf{H}_{e1}, \tag{5.14}$$

where $\mathbf{H}_{e1} \in \mathbb{C}^{N \times K}$ is independent of $\hat{\mathbf{H}}_1$ and its entries are independent $\mathcal{CN}(0,1)$ random variables. Furthermore, $\mathbf{U}_{e_1}$ and $\mathbf{U}_1$ are the error correlation matrix and estimation correlation matrix, given respectively by

$$\mathbf{U}_{e_1} \triangleq \left(\mathbf{R}_1^{-1} + \tau_p P_p \mathbf{F}_1^H \mathbf{F}_1 \right)^{-1}, \tag{5.15}$$

$$\mathbf{U}_1 \triangleq \mathbf{R}_1 - \mathbf{U}_{e_1}. \tag{5.16}$$

5.3 Achievable Rate Analysis

In this section, we analytically evaluate the achievable rate of hybrid configuration paradigm to get better insights into how analog beamformers affect the system performance. The achievable rate is derived via the "use and forget" technique which is commonly used in the context of massive MIMO [20]. Let us recall (5.4) and take the channel estimations (5.10), (5.11) into account, then we have

$$y_{D_k} = \underbrace{\mathbb{E}\left[\sqrt{P_u}\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,k}\right] x_k}_{\text{desired signal}} + \underbrace{\sqrt{P_u}\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,k} x_k - \mathbb{E}\left[\sqrt{P_u}\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,k}\right] x_k}_{\text{desired signal fluctuation}}$$

$$+ \underbrace{\sqrt{P_u} \sum_{i \neq k}^{K} \hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,i} x_i}_{\text{inter-user interference}} + \underbrace{\sqrt{P_u}\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \mathbf{E}_1 \mathbf{x}}_{\text{estimation error}} + \underbrace{\sqrt{P_u}\mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{G}}_1 \mathbf{x}}_{\text{estimation error}}$$

$$+ \underbrace{\sqrt{P_u}\mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \mathbf{E}_1 \mathbf{x}}_{\text{estimation error}} + \underbrace{\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 2 + \mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 2 + n_{D_k}}_{\text{aggregated noise \& estimation error}}. \tag{5.17}$$

We introduce the first term as a desirable signal due to the fact that the channel estimates are only available at the relay station. However, it is reasonable to assume that the long-term statistic of the channels are available at the destinations. The sec-

ond term represents a fluctuation over the desired signal which should be treated as a new source of ambiguity at the destination. The third term points out to the inter-user interference, while the rest describe the compound noise which includes AWGN and channel estimation errors. Thus, we can obtain a sum achievable rate as

$$R = \frac{T - \tau_p}{2T} \sum_{k=1}^{K} \log_2 \left(1 + \text{SINR}_k \right), \tag{5.18}$$

where,

$$\text{SINR}_k = \frac{P_u t_0}{P_u \left(t_1 + t_2 + t_3 + t_4 + t_5 \right) + \left(t_6 + t_7 \right) + \sigma_{n_D}^2}. \tag{5.19}$$

In (5.18), the factor $\frac{T-\tau_p}{2T}$ denotes the penalty loss due to the half-duplex relaying operation and the channel estimation overhead. In (5.19), $t_0, t_1, \ldots, t_7$ are defined as

$$t_0 = \left| \mathbb{E} \left[\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,k} \right] \right|^2, \tag{5.20}$$

$$t_1 = \text{Cov} \left(\hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,k} \right), \tag{5.21}$$

$$t_2 = \mathbb{E} \left[\left| \sum_{i \neq k} \hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{g}}_{1,i} x_i \right|^2 \right], \tag{5.22}$$

$$t_3 = \mathbb{E} \left[\left| \hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \mathbf{E}_1 \mathbf{x} \right|^2 \right], \tag{5.23}$$

$$t_4 = \mathbb{E} \left[\left| \mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \hat{\mathbf{G}}_1 \mathbf{x} \right|^2 \right], \tag{5.24}$$

$$t_5 = \mathbb{E} \left[\left| \mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \mathbf{E}_1 \mathbf{x} \right|^2 \right], \tag{5.25}$$

$$t_6 = \mathbb{E} \left[\left| \hat{\mathbf{g}}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \tilde{\mathbf{n}}_R \right|^2 \right], \tag{5.26}$$

$$t_7 = \mathbb{E} \left[\left| \mathbf{e}_{2,k}^H \mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 \tilde{\mathbf{n}}_R \right|^2 \right], \tag{5.27}$$

where for the sake of simplicity we omitted index k that indicates these terms that correspond to the k-th user pair. It is also notable that in the above equations t_0–t_7, the expectation is over all the random variables: information symbols, estimated channels, channels estimation errors, and AWGN. Since finding the optimal $\mathbf{W}$ is a demanding task due to the non-convex nature of the problem, in the subsections that follow we

adopt MRC/MRT and ZF digital processors to simplify the achievable rate under these simple but good linear schemes. Then, given the closed-form expression in this section, we can design our analog beamformer in the next section (Section 5.5) to improve the achievable rate.

5.4 MRC/MRT Digital Processor

MRC/MRT strategy is a simple choice that coherently combines the received signals, and then sends them toward the destinations. With the MRC/MRT scheme, $\mathbf{W}$ is chosen so that $\mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 = \alpha_{\mathrm{m}} \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H$, where α_{m} is a relay amplification factor chosen to satisfy the long-term constraint in (5.5). Next, we provide some propositions that further simplify the SINR terms $t_0 - t_7$, and provide a closed-form expression for the achievable rate which only involves the long-term statistics of the propagation channels. For ease of exposition, we use superscript "mrc" to denote the MRC/MRT scheme.

Proposition 5.1. *With MRC/MRT, the mathematical term corresponding to the desired signal and its fluctuations, t_0 and t_1, can be expressed in closed-form, respectively, as follows*

$$t_0^{\mathrm{mrc}} = \alpha_{\mathrm{m}}^2 \mathrm{Tr}^2(\mathbf{U}_1) \mathrm{Tr}^2(\mathbf{U}_2), \tag{5.28}$$

$$t_1^{\mathrm{mrc}} = \alpha_{\mathrm{m}}^2 \left(\mathrm{Tr}^2(\mathbf{U}_1) \|\mathbf{U}_2\|^2 + \mathrm{Tr}^2(\mathbf{U}_2) \|\mathbf{U}_1\|^2 + K \|\mathbf{U}_1\|^2 \|\mathbf{U}_2\|^2 \right). \tag{5.29}$$

Proof. The proof of the first part is trivial by applying Lemma 1 in Appendix C.1. For the second part let us define an auxiliary variable namely $t_{\mathrm{aux}}^{\mathrm{mrc}}$, and then simplify it by expanding the matrix product around its columns. Then, considering this fact that the columns of the estimation channels are independent, and by recalling Lemma 2 we have

$$
\begin{aligned}
t_{\mathrm{aux}}^{\mathrm{mrc}} &= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\left|\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \hat{\mathbf{g}}_{1,k}\right|^2\right] \\
&= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\left|\hat{\mathbf{g}}_{2,k}^H \left(\sum_{m=1}^{K} \hat{\mathbf{g}}_{2,m} \hat{\mathbf{g}}_{1,m}^H\right) \hat{\mathbf{g}}_{1,k}\right|^2\right] \\
&= \alpha_{\mathrm{m}}^2 \sum_{m=1}^{K} \mathbb{E}\left[\left|\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{g}}_{2,m} \hat{\mathbf{g}}_{1,m}^H \hat{\mathbf{g}}_{1,k}\right|^2\right] \\
&\stackrel{\text{(L2)}}{=} \alpha_{\mathrm{m}}^2 \left(\mathrm{Tr}^2(\mathbf{U}_2) + \|\mathbf{U}_2\|^2\right)\left(\mathrm{Tr}^2(\mathbf{U}_1) + \|\mathbf{U}_1\|^2\right) + \alpha_{\mathrm{m}}^2 (K-1)\|\mathbf{U}_1\|^2 \|\mathbf{U}_2\|^2.
\end{aligned}
$$

$$(5.30)$$

The proof is completed by using $t_1^{\mathrm{mrc}} = t_{\mathrm{aux}}^{\mathrm{mrc}} - t_0^{\mathrm{mrc}}$. For the sake of notational simplicity, hereafter, we use (L1), (L2), and (L3) to denote Lemmas 1–3. $\qquad\square$

Proposition 5.2. *With MRC/MRT, the mathematical term corresponding to inter-user interference, t_2, can be expressed in closed-form as*

$$
t_2^{\mathrm{mrc}} = \alpha_{\mathrm{m}}^2 (K-1)\left(\mathrm{Tr}^2(\mathbf{U}_1)\|\mathbf{U}_2\|^2 + \mathrm{Tr}^2(\mathbf{U}_2)\|\mathbf{U}_1\|^2 + K\|\mathbf{U}_1\|^2\|\mathbf{U}_2\|^2\right). \quad (5.31)
$$

Proof. It can be proved in a similar fashion as in Proposition 5.1. $\qquad\square$

Proposition 5.3. *With MRC/MRT, the mathematical term t_3 is given by*

$$
t_3^{\mathrm{mrc}} = K\alpha_{\mathrm{m}}^2 \mathrm{Tr}(\mathbf{U}_1 \mathbf{U}_{e_1})\left(\mathrm{Tr}^2(\mathbf{U}_2) + K\|\mathbf{U}_2\|^2\right). \qquad (5.32)
$$

Proof. By leveraging the property of Gaussian random matrices from Appendix C.1, we expand t_3^{mrc} as follows

$$
\begin{aligned}
t_3^{\mathrm{mrc}} &= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\left|\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \mathbf{E}_1 \mathbf{x}\right|^2\right] \\
&= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \mathbf{E}_1 \mathbf{E}_1^H \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_2^H \hat{\mathbf{g}}_{2,k}\right] \\
&\stackrel{\text{(L1)}}{=} K\alpha_{\mathrm{m}}^2 \mathbb{E}\left[\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \mathbf{U}_{e_1} \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_2^H \hat{\mathbf{g}}_{2,k}\right] \\
&\stackrel{\text{(L1)}}{=} K\alpha_{\mathrm{m}}^2 \mathrm{Tr}(\mathbf{U}_1 \mathbf{U}_{e_1}) \mathbb{E}\left[\hat{\mathbf{g}}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_2^H \hat{\mathbf{g}}_{2,k}\right] \\
&\stackrel{\text{(L3)}}{=} K\alpha_{\mathrm{m}}^2 \mathrm{Tr}(\mathbf{U}_1 \mathbf{U}_{e_1})\left(\mathrm{Tr}^2(\mathbf{U}_2) + K\|\mathbf{U}_2\|^2\right). \qquad (5.33)
\end{aligned}
$$

□

Proposition 5.4. *With MRC/MRT, the mathematical term t_4 can be expressed in closed-form as*

$$t_4^{\mathrm{mrc}} = K\alpha_{\mathrm{m}}^2 \mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right)\left(\mathrm{Tr}^2\left(\mathbf{U}_1\right) + K\left\|\mathbf{U}_1\right\|^2\right). \tag{5.34}$$

Proof. We trivially conclude the proof, by invoking Lemma 1 and Lemma 3

$$
\begin{aligned}
t_4^{\mathrm{mrc}} &= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\left|\mathbf{e}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1 \mathbf{x}\right|^2\right] \\
&= \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\mathbf{e}_{2,k}^H \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_2^H \mathbf{e}_{2,k}\right] \\
&\overset{(\mathrm{L1})}{=} \alpha_{\mathrm{m}}^2 \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{U}_{e_2} \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1 \hat{\mathbf{G}}_2^H\right)\right] \\
&\overset{(\mathrm{L3})}{=} \alpha_{\mathrm{m}}^2 \left(\mathrm{Tr}^2\left(\mathbf{U}_1\right) + K\left\|\mathbf{U}_1\right\|^2\right)\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{U}_{e_2} \hat{\mathbf{G}}_2 \hat{\mathbf{G}}_2^H\right)\right] \\
&\overset{(\mathrm{L1})}{=} \alpha_{\mathrm{m}}^2 K \mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right)\left(\mathrm{Tr}^2\left(\mathbf{U}_1\right) + K\left\|\mathbf{U}_1\right\|^2\right).
\end{aligned}
\tag{5.35}
$$

□

The similar methodology can be applied for the rest of terms t_5–t_7 to finally obtain the following results

$$t_5^{\mathrm{mrc}} = K^2 \alpha_{\mathrm{m}}^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)\mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right), \tag{5.36}$$

$$t_6^{\mathrm{mrc}} = \alpha_{\mathrm{m}}^2 \sigma_{n_R}^2 \mathrm{Tr}\left(\mathbf{U}_1\right)\left(\mathrm{Tr}^2\left(\mathbf{U}_2\right) + K\left\|\mathbf{U}_2\right\|^2\right), \tag{5.37}$$

$$t_7^{\mathrm{mrc}} = \alpha_{\mathrm{m}}^2 \sigma_{n_R}^2 K \mathrm{Tr}\left(\mathbf{U}_1\right)\mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right). \tag{5.38}$$

As was pointed out earlier, the power amplification factor α_{m} is enforced by the long-term power constraint at the AF relay station. Hence, starting from (5.5), recalling the relay transformation matrix in (5.2), and then proceeding with the same strategy that we used to simplify the SINR terms, this amplification gain can be calculated by

$$\alpha_{\mathrm{m}} = \left(\frac{P_r}{P_u K \mathrm{Tr}\left(\mathbf{U}_2\right)\left(\mathrm{Tr}^2\left(\mathbf{U}_1\right)+K\left\|\mathbf{U}_1\right\|^2\right)+P_u K^2 \mathrm{Tr}\left(\mathbf{U}_1\mathbf{U}_{e_1}\right)\mathrm{Tr}\left(\mathbf{U}_2\right)+\sigma_{n_R}^2 K \mathrm{Tr}\left(\mathbf{U}_1\right)\mathrm{Tr}\left(\mathbf{U}_2\right)} \right)^{\frac{1}{2}}.$$

(5.39)

5.4.1 ZF Digital Processor

As previously shown in (5.31), the MRC/MRT scheme suffers from relatively high in-terference compared to other components: estimation error and AWGN, especially at large K. Motivated by this observation, we employ a ZF scheme at the DSP unit, and then derive an approximation of the achievable rate. The ZF receiver can be mathemat-ically expressed based on (5.3) so that $\mathbf{F}_2^H \mathbf{W} \mathbf{F}_1 = \alpha_{\mathrm{z}} \hat{\mathbf{G}}_2 \left(\hat{\mathbf{G}}_2^H \hat{\mathbf{G}}_2\right)^{-1}\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H$, where α_{z} is the relay amplification factor in ZF DSP. For the ease of exposition, we use the superscript "zf" denote the expressions related with the ZF digital processor. Considering the ZF strategy we have

$$t_0^{\mathrm{zf}} = \alpha_{\mathrm{z}}^2, \qquad t_1^{\mathrm{zf}} = 0, \qquad t_2^{\mathrm{zf}} = 0,$$

(5.40)

$$t_3^{\mathrm{zf}} = \left[\mathbf{T}_3\right]_{k,k} = \alpha_{\mathrm{z}}^2 \left[\mathrm{Cov}\left(\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{E}_1 \mathbf{x}\right)\right]_{k,k},$$

(5.41)

$$t_4^{\mathrm{zf}} = \left[\mathbf{T}_4\right]_{k,k} = \alpha_{\mathrm{z}}^2 \left[\mathrm{Cov}\left(\mathbf{E}_2^H \hat{\mathbf{G}}_2 \left(\hat{\mathbf{G}}_2^H \hat{\mathbf{G}}_2\right)^{-1}\mathbf{x}\right)\right]_{k,k},$$

(5.42)

$$t_5^{\mathrm{zf}} = \left[\mathbf{T}_5\right]_{k,k} = \alpha_{\mathrm{z}}^2 \left[\mathrm{Cov}\left(\mathbf{E}_2^H \hat{\mathbf{G}}_2 \left(\hat{\mathbf{G}}_2^H \hat{\mathbf{G}}_2\right)^{-1}\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{E}_1 \mathbf{x}\right)\right]_{k,k},$$

(5.43)

$$t_6^{\mathrm{zf}} = \left[\mathbf{T}_6\right]_{k,k} = \alpha_{\mathrm{z}}^2 \left[\mathrm{Cov}\left(\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{n}_R\right)\right]_{k,k},$$

(5.44)

$$t_7^{\mathrm{zf}} = \left[\mathbf{T}_7\right]_{k,k} = \alpha_{\mathrm{z}}^2 \left[\mathrm{Cov}\left(\mathbf{E}_2^H \hat{\mathbf{G}}_2 \left(\hat{\mathbf{G}}_2^H \hat{\mathbf{G}}_2\right)^{-1}\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{n}_R\right)\right]_{k,k},$$

(5.45)

which can be approximated in the following manner.

Proposition 5.5. *With ZF, the mathematical expression t_3 can be approximated by*

$$t_3^{\mathrm{zf}} \xrightarrow{\text{a.s.}} \frac{K\alpha_{\mathrm{z}}^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)}{\mathrm{Tr}^2\left(\mathbf{U}_1\right)}.$$

(5.46)

Proof. The proof is straightforward by invoking Corollary 1 ($C1$)

$$\mathbf{T}_3 = \alpha_z^2 \mathbb{E}\left[\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{E}_1 \mathbf{x}\mathbf{x}^H \mathbf{E}_1^H \hat{\mathbf{G}}_1 \left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\right]$$

$$= K\alpha_z^2 \mathbb{E}\left[\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\hat{\mathbf{G}}_1^H \mathbf{U}_{e_1} \hat{\mathbf{G}}_1 \left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-1}\right]$$

$$\overset{(C1)}{\underset{\text{a.s.}}{\longrightarrow}} K\alpha_z^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)\mathbb{E}\left[\left(\hat{\mathbf{G}}_1^H \hat{\mathbf{G}}_1\right)^{-2}\right]$$

$$\overset{(C1)}{\underset{\text{a.s.}}{\longrightarrow}} \frac{K\alpha_z^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)}{\mathrm{Tr}^2\left(\mathbf{U}_1\right)}\mathbf{I}_K, \tag{5.47}$$

which concludes the proof. $\qquad\square$

The other terms can be obtained in a similar spirit, hence for the brevity, we omit their proofs and just point out to their final results:

$$t_4^{\mathrm{zf}} \overset{\text{a.s.}}{\longrightarrow} \frac{K\,\alpha_z^2 \mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right)}{\mathrm{Tr}^2\left(\mathbf{U}_2\right)}, \tag{5.48}$$

$$t_5^{\mathrm{zf}} \overset{\text{a.s.}}{\longrightarrow} \frac{K^2\alpha_z^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)\mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right)}{\mathrm{Tr}^2\left(\mathbf{U}_1\right)\mathrm{Tr}^2\left(\mathbf{U}_2\right)}, \tag{5.49}$$

$$t_6^{\mathrm{zf}} \overset{\text{a.s.}}{\longrightarrow} \frac{\alpha_z^2 \sigma_{n_R}^2}{\mathrm{Tr}\left(\mathbf{U}_1\right)}, \tag{5.50}$$

$$t_7^{\mathrm{zf}} \overset{\text{a.s.}}{\longrightarrow} \frac{K\alpha_z^2 \sigma_{n_R}^2 \mathrm{Tr}\left(\mathbf{U}_2 \mathbf{U}_{e_2}\right)}{\mathrm{Tr}\left(\mathbf{U}_1\right)\mathrm{Tr}^2\left(\mathbf{U}_2\right)}, \tag{5.51}$$

$$\alpha_z = \left(\frac{P_r \mathrm{Tr}\left(\mathbf{U}_1\right)\mathrm{Tr}\left(\mathbf{U}_2\right)}{P_u K \mathrm{Tr}\left(\mathbf{U}_1\right) + \frac{P_u K^2 \mathrm{Tr}\left(\mathbf{U}_1 \mathbf{U}_{e_1}\right)}{\mathrm{Tr}\left(\mathbf{U}_1\right)} + K\sigma_{n_R}^2}\right)^{\frac{1}{2}}. \tag{5.52}$$

5.5 Analog Beamformer Design

5.5.1 Analog Beamformer Design

Massive MIMO can significantly increase the achievable rate thanks to its ability to provide a large multiplexing gain. However, this ability is somewhat restricted in highly correlated channels or in networks with few number of active users, where either ill-conditioned channel matrices or the limited number of users restrict the huge degrees of freedom offered by massive MIMO. Therefore, in these cases, exploiting the multiplexing gain is not the main concern, and consequently the idea of the use of hybrid structure with reduced number of RF chains suits to these scenarios. However, a

limited number of RF chains restricts the control of the DSP unit on the antenna arrays to finely steer the beams and place the nulls in predefined directions especially when perfect CSI is not available at the relay station. Although power and multiplexing gain are restricted in hybrid structures, we will show that this topology still can deliver a reasonable and reliable achievable rate by extracting the best eigenmodes of the channel correlation matrix. Therefore, in this section we design an analog beamformer based on the statistics of the propagation channel. According to the signal model in (5.17) the analog beamformers, i.e. $\mathbf{F}_1$ and $\mathbf{F}_2$, are only involved in the channel estimates $\hat{\mathbf{G}}_1$ and $\hat{\mathbf{G}}_2$. Motivated by this observation, we design an analog beamformer in order to reduce the estimation errors, which in turn boosts the desired signal power, and finally improves the achievable rate. For the sake of simplicity, we confine our focus on designing the matrix $\mathbf{F}_1$, while the same results can be derived for the matrix $\mathbf{F}_2$. Borrowing the channel estimation error from (5.14), the total estimation error is given by

$$
\begin{aligned}
\varepsilon_1 &= \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{E}_1 \mathbf{E}_1^H\right)\right] \\
&= K\mathrm{Tr}\left(\mathbf{R}_1 - \tau_p P_p \mathbf{R}_1 \mathbf{F}_1^H \left(\tau_p P_p \mathbf{F}_1 \mathbf{R}_1 \mathbf{F}_1^H + \mathbf{I}_{K_a}\right)^{-1} \mathbf{F}_1 \mathbf{R}_1\right) \\
&= K\mathrm{Tr}\left(\left(\mathbf{R}_1^{-1} + \tau_p P_p \mathbf{F}_1^H \mathbf{F}_1\right)^{-1}\right) \\
&= K\mathrm{Tr}\left(\mathbf{U}_{e_1}\right).
\end{aligned}
\tag{5.53}
$$

Taking the practical constraints of analog beamformer into account, we can design the analog beamformer from the following optimization problem

$$
\begin{aligned}
\min_{\mathbf{F}_1} \quad & \varepsilon_1 \\
\text{s.t.} \quad & \mathrm{Tr}\left(\mathbf{F}_1 \mathbf{F}_1^H\right) \leq K_a, \\
& \mathbf{F}_1 \in \mathbb{C}^{K_a \times N}, \\
& \mathbf{F}_1 \in \mathcal{F},
\end{aligned}
\tag{5.54}
$$

where $\mathcal{F}$ denotes the set of $K_a \times N$ complex matrices with equal modulus. Since $\mathcal{F}$ is

not a convex set, we ignore this constraint at this stage, but we will show the impact of this assumption in simulation results. Under this assumption, we can rewrite the optimization problem as

$$\min_{\mathbf{F}_1} \ \mathrm{Tr}\left(\left(\mathbf{R}_1^{-1} + \tau_p P_p \mathbf{F}_1^H \mathbf{F}_1 \right)^{-1} \right)$$

$$\text{s.t. } \mathrm{Tr}\left(\mathbf{F}_1 \mathbf{F}_1^H \right) \leq K_a,$$

$$\mathbf{F}_1 \in \mathbb{C}^{K_a \times N}. \tag{5.55}$$

Let $\mathbf{F}_1 = \mathbf{U}_{F_1} \mathbf{\Sigma}_{F_1} \mathbf{V}_{F_1}^H$ and $\mathbf{R}_1 = \mathbf{U}_{R_1} \mathbf{\Lambda}_{R_1} \mathbf{U}_{R_1}^H$ denote the SVD and eigen value decomposition of matrices $\mathbf{F}_1$ and $\mathbf{R}_1$, respectively. The minimum value of the objective function in (5.55) can be achieved if the eigenvectors of $\mathbf{F}_1^H \mathbf{F}_1$ are chosen along with the eigenmodes of $\mathbf{R}_1^{-1}$, i.e. $\mathbf{V}_{F_1} = \mathbf{U}_{R_1}$. By invoking this property, the optimization problem on hand can be further simplified to

$$\min_{\mathbf{\Sigma}_{F_1}} \ \mathrm{Tr}\left(\left(\mathbf{\Lambda}_{R_1}^{-1} + \tau_p P_p \mathbf{\Sigma}_{F_1}^H \mathbf{\Sigma}_{F_1} \right)^{-1} \right) \tag{5.56}$$

$$\text{s.t. } \mathrm{Tr}\left(\mathbf{\Sigma}_{F_1}^H \mathbf{\Sigma}_{F_1} \right) \leq K_a,$$

$$\mathbf{\Sigma}_{F_1} \in \mathbb{C}^{K_a \times N}.$$

Now, let x_i and γ_i denote the i-th biggest eigenvalue of the matrices $\mathbf{F}_1^H \mathbf{F}_1$ and $\mathbf{R}_1^{-1}$, respectively. Then, this problem can be reduced to a conventional water-filling optimization as follows [100]

$$\min_{x_i} \ \sum_{i=1}^{N} \left(\frac{1}{\gamma_i + \tau_p P_p x_i} \right)$$

$$\text{s.t. } \sum_{i=1}^{N} x_i \leq K_a,$$

$$x_i \geq 0, \ \text{for } i = 1, \ldots, K_a,$$

$$x_i = 0, \ \text{for } i = K_a + 1, \ldots, N. \tag{5.57}$$

We note that this optimization problem is a convex optimization problem as it can be

rewritten in a format of a standard optimization problem

$$\min_{x_i} \ \sum_{i=1}^{N} \left(\frac{1}{\gamma_i + \tau_p P_p x_i} \right)$$

$$\text{s.t.} \ \sum_{i=1}^{N} x_i - K_a \leq 0,$$

$$- x_i \leq 0, \ \text{for } i = 1, \ldots, K_a,$$

$$x_i = 0, \ \text{for } i = K_a + 1, \ldots, N, \tag{5.58}$$

where all the constraint functions are affine, so this problem is subject to the convex constrains. Also, the objective function is the sum of convex functions which means the objective function is a convex function. By utilizing the Lagrangian duality, and considering Karush-Kuhn-Tucker (KKT) conditions for optimality, we obtain that the first K_a' bins should be filled up to a certain level as illustrated in Fig. 5.2 so that

$$\sqrt{\frac{\tau_p P_p}{\nu_1}} = \gamma_i + \tau_p P_p x_i, \tag{5.59}$$

where ν_1 is a constant number to satisfy the constraint $\sum_{i=1}^{N} x_i \leq K_a$ with equality. Then, after some simple mathematics manipulations we can find that

$$\sqrt{\frac{\tau_p P_p}{\nu_1}} = \frac{\tau_p P_p K_a + \sum_{i=1}^{K_a'} \gamma_i}{K_a'}, \tag{5.60}$$

$$x_i = \frac{1}{\tau_p P_p} \left(\frac{\tau_p P_p K_a + \sum_{i=1}^{K_a'} \gamma_i}{K_a'} - \gamma_i \right), \qquad \text{for } i = 1, \ldots, K_a', \tag{5.61}$$

$$x_i = 0, \qquad \text{for } i = K_a', \ldots, N. \tag{5.62}$$

It is noteworthy that the maximum rank of matrix $\mathbf{F}_1$ is K_a, hence the water filling algorithm is applied for the first K_a bins, whereas $K_a' \leq K_a$ bins can be filled by this algorithm as illustrated in Fig. 5.2. We also note that the design of matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ only depends on the long-term statistics of the channel according to the above discussion, therefore, we just need to design the analog beamformer once over each
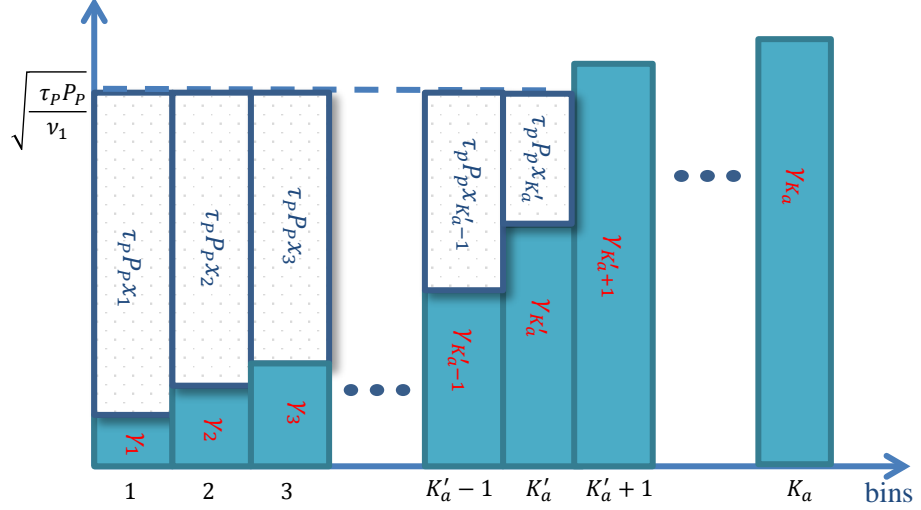
Figure 5.2: Water filling structure

coherence time T. The final design of the analog beamformer can be summarized as

$$\mathbf{F}_1 = \mathbf{U}_{F_1} \boldsymbol{\Sigma}_{F_1} \mathbf{U}_{R_1}^H, \tag{5.63}$$

where $\boldsymbol{\Sigma}_{F_1}$ is a $K_a \times N$ rectangular diagonal matrix with $\left[\sqrt{x_1}, \sqrt{x_2}, \ldots, \sqrt{x_{K_a'}}, 0, \ldots, 0\right]^T$ on its main diagonal. The matrix $\mathbf{U}_{F_1}$ does not play any role in the optimization problem (5.55), hence any unitary matrix can be chosen. For convenience in our subsequent developments, we just assume $\mathbf{U}_{F_1} = \mathbf{I}_N$ which leads to[1]

$$\mathbf{F}_1 = \boldsymbol{\Sigma}_{F_1} \mathbf{U}_{R_1}^H. \tag{5.64}$$

5.5.2 Discussion

Altogether, we were able find a closed-form lower bound for the achievable rate for the MRC/MRT scheme, and also an approximation of the achievable rate for the ZF scheme which depend on the long-term components of the propagation channels as reflected by $\mathrm{Tr}(\mathbf{U}_i)$, $\mathrm{Tr}(\mathbf{U}_i \mathbf{U}_{e_i})$, and $\|\mathbf{U}_i\|$ for $i = 1, 2$. Proposition 5.6 will reduce

---

[1] Precisely speaking, $\mathbf{U}_{F_1}$ can affect the original optimization problem (5.54) where we need to design $\mathbf{U}_{F_1}$ so that the matrix $\mathbf{F}_1$ falls within the set $\mathcal{F}$. It can also contribute to the system robustness, however, Monte Carlo simulations will confirm that the results are still good enough by considering $\mathbf{U}_{F_1} = \mathbf{I}_N$.

the number of these components by finding a direct connection between these factors.

Proposition 5.6. *It can be shown that* $\mathrm{Tr}\big(\mathbf{U}_i\big)$ *and* $\mathrm{Tr}\big(\mathbf{U}_i\mathbf{U}_{e_i}\big)$ *for* $i = 1, 2$ *are directly connected to each other by the level of water-filling ceiling (See also Fig. 5.2)*

$$\mathrm{Tr}\big(\mathbf{U}_i\big) = \sqrt{\frac{\tau_p P_p}{\nu_i}}\mathrm{Tr}\big(\mathbf{U}_i\mathbf{U}_{e_i}\big), \qquad \text{for } i = 1, 2. \tag{5.65}$$

Proof. See Appendix C.2. $\square$

It is notable that the impact of $\mathrm{Tr}\big(\mathbf{U}_i\big)$ for $i = 1, 2$ is more pronounced at the numerator of the SINR in (5.19), and hence, it can be treated as a desirable value that boosts the achievable rate. In contrast, $\mathrm{Tr}\big(\mathbf{U}_i\mathbf{U}_{e_i}\big)$, which can be interpreted as the inner product between the error correlation-matrix $\mathbf{U}_{e_i}$ and estimation correlation-matrix $\mathbf{U}_i$, deteriorates the achievable rate. Hence, Proposition 5.6 is seemingly useful to both mathematically and intuitively explain the relation between these two factors. This proposition numerically explains that for a given MMSE channel estimate, the impact of error estimation can be harnessed by a scalar, i.e. $\sqrt{\frac{\tau_p P_p}{\nu_i}}$ for $i = 1, 2$, which basically depends on the number of RF chains, pilot power, and the duration of pilot sequences. This is, of course, in line with the canonical concept of the pilot-based channel estimation originally observed in [101, 102].

To obtain a clear understanding how the different parameters affect the achievable rate in our analog beamformer design followed by the ZF DSP, we consider a special case with the same channel correlation matrix and the same amount of RF chains at both sides of the relay station $\mathbf{R}_1 = \mathbf{R}_2$ and $K_a = K_b$. Also, we note that t_5^{zf}, t_7^{zf} and a part of α_{z} are negligible compared to other terms.[1] Hence, we can derive a new simplified approximation for the achievable rate namely $\bar{R}^{\mathrm{zf}}$

$$\bar{R}^{\mathrm{zf}} = \frac{K\left(T - \tau_p\right)}{2T} \log_2\left(1 + \frac{\mathrm{Tr}\left(\mathbf{U}\right)}{2K\zeta + \frac{\sigma_{n_R}^2}{P_u} + \frac{\sigma_{n_D}^2}{\frac{P_r}{K}}}\right), \tag{5.66}$$

where we define $\mathbf{U} \triangleq \mathbf{U}_1 = \mathbf{U}_2$, $\zeta \triangleq \sqrt{\frac{\nu}{\tau_p P_p}}$, and $\nu \triangleq \nu_1 = \nu_2$ for notational

---

[1]By invoking Proposition 5.6 in (5.52), we can readily conclude that the last two terms of the denominator of α_{z} are much smaller than its first term.

simplicity. This closed-form approximation can shed further light on how the system parameters contribute to the achievable rate. Precisely speaking, this result implies that exploiting the long-term statistics of the channel, leads to a diversity gain $\mathrm{Tr}\left(\mathbf{U}\right)$ that linearly scales the SINR. On the other hand, only the inverse of the uplink SNR $\frac{\sigma_{n_R}^2}{P_u}$, and downlink SNR per user $\frac{\sigma_{n_D}^2}{\frac{P_r}{K}}$ participate as an AWGN in the achievable rate. In addition, the impact of channel estimation errors simply appears as $2K\sqrt{\frac{\nu}{\tau_p P_p}}$ which can be controlled by the pilot specifications and the level of water-filling.

All in all, we developed a correlation-based analog beamformer followed by simple linear digital processors, i.e. MRC/MRT and ZF scheme. Then, we provided mathematical closed-form expressions which only involve the long-term characteristics of propagation channels. These results reveal that even in the worst-case scenario that multiplexing gain is restricted due to the limited number of RF chains, and array gain is also confined due to the less control of the DSP unit on the analog beamformers, the proposed system can still leverage a substantial diversity gain. This diversity gain is related to the statistical properties of the channels, where analog beamformers exploit the strongest eigenmodes of the channel correlation matrix. This analysis also shows that the dominant statistical terms, i.e. $\mathrm{Tr}\left(\mathbf{U}_i\right)$ for $i = 1, 2$ have a tendency to boost the desired signal. On the other hand, the moderate statistical terms like $\left\|\mathbf{U}_i\right\|$ for $i = 1, 2$ mainly contribute to the aggregated noise and interference.

5.6 Numerical Results

In this section, we provide numerical results to evaluate the performance of hybrid multipair massive relaying with the channel estimation at the relay station. We model the two deterministic correlation matrices, $\mathbf{R}_1$ and $\mathbf{R}_2$, as follows:

$$\left[\mathbf{R}_1\right]_{m,n} = e^{-j2\pi(n-m)\Delta_r \cos(\theta_r)} e^{-\frac{1}{2}(2\pi(n-m)\Delta_r \sin(\theta_r)\sigma_r)^2}, \tag{5.67}$$

$$\left[\mathbf{R}_2\right]_{m,n} = e^{-j2\pi(m-n)\Delta_t \cos(\theta_t)} e^{-\frac{1}{2}(2\pi(m-n)\Delta_t \sin(\theta_t)\sigma_t)^2}, \tag{5.68}$$

Parameter	Value	Parameter	Value
Number of antennas	$N = 128$	Coherence time	$T = 196$ symbols
Number of RF chains	$K_a = K_b = 50$	Pilot length	$\tau_p = 2K$ symbols
Number of users	$K = 10$	Mean AoA, AoD	$\theta_r = \theta_t = 0.4\pi$ radian
Relay power	$P_r = 10^{0.5}$ W	Angle spread	$\sigma_r = \sigma_t = 0.25$ radian
User's power	$P_u = 1$ W	Noise variance	$\sigma_{n_R}^2 = 1$
Pilot power	$P_p = 1$ W	Noise variance	$\sigma_{n_D}^2 = 1$

Table 5.1: Simulation parameters of system model with channel correlation at the relay side.

where Δ_r, Δ_t denote the antenna spacing, θ_r, θ_t, σ_r^2, and σ_r^2 represent the mean AoA to the relay station, mean AoD from the relay station, receive angle spread and transmit angle spread, respectively [64, 68]. These correlation models basically represent Gaussian matrices with a spread inversely proportional to the product of the antenna spacing and angle spread. We note that a smaller angel spreads signify higher levels of spatial correlation. It is also noteworthy that this model can be easily expanded to the clustered channels where each cluster corresponds to a specific AoA and AoD, and there are sufficient scatterers in each cluster: approximately 10 or more [103]. For the sake of simplicity, we assume that the channel matrices are normalized so that $\sigma_{n_R}^2$ and $\sigma_{n_D}^2$ contain both noise variance and pathloss. Throughout the simulation, unless otherwise specified, we utilize the detailed parameter settings as summarized in Table 5.1.

All analytical results obtained in this chapter assume that phase shifters are designed in the digital domain, and they are able to take any modulus and phase. However, as we described earlier in (5.55), this so-called unconstrained analog beamforming (UAB) design, is not quite practical due to the fact that analog beamformers are implemented by means of phase shifters with a constant modulus constraint. To circumvent this problem, we first design the phase shifters as before, and then will normalize their modulus to satisfy the aforementioned constraint. For the sake of presentation clarity, we call this methodology as constrained analog beamforming (CAB) design which satisfies (5.54).

Figure 5.3 shows a great congruency between Monte Carlo simulation and our analytical result. Moreover, it showcases that the gap between UAB and CAB design is very small. This observation implies that our methodology is quite robust with respect
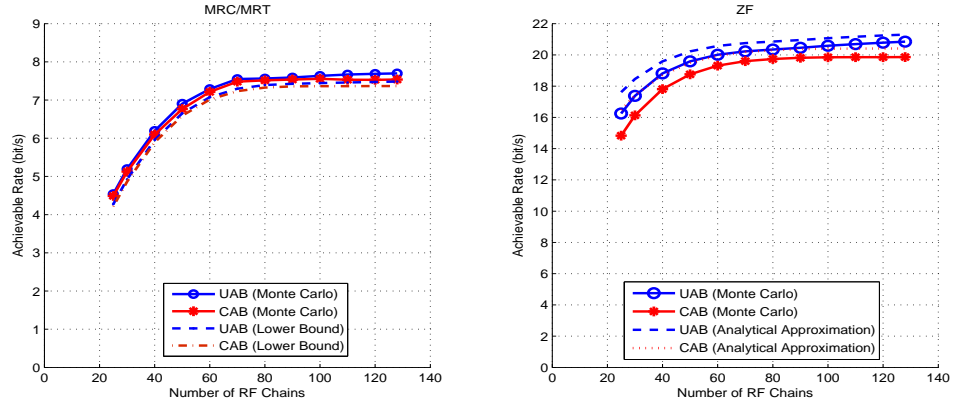
Figure 5.3: Performance of the proposed hybrid beamformer under channel correlation.
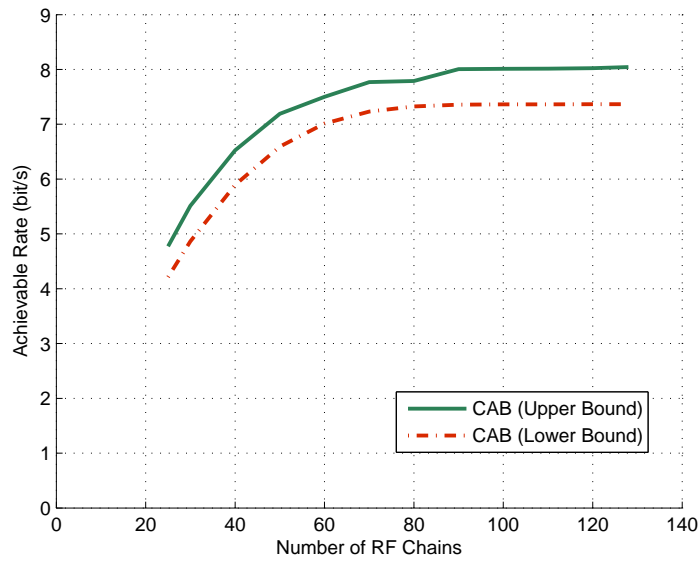
Figure 5.4: Lower bound and upper bound on the sum achievable rate in hybrid structure equipped with MRC/MRT DSP unit under channel correlation.

to the phase shifters constraint. In addition, this figure illustrates that the ZF processing can greatly enhance the achievable rate compared to the MRC/MRT processing by nulling out the inter-user interference. It signifies that by deploying only 50 RF chains the hybrid configuration paradigm can capture more than 90% of the achievable rate offered by the fully digital structure with 128 RF chains.

In order to understand how tight is our lower bound, Fig. 5.4 compares our suggested bound with an idealistic case namely *upper bound*. In this case, we assume that the destinations know perfectly the channels estimates $\hat{\mathbf{G}}_1$, $\hat{\mathbf{G}}_2$. Strictly speaking, under this assumption, the received baseband signal in (5.17) can be substituted by the

Figure 5.5: Performance of the proposed analog beamformer with different levels of correlation. Left: $N = 256$, Right: $N = 128$.

more accurate model given as follows

$$
\begin{aligned}
y_{D_k} = &\underbrace{\sqrt{P_u}\alpha_{\mathrm{m}}\hat{\mathbf{g}}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H\hat{\mathbf{g}}_{1,k}x_k}_{\text{desired signal}} \\
&+\underbrace{\sqrt{P_u}\alpha_{\mathrm{m}}\sum_{i\neq k}^{K}\hat{\mathbf{g}}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H\hat{\mathbf{g}}_{1,i}x_i}_{\text{inter}-\text{user interference}}+\underbrace{\sqrt{P_u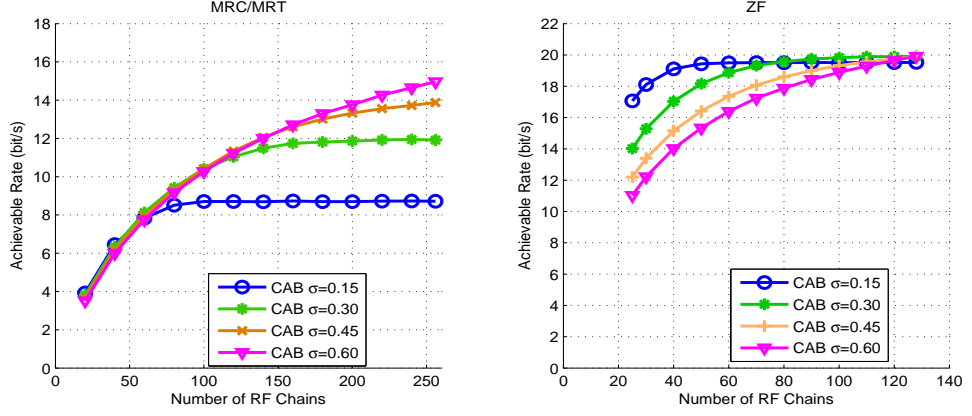}\alpha_{\mathrm{m}}\hat{\mathbf{g}}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H\mathbf{E}_1x_k}_{\text{estimation error}}+\underbrace{\sqrt{P_u}\alpha_{\mathrm{m}}\mathbf{e}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H\hat{\mathbf{G}}_1x_k}_{\text{estimation error}} \\
&+\underbrace{\sqrt{P_u}\alpha_{\mathrm{m}}\mathbf{e}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H\mathbf{E}_1x_k}_{\text{estimation error}}+\underbrace{\alpha_{\mathrm{m}}\hat{\mathbf{g}}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H2+\alpha_{\mathrm{m}}\mathbf{e}_{2,k}^H\hat{\mathbf{G}}_2\hat{\mathbf{G}}_1^H2+n_{D_k}}_{\text{aggregated noise \& estimation error}}, \qquad (5.69)
\end{aligned}
$$

where the fluctuations of desired signal is not a matter anymore. Figure 5.4 showcases a very small gap between the lower and upper bounds which concludes that the analytical lower bound approximates very well the actual achievable achievable rate, particularly, for a limited number of RF chains.

Fig. 5.5 illustrates the performance of hybrid topology with different levels of channel spatial correlation. It implies that with a limited number of RF chains, the proposed analog beamformer followed by ZF scheme can highly avail of the channel correlation. It is due to this fact that, in highly correlated channels, e.g., smaller σ, the eigenvalues of the correlation matrix are decentralized and will be spread on a wider range (See Fig. 5.6). Therefore, there exist few strong modes (paths) that can be captured by the limited available number of RF chains. On the other hand, MRC/MRT scheme is unable to decorrelate and extract the strong eigenmodes. Thus,
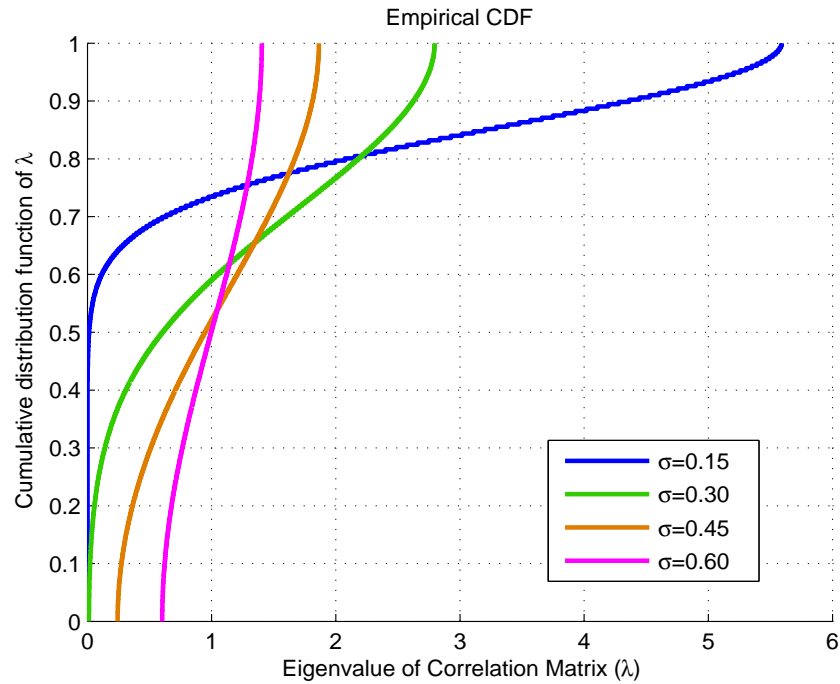
Figure 5.6: Empirical cumulative distribution function.

the MRC/MRT scheme showcases a poor performance with a limited number of RF chains.

Finally, Fig. 5.7 evaluates the impact of deployed antennas at the relay station, where we increase the number of antennas while the ratio N/K_a is constant. In fact, the more antennas we deploy at the relay station, the more diversity gain we can achieve. Therefore, the achievable rate will increase, although the ratio of the service antennas and RF chains is still constant. More interestingly, Fig. 5.8 demonstrates that the system performance improves by deploying more antennas, even though the number of RF chains is fixed $K_a = K_b = 50$. This is due to the fact that by deploying more antennas, the number of eigenvalues for correlation matrices ($\mathbf{R}_1$, $\mathbf{R}_2$) will increase, and the proposed analog beamformers smartly collect only the strongest ones. In other words, the water-filling algorithm enhances the system performance by exploiting the diversity gain from the long-term characteristics of propagation channels.
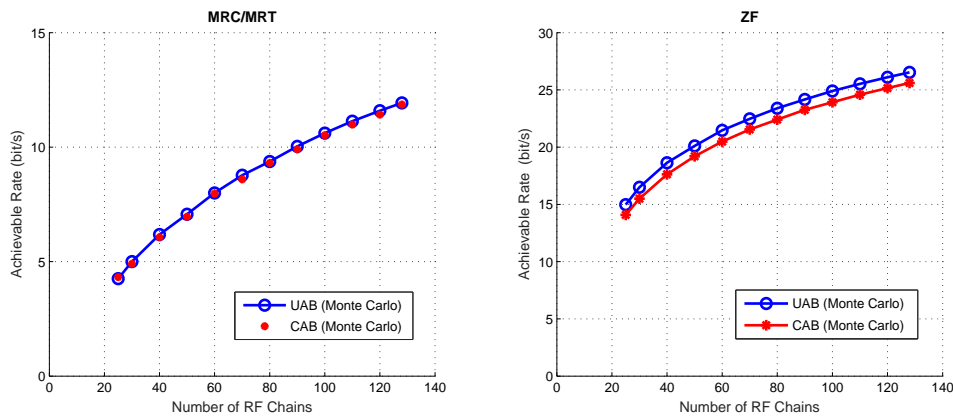
Figure 5.7: The impact of the number of relay antennas with $\frac{N}{K_a} = \frac{N}{K_b} = 3$.
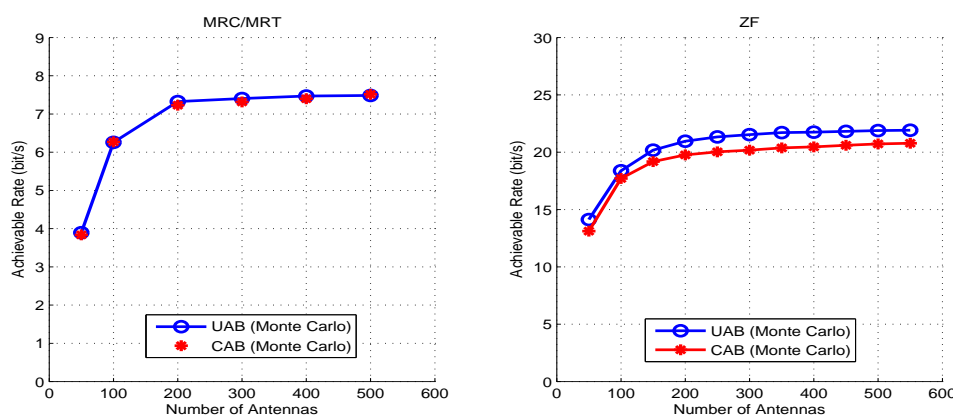
Figure 5.8: The impact of the number of relay antennas with constant RF chains $K_a = K_b = 50$.

5.7 Conclusion

Massive MIMO is a promising technique for the next generation of wireless communication systems which addresses most of the critical challenges associated with concurrent relaying systems, such as DSP complexity, long processing delay and low SINR at the cell edges. However, massive relaying experiences a high fabrication/implementation cost and power consumption due to the large number of RF chains. To avoid this, we greatly reduced the number of RF chains in a viable analog/digital configuration which is usually referred to hybrid structure. It is well-known that this structure reduces the multiplexing gain and also restricts the power gain due to the less flexibility of analog beamformers, particularly, with imperfect CSI at the relay station. Thus, we assumed a correlated channel and then designed a novel analog beamformer which exploits the long-term channel statistics and results in a high achievable rate. We also derived an

approximation and lower-bound on the achievable rate which involves the long-term parameters of the propagation channels.

84

Chapter 6

Performance Limits of MIMO Systems with Nonlinear Power Amplifiers

The development of 5G enabling technologies brings new challenges to the design of PAs. In particular, there is a strong demand for low-cost, nonlinear PAs which, however, introduce nonlinear distortions. On the other hand, contemporary expensive PAs show great power efficiency in their nonlinear region. Inspired by this trade-off between nonlinearity distortions and efficiency, finding an optimal operating point is highly desirable. Hence, it is first necessary to fully understand how and how much the performance of MIMO systems deteriorates with PA nonlinearities. In this chapter, we first reduce the ergodic achievable rate (EAR) optimization problem from a power allocation to a power control problem with only one optimization variable, i.e. total input power. Then, we develop a closed-form expression for the EAR, where this variable is fixed. Since this expression is complicated for further analysis, two simple lower bounds and one upper bound are proposed. These bounds enable us to find the best input power and approach the channel capacity. Finally, our simulation results evaluate the EAR of MIMO channels in the presence of nonlinearities. An important observation is that the MIMO performance can be significantly degraded if we utilize the whole power budget.

6.1   Introduction

MIMO wireless communication systems have been well investigated over the last two decades thanks to their ability to enhance the spectral efficiency and reliability [15], [14]. It is also well known that in a MIMO system, power is mainly consumed by the last parts of the transmitter chain, and in particular, by the PAs. Most contemporary MIMO systems deploy expensive, linear PAs although these components are intimately limited in terms of power efficiency. Yet, with the current strive towards network densification (i.e. the massive MIMO paradigm [80]), future systems are anticipated to deploy inexpensive, nonlinear PAs of high power efficiency.

In this context, only a few publications have studied the impact of nonlinear PAs on the MIMO capacity from a communication theory prospective. For instance, [104] and [105] considered a power consumption model of the PAs, and then introduced a low complexity algorithm to maximize the sum rate of multiple-input single-output systems. They also extended their beamforming method to the case of parallel MIMO by utilizing a dynamic programming language algorithm. Although the presented model is relatively precise, it does not go beyond the linear region where we can enjoy high power efficiency. In [49], the authors considered the impact of PA non-linearities on channel estimation and then proposed a quantized method to optimize the bit-error-rate and mutual information. Moreover, they proposed a constellation-based compensation method for high-power amplifier nonlinearities in [50]. There are also some other research efforts like [51–56], which have dealt with PA nonlinearities in different wireless applications, but they either do not focus on MIMO systems or do not evaluate the achievable rate.

Motivated by the above discussion, this work focuses on studying the performance of a MIMO system deploying nonlinear PAs. In particular, we start our analysis by optimizing the EAR over the input power allocation matrix. Then, we simplify this power allocation problem to a power control problem where the only variable is the total consumed power. Finally, we propose lower and upper bounds on the EAR, which can be easily optimized by conventional optimization methods. These analytical

results followed by simulations showcase that using the full power budget will reduce the EAR to zero, since PAs nonlinearities become dominant.

6.2   Signal and System Models

The traditional model of flat-fading point-to-point MIMO channels with N_t transmit antennas and N_r receive antennas is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}, \tag{6.1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_{N_t}]^T \in \mathbb{C}^{N_t \times 1}$ represents the complex Gaussian distributed transmitted signal with zero mean and covariance matrix $\mathbf{K_s} = \mathbb{E}[\mathbf{s}\mathbf{s}^H]$. The received signal is denoted by $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$, while the N_r-dimensional vector $\mathbf{w}$ models the additive circularly symmetric complex Gaussian noise $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I}_{N_r})$. Throughout this chapter, we assume that the propagation channel coefficients are independently circular symmetric complex Gaussian variable with unit variance. Additionally, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is assumed to be known to the receiver, but not at the transmitter; however, its statistical characteristics are available at the transmitter. Note that the channel matrix $\mathbf{H}$ is normalized so that N_0 contains both noise variance and pathloss.

Unfortunately, the above canonical model falls short of describing the nonlinear behavior of PAs and its impact on the end-to-end performance. To this end, in the following we extend the canonical model to take the PA nonlinearities into account. Note that our analysis remains agnostic to any type of nonlinearity that may be induced by mixers, filters and DACs.

In order to incorporate the impact of transceiver impairments, we first need to explain the input/output relation of the PA. In general, for a complex baseband input signal represented as $u_{in}(t) = A(t)e^{j\phi(t)}$, the signal at the output of a PA with amplitude gain $g_A(A(t))$ and phase gain $g_\phi(A(t))$ is [106]

$$u_{out}(t) = g(u_{in}(t)) = g_A(A(t))e^{j(\phi(t)+jg_\phi(A(t)))}. \tag{6.2}$$

There are various types of PA models for the amplitude and phase gains, such as ideal clipping, traveling wave tube, and solid-state amplifier model [106]. Fortunately, all the models can be encompassed under the umbrella of a polynomial PA model. For the sake of simplicity, we hereafter assume that all PAs have the same nonlinear conversion functions which are known at the transceivers [49], [50]. In general, a polynomial PA model can be easily determined by the following curve fitting of degree N:

$$g_A(A_i) = \sum_{n=0}^{N-1} \beta_{n+1} A_i^{n+1},$$ (6.3)

where A_i is the voltage of the input signal in the i-th PA, $i = 1, 2, \ldots, N_t$. Herein, we also consider memoryless PAs and ignore the phase distortion [55].[1] Thus, the coefficients, β_{n+1} for $n = 0, 1, \ldots, N-1$, are real constant numbers. By this preamble and regarding the Bussgang's theorem [107], the i-th PA output can be expressed in the form of

$$s_i = \alpha_i x_i + d_i \qquad i = 1, 2, \ldots, N_t,$$ (6.4)

where d_i represents the distortion noise which is uncorrelated with the input signal, x_i. Note that, d_i is a zero-mean (not necessarily Gaussian) distribution with power density $\sigma_{d_i}^2$. Since the input signal has a complex Gaussian distribution, its magnitude (A_i) follows a Rayleigh probability density function, so that

$$P(A_i) = \frac{A_i}{\sigma_i^2} \exp\left(-\frac{A_i^2}{2\sigma_i^2}\right),$$ (6.5)

where $\mathbb{E}[x_i^* x_i] = 2\sigma_i^2$. Furthermore, in (6.4), s_i stands for the PA output which will be emitted from the transmit antennas, and α_i is a constant affected by the PA gain function and its input power. In general, it can be shown that [56]

---

[1]This assumption is widespread in the literature especially for solid state PAs. In other words, memory leads to delay and consequently phase distortion. Therefore, we can ignore the phase distortion in these PAs as they are memoryless.

$$\alpha_i = \frac{\mathbb{E}\left[x_i{}^* s_i\right]}{\mathbb{E}\left[x_i{}^* x_i\right]}, \tag{6.6}$$

$$\sigma_{d_i}^2 = \mathbb{E}\left[s_i{}^* s_i\right] - \alpha_i^2 \mathbb{E}\left[x_i{}^* x_i\right], \tag{6.7}$$

where we use the fact that $\mathbb{E}\left[x_i^* d_i\right] = 0$. Subsequently, these parameters can be easily expressed as

$$\alpha_i = \frac{1}{2\sigma_i^2} \int_0^\infty A_i g_A(A_i) P(A_i) dA_i, \tag{6.8}$$

$$\sigma_{d_i}^2 = \int_0^\infty g_A^2(A_i) P\left(A_i\right) dA_i - \frac{1}{2\sigma_i^2} \left(\int_0^\infty A_i g_A\left(A_i\right) P\left(A_i\right) dA_i\right)^2. \tag{6.9}$$

Accordingly, it can be shown that the Bussgang's parameters for the polynomial model, can be obtained as [56]

$$\alpha_i = \sum_{n=0}^{N-1} \beta_{n+1} 2^{n/2} \sigma_i^n \Gamma\left(2 + \frac{n}{2}\right), \tag{6.10}$$

$$\sigma_{d_i}^2 = \sum_{n=2}^{2N} \left(\gamma_n 2^{n/2} \sigma_i^n \Gamma\left(1 + \frac{n}{2}\right)\right) - 2\sigma_i^2 \left(\sum_{n=0}^{N-1} \beta_{n+1} 2^{n/2} \sigma_i^n \Gamma\left(2 + \frac{n}{2}\right)\right)^2, \tag{6.11}$$

where $\Gamma(.)$ is the Gamma function [108, Eq. (8.310.1)], and γ_n can be defined as follows

$$\gamma_n \triangleq \sum_{k=1}^{n-1} \breve{\beta}_k \breve{\beta}_{n-k}^*, \tag{6.12}$$

and

$$\breve{\beta}_k \triangleq \begin{cases} \beta_k, & 1 \leq k \leq N \\ 0, & \text{otherwise.} \end{cases} \tag{6.13}$$

Now, the impact of PA nonlinearities can be incorporated into canonical system model based on the Bussgang's theorem (6.4):

$$\mathbf{y} = \mathbf{H} \left(\mathbf{\Lambda} \mathbf{x} + \mathbf{d} \right) + \mathbf{w}, \tag{6.14}$$

where $\mathbf{\Lambda} = \text{diag}\left\{ \alpha_1, \alpha_2, \ldots, \alpha_{N_t} \right\}$ and $\mathbf{d} = \left[d_1, d_2, \ldots, d_{N_t} \right]^T$. This equation can be reorganized as

$$\mathbf{y} = \mathbf{H}\mathbf{\Lambda}\mathbf{x} + \underbrace{\left(\mathbf{H}\mathbf{d} + \mathbf{w} \right)}_{\mathbf{n}}, \tag{6.15}$$

in which the vector $\mathbf{n}$ denotes the aggregated noise at the receiver with covariance matrix

$$= \mathbf{R}_{\mathbf{n}} = \mathbb{E}\left[\mathbf{n}\mathbf{n}^H \right] = \mathbf{H}\mathbf{D}\mathbf{H}^H + N_0 \mathbf{I}_{N_r}, \tag{6.16}$$

where we define $\mathbf{D} \triangleq \mathbb{E}\left[\mathbf{d}\mathbf{d}^H \right]$.

Remark 6.1. *We note that the aggregated noise at the receiver $\mathbf{n}$ is not necessarily a Gaussian random vector due to the non-Gaussian distribution of distortion vector $\mathbf{d}$. However, for the sake of analysis simplicity, and also considering the worst case scenario, we assume that the vector $\mathbf{n}$ follows a Gaussian distribution such that $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\mathbf{n}})$.*

6.3   Ergodic Achievable Rate Analysis

In this section, we analyze the EAR in the presence of PA nonlinearities by simplifying the power allocation problem to a power control problem.

6.3.1   Ergodic Achievable Rate

Based on the channel impairment model in Section 6.2, we are now ready to determine the MIMO achievable rate under Gaussian signaling for an arbitrary number of antennas. When Gaussian symbols are transmitted over the MIMO channel, an achievable

rate (in bits/s) is given by

$$R = \sup_{\mathrm{Tr}(\mathbf{Q}) \leq P_t, \mathbf{Q} \succeq \mathbf{0}} \mathbb{E}\left[B \log_2 \left(\det \left(\mathbf{I}_{N_r} + \mathbf{R_n}^{-1} \mathbf{H} \mathbf{\Lambda} \mathbf{Q} \mathbf{\Lambda}^H \mathbf{H}^H \right) \right) \right], \qquad (6.17)$$

where B denotes the bandwidth. For the sake of clarity, we will drop B from our subsequent analytical results, but in the numerical results section, we do include the impact of bandwidth. Moreover, $\mathbf{Q} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H \right]$ and $\mathrm{Tr}(\mathbf{Q}) \leq P_t$ indicates that the transmitter is constrained to its total power. For the purpose of simplification we define $\mathrm{Tr}(\mathbf{Q}) = \sum_{i=1}^{N_t} 2\sigma_i^2 \triangleq P$, and also define the instantaneous MIMO channel correlation matrix as

$$\mathbf{W} \triangleq \begin{cases} \mathbf{H}\mathbf{H}^H, & N_r \leq N_t \\ \mathbf{H}^H\mathbf{H}, & N_r > N_t, \end{cases} \qquad (6.18)$$

since its eigenvalues, λ_i, will be often used in our calculations.

Remark 6.2. *In order to reach the ergodic capacity, we need to assume a propagation channel with the maximum uncertainty [102]. For channel with known finite energy, the i.i.d Gaussian channel provides the maximum entropy. On the other hand, capacity can be achieved by jointly Gaussian input signals [15]. Although we consider the inputs to the PAs to be jointly Gaussian, the outputs of the PAs are not strictly jointly Gaussian. In light of this fact, in the remainder of this chapter, we will be referring to (6.17) as the maximum EAR.*

Proposition 6.3. *The EAR is a concave function in its domain with respect to the covariance matrix $\mathbf{Q}$.*

Proof. Note that $\log\left(\det\left(\cdot \right) \right)$ is a concave function in the cones of positive semi-definite matrices [109]. Also $\mathbf{R}_n^{-1}$, $\mathbf{\Lambda}$, and $\mathbf{Q}$ are all positive semi-definite matrices. Thus, the EAR is a concave function in its domain. $\qquad \square$

Proposition 6.4. *The maximum EAR is achieved when $\mathbf{Q}$ is a scaled identity matrix, i.e. $\mathbf{Q} = \frac{P}{N_t}\mathbf{I}_{N_t}$.*

Proof. See Appendix D.1. □

It is noteworthy that the PAs nonlinearities are affected by two factors: (i) PA transition function $g_A(\cdot)$, which is assumed to be the same for all the PAs; and (ii) the input power of each PA, i.e. $2\sigma_i^2$. As a consequence, Proposition 6.4 simplifies extensively our analysis as it allocates equal powers to each PA. In other words, we can conclude that $\sigma_1 = \sigma_2 = \ldots = \sigma_{N_t} \triangleq \sigma$, then $\alpha_1 = \alpha_2 = \ldots = \alpha_{N_t} \triangleq \alpha$, and $\sigma_{d_1} = \sigma_{d_2} = \ldots = \sigma_{d_{N_t}} \triangleq \sigma_d$. Therefore, the power allocation optimization upon the covariance matrix will be reduced to a power control over the power, i.e. P. Correspondingly, the noise covariance and EAR are respectively simplified as follows

$$\mathbf{R_n} = \sigma_d^2 \mathbf{HH}^H + N_0 \mathbf{I}_{N_r}, \tag{6.19}$$

$$R = \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2 \left(\det \left(\mathbf{I}_{N_r} + \frac{P\alpha^2}{N_t}(N_0 \mathbf{I}_{N_r} + \sigma_d^2 \mathbf{HH}^H)^{-1} \mathbf{HH}^H \right) \right) \right]. \tag{6.20}$$

Note that, in contrast to Telatar's methodology, our objective function in (6.20) is not necessarily a strictly ascending or descending function of the power P. This is due to the presence of σ_d^2 in (6.20) which also increases with P.

Corollary 6.5. *The maximum EAR can be rewritten as*

$$R = \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2 \det \left(\mathbf{I}_{N_t} + \mathbf{Z} \right) \right], \tag{6.21}$$

where $\mathbf{Z}$ is a diagonal $N_t \times N_t$ square matrix whose entries are $\zeta_{i,i} = \left(\frac{P\alpha^2 \lambda_i}{N_t \sigma_d^2 \lambda_i + N_t N_0} \right)$.

Proof. See Appendix D.2. □

Proposition 6.6. *Assuming i.i.d. Rayleigh fading channels, the maximum EAR under the proposed PAs nonlinearities model is analytically given by*

$$R = \sup_{0 \leq P \leq P_t} \frac{r}{\ln 2} K \sum_{m=1}^{r} \sum_{n=1}^{r} (-1)^{n+m} \det\left(\boldsymbol{\Omega}\right) \Gamma\left(t+1\right)$$
$$\sum_{k=1}^{t+1} \left(e^{1/f} E_{t+2-k}\left(\frac{1}{f}\right) - e^{1/g} E_{t+2-k}\left(\frac{1}{g}\right) \right), \tag{6.22}$$

where we define $q \triangleq \max\{N_t, N_r\}$, $r \triangleq \min\{N_t, N_r\}$, $t \triangleq n + m + q - r - 2$, and $K \triangleq \left(\prod_{i=1}^{r}(q-i)! \prod_{j=1}^{r}(r-j)!\right)^{-1}$ is a constant. Moreover, $f \triangleq \frac{P\alpha^2 + N_t\sigma_d^2}{N_0 N_t}$, $g \triangleq \frac{\sigma_d^2}{N_0}$, and also $E_n(x) = x^{n-1}\Gamma(1-n, x)$. Here, $\Gamma(s, x) = \int_x^\infty t^{s-1}e^{-t}dt$ is incomplete Gamma function [108, Eq. (8.350.2)]. Finally, $\boldsymbol{\Omega}$ is an $(r-1) \times (r-1)$ matrix whose (i, j)-th entry is given by

$$\boldsymbol{\Omega} = (\phi_{ij}^{(n)(m)} + q - r)! r^{-\frac{1}{r-1}}, \tag{6.23}$$

for which,

$$\phi_{ij}^{(n)(m)} \triangleq \begin{cases} i + j - 2, & \text{if } i \leq n, \text{ and } j \leq m \\ i + j, & \text{if } i \geq n, \text{ and } j \geq m \\ i + j - 1, & \text{otherwise.} \end{cases} \tag{6.24}$$

Proof. Assuming i.i.d. Rayleigh fading channels, $\mathbf{W}$ is full-rank with probability one [70]. Recalling Corollary 6.5, the maximum EAR can be expressed as

$$\begin{aligned} R &= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2 \prod_{i=1}^{\text{rank}(\mathbf{Z})} \left(1 + \zeta_{i,i}\right) \right] \\ &= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\sum_{i=1}^{r} \log_2 \left(1 + \left(\frac{P\alpha^2\lambda_i}{N_t\sigma_d^2\lambda_i + N_t N_0}\right)\right) \right] \\ &= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[r \log_2 \left(1 + \frac{P\alpha^2\lambda}{N_t\sigma_d^2\lambda + N_t N_0}\right) \right]. \end{aligned} \tag{6.25}$$

Note that $\mathbf{W}$ is an $r \times r$ random, positive semi-definite matrix following the complex Wishart distribution. Therefore, it has real non-negative eigenvalues, and the probabil-

ity density function of its unordered eigenvalue, λ, can be found in [110]

$$P_\lambda(\lambda) = K \sum_{m=1}^{r} \sum_{n=1}^{r} (-1)^{n+m} \lambda^{n+m+q-r-2} e^{-\lambda} \det(\mathbf{\Omega}). \qquad (6.26)$$

We can now define $a \triangleq P\alpha^2$, $b \triangleq N_t\sigma_d^2$, $c \triangleq N_0N_t$, and then proceed along with some integral techniques

$$
\begin{aligned}
R &= \sup_{0 \le P \le P_t} \int_0^\infty r \log_2\left(1 + \frac{a\lambda}{b\lambda + c}\right) P_\lambda(\lambda) d\lambda \\
&= \sup_{0 \le P \le P_t} \int_0^\infty r \log_2(a\lambda + b\lambda + c) P_\lambda(\lambda) d\lambda - \int_0^\infty r \log_2(b\lambda + c) P_\lambda(\lambda) d\lambda \\
&= \sup_{0 \le P \le P_t} \int_0^\infty r \log_2\left(1 + \frac{a+b}{c}\lambda\right) P_\lambda(\lambda) d\lambda - \int_0^\infty r \log_2\left(1 + \frac{b}{c}\lambda\right) P_\lambda(\lambda) d\lambda. \quad (6.27)
\end{aligned}
$$

Then, the final result can be easily obtained following the methodology of [111]. □

6.3.2 Asymptotic Analysis

Intuitively, the maximum EAR may behave as a non-increasing (non-monotonic) function of the input power. This is due to the presence of the nonlinearity distortion power σ_d^2 in (6.20), which also scales with the input power (see (6.11)). Thus, we seek to work out the EAR in the asymptotic regime. First, we obtain the EAR when we use the whole power budget, i.e. $P = P_t \to \infty$. After some manipulations, it can be shown that the EAR approaches to a saturation point, according to

$$\lim_{\sigma \to \infty} \alpha = \beta_N 2^{N-1/2} \sigma^{N-1} \Gamma\left(\frac{3+N}{2}\right), \qquad (6.28)$$

$$\lim_{\sigma \to \infty} \sigma_d^2 = \gamma_{2N} 2^N \sigma^{2N} \Gamma(1+N) - N_t 2^N \beta_N^2 \sigma^{2N} \Gamma^2\left(\frac{3+N}{2}\right), \qquad (6.29)$$

where $\sigma^2 = \frac{P}{2N_t}$ is the power of input signal (real/imaginary part). Thus,

$$R_{\text{high}} = \lim_{P \to \infty} R = r \log_2\left(1 + \frac{\beta_N^2 \Gamma^2\left(\frac{3+N}{2}\right)}{\gamma_{2N}\Gamma(1+N) - N_t\beta_N^2\Gamma^2\left(3 + \frac{n}{2}\right)}\right). \qquad (6.30)$$

The high-power asymptote in (6.30) reveals that increasing the input power with no bound, leads to a saturated value for the EAR. This observation is in sharp contrast with the classical MIMO results [15], [14], which consider perfectly linear PAs.[1]

At the other extreme, we find a closed-form expression for the EAR in the low SNR regime where $P \to 0$. In this case, we use the approximation $\ln(1 + u) \approx u$ for small u to get

$$
\begin{aligned}
R_{\text{low}} = \lim_{P \to 0} R &= \sup_{0 \le P \le P_t} \beta_1^2 \frac{(\log_2 e) P}{N_0 N_t} \mathbb{E}\big[\text{Tr}(\mathbf{W})\big] \\
&= \sup_{0 \le P \le P_t} \beta_1^2 \frac{(\log_2 e) P}{N_0 N_t} \mathbb{E}\left[\sum_{i=1}^{N_r} \sum_{j=1}^{N_t} |h_{ij}|^2\right] \\
&= \sup_{0 \le P \le P_t} \beta_1^2 \frac{(\log_2 e) P}{N_0 N_t} \sum_{i=1}^{r} \lambda_i \\
&= \sup_{0 \le P \le P_t} \beta_1^2 \frac{(\log_2 e) P N_r}{N_0} \\
&= \beta_1^2 \frac{(\log_2 e) P N_r}{N_0}.
\end{aligned}
\tag{6.31}
$$

The result in (6.31) is consistent with the classical MIMO results where PAs are assumed to be perfectly linear [70], [113]. This is due to the fact that in the low SNR regime, PAs are still operating in their linear regime.

6.3.3 Bounds

In the previous subsections, we have reduced the problem from a power allocation optimization to a power control based on only one variable, P. However, (6.22) and (6.25) are complicated formulas. Motivated by this, we confine the EAR between an upper and two lower bounds such that $R_{\text{lower}} \le R \le R_{\text{upper}}$. We start with the upper bound using Jensen's inequality for the concave function $\log_2 (1 + x)$ in (a) and for the concave function $\frac{1}{1+\frac{1}{x}}$ in (b):

---

[1]Interestingly, our result is in line with [112] which used a Gaussian model for the hardware residual distortions.

$$R = \sup_{0 \leq P \leq P_t} \mathbb{E}\left[r \log_2 \left(1 + \frac{P\alpha^2 \lambda}{N_t \sigma_d^2 \lambda + N_t N_0} \right) \right]$$

$$\overset{(a)}{\leq} \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \mathbb{E}\left[\frac{P\alpha^2 \lambda}{N_t \sigma_d^2 \lambda + N_t N_0} \right] \right)$$

$$\overset{(b)}{\leq} \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2}{N_t \sigma_d^2 \left(1 + \frac{N_0}{\sigma_d^2 \mathbb{E}[\lambda]} \right)} \right)$$

$$= \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2 q}{N_t \sigma_d^2 q + N_t N_0} \right)$$

$$\overset{\Delta}{=} R_{\text{upper}}. \tag{6.32}$$

The EAR can also be lower bounded by recalling that $\mathbb{E}\left[\log_2 \left(1 + \rho(\lambda) \right) \right] \geq \log_2 \left(1 + \exp^{\mathbb{E}\left[\ln \rho(\lambda) \right]} \right)$ in (c), to get

$$R = \sup_{0 \leq P \leq P_t} \mathbb{E}\left[r \log_2 \left(1 + \frac{P\alpha^2 \lambda}{N_t \sigma_d^2 \lambda + N_t N_0} \right) \right]$$

$$= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[r \log_2 \left(1 + \exp\left(\ln\left(\frac{P\alpha^2}{N_t \sigma_d^2} \frac{\lambda}{\lambda + \frac{N_0}{\sigma_d^2}} \right) \right) \right) \right]$$

$$\overset{(c)}{\geq} \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2}{N_t \sigma_d^2} \exp\left(\mathbb{E}\left[\ln\left(\frac{\lambda}{\lambda + \frac{N_0}{\sigma_d^2}} \right) \right] \right) \right)$$

$$= \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2}{N_t \sigma_d^2} \exp\left(\mathbb{E}\left[\ln(\lambda) \right] - \mathbb{E}\left[\ln\left(\lambda + \frac{N_0}{\sigma_d^2} \right) \right] \right) \right)$$

$$\geq \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2}{N_t \sigma_d^2} \exp\left(\mathbb{E}\left[\ln(\lambda) \right] - \ln\left(\mathbb{E}[\lambda] + \frac{N_0}{\sigma_d^2} \right) \right) \right)$$

$$= \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2 \exp\left(\mathbb{E}[\ln(\lambda)] \right)}{N_t \sigma_d^2} \exp\left(-\ln\left(q + \frac{N_0}{\sigma_d^2} \right) \right) \right)$$

$$= \sup_{0 \leq P \leq P_t} r \log_2 \left(1 + \frac{P\alpha^2 \exp\left(\mathbb{E}[\ln(\lambda)] \right)}{N_t \sigma_d^2 q + N_t N_0} \right) \overset{\Delta}{=} R_{\text{lower1}}. \tag{6.33}$$

It is known that $\mathbb{E}\left[\ln\left(\det(\mathbf{W}) \right) \right] = \sum_{l=0}^{r-1} \psi(q - l)$, where $\psi(\cdot)$ represents the Euler digamma function [114]. Hence, we can represent $\mathbb{E}\left[\ln(\lambda) \right]$ in the following

way

$$\mathbb{E}\big[\ln(\lambda)\big] = \frac{1}{r}\mathbb{E}\left[\sum_{i=1}^{r}\ln(\lambda_i)\right] = \frac{1}{r}\mathbb{E}\left[\ln\big(\det(\mathbf{W})\big)\right] = \frac{1}{r}\sum_{l=0}^{r-1}\psi(q-l). \quad (6.34)$$

To sum up, the EAR can be lower- and upper-bounded as follows

$$\sup_{0\leq P\leq P_t} r\log_2\left(1 + \frac{P\alpha^2\exp\left(\frac{1}{r}\sum_{l=0}^{r-1}\psi(q-l)\right)}{N_t\sigma_d^2 q + N_t N_0}\right) \leq R \leq \sup_{0\leq P\leq P_t} r\log_2\left(1 + \frac{P\alpha^2 q}{N_t\sigma_d^2 q + N_t N_0}\right).$$

$$(6.35)$$

Hereafter, we also use another lower bound, named second lower bound, that can be especially useful when the number of receive antennas is much higher than the number of transmit antennas.[1] We follow the same approach to derive this bound:

$$\begin{aligned}
R &\geq \sup_{0\leq P\leq P_t} \mathbb{E}\left[r\log_2\left(1 + \exp\left(\ln\left(\frac{P\alpha^2}{N_t\sigma_d^2}\frac{\lambda}{\lambda + \frac{N_0}{\sigma_d^2}}\right)\right)\right)\right] \\
&\geq \sup_{0\leq P\leq P_t} r\log_2\left(1 + \frac{P\alpha^2}{N_t\sigma_d^2}\exp\left(\mathbb{E}\left[\ln\left(\frac{\lambda}{\lambda + \frac{N_0}{\sigma_d^2}}\right)\right]\right)\right) \\
&= \sup_{0\leq P\leq P_t} r\log_2\left(1 + \frac{P\alpha^2}{N_t\sigma_d^2}\exp\left(\ln\left(\frac{1}{1 + \frac{N_0}{\sigma_d^2}\mathbb{E}\left[\frac{1}{\lambda}\right]}\right)\right)\right) \\
&= \sup_{0\leq P\leq P_t} r\log_2\left(1 + \frac{P\alpha^2}{N_t\sigma_d^2 + N_t N_0\frac{1}{N_r - N_t}}\right) \\
&= R_{\text{lower2}},
\end{aligned} \quad (6.36)$$

where we have invoked the central Wishart matrix property

$$\mathbb{E}\left[\text{Tr}\left(\mathbf{W}^{-1}\right)\right] = \frac{N_t}{N_r - N_t} \qquad N_r \geq N_t + 1. \quad (6.37)$$

---

[1]According to (6.37), the second lower bounds becomes loose whenever the numbers of transmit and receive antennas are close to each other. However, we can introduce a new lower bound like $R_{\text{lower}} = \max\{R_{\text{lower1}}, R_{\text{lower2}}\}$ to always guarantee a tight lower bound.

6.4 Simulation Results

In this section, we present simulation results illustrating the EAR of MIMO systems in the presence of PA nonlinearities by generating 10^4 Monte Carlo realizations of the flat fading matrix $\mathbf{H}$. Hereafter, we choose solid state PAs, whose AM/AM and AM/PM functions are specified by [115]

$$g_A(A) = \frac{A}{\left[1 + \left(\frac{A}{A_{os}}\right)^{2v}\right]^{\frac{1}{2v}}},$$

(6.38)

$$g_\phi(A) = 0,$$

(6.39)

where A_{os} denotes the output saturation voltage and v sets the smoothness of transition from the linear region to the saturation part. In particular, for large v this model approaches the ideal clipping PA model which is commonly used to represent the hard clipping effect [49]. Furthermore, we fit this PA conversion function with a polynomial of degree 9 in a least-squares sense, and ignore the even order terms, since they
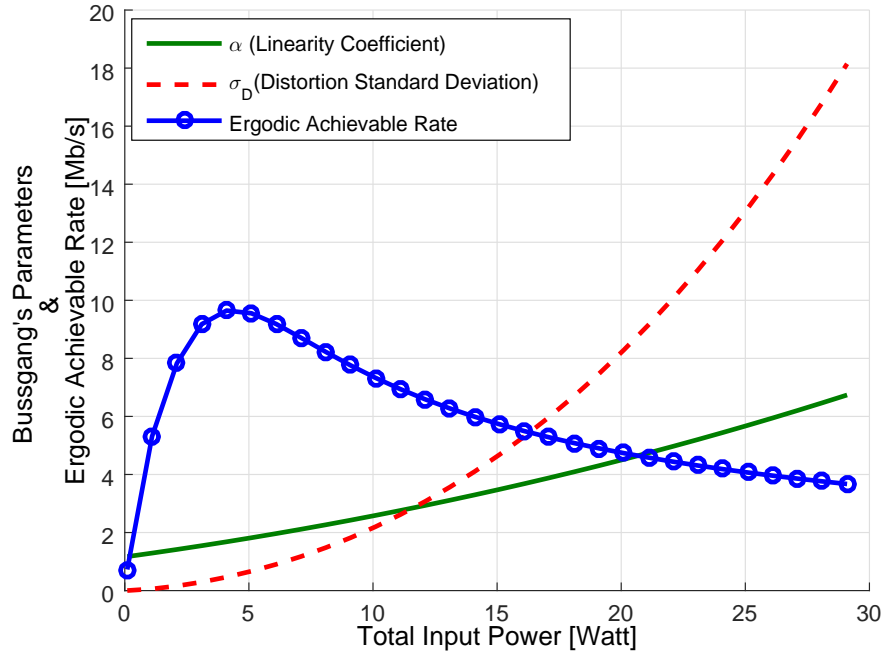

Figure 6.1: Impact of Bussgang's theorem parameters on the EAR ($N_t = 3$, $N_r = 4$, $A_{os} = 1$, $v = 1$, $N_0 = 1$, $B = 1$ MHz).
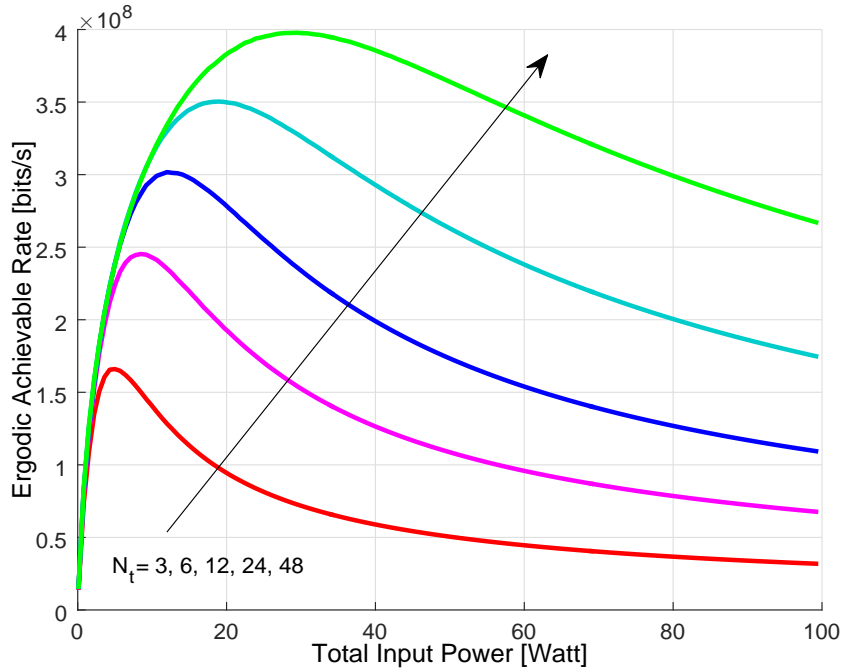
Figure 6.2: Impact of the number of transmit antennas on EAR ($N_r = 4$, $A_{os} = 1$, $v = 1$, $N_0 = 1$, $B = 20$ MHz).

contribute with only out-band distortion [106]. Figure 6.1 depicts the role of the Bussgang's parameters on the EAR. Although, the linearity coefficient (α), is dominant in the low-power regime, distortion (σ_d) dominates in the higher input power. By this observation, a non-monotonic behavior of EAR function is anticipated.

The performance of MIMO system under PA nonlinearities model for different transceiver antennas is shown in Fig. 6.2 and 6.3, respectively. It can be easily observed that an increase in the number of transmit antennas leads to an array gain and pushes the maximum of the EAR into higher input powers. On the other hand, when we increase the number of antennas on the receiver side, the maximum point is obtained in the lower input power regime. It is best suitable to justify this behavior by the fact that PAs, as a major source of nonlinearities, only exist in the transmitter side. So, as a practical result, we are interested in utilizing more antennas on the receiver side to achieve a high EAR by a lower input power. Finally, Fig. 6.4 illustrates how tight the suggested bounds are. These bounds seem to offer a very good approximation of the EAR for any total input power. However, approximating the best total input power that leads to the maximum EAR is more important. Figure 6.4 confirms indeed that the
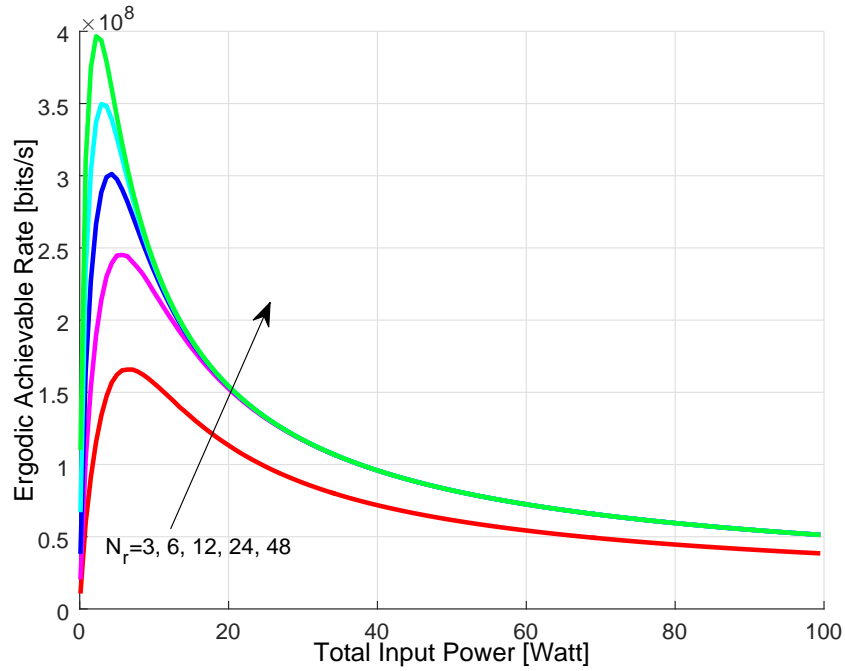
Figure 6.3: Impact of the number of receive antennas on ergodic achievable rate ($N_t = 4$, $A_{os} = 1$, $v = 1$, $N_0 = 1$, $B = 20$ MHz).
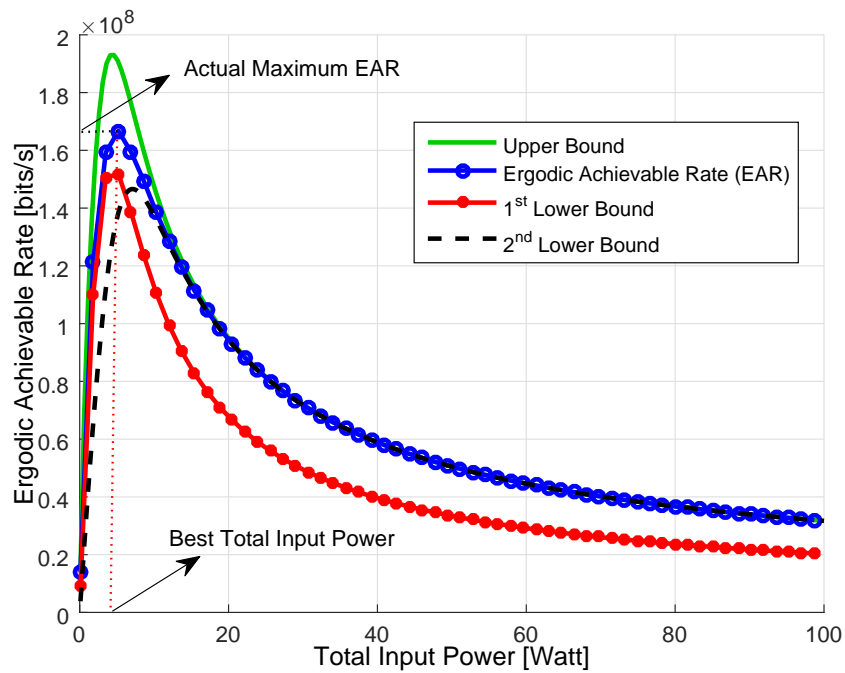
Figure 6.4: Lower and upper bounds for the ergodic achievable rate of MIMO systems under PA nonlinearity assumption ($N_t = 3$, $N_r = 4$, $A_{os} = 1$, $v = 1$, $N_0 = 1$, $B = 20$ MHz).

best operating power of lower/upper bounds leads to the maximum EAR.

6.5 Concluding Remarks

Working with inexpensive nonlinear PAs seems to be a viable solution for the next generation of wireless systems. This nonlinear behavior distorts the transmitted signal and unfortunately reduces the achievable rate in any communication system. This performance degradation becomes substantial when the power fed into the PAs is high. On the other hand, PAs offer their best efficiency in their nonlinear regime. Motivated by the above fundamental tradeoff, we have analytically quantified the impact of PAs nonlinearities on the achievable rate of MIMO systems. Our analysis derived closed-form exact expressions along with tractable asymptotic approximations.

Chapter 7

Conclusion and Future Work

In this thesis, we considered a MIMO relay system with a large number of antennas at both sides, but with limited number of RF chains to keep the cost, power consumption, and complexity of the system to affordable levels. To do this, we proposed a low-cost hybrid configuration paradigm at the relay station which consists of an RF analog beamformer followed by a linear baseband digital signal processors, either MRC/MRT or ZF. We assumed a fully connected structure for the analog beamformer, where each RF chain is connected to all the service antennas via a bank of phase shifters. We also took into account that phase shifters contribute to generate the beam patterns instead of adjusting both modulus and phase. Then, we evaluated the performance of the system in terms of achievable rate under a wide range of assumptions such as: Given perfect CSI at the relay station, given only imperfect CSI at the relay station, having ideal or quantized phase shifters, correlated and uncorrelated channels. Also, in order to have a low-cost architecture we assumed another scenario, but in the context of MIMO systems, where we deployed inexpensive nonlinear PAs at the transmitter side. By making this assumption, we further analyzed the performance of system and showed that the linearity coefficient is dominant in the low-power regime, while in-band distortion dominates in the higher input power. Then, we developed upper and lower bounds on the EAR which describe rigorously the behavior of system for a wide range of input powers.

7.1 Conclusion

In Chapter 3, we proposed a hybrid structure at the relay station for multipair massive MIMO relaying, where perfect CSI is known at the relay station. Then, we derived the asymptotic achievable rate under three different power regimes and showed that the power at either the relay or user nodes can scale down inversely proportional to the number of service antennas at the relay station, while the system can still achieve the same achievable rate as before. To have a case of practical interest, we relaxed the accuracy of phase shifters to q-bit resolution, and then analytically showed that the performance of system will decrease by the function of $\mathrm{sinc}^2\left(\frac{\pi}{2^q}\right)$. Our simulation results illustrated that only 2 quantization bits are sufficient to capture near 90% of the achievable rate offered by an unrealistic system with ideal phase shifters.

In Chapter 4, we assumed that perfect CSI is not available at the relay station, hence we estimated the channels with uplink orthogonal pilots. Then, we deployed a Hadamard analog beamformer with phases $\pm\pi$ to further reduce the cost and complexity of the system. Our analytical results followed by Monte Carlo simulation imply that the system can capture 75% of the achievable rate offered by a fully digital structure where the number of RF chains are equal to the number of antennas.

In Chapter 5, we generalized our system model to the case that the propagation channels are correlated at the relay side. Next, we estimated the channel at the relay node and designed the analog beamformer to minimize the channel estimation error and, in turn, optimize the achievable rate. This optimization finally resulted in a conventional water-filling algorithm, where the analog beamformer chose the strongest eigenmodes of the channel correlation matrix. Our analytical and simulation results showcased that by removing half of the RF chains, the hybrid structure can achieve 95% of the achievable rate of the fully digital structure. Moreover, the performance of the system is robust to the practical constraints on the phase shifters. Furthermore, the results intuitively imply that although multiplexing gain is confined due to the limited number of RF chains at the relay station, and also the power gain is restricted due to the less control of the DSP unit on the antennas, the system avails of a diversity gain

by exploring the strongest paths.

Finally, in Chapter 6, we assumed another low-cost architecture in the context of point-to-point MIMO systems, where there exists a crucial demand for inexpensive but nonlinear PAs. We first brought all the PA modelings under the same umbrella by utilizing the polynomial model. Then, we reduced the optimization of the EAR from a power allocation optimization into a simple power control with a closed-form expression based on the total input power. To have a tractable and insightful result, we developed two tight lower bounds and one upper bound on the EAR. In addition, we showed that in the low power regime the system performance is similar to the systems with linear PAs, while in the very high power regime the performance of the system deteriorates due to the in-band distortion.

7.2 Future Work

This thesis has provided important contributions to our understanding of low-cost architectures for future MIMO systems. It has proposed a multipair massive MIMO relaying with limited number of RF front-ends, which can achieve most of the achievable rate offered by a fully RF chain structure. As well as this, the thesis has studied the impact of PA nonlinearities in the context of MIMO systems. Our results revealed that increasing the input power with no bound, may substantially reduce the EAR. Nonetheless, there are still some issues that warrant investigation in the future. These include:

- **The number of RF chains which achieve the maximum energy efficiency:** In general, there is a trade-off between the achievable rate and power consumption in wireless communication systems. This trade-off can be best captured by energy efficiency which signifies the total number of bits sent per Joule of energy. Although this thesis has shown a great power saving in the proposed hybrid topology, the energy efficiency of the system should be analytically derived to understand how many RF chains is required to meet the maximum energy efficiency.

- **Hybrid topology for other types of massive MIMO relaying:** In this thesis, we explored the performance of multipair massive MIMO relaying with the hybrid structure at the AF relay station. However, there is a dearth of literature for other types of relying schemes such as hybrid topology for decode-and-forward massive relaying, full duplex massive relaying, and two-way massive relaying. These systems can be analyzed from different view points such as outage probability, latency, spectral efficiency, and energy efficiency.

- **Millimetre-wave hybrid relaying systems:** In general, hybrid massive MIMO structure is beneficial in microwave wireless communications, but this topology is seemingly essential in millimetre-wave communications. Since the RF technology is pushed to its operation boundaries, the intrinsic imperfections and power consumption of the RF chains are more and more governing the system performance of wireless systems, particularly in millimetre wave bands. On the other hand, the sparse-scattering structure of millimetre-wave channels enables us to employ low-complex precoding/decoding algorithms, such as orthogonal matching pursuit and compressive sensing schemes.

Appendix A

Proofs for Chapter 3

A.1 Proof of Lemma 3.2

Having discussed how to construct $\mathbf{F}_1$, we can write each entry of this matrix as $\frac{1}{\sqrt{N}} \exp\left(j\theta_{m,n}\right)$, where $\theta_{m,n}$ is a uniform random variable, i.e. $\theta_{m,n} \sim U\left[0, 2\pi\right)$. Now, let the vectors $\mathbf{f}_{1_p}$ and $\mathbf{f}_{1_q}$ contain the elements of the p-th and q-th rows in matrix $\mathbf{F}_1$, respectively. Then, $\mathbf{f}_{1_p}^H \mathbf{f}_{1_p} = 1$ since the phases cancel out each other. On the other hand, if $N \to \infty$, due to the central limit theorem for any $p \neq q$ we have

$$\mathbf{f}_{1_p}^H \mathbf{f}_{1_q} = \frac{1}{N} \sum_{l=1}^{N} \mathrm{e}^{j\left(\theta_{p,l} - \theta_{q,l}\right)} \to \mathbb{E}\left[\mathrm{e}^{j\theta_p}\right] \mathbb{E}\left[\mathrm{e}^{-j\theta_q}\right] = 0, \tag{A.1}$$

where the distribution of θ_p and θ_q are defined similar to $\theta_{m,n}$. The same result is analogous for matrix $\mathbf{F}_2$.

A.2 Proof of Lemma 3.4

Let us rewrite the (m,n)-th entry of matrix $\mathbf{H}_1$ like $r_{m,n}\mathrm{e}^{j\phi_{m,n}}$, where the modulus and phase have a Rayleigh and uniform distribution, respectively. In other words, $r_{m,n} \sim R\left(0, \frac{1}{2}\right)$, and $\phi_{m,n} \sim U\left[0, 2\pi\right)$. Now, let the vectors $\mathbf{f}_{1_p}$ and $\mathbf{h}_{1_p}$ denote the p-th row of matrix $\mathbf{F}_1$ and $\mathbf{H}_1$, respectively. Then, since phases cancel out each other, for any $p \leq r_1$ we have that

$$\mathbf{f}_{1_p}^H \mathbf{h}_{1_p} = \frac{1}{\sqrt{N}} \sum_{l=1}^{N} r_{p,l} \overset{(a)}{\to} \sqrt{N} \mathbb{E}\left[r_p\right] = \sqrt{\frac{N\pi}{4}}, \tag{A.2}$$

where we have used the central limit theorem in (a), and the fact that r_p is a Rayleigh random variable with parameter $\frac{1}{2}$. In a similar way, we can prove the second part.

A.3 Proof of Lemma 3.5

The results follow trivially by using the same methodology outlined in Lemma 3.4. Let vectors $\hat{\mathbf{f}}_{1_p}$ and $\mathbf{h}_{1_p}$ denote the p-th row of matrix $\hat{\mathbf{F}}_1$ and $\mathbf{H}_1$, respectively. Then, for any $p \leq r_1$ we have

$$\hat{\mathbf{f}}_{1_p}^H \mathbf{h}_{1_p} = \frac{1}{\sqrt{N}} \sum_{l=1}^{N} r_{p,l} \mathrm{e}^{\left(\phi_{p,l} - \hat{\phi}_{p,l}\right)} = \sqrt{N} \mathbb{E}\left[r_p\right] \mathbb{E}\left[\mathrm{e}^{\epsilon}\right] = \sqrt{\frac{N\pi}{4}} \operatorname{sinc}(\delta), \qquad (A.3)$$

where r_p is a Rayleigh random variable with parameter $\frac{1}{2}$, and ϵ is an uniform random variable $\epsilon \sim U\left[-\delta, +\delta\right)$.

Appendix B

Proofs for Chapter 4

B.1   Prerequisite Lemmas for Chapter 4

In this appendix, we derive some prerequisite lemmas for the random matrix analysis which are useful in our analysis.

Lemma 1. *Let* $\mathbf{M} \in \mathbb{C}^{M \times N}$ *be a zero-mean Gaussian random matrix with i.i.d. entries of power* σ^2 *. Also, assume that* $\mathbf{V} \in \mathbb{C}^{N \times N}$ *is a deterministic matrix. Then, we have*

$$\mathbb{E}\big[\mathbf{M}\mathbf{V}\mathbf{M}^H\big] = \sigma^2 \mathrm{Tr}\big(\mathbf{V}\big)\mathbf{I}_M. \tag{B.1}$$

Proof. Let us expand this multiplication across the columns of $\mathbf{M}$, and then simplify the expression using the fact that these columns are mutually uncorrelated.

$$\mathbb{E}\Big[\mathbf{M}\mathbf{V}\mathbf{M}^H\Big] = \sum_{i=1}^{N}\sum_{j=1}^{N}\big[\mathbf{V}\big]_{ij}\mathbb{E}\Big[\mathbf{m}_i\mathbf{m}_j^H\Big] = \sum_{i=1}^{N}\big[\mathbf{V}\big]_{ii}\mathbb{E}\Big[\mathbf{m}_i\mathbf{m}_i^H\Big] = \sigma^2\mathrm{Tr}\big(\mathbf{V}\big)\mathbf{I}_M. \tag{B.2}$$

$\square$

Lemma 2. *Let* $\mathbf{v} \in \mathbb{C}^N$ *denote a complex Gaussian random vector with i.i.d. entries* $\mathcal{CN}(0, 1)$. *Then, we can conclude that*

$$\mathbb{E}\Big[\|\mathbf{v}\|_2^4\Big] = N^2 + N. \tag{B.3}$$

Proof. Let $(v)_i$ denote the i-th element of vector $\mathbf{v}$. Then, we have

$$\mathbb{E}\Big[\|\mathbf{v}\|_2^4\Big] = \mathbb{E}\Bigg[\Big[\sum_{i=1}^{N}\,|\,(v)_i\,|^2\Big]^2\Bigg] = \mathbb{E}\Bigg[\Big[\sum_{i=1}^{N}2Y_i\Big]^2\Bigg] = 4\mathbb{E}\big[Z^2\big] = N^2 + N, \quad \text{(B.4)}$$

where $Y_i \sim \mathrm{Exp}(\lambda = 2)$ is an exponential random variable, and consequently Z follows an Erlang distribution. $\square$

Lemma 3. *Let* $\mathbf{A} = \mathbf{H}\mathbf{D}^{\frac{1}{2}}$*, where* $\mathbf{H}$ *is an* $M \times N$ *matrix with independent zero-mean unit variance complex Gaussian random entries, and* $\mathbf{D}$ *is an* $N \times N$ *deterministic*

diagonal matrix $\mathrm{diag}\{\beta_1, \ldots, \beta_N\}$. *Then,*

$$\mathbb{E}\big[\mathbf{A}^H\mathbf{A}\mathbf{A}^H\mathbf{A}\big] = M^2\mathbf{D}^2 + M\mathrm{Tr}\big(\mathbf{D}\big)\mathbf{D}. \tag{B.5}$$

Proof. First, let us expand $\mathbf{A}$ based on its definition and then define $\mathbf{Q}$ in a following way

$$\mathbb{E}\big[\mathbf{A}^H\mathbf{A}\mathbf{A}^H\mathbf{A}\big] = \mathbf{D}^{\frac{1}{2}}\underbrace{\mathbb{E}\big[\mathbf{H}^H\mathbf{H}\mathbf{D}\mathbf{H}^H\mathbf{H}\big]}_{\mathbf{Q}}\mathbf{D}^{\frac{1}{2}}, \tag{B.6}$$

where the (i,j)-th element of $\mathbf{Q}$ can be obtained by

$$\big[\mathbf{Q}\big]_{ij} = \mathbb{E}\big[\mathbf{h}_i^H\mathbf{H}\mathbf{D}\mathbf{H}^H\mathbf{h}_j\big] = \sum_{n=1}^{N}\beta_n\mathbb{E}\big[\mathbf{h}_i^H\mathbf{h}_n\mathbf{h}_n^H\mathbf{h}_j\big]. \tag{B.7}$$

The off-diagonal entries $i \neq j$ are zero due to the orthogonality of columns, however for the diagonal entries we can obtain

$$\big[\mathbf{Q}\big]_{ii} = \sum_{n \neq i}^{N}\beta_n\mathbb{E}\big[\mathbf{h}_i^H\mathbf{h}_n\mathbf{h}_n^H\mathbf{h}_i\big] + \beta_i\mathbb{E}\big\|\mathbf{h}_i\big\|_2^4 = M^2\beta_i + M\sum_{n=1}^{N}\beta_n, \tag{B.8}$$

where the last equation followed directly by Lemma 2. Therefore, we conclude that

$$\mathbf{Q} = M^2\mathbf{D} + M\mathrm{Tr}\big(\mathbf{D}\big)\mathbf{I}_N. \tag{B.9}$$

$\square$

B.2 Proof of Proposition 4.2

Considering that the effective channels are independent of each other, and also according to Proposition 4.1, we can rewrite $\mathrm{var}\big(\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big)$ based on its power and average such that

$$\mathrm{var}\big(\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\big) = \mathbb{E}\Big[\hat{\mathbf{a}}_{2,k}^H\hat{\mathbf{A}}_2\mathbf{B}\hat{\mathbf{A}}_2^H\hat{\mathbf{a}}_{2,k}\Big] - K_a^2K_b^2\hat{\beta}_{1,k}^2\hat{\beta}_{2,k}^2, \tag{B.10}$$

where the matrix $\mathbf{B}$ is defined as

$$\mathbf{B} \triangleq \mathbb{E}\big[\hat{\mathbf{A}}_1^H\hat{\mathbf{a}}_{1,k}\hat{\mathbf{a}}_{1,k}^H\hat{\mathbf{A}}_1\big]. \tag{B.11}$$

We show that the matrix $\mathbf{B}$ is a diagonal matrix, like $\mathbf{B} \triangleq \mathrm{diag}\big\{\hat{\beta}_{b,1}, \cdots, \hat{\beta}_{b,k}, \cdots, \hat{\beta}_{b,K}\big\}$. The off diagonal elements of this matrix are zero due to the orthogonal columns of $\hat{\mathbf{A}}_1$. Precisely speaking, we have

$$\left[\mathbf{B}\right]_{ij} = \mathbb{E}\left[\hat{\mathbf{a}}_{1,i}^H \hat{\mathbf{a}}_{1,k} \hat{\mathbf{a}}_{1,k}^H \hat{\mathbf{a}}_{1,j}\right] = 0, \quad \text{for } \forall i \neq j. \tag{B.12}$$

On the other hand, by taking Lemma 1 into account in $(L1)$ below, the i-th entry on the main diagonal vector, is given by

$$\left[\mathbf{B}\right]_{ii} = \mathbb{E}\left[\hat{\mathbf{a}}_{1,i}^H \mathbb{E}\left[\hat{\mathbf{a}}_{1,k} \hat{\mathbf{a}}_{1,k}^H\right] \hat{\mathbf{a}}_{1,i}\right] \stackrel{(L1)}{=} K_a \hat{\beta}_{1,k} \hat{\beta}_{1,i} \stackrel{\Delta}{=} \hat{\beta}_{b,i}, \quad \text{for } i \neq k. \tag{B.13}$$

In a similar manner and using Lemma 2, we obtain the k-th entry of the main diagonal vector

$$\left[\mathbf{B}\right]_{kk} = \mathbb{E}\left[\left(\hat{\mathbf{a}}_{1,k}^H \hat{\mathbf{a}}_{1,k}\right)^2\right] = (K_a^2 + K_a)\hat{\beta}_{1,k}^2 \stackrel{\Delta}{=} \hat{\beta}_{b,k}. \tag{B.14}$$

By using the above, we can express the matrix $\mathbf{B}$ as

$$\mathbf{B} = K_a \hat{\beta}_{1,k} \hat{\mathbf{D}}_1 + K_a^2 \hat{\beta}_{1,k}^2 \mathbf{J}^{kk} = \text{diag}\left\{\hat{\beta}_{b,1}, \cdots, \hat{\beta}_{b,K}\right\}, \tag{B.15}$$

where $\mathbf{J}^{kk}$ is a single-entry matrix, in which $\left[\mathbf{J}\right]_{kk} = 1$ and the rest of the elements are zero. Now, considering that $\mathbf{B}$ is a diagonal deterministic matrix we can further simplify $\mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \mathbf{B} \hat{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right]$ by expanding the matrix product across the columns, and then after some straightforward mathematical manipulations we get

$$\mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \mathbf{B} \hat{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right] = K_b^2 \hat{\beta}_{2,k}^2 \hat{\beta}_{b,k} + K_b \hat{\beta}_{2,k} \sum_{i=1}^{K} \hat{\beta}_{2,i} \hat{\beta}_{b,i}. \tag{B.16}$$

This finishes the proof.

B.3 Proof of Proposition 4.3

In the first stage, we expand the matrix multiplication across the columns, and then we use the column orthogonality in (a) below to get a simple expression.

$$\sum_{i \neq k} \mathbb{E}\left[\left|\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{a}}_{1,i}\right|^2\right] = \sum_{i \neq k} \mathbb{E}\left[\left(\hat{\mathbf{a}}_{2,k}^H \sum_{m=1}^K \hat{\mathbf{a}}_{2,m} \hat{\mathbf{a}}_{1,m}^H \hat{\mathbf{a}}_{1,i}\right)\left(\hat{\mathbf{a}}_{2,k}^H \sum_{n=1}^K \hat{\mathbf{a}}_{2,n} \hat{\mathbf{a}}_{1,n}^H \hat{\mathbf{a}}_{1,i}\right)^H\right]$$

$$= \sum_{i \neq k} \sum_{m=1}^K \sum_{n=1}^K \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{a}}_{2,m} \mathbb{E}\left[\hat{\mathbf{a}}_{1,m}^H \hat{\mathbf{a}}_{1,i} \hat{\mathbf{a}}_{1,i}^H \hat{\mathbf{a}}_{1,n}\right] \hat{\mathbf{a}}_{2,n}^H \hat{\mathbf{a}}_{2,k}\right]$$

$$\stackrel{(a)}{=} \sum_{i \neq k} \sum_{m=1}^K \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{a}}_{2,m} \mathbb{E}\left[\hat{\mathbf{a}}_{1,m}^H \hat{\mathbf{a}}_{1,i} \hat{\mathbf{a}}_{1,i}^H \hat{\mathbf{a}}_{1,m}\right] \hat{\mathbf{a}}_{2,m}^H \hat{\mathbf{a}}_{2,k}\right], \tag{B.17}$$

for which,

$$\mathbb{E}\left[\hat{\mathbf{a}}_{1,m}^H \hat{\mathbf{a}}_{1,i} \hat{\mathbf{a}}_{1,i}^H \hat{\mathbf{a}}_{1,m}\right] = \begin{cases} \left(K_a^2 + K_a\right) \hat{\beta}_{1,i}^2, & \text{if } k \neq m = i \\[2mm] K_a \hat{\beta}_{1,i} \hat{\beta}_{1,k}, & \text{if } k = m \neq i \\[2mm] K_a \hat{\beta}_{1,i} \hat{\beta}_{1,m}, & \text{if } k \neq m \neq i, \end{cases} \tag{B.18}$$

which finally after some manipulations leads to (4.25).

B.4 Proof of Proposition 4.4

Let $(L1)$ and $(L2)$ refer to Lemma 1 and Lemma 2, respectively. Then, using these lemmas we reach to

$$t_3 = \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \mathbb{E}\left[\mathbf{E}_1 \mathbf{E}_1^H\right] \hat{\mathbf{A}}_1 \hat{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right]$$

$$\stackrel{(L1)}{=} \text{Tr}\left(\mathbf{D}_{e1}\right) \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \mathbb{E}\left[\hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1\right] \hat{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right]$$

$$\stackrel{(L1)}{=} K_a \text{Tr}\left(\mathbf{D}_{e1}\right) \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{D}}_1 \hat{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right]$$

$$\stackrel{(a)}{=} K_a \text{Tr}\left(\mathbf{D}_{e1}\right) \mathbb{E}\left[\hat{\mathbf{a}}_{2,k}^H \tilde{\mathbf{A}}_2 \tilde{\mathbf{A}}_2^H \hat{\mathbf{a}}_{2,k}\right]$$

$$= K_a \text{Tr}\left(\mathbf{D}_{e1}\right) \mathbb{E}\left[\sum_{i \neq k}^K \left|\hat{\mathbf{a}}_{2,k}^H \tilde{\mathbf{a}}_{2,i}\right|^2 + \left|\hat{\mathbf{a}}_{2,k}^H \tilde{\mathbf{a}}_{2,k}\right|^2\right]$$

$$\stackrel{(L2)}{\underset{(b)}{=}} K_a \text{Tr}\left(\mathbf{D}_{e1}\right)\left(K_b \hat{\beta}_{2,k} \sum_{i \neq k}^K \hat{\beta}_{1,i} \hat{\beta}_{2,i} + \left(K_b^2 + K_b\right) \hat{\beta}_{1,k} \hat{\beta}_{2,k}^2\right), \tag{B.19}$$

where $\tilde{\mathbf{A}}_2$ in (a) is given as

$$\tilde{\mathbf{A}}_2 = \hat{\mathbf{A}}_2 \hat{\mathbf{D}}_1^{\frac{1}{2}} = \hat{\mathbf{H}}_2 \hat{\mathbf{D}}_2^{\frac{1}{2}} \hat{\mathbf{D}}_1^{\frac{1}{2}}. \tag{B.20}$$

Finally, the result in (b) is derived invoking the following technique, where we apply this fact that all entries of effective channel matrices are independent:

$$
\mathbb{E}\left[\left|\hat{\mathbf{a}}_{2,k}^H \tilde{\mathbf{a}}_{2,i}\right|^2\right] = \mathbb{E}\left[\left|\sum_{j=1}^{K_b} \left[\hat{\mathbf{A}}_2\right]_{jk}^* \left[\tilde{\mathbf{A}}_2\right]_{ji}\right|^2\right] = \mathbb{E}\left[\sum_{j=1}^{K_b}\sum_{l=1}^{K_b} \left(\hat{\mathbf{A}}_2\right)_{jk}^* \left(\tilde{\mathbf{A}}_2\right)_{ji} \left(\tilde{\mathbf{A}}_2\right)_{li}^* \left(\hat{\mathbf{A}}_2\right)_{lk}\right]
$$

$$
= \sum_{j=1}^{K_b} \mathbb{E}\left|\left[\hat{\mathbf{A}}_2\right]_{jk}\right|^2 \mathbb{E}\left|\left[\tilde{\mathbf{A}}_2\right]_{ji}\right|^2 = K_b \hat{\beta}_{2,k} \hat{\beta}_{2,i} \hat{\beta}_{1,i}. \tag{B.21}
$$

B.5 Proof of Proposition 4.5

Applying Lemma 3 $(L3)$, followed by Lemma 1 $(L1)$, we find that

$$
\mathbb{E}\left[\left\|\mathbf{e}_{2,k}^H \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1\right\|^2\right] = \mathbb{E}\left[\mathbf{e}_{2,k}^H \hat{\mathbf{A}}_2 \mathbb{E}\left[\hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1 \hat{\mathbf{A}}_1^H \hat{\mathbf{A}}_1\right] \hat{\mathbf{A}}_2^H \mathbf{e}_{2,k}\right]
$$

$$
\overset{(L3)}{=} K_a^2 \mathbb{E}\left[\mathbf{e}_{2,k}^H \mathbb{E}\left[\hat{\mathbf{H}}_2 \hat{\mathbf{D}}_1^2 \hat{\mathbf{D}}_2 \hat{\mathbf{H}}_2^H\right] \mathbf{e}_{2,k}\right] + K_a \text{Tr}(\hat{\mathbf{D}}_1) \mathbb{E}\left[\mathbf{e}_{2,k}^H \mathbb{E}\left[\hat{\mathbf{H}}_2 \hat{\mathbf{D}}_2 \hat{\mathbf{D}}_1 \hat{\mathbf{H}}_2^H\right] \mathbf{e}_{2,k}\right]
$$

$$
\overset{(L1)}{=} K_a^2 \text{Tr}(\hat{\mathbf{D}}_1^2 \hat{\mathbf{D}}_2) \mathbb{E}\left[\mathbf{e}_{2,k}^H \mathbf{e}_{2,k}\right] + K_a \text{Tr}(\hat{\mathbf{D}}_1) \text{Tr}(\hat{\mathbf{D}}_1 \hat{\mathbf{D}}_2) \mathbb{E}\left[\mathbf{e}_{2,k}^H \mathbf{e}_{2,k}\right], \tag{B.22}
$$

from which the proposition is obtained.

Appendix C

Proofs for Chapter 5

C.1 Prerequisite Lemmas for Chapter 5

In this appendix, we provide some prerequisite lemmas which are useful in our analysis.

Lemma 1. *Let* $\mathbf{H} \in \mathbb{C}^{N \times K}$ *be a zero-mean Gaussian random matrix with i.i.d. entries with variance* σ^2. *Also, assume that* $\mathbf{V} \in \mathbb{C}^{N \times N}$ *is a deterministic matrix. Then, we have*

$$\mathbb{E}\left[\mathbf{H}^H \mathbf{V} \mathbf{H}\right] = \sigma^2 \mathrm{Tr}\left(\mathbf{V}\right) \mathbf{I}_K. \tag{C.1}$$

Lemma 2. *Let the vectors* $\mathbf{h}$, $\mathbf{g} \in \mathbb{C}^N$ *be two independent zero-mean circular Gaussian random vectors such that* $\mathbf{h}$, $\mathbf{g} \sim \mathcal{CN}\left(0, \mathbf{I}_N\right)$. *Also, consider the deterministic matrix* $\mathbf{U}$, *then it holds that*

$$\mathbb{E}\left[\left|\mathbf{h}^H \mathbf{U} \mathbf{g}\right|^2\right] = \|\mathbf{U}\|^2, \tag{C.2}$$

$$\mathbb{E}\left[\left|\mathbf{h}^H \mathbf{U} \mathbf{h}\right|^2\right] = \left|\mathrm{Tr}\left(\mathbf{U}\right)\right|^2 + \|\mathbf{U}\|^2. \tag{C.3}$$

Proof. See [116, Lemma 2]. $\square$

Lemma 3. *Let* $\mathbf{G} = \mathbf{U}^{\frac{1}{2}} \mathbf{H}$, *where* $\mathbf{H} \in \mathbb{C}^{N \times K}$ *is a zero-mean Gaussian random matrix with i.i.d. entries of unit variance. Then, for any Hermitian deterministic matrix* $\mathbf{U} \in \mathbb{C}^{N \times N}$ *we have*

$$\mathbb{E}\left[\mathbf{G}^H \mathbf{G} \mathbf{G}^H \mathbf{G}\right] = \left(\mathrm{Tr}^2(\mathbf{U}) + K\|\mathbf{U}\|^2\right) \mathbf{I}_K. \tag{C.4}$$

Proof. Let us define the matrix $\mathbf{Q}$ as it is shown in below,

$$\mathbf{Q} \triangleq \mathbb{E}\left[\mathbf{G}^H \mathbf{G} \mathbf{G}^H \mathbf{G}\right], \tag{C.5}$$

then the (i, i)-th element of matrix $\mathbf{Q}$ can be obtained as

$$\begin{aligned}
\left[Q\right]_{i,i} &= \mathbb{E}\left[\mathbf{g}_i^H \left(\sum_{m=1}^K \mathbf{g}_m \mathbf{g}_m^H\right)\mathbf{g}_i\right] \\
&= \mathbb{E}\left[\mathbf{g}_i^H \mathbf{g}_i \mathbf{g}_i^H \mathbf{g}_i + \mathbf{g}_i^H \left(\sum_{m\neq i}^K \mathbf{g}_m \mathbf{g}_m^H\right)\mathbf{g}_i\right] = \mathbb{E}\left[\left|\mathbf{g}_i^H \mathbf{g}_i\right|^2\right] + \sum_{m\neq i}^K \mathbb{E}\left[\left|\mathbf{g}_i^H \mathbf{g}_m\right|^2\right] \\
&= \text{Tr}^2\left(\mathbf{U}\right) + K\left\|\mathbf{U}\right\|^2 .
\end{aligned} \tag{C.6}$$

We note that in the last equation we applied Lemma 2. In a similar spirit we can show the off-diagonal elements of matrix $\mathbf{Q}$ are zero. This finishes the proof. $\qquad\square$

Lemma 4. *Let the vector* $\mathbf{h} \in \mathbb{C}^N$ *be a zero-mean circular Gaussian random vector such that* $\mathbf{h} \sim \mathcal{CN}\left(0, \mathbf{I}_N\right)$. *Also let us assume that* $\mathbf{U} \in \mathbb{C}^{N\times N}$ *is a positive definite matrix with a limited spectral norm,* $\lambda_{\max}(\mathbf{U}) < \infty$, *where* $\lambda_{\max}(\mathbf{U})$ *denotes the largest eigenvalue of matrix* $\mathbf{U}$. *Then, it can be shown that for* $N \to \infty$

$$\frac{\mathbf{h}^H \mathbf{U}\mathbf{h}}{\text{Tr}\left(\mathbf{U}\right)} \xrightarrow{\text{a.s.}} 1. \tag{C.7}$$

Proof. This Lemma can be directly proved invoking the Chebyshev's inequality for the random variable $X = \frac{\mathbf{h}^H \mathbf{U}\mathbf{h}}{\text{Tr}\left(\mathbf{U}\right)}$

$$\Pr\left(\left|X - \mathbb{E}\left[X\right]\right| \geq \epsilon\right) \leq \frac{\text{var}\left(X\right)}{\epsilon^2}, \tag{C.8}$$

where we assume that ϵ is a very small real number. The expectation and power of the random variable X are given by (see equation (C.3))

$$\mathbb{E}\left[X\right] = 1, \tag{C.9}$$

$$\mathbb{E}\left[X^2\right] = \frac{\text{Tr}^2\left(\mathbf{U}\right) + \left\|\mathbf{U}\right\|^2}{\text{Tr}^2\left(\mathbf{U}\right)}. \tag{C.10}$$

Let us assume that λ_i denotes the i-th biggest eigenvalue of matrix $\mathbf{U}$, then

$$\begin{aligned}
\text{var}\left(X\right) &= \frac{\left\|\mathbf{U}\right\|^2}{\text{Tr}^2\left(\mathbf{U}\right)} = \frac{\sum_{i=1}^N \lambda_i^2}{\sum_{i=1}^N \lambda_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^N \sum_{j=1}^N \lambda_i \lambda_j} \leq \frac{\sum_{i=1}^N \lambda_i^2}{\sum_{i=1}^N \lambda_i^2 + (N^2 - N)\lambda_N} \\
&\leq \frac{1}{1 + \frac{(N-1)\lambda_N^2}{\lambda_1^2}},
\end{aligned} \tag{C.11}$$

and considering that $\lambda_1 < \infty$ we can conclude that for any positive real number ϵ

$$\lim_{N \to \infty} \frac{\text{var}(X)}{\epsilon^2} = 0, \tag{C.12}$$

which implies that

$$\lim_{N \to \infty} \Pr\left(\left| X - \mathbb{E}[X] \right| \geq \epsilon \right) = 0. \tag{C.13}$$

Thus, the proof is complete. $\qquad\square$

Lemma 5. *Let the vectors* $\mathbf{h}$, $\mathbf{g} \in \mathbb{C}^N$ *be two independent zero-mean circular Gaussian random vectors such that* $\mathbf{h}$, $\mathbf{g} \sim \mathcal{CN}(0, \mathbf{I}_N)$. *Also, let us assume that* $\mathbf{U} \in \mathbb{C}^{N \times N}$ *is a positive definite matrix with a limited spectral norm. Then, it can be shown that for* $N \to \infty$

$$\frac{\mathbf{h}^H \mathbf{U} \mathbf{g}}{\text{Tr}(\mathbf{U})} \xrightarrow{\text{a.s.}} 0. \tag{C.14}$$

Proof. The proof is in a similar manner as Lemma 4, by defining a zero-mean random variable $Y = \frac{\mathbf{h}^H \mathbf{U} \mathbf{g}}{\text{Tr}(\mathbf{U})}$ and considering that

$$\text{var}(Y) = \frac{\|\mathbf{U}\|^2}{\text{Tr}^2(\mathbf{U})}. \tag{C.15}$$

$\qquad\square$

Corollary 1. *Let* $\mathbf{H} \in \mathbb{C}^{K \times N}$ *be a Gaussian random matrix with i.i.d* $\mathcal{CN}(0, \mathbf{I}_N)$ *entries. Also let us assume that* $\mathbf{U} \in \mathbb{C}^{N \times N}$ *is a positive definite matrix with a limited spectral norm. Then, it can be shown that for* $N \to \infty$

$$\frac{\mathbf{H}^H \mathbf{U} \mathbf{H}}{\text{Tr}(\mathbf{U})} \xrightarrow{\text{a.s.}} \mathbf{I}_K. \tag{C.16}$$

Remark: Lemma 4, 5 and consequently Corollary 1 remain agnostic to semi-definite (and ill-conditioned) matrices with large number of non-zero (and non-trivial) eigenvalues. To prove this, just define the number of non-trivial and non-zero eigenvalues of matrix $\mathbf{U}$ as N'. Then, the proof is trivially followed by substituting N' instead of N in (C.11).

C.2 Proof of Proposition 5.6

Without loss of generality, we just prove the proposition for the case $i = 1$, and the same result can be similarly deduced for the case $i = 2$. Let $\mathbf{R}_1 = \mathbf{U}_{R_1} \mathbf{\Lambda}_{R_1} \mathbf{U}_{R_1}^H$

represent the eigenvalue decomposition of the channel correlation matrix, then we have

$$
\mathbf{\Lambda}_{R_1} = \left[\begin{array}{cc} \mathbf{\Lambda}'_{R_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}''_{R_1} \end{array} \right],
\tag{C.17}
$$

where $\mathbf{\Lambda}_{R_1} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ includes the eigenvalues of the correlation matrix in descending order such that $\lambda_1 \geq \ldots \geq \lambda_N$. We also note that the size of matrices $\mathbf{\Lambda}'_{R_1} \in \mathcal{C}^{K'_a \times K'_a}$ and $\mathbf{\Lambda}''_{R_1} \in \mathcal{C}^{(N-K'_a) \times (N-K'_a)}$ depends on K'_a, i.e. the number of bins which is filled by water (See Fig. 5.2). Based on the water-filling algorithm in Subsection 5.5, we readily conclude that

$$
\mathbf{U}_{e_1} = \mathbf{U}_{R_1} \left[\begin{array}{cc} \sqrt{\frac{\nu}{\tau_p P_p}}\mathbf{I}_{K_{a'}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}''_{R_1} \end{array} \right] \mathbf{U}^H_{R_1},
\tag{C.18}
$$

$$
\mathbf{U}_1 = \mathbf{R}_1 - \mathbf{U}_{e_1} = \mathbf{U}_{R_1} \left[\begin{array}{cc} \mathbf{\Lambda}'_{R_1} - \sqrt{\frac{\nu}{\tau_p P_p}}\mathbf{I}_{K_{a'}} & \mathbf{0}. \\ \mathbf{0} & \mathbf{0} \end{array} \right] \mathbf{U}^H_{R_1}.
\tag{C.19}
$$

Thus, after some simple manipulations we can complete the proof by

$$
\begin{aligned}
\mathrm{Tr}\big(\mathbf{U}_1 \mathbf{U}_{e_1}\big) &= \sqrt{\frac{\nu}{\tau_p P_p}} \mathrm{Tr}\left(\mathbf{U}_{R_1} \Big(\mathbf{\Lambda}'_{R_1} - \sqrt{\frac{\nu}{\tau_p P_p}}\mathbf{I}_{K_{a'}} \Big) \mathbf{U}^H_{R_1} \right) \\
&= \sqrt{\frac{\nu}{\tau_p P_p}} \mathrm{Tr}\big(\mathbf{U}_1\big).
\end{aligned}
\tag{C.20}
$$

Appendix D

Proofs for Chapter 6

D.1 Proof of Proposition 6.4

The optimality of this result is an advanced consequence of [112, Corollary 1] or following Telatar's methodology in [15]. A standard Gaussian random matrix, $\mathbf{H}$, is a bi-unitarily invariant matrix. It means that the joint distribution of its entries equals that of $\mathbf{UHV}^H$ for any unitary matrices $\mathbf{U}$ and $\mathbf{V}$ independent of $\mathbf{H}$. Now, by following Telatar's approach, we can limit our attention only to a diagonal $\mathbf{D}$, and diagonal $\mathbf{\Lambda Q \Lambda}^H$ in (6.16), (6.17) which implies that

$$\mathbf{R_n} = \mathbb{E}\left[\mathbf{nn}^H\right] = \mathbf{HDH}^H + N_0\mathbf{I}_{N_r}, \tag{D.1}$$

$$R = \sup_{\mathrm{Tr}(\mathbf{Q}) \leq P_t, \mathbf{Q} \succeq \mathbf{0}} \mathbb{E}\left[B \log_2\left(\det\left(\mathbf{I}_{N_r} + \mathbf{R_n}^{-1}\mathbf{H\Lambda Q \Lambda}^H\mathbf{H}^H\right)\right)\right]. \tag{D.2}$$

According to the diagonal structure of $\mathbf{\Lambda}$, we simply conclude that $\mathbf{Q}$ must be a diagonal matrix as well. Assume that $\hat{\mathbf{Q}}$ is the best power allocation, and also $\mathbf{\Pi}_i$ $i = 1, 2, ..., N_t!$, is a permutation matrix which has exactly one "1" in each row and each column and zeros elsewhere. Since $\hat{\mathbf{Q}}$ is a diagonal matrix which satisfies $\mathrm{Tr}(\hat{\mathbf{Q}}) \leq P_t$ and $\hat{\mathbf{Q}} \succeq \mathbf{0}$, it can be easily derived that $\tilde{\mathbf{Q}} = \frac{1}{N_t!}\sum_{i=1}^{N_t!}\mathbf{\Pi}_i\hat{\mathbf{Q}}\mathbf{\Pi}_i^H$ satisfies both constraints as well. Now, given the following function

$$\Psi(\mathbf{Q}) \triangleq \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_r} + \left(\mathbf{HDH}^H + N_0\mathbf{I}_{N_r}\right)^{-1}\mathbf{H\Lambda Q \Lambda}^H\mathbf{H}^H\right)\right)\right], \tag{D.3}$$

we can demonstrate that $\Psi(\tilde{\mathbf{Q}}) \geq \Psi(\hat{\mathbf{Q}})$. By starting from the left hand side, and by taking into account the concavity of the function Ψ in (d) below, we get

$$
\begin{aligned}
\Psi(\tilde{\mathbf{Q}}) = & \Psi\left(\frac{1}{N_t!} \sum_{i=1}^{N_t!} \mathbf{\Pi_i} \hat{\mathbf{Q}} \mathbf{\Pi}_i^H \right) \overset{(d)}{\geq} \frac{1}{N_t!} \sum_{i=1}^{N_t!} \Psi\left(\mathbf{\Pi_i} \hat{\mathbf{Q}} \mathbf{\Pi}_i^H \right) \\
= & \frac{1}{N_t!} \sum_{i=1}^{N_t!} \mathbb{E}\Bigg[\log_2 \Bigg(\det \Bigg(\mathbf{I}_{N_r} + \Big(\mathbf{H}\Big(\mathbf{\Pi}_i \mathbf{D} \mathbf{\Pi}_i^H\Big) \mathbf{H}^H + N_0 \mathbf{I}_{N_r} \Big)^{-1} \\
& \Big(\mathbf{H}\Big(\mathbf{\Pi}_i \mathbf{\Lambda} \mathbf{\Pi}_i^H\Big) \Big(\mathbf{\Pi}_i \hat{\mathbf{Q}} \mathbf{\Pi}_i^H\Big) \Big(\mathbf{\Pi}_i \mathbf{\Lambda}^H \mathbf{\Pi}_i^H\Big) \mathbf{H}^H \Big) \Bigg) \Bigg) \Bigg].
\end{aligned}
\tag{D.4}
$$

Note that any permutation on the input covariance matrix, leads to the same permutation on the matrices $\mathbf{D}$ and $\mathbf{\Lambda}$ as we have already done in (D.4). Considering the fact that $\mathbf{\Pi}\mathbf{\Pi}^H = \mathbf{I}$, the last equation is simplified to

$$
\Psi(\tilde{\mathbf{Q}}) \geq \frac{1}{N_t!} \sum_{i=1}^{N_t!} \mathbb{E}\Bigg[\log_2 \Bigg(\det \Big(\mathbf{I}_{N_r} + \Big(\mathbf{H}\mathbf{D}\mathbf{H}^H + N_0 \mathbf{I}_{N_r}\Big)^{-1} \mathbf{H}\mathbf{\Lambda}\hat{\mathbf{Q}}\mathbf{\Lambda}^H \mathbf{H}^H \Big) \Bigg) \Bigg] = \Psi(\hat{\mathbf{Q}}).
\tag{D.5}
$$

As we assumed $\hat{\mathbf{Q}}$ is the best power allocation, so $\Psi(\tilde{\mathbf{Q}}) \geq \Psi(\hat{\mathbf{Q}})$ would be valid only by equality. In other words, $\tilde{\mathbf{Q}}$ is the best power allocation which can be written in terms of the following matrix transformation

$$
\tilde{\mathbf{Q}} = \frac{1}{N_t!} \sum_{i=1}^{N_t!} \mathbf{\Pi}_i \hat{\mathbf{Q}} \mathbf{\Pi}_i^H = \frac{1}{N_t} \mathrm{Tr}\big(\hat{\mathbf{Q}}\big) \mathbf{I}_{N_t} \triangleq \frac{P}{N_t} \mathbf{I}_{N_t},
\tag{D.6}
$$

which is indeed a scaled identity matrix. It is obvious that P can be interpreted as the total input power.

D.2 Proof of Corollary 6.5

Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ be a SVD of the Rayleigh fading channel. By applying this decomposition, and using the fact that $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$ we derive a simpler closed-form expression for the maximum EAR

$$
\begin{aligned}
R &= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_r} + \frac{P\alpha^2}{N_t}\left(N_0\mathbf{I}_{N_r} + \sigma_d^2\mathbf{H}\mathbf{H}^H\right)^{-1}\mathbf{H}\mathbf{H}^H\right)\right)\right] \\
&= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_r} + \frac{P\alpha^2}{N_t}\left(N_0\mathbf{I}_{N_r} + \sigma_d^2\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H\right)^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H\right)\right)\right] \\
&= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_r} + \frac{P\alpha^2}{N_t}\left(\mathbf{U}\left(N_0\mathbf{I}_{N_r} + \sigma_d^2\boldsymbol{\Sigma}\right)\mathbf{U}^H\right)^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H\right)\right)\right] \\
&= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_t} + \frac{P\alpha^2}{N_t}\mathbf{U}^H\mathbf{U}\left(N_0\mathbf{I}_{N_r} + \sigma_d^2\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}\mathbf{U}^H\mathbf{U}\right)\right)\right] \\
&= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_t} + \frac{P\alpha^2}{N_t}\left(N_0\mathbf{I}_{N_r} + \sigma_d^2\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}\right)\right)\right] \\
&= \sup_{0 \leq P \leq P_t} \mathbb{E}\left[\log_2\left(\det\left(\mathbf{I}_{N_t} + \mathbf{Z}\right)\right)\right]. \tag{D.7}
\end{aligned}
$$

Appendix E

Author's Publications

- **M. Fozooni**, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation," submitted in Aug. 2017 to *IEEE Trans. Commun.*

- **M. Fozooni**, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Spectral efficiency of multipair massive MIMO relaying with offline hybrid processing," submitted in Aug. 2017 to *IEEE Wireless Commun. Lett.*

- **M. Fozooni**, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Massive MIMO relaying with hybrid processing" in *Proc. IEEE ICC*, May 2016.

- **M. Fozooni**, M. Matthaiou, E. Bjornson, and T. Q. Duong, "Performance limits of MIMO systems with nonlinear power amplifiers," in *Proc. IEEE GLOBE-COM*, Dec. 2015.

References

[1] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO- A survey," [Online]. Available: https://arxiv.org/pdf/1609.05078v1.pdf.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part i. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[4] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[5] G. J. Bradford and J. N. Laneman, "A survey of implementation efforts and experimental design for cooperative communications," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 5602–5605.

[6] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis, and X. Shen, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 156–162, Feb. 2012.

[7] J. Zhao, M. Kuhn, A. Wittneben, and G. Bauch, "Cooperative transmission schemes for decode-and-forward relaying," in *Proc. IEEE PIMRC*, Sept. 2007.

[8] S. H. Lim, Y. H. Kim, A. E. Gamal, and S. Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[9] L.-L. Xie, "An improvement of Cover/El Gamal's compress-and-forward relay scheme," [Online]. Available: http://arxiv.org/abs/0908.0163.

[10] X. Wu and L. L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, May 2013.

[11] J. Li, L. J. Cimini, J. Ge, C. Zhang, and H. Feng, "Optimal and suboptimal joint relay and antenna selection for two-way amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 980–993, Feb. 2016.

[12] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sept. 2012.

[13] J. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Trans. Commun.*, vol. 35, no. 11, pp. 1222–1230, Nov. 1987.

[14] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[15] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecom*, vol. 10, no. 6, pp. 585–595, 1999.

[16] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[17] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[18] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[19] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[20] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*.   Cambridge University Press, 2016.

[21] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[22] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.

[23] D. Gesbert, M. Kountouris, R. W. H. Jr., C. b. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[24] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. ACSSC*, Oct. 2006, pp. 359–363.

[25] ——, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[26] 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.

[27] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[28] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[29] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. EUSIPCO*, Sept. 2014, pp. 76–80.

[30] H. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE ICC*, June 2013, pp. 4635–4640.

[31] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, May 2014.

[32] Y. Dai and X. Dong, "Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5932–5946, Sept. 2016.

[33] H. Q. Ngo and E. G. Larsson, "Large-scale multipair two-way relay networks with distributed AF beamforming," *IEEE Wireless Commun. Lett.*, vol. 17, no. 12, pp. 1–4, Dec. 2013.

[34] H. Q. Ngo, H. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sept. 2014.

[35] Z. Zhang, Z. Chen, M. Shen, and B. Xia, "Spectral and energy efficiency of multipair two-way full-duplex relay systems with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 848–863, Apr. 2016.

[36] X. Xia, Y. Xu, K. Xu, D. Zhang, and W. Ma, "Full-duplex massive MIMO AF relaying with semiblind gain control," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5797–5804, July 2016.

[37] M. Mohammadi, B. K. Chalise, H. A. Suraweera, C. Zhong, G. Zheng, and I. Krikidis, "Throughput analysis and optimization of wireless-powered multiple antenna full-duplex relay systems," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1769–1785, Apr. 2016.

[38] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, Dec. 2006.

[39] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Oct. 2005.

[40] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal. Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[41] C. S. Lee and W. H. Chung, "Hybrid RF-baseband precoding for cooperative multiuser massive MIMO systems with limited RF chains," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1575–1589, Apr. 2017.

[42] S. Han, C. l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[43] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[44] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[45] J. Zhu, W. Xu, and N. Wang, "Secure massive MIMO systems with limited RF chains," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5455–5460, June 2017.

[46] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *Proc. IEEE ICC*, June 2014, pp. 5838–5842.

[47] W. Xu, J. Liu, S. Jin, and X. Dong, "Spectral and energy efficiency of multi-pair massive MIMO relay network with hybrid processing," vol. 65, no. 9, pp. 3794–3809, Sept. 2017.

[48] H. V. Cheng, D. Persson, and E. G. Larsson, "MIMO capacity under power amplifiers consumed power and per-antenna radiated power constraints," in *Proc. IEEE SPAWC*. IEEE, June 2014, pp. 179–183.

[49] J. Qi and S. Aïssa, "On the power amplifier nonlinearity in MIMO transmit beamforming systems," *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 876–887, Mar. 2012.

[50] ——, "Analysis and compensation of power amplifier nonlinearity in MIMO transmit diversity systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2921–2931, July 2010.

[51] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1755–1764, Oct. 2000.

[52] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.

[53] P. Banelli, "Theoretical analysis and performance of OFDM signals in nonlinear fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 284–293, Mar. 2003.

[54] G. K. Psaltopoulos and A. Wittneben, "Nonlinear MIMO: Affordable MIMO technology for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 824–832, Feb. 2010.

[55] F. H. Gregorio, *Analysis and compensation of nonlinear power amplifier effects in multi-antenna OFDM systems*.   Helsinki University of Technology, Nov. 2007.

[56] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*.   Springer, 2008.

[57] M. Fozooni, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Massive MIMO relaying with hybrid processing," in *Proc. IEEE ICC*, May 2016.

[58] M. Fozooni, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Spectral efficiency of multipair massive MIMO relaying with offline hybrid processing," *IEEE Wireless Commun. Lett., under review.*

[59] ——, "Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation," *IEEE Trans. Commun., under review.*

[60] M. Fozooni, M. Matthaiou, E. Bjornson, and T. Q. Duong, "Performance limits of MIMO systems with nonlinear power amplifiers," in *Proc. IEEE GLOBE-COM*, Dec. 2015.

[61] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[62] A. Goldsmith, *Wireless Communications*.   Cambridge university press, 2005.

[63] G. L. Stüber, *Principles of Mobile Communication*.   Springer, 2001, vol. 2.

[64] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, Feb. 1998.

[65] J. Choi and D. J. Love, "Bounds on eigenvalues of a spatial correlation matrix," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1391–1394, Aug. 2014.

[66] H. Lim, Y. Jang, and D. Yoon, "Bounds for eigenvalues of spatial correlation matrices with the exponential model in MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1196–1204, Feb. 2017.

[67] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[68] H. Bolcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.

[69] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[70] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[71] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[72] S. Jin, X. Liang, K.-K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.

[73] J. Cao, Z. Zhong, and F. Wang, "Regenerative multi-way relaying: Relay precoding and ordered MMSE-SIC receiver," in *Proc. IEEE VTC*, May 2012.

[74] K. Phan, T. Le-Ngoc, S. Vorobyov, and C. Tellambura, "Power allocation in wireless multi-user relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2535–2545, May 2009.

[75] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[76] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[77] W. Zhang, X. Ma, B. Gestner, and D. V. Anderson, "Designing low-complexity equalizers for wireless systems," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 56–62, Jan. 2009.

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

[79] C. Kong, C. Zhong, M. Matthaiou, E. Björnson, and Z. Zhang, "Multi-pair two-way AF relaying systems with massive arrays and imperfect CSI," in *Proc. IEEE ICASSP*, Mar. 2016, pp. 3651–3655.

[80] 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.

[81] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[82] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," [Online]. Available: http://arxiv.org/pdf/1504.03777v1.pdf.

[83] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[84] D. Ying, F. W. Vook, T. A. Thomas, and D. J. Love, "Hybrid structure in massive MIMO: Achieving large sum rate with fewer RF chains," in *Proc. IEEE ICC*, June 2015, pp. 2344–2349.

[85] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. ITA*, San Diego, USA, Feb. 2013.

[86] J. C. Roh and B. D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[87] A. Hajimiri, H. Hashemi, A. Natarajan, X. Guan, and A. Komijani, "Integrated phased array systems in silicon," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1637–1655, Aug. 2005.

[88] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sept. 2014.

[89] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3679–3684.

[90] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[91] J. J. Sylvester, "Thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers," *Philos. Mag.*, vol. 34, no. 232, pp. 461–475, 1867.

[92] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[93] D. Park and S. Y. Park, "Effect of transmit antenna correlation on multiuser diversity," in *Proc. IEEE ISIT*, Sept. 2005, pp. 1421–1425.

[94] G. C. Alexandropoulos and M. Kountouris, "Maximal ratio transmission in wireless Poisson networks under spatially correlated fading channels," in *Proc. IEEE GLOBECOM*, Dec. 2015.

[95] H. Kim, W. Choi, and H. Park, "Effects of antenna correlation on spatial diversity and multiuser diversity," in *Proc. IEEE WCNC*, Mar. 2008, pp. 65–69.

[96] H. Yang and T. L. Marzetta, "A macro cellular wireless network with uniformly high user throughputs," in *Proc. IEEE VTC Fall*. IEEE, Sept. 2014.

[97] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, June 2015.

[98] E. Björnson, L. Sanguinetti, and M. Kountouris, "Energy-efficient future wireless networks: A marriage between massive MIMO and small cells," in *Proc. IEEE SPAWC*. IEEE, June 2015, pp. 211–215.

[99] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[100] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, Sept. 2004.

[101] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, Dec. 2014.

[102] 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.

[103] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array system for spatially selective transmission," in *Proc. IEEE VTC*, June 1994, pp. 1517–1521.

[104] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input multiple-output power allocation," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1112–1115, June 2013.

[105] ——, "Amplifier-aware multiple-input single-output capacity," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 913–919, Mar. 2014.

[106] S. C. Cripps, *Advanced Techniques in RF Power Amplifier Design*. Artech House, 2002.

[107] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," 1952.

[108] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 2007.

[109] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[110] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of Wishart matrices," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1050–1060, Apr. 2009.

[111] X. Zhang, M. Matthaiou, E. Björnson, M. Coldrey, and M. Debbah, "On the MIMO capacity with residual transceiver hardware impairments," in *Proc. IEEE ICC*, June 2014, pp. 5299–5305.

[112] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 91–94, Jan. 2013.

[113] A. J. Paulraj, D. Gore, R. U. Nabar, and H. Bölcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[114] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Communications and Information theory*, vol. 1, no. 1, pp. 1–182, 2004.

[115] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting signal," in *European Conf. Satellite Commun.*, vol. 1, Oct. 1991, pp. 179–184.

[116] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.