Performance of Massive MIMO with Interference Decoding

by

Meysam Shahrbaf Motlagh

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Raviraj Adve Professor, Dept. of Electrical and Computer Engineering, University of Toronto, Canada
Supervisor:	Dr. Patrick Mitran Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal Members:	Dr. Ravi R. Mazumdar Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
	Dr. Guang Gong Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal-External Member:	Dr. Henry Wolkowicz Professor, Dept. of Combinatorics and Optimization, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In a massive MIMO system, base stations (BS) utilize a large number of antennas to simultaneously serve several (single or multi-antenna) users at once, where the number of BS antennas is normally assumed to be significantly larger than the number of users. In massive MIMO systems operating in time division duplex (TDD) mode, the channel state information (CSI) is estimated via uplink pilot sequences that are orthogonal in a cell but re-used in other cells. Re-using the pilots, however, contaminates the CSI estimate at BSs by the channel of the users sharing the same pilot in other cells; thus causing pilot contamination which creates coherent interference that, as the number of BS antennas grows, scales at the same rate as the desired signal. Hence, in the asymptotic limits of large antennas, the effects of non-coherent interference terms and noise disappear, except for the pilot contamination interference. A common technique used in the literature to deal with this interference is to treat it as noise (TIN). When using TIN, users' throughput will converge to a constant and thus the benefits of using an ever greater number of BS antennas saturate. However, it is known that the use of TIN in interference networks is only preferred in the weak interference regime, and it is sub-optimal in other regimes (e.g., moderate or strong interference). In this thesis, we show that as the number of BS antennas increases, the pilot contamination interference is no longer weak, and therefore it is beneficial to treat it differently (e.g., decode it jointly with the desired signal) to improve users' throughput.

In the first part of the thesis, we study the performance of interference decoding schemes based on simultaneous unique decoding (SD) and simultaneous non-unique decoding (SND), and show that by doing so the rate saturation effect is eliminated as the number of antennas increases; hence, the per-user rates grow unbounded. We analytically study the performance of two well-known linear combining/precoding methods, namely, MRC/MRT and ZF, for spatially correlated/uncorrelated Rayleigh fading channel models, and obtain closed-form expressions of rate lower bounds for these using a worst-case uncorrelated noise technique for multi-user channels. We compare the performance of the different interference management schemes, TIN/SD/SND, based on the maximum symmetric rate they can offer to the users. Specifically, we first obtain structural results for a symmetric two-cell setting as well as the high SINR regime, that provide insights into the

benefits of using interference decoding schemes in different regimes of number of BS antennas. We numerically illustrate the performance of the different schemes and show that with a practical number of antennas, SND strictly outperforms TIN. This gain improves with increasing the number of antennas, and also ZF performs significantly better than MRC/MRT due to better mitigation of multi-user interference. Furthermore, we study the performance of regularized ZF (RZF) via Monte Carlo simulations, and observe that it achieves better rates than ZF for moderately small number of antennas only. Lastly, we numerically investigate the impact of increasing the number of cells, the cell radius, the number of users, the correlation of the channel across antennas and the degree of shadow fading on system performance.

In the second part of the thesis, we study the performance of partial interference decoding based on rate splitting (RS) and non-unique decoding. Specifically, we propose to partition each user's message into two independent layers, and partially decode the pilot contamination interference while treating the remaining part as noise based on a power splitting strategy. In particular, for a two-cell system, we investigate the benefits of an RS scheme based on the celebrated Han-Kobayashi (HK) region, which provides the best known achievable performance for a two-user interference channel (IC). In the case of more than two cells, we propose a generalized RS scheme that non-uniquely decodes each layer of the pilot contamination interference and uses only one power splitting coefficient per IC. In addition, we establish an achievable region for this generalized RS scheme using the non-unique decoding technique. In both cases of two cells and more than two cells and for a practical number of antennas, we numerically study the performance of the proposed RS schemes by numerically optimizing the power splitting coefficients, and show that they achieve significantly higher rates than TIN/SD/SND in all scenarios. Similar to the first part of the thesis, we also numerically examine the impact of increasing the number of cells, the cell radius, the number of users, the correlation of the channel across antennas and the degree of shadow fading on the performance of the RS schemes. Lastly, our simulation results reveal that by replacing the numerically optimized values of the power splitting coefficients with their pre-computed average values (over a large number of realizations), the performance loss is quite negligible, thus reducing the optimization complexity.

Acknowledgements

First, foremost and solely, I thank God, the most merciful, the all-knowing and the almighty, for everything.

I wish to extend my sincere gratitude to all the kind people who helped me throughout my journey as a PhD student. I would like to sincerely thank my supervisor, Prof. Patrick Mitran, for his priceless help and support. Prof. Mitran supported me intellectually and financially, challenged me to unleash my potentials by his insightful observations, helpful comments and undivided attention during countless meetings. I would like to extend my appreciation to all members of my PhD defense committee, Prof. Ravi R. Mazumdar, Prof. Guang Gong, and Prof. Henry Wolkowicz. Furthermore, I would like to thank Prof. Raviraj Adve for accepting to serve as the external examiner for my thesis defense.

Time would fail me to tell how I am indebted to my parents, Akram and Mohammad Bagher, who are simply the best parents one could dream of. May God endow them with best of rewards, for their patience, endurance, and never ending love. My heartfelt thanks go to my lovely wife, Fatemeh, who made my stay in the past two and a half years the most pleasant and memorable days in Canada, through which we shared moments of delights, happiness and grieves together. The last 16 months have been tough and challenging; and only God knows how I would have been able to keep up without Fatemeh. I would also express my deepest regards to my sisters and brother Maryam, Mahboubeh and Mohammad Masoud, for filling my empty space in the family, and apologize them for not being with them for this long.

Moreover, I would like to thank all the professors at the University of Waterloo, from whom I have learned a lot during various graduate courses I took along this journey. I also appreciate the time I spent and the fruitful discussions I had with my friends at the University of Waterloo: Subhajit Majhi, Ehsan Alian, Ahmed Wagdy Shaban and Kazem Izadinasab.

Dedication

This is dedicated to my wonderful parents, Akram and Mohammad Bagher, and to my lovely wife, Fatemeh.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 5G Cellular Systems

With the ever increasing demand for significantly higher data rates in cellular communication systems as well as the development of new services such as Device-to-Device (D2D), Internet of Things (IoT), High Speed Train (HST) communications, e-banking, e-health and e-learning, the need for new technologies that have a great potential to provide higher cellular system capacity is inevitable. In particular, it is predicted that by 2023 approximately two-thirds of the world population will have access to the Internet, i.e., about 5.3 billion Internet users [3]. In 2018, this number was about 3.9 billion (i.e., 51 percent of the world population) [3]. According to [3], it is also anticipated that there will be a total of 29.3 billion connected devices by 2023 (more than three times the world population). The explosive increase of mobile data traffic is real, driven primarily due to the increased use of smart phones, tablets, video streaming services as well as machine-to-machine (M2M) connections. Hence, the need for a new generation of highly scalable cellular networks is inescapable. In particular, this new generation should:

- have a highly scalable and flexible architecture to support various services and applications, such as massive device connectivity in IoT and M2M communications;

- provide larger capacity and improved coverage, while reducing the complexity and cost of ultra dense network deployment;
- be able to make efficient use of different spectrum resources, including both licensed and unlicensed bands as well as high frequency and low frequency bands;
- improve network energy efficiency to adapt to the performance requirements of different applications and services;
- reduce the complexity incurred due to the co-existence of multi radio access technologies (multi-RATs), and improve quality of users' experience.

To meet the above requirements and to cope with the challenges of ever increasing mobile data traffic as well as demands for much higher data throughput, the fifth generation of cellular networks, also known as 5G new radio or 5G NR, is now becoming a reality. In 2018, the first full set of 5G NR standards were announced by 3rd Generation Partnership Project (3GPP) in release 15 (5G NR phase 1) [4], followed by release 16 in 2020 which drove 5G NR phase 2 expansion [5]. Currently, 5G NR networks are being deployed globally, and many 5G-ready smart devices (e.g., cell phones and tablets) are starting to come to market. Unsurprisingly, it is also predicted that approximately 1.4 billion worldwide smart devices will be 5G-ready by 2023, with a speed that will be 13 times faster (i.e., 575 Mbps) than the current average mobile connection speed [3].

The International Telecommunication Union (ITU) has defined the following three requirements for 5G NR, each to be fulfilled for one of the three 5G usage scenarios [6]

- peak data rate of 10-20 Gbps (required for the enhanced mobile broadband, eMBB [7]);
- 1 million connected devices per square kilometer (required for massive machine type communication, mMTC [8]);
- less than 1 ms latency (required for ultra reliable low latency communications, URLLC [9]).

URLLC renders time-sensitive communications possible for applications such as autonomous driving [10] or remote medical surgeries [11], whereas mMTC defines connectivity between a massive number of IoT devices [12]. eMBB is undoubtedly the usage scenario that wireless communication companies have been mainly waiting for since the start of 5G development.

User throughput is defined as the amount of information bits that can be successfully transmitted to a destination per unit time, i.e.,

$$\text{Throughput } [bit/sec] \;=\; \text{Bandwidth } [Hz] \;\times\; \text{Spectral efficiency } [bit/sec/Hz]. \qquad (1.1)$$

Therefore, throughput can be improved by either increasing the bandwidth, the spectral efficiency or both. While increasing bandwidth seems the easiest way of increasing throughput, it is costly as the frequency spectrum is a limited resource shared by many wireless technologies and applications, especially in sub-6 GHz bands. Nevertheless, making good use of the vast spectrum available in mmWave bands (30-300 GHz) is considered an interesting option for 5G NR. However, one should bear in mind that due to the large attenuation as well as the high blockage sensitivity in higher frequencies, transmission in mmWave bands can typically only be carried out for short range communications or line of sight (LoS). Increasing spectral efficiency, which is the focus of this thesis, may be achieved via multiple transmit/receive antennas at both the base station (BS) and the mobile terminal.

The following key technology components are currently being considered attractive solutions to achieve improved user throughput in 5G cellular networks:

- **Multi-node/Multi-antenna transmission:** development of advanced inter-node coordination, relaying and multi-hop techniques as well as the study of innovative transmission/reception schemes enabled by massive multi-antenna BSs [13].
- **Heterogeneous multi-RAT and multi-layer networks:** development of novel and proactive demand, interference and mobility management techniques that are adapted to the co-existence of multi-RATs, cells of different sizes and heterogeneous deployments [14].

- **Spectrum usage**: development of advanced spectrum sharing techniques as well as investigation of new spectrum resources, e.g., making efficient use of mmWave bands [15].

With respect to the multi-node multi-antenna transmission, massive multi-input multi-output (MIMO) technology is currently under active research investigations with the promise of significantly improving data rates, spectral efficiency, energy efficiency and coverage [13].

1.2 Brief Review of multi-user MIMO Technology

In multi-user MIMO (MU-MIMO), the BS can simultaneously transmit to several single/multi antenna terminals spatially sharing the same channel, and thus large improvements in terms of spectral efficiency can be achieved [16]. More specifically, BS antennas will be used to direct a signal towards each of the desired terminals in the downlink, and to separate signals received in the uplink. However, prior knowledge of the channel state is essential in MU-MIMO, as effective signal processing techniques that are adapted to the instantaneous channel state information (CSI) are required to eliminate the interference. Over the past decades, an enormous number of papers have been published in the area of MU-MIMO, among which are the pioneering works of [17–20] on array processing, and [21–25] on characterizing achievable multi-user capacity assuming known CSI.

The main advantages brought by MU-MIMO systems are:

- improved coverage resulting from increased received signal power via beamforming;
- improved link reliability resulting from diversity schemes that reduce the effects of fading;
- larger throughput resulting from spatial multiplexing of several data streams over the same time-frequency resource;
- reduced delay dispersion resulting from channel shortening effects via beamforming.

In theory, when a transmitter/receiver is equipped with more and more antennas, MU-MIMO technology has a greater potential to enlarge the scale on which the spatial domain is utilized. As a result, with more antennas one can expect to achieve better performance in terms of the MIMO advantages mentioned above.

1.3 Massive MIMO

In a massive MIMO communications system, each BS utilizes a large number of antennas, which allows for the simultaneous serving of several (single or multi-antenna) users over the same time-frequency resource, where the number of BS antennas is normally assumed to be significantly larger than the number of users.

The introduction of massive MIMO technology dates back to the seminal work of Marzetta in [26]. Therein, it was shown that when the number of BS antennas grows without limit, due to the channel hardening and favorable propagation, the effects of small-scale fading, additive noise and non-coherent interference asymptotically disappear. The only remaining impediment is the inter-cell interference that results from users in other cells utilizing the same pilot sequences for channel estimation. This effect will be discussed in detail in the next chapter of the thesis. Even though the striking results reported by Marzetta rely heavily on his choice of propagation and system model, Marzetta opened the gate to an important path in which future cellular networks may significantly evolve.

The key advantages brought by massive MIMO systems are:

- **Multiplexing gain:** It is theoretically possible to increase the capacity by 10x using aggressive spatial multiplexing [27].
- **Spectral efficiency:** Using a massive number of BS antennas to simultaneously multiplex several data streams to many users was shown to significantly improve spectral efficiency [28, 29].
- **Energy efficiency:** Analytically, it was shown that the uplink transmit power of

each user can be scaled inversely proportional to the number of BS antennas, without any performance loss [28, 30].

- **Improved link reliability:** By increasing the number of BS antennas, diversity gains increase and the effects of additive noise and intra-cell interference all vanish; hence improving link reliability and data rate [26, 31].

- **Simple linear processing:** It has been shown that linear matched filtering is optimal when the number of BS antennas is much larger than the number of user antennas [26, 32].

- **Low cost RF power components:** Reduced energy consumption due to the large number of antennas makes the use of low cost milli-Watt RF amplifiers practical [31].

Since the introduction of the massive MIMO concept in 2010, many theoretical and experimental studies have been carried out to understand its benefits as well as limitations and challenges [32–42]. Additionally, in response to the concerns regarding the theoretical channel models used in massive MIMO literature, it was reported in [43] that based on channel measurements for a large number of BS antennas, achieved array and multiplexing gains are relatively close to the ones provided by theory.

Furthermore, implementation and demonstration of several real-time massive MIMO testbeds have been reported in [44–48], and today massive MIMO is considered a key integral part of 5G. One of the very first massive MIMO products was the AIR 6468, built by Ericsson in 2017, which uses 64 antennas in both uplink and downlink operating in sub-6 GHz band [47]. It should also be pointed out that Ericsson developed this product for 4G LTE-A, and it is thus considered a pre-5G product. In 2018, the first line of massive MIMO products for 5G NR was approved by the federal communications commission (FCC), among which was the Ericsson AIR 6468. Since then, many other massive MIMO products, including the Huawei AAU and Nokia Airscale, have been developed for massive MIMO BSs equipped with 128 antennas. This evolution has made massive MIMO technology, operating with a large but finite number of antennas in the conventional sub-6 GHz band, a reality. Moreover, many communication carriers (e.g., the massive MIMO deployment

by the US-based carrier Sprint [13]) has been using the term "massive MIMO" in their marketing and advertisements. In addition, during the Mobile World Congress (MWC) in 2019, Huawei announced that 95 percent of their current commercial products will support either 32 or 64 antennas [49]. More recently, a new research direction has been introduced in the massive MIMO literature, known as extremely large aperture arrays (ELAA), allowing new designs of massive antenna arrays that can support thousands of antennas [13].

On the other hand, at higher carrier frequencies such as mmWave bands, massive MIMO can be leveraged to reduce the effects of high propagation loss [50]. In recent years, there has been a significant progress in the development of massive MIMO arrays for communications in mmWave bands (i.e., beyond 30 GHz) [15, 51–54].

1.4 Some of the Challenges associated with Massive MIMO

Although, massive MIMO communication systems offer many advantages and performance gains, there still exist several challenges that must be taken into account. Below we discuss some of these challenges in more details.

1.4.1 Frequency Division Duplexing (FDD)

To fully benefit from the capacity gains brought by massive MIMO systems, the BS requires instantaneous CSI. It is, however, known that as the number of BS antennas increases the overhead incurred from obtaining CSI using feedback in FDD mode also increases. On the other hand, by exploiting the channel reciprocity assumption in TDD mode, the overhead becomes independent of the number of BS antennas [31]. As a result, most research in the context of massive MIMO are focused on TDD systems that induce much less CSI overhead compared to FDD systems.

Due to its efficient utilization of radio resources, the use of TDD mode has been suggested as the "canonical form" of massive MIMO [13]. Nevertheless, the use of FDD mode

is also widely studied in the literature [55–58]. Thus, the possibility of using FDD mode in massive MIMO systems is still under investigation [59–63].

1.4.2 Pilot Contamination

It is evident that the spectral efficiency gains and achievable throughput in MIMO systems rely heavily on the type of linear precoding and combining techniques used at the BS on the downlink and uplink, respectively. As mentioned earlier, however, the performance of linear processing at the BS is itself directly affected by the accuracy of CSI estimation. Therefore, the performance gains offered by massive MIMO systems depend on how accurately CSI is obtained at the BS using the transmission of uplink pilots in TDD mode. In addition, it should be noted that the number of pilot sequences that can be used within each coherence interval is limited, as the length of the coherence interval is finite due to the movement of mobile users. Hence, a finite number of orthogonal pilot sequences may be re-used across the cells, e.g., the k^{th} pilot sequence is assigned to the k^{th} user of all cells [38, 64, 65]. Consequently, the channel estimation of an arbitrary user at the BS will be contaminated by the channel of users in other cells using the same pilot sequence. This phenomenon, known as pilot contamination in TDD mode, will inevitably cause inter-cell interference that does not vanish even when the number of BS antennas grows unboundedly, and thus saturates the system throughput [28]. As a result, pilot contamination is deemed a major challenge that limits the performance of multi-cell massive MIMO systems.

1.5 Motivation and Overview of Contributions

Pilot contamination is a source of interference that precludes the unbounded growth of users' rate with the number of BS antennas. Although, many novel techniques based on treating interference as noise (TIN) have been proposed in the literature to eliminate or alleviate the effects of pilot contamination interference [38–40, 65–70], it is still an active area of research under investigation. In addition, TIN is known to be a sub-optimal decoding strategy in some scenarios [71]. Therefore, a natural question to ask is whether

one can propose alternative solutions that treat pilot contamination interference differently, and can achieve, in principle, unbounded throughput when the number of antennas is sufficiently large?

In this thesis, we aim to answer the above question by replacing TIN with more advanced interference-aware schemes based on *full/partial* interference decoding, and it is observed that by doing so the rate saturation effects are eliminated. We summarize the major contributions of this thesis as follows:

- In Chapter 3, we study the performance of full interference decoding techniques. Specifically, we first intuitively show that using the capacity region obtained by simultaneous unique decoding (SD) of the desired signal and pilot contamination interference (as opposed to TIN), when linear combining/precoding techniques are applied in uplink/downlink, noise-free channels are obtained as $M \to \infty$, and thus the per-user rates tend to infinity. We also establish a worst-case uncorrelated noise technique for multiple access channels (MAC) that yields new expressions of achievable rate lower bounds. Using this worst-case uncorrelated noise technique as well as the derivations of the minimum mean squared error (MMSE) channel estimate for a spatially correlated channel model, we derive new achievable lower bounds for the uplink and downlink of a multi-cell massive MIMO system that applies joint decoding to each set of pilot sharing users across all cells. These lower bounds are valid in general, do not depend on specific linear combining/precoding techniques, and help to evaluate the performance of combining/precoding techniques such as regularized zero forcing (RZF) via Monte Carlo simulation. We then specialize these lower bounds to the cases of maximum ratio combining/transmission (MRC/MRT), and obtain closed-form expressions for both uplink and downlink. Furthermore, for the cases of MRC/MRT as well as zero forcing (ZF), we simplify these lower bounds assuming that a spatially uncorrelated channel model is used.

- In Chapter 3, it is also shown that when decoding interference due to pilot contamination, reusing the same orthogonal pilots across all cells (as opposed to using different rotated versions of pilots) is preferable as it requires decoding significantly

fewer interference terms. In addition, the benefits of using simultaneous non-unique decoding (SND) [72] is investigated in Section 3.2.6, which strictly contains regions SD and TIN and thus outperforms these schemes. Moreover, a simplified subset of SND (S-SND) is studied in Chapter 3, which is shown to be strictly larger than SD and also provides a lower bound to SND.

- To evaluate the performance of the different schemes TIN/SD/SND, the problem of maximum symmetric rate allocation (i.e., maximizing the minimum achievable rate) is investigated in Chapter 3. The full interference decoding schemes of this chapter only require a mechanism to find a rate vector (e.g., the maximum symmetric rate vector) inside the achievable region. This can be done, for example, using rate adaptation with feedback [73]. Another approach would be to make the average effective channel gains for each set of pilot-sharing users available at the BSs. These gains vary slowly (i.e., stay constant over many channel coherence intervals), and can thus be estimated and tracked efficiently. In Chapter 3, it is shown that for the special case of an uncorrelated Rayleigh fading channel, these gains are a function of large-scale fading coefficients only, for which it has been argued previously that these can be estimated and tracked efficiently [38, 64, 74].

- In Chapter 3, structural results are also presented for the high SINR regime. In particular, it is found that when the number of BS antennas M is truly large, the full interference decoding schemes SD/SND achieve the same performance and also strictly outperform TIN. For the special case of a two-cell system and assuming a symmetric geometry, it is shown that for relatively small values of M, pilot contamination interference is "weak" in that SND and TIN achieve the same rate and both of these strictly outperform SD. Hence, one may choose TIN which is simpler to implement. Nevertheless, for large values of M, pilot contamination interference becomes "strong" so that the full interference decoding schemes SD/SND provide the same performance and both of these strictly outperform TIN. Analytical conditions in terms of mutual information expressions under which these results hold are also found. We also numerically study the performance of networks with three, four and seven cells, and show that with a practical number of antennas M, SND

outperforms *all other schemes*, and further show that RZF and ZF achieve higher rates than MRC/MRT, while RZF also outperforms ZF for small values of M (e.g., $M < 64$) only.

- Moreover, in Chapter 3, we numerically study the impact of increasing the number of users, the number of cells, cell radius, correlation magnitude as well as shadow fading on the performance of the proposed schemes. In particular, it is found that while increasing the number of users, cell radius and shadow fading degrade the performance of all schemes, they lead to improving the gain provided by SND over other schemes. Also, increasing the correlation magnitude gives rise to improving the performance of both TIN and SND, and thus results in reducing the gain provided by SND. Therefore, the maximum gain is obtained in scenarios with a spatially uncorrelated channel model. Nevertheless, for the case of moderate spatial correlation, it is observed that SND still provides a significant gain over the other schemes. Lastly, it is found that increasing the number of cells improves the gain offered by SND over the other schemes; hence showing the importance of the proposed scheme in practical implementations.

- In Chapter 4, we study the performance of partial interference decoding schemes for the cases of two cells and more than two cells. In particular, we observe that decoding part of the pilot contamination interference while treating the remaining part as noise (according to a power splitting strategy that is optimized), yields performance that is considerably better than the schemes of Chapter 3. Specifically, for the case of a two-cell system, we study the performance of a rate splitting (RS) technique based on the celebrated Han-Kobayashi (HK) [75] scheme by numerically optimizing the power splitting coefficients, and show that this scheme strictly outperforms SND for a practical range of M, while providing a significant gain that improves with increasing the number of antennas M.

- An extension of the RS scheme to the case of more than two cells (i.e., three, four and seven cells) is also proposed in Chapter 4, where the corresponding achievable rate region based on non-unique decoding is derived. Since the rates of individual

layers for each set of pilot-sharing users needs to be adjusted globally across the entire network, the partial interference decoding schemes of this chapter are implemented in a centralized manner. Specifically, in Chapter 4, it is shown that by centrally optimizing the rates of all individual layers (e.g., with the help of a centralized network controller, where the mean of the effective channel gains for each set of pilot-sharing users are available), for a practical number of BS antennas one can achieve higher rates compared to full interference decoding schemes.

- In Chapter 4, for the case of three cells the true performance of the RS scheme is studied, while for the cases of four and seven cells an achievable sub-region of the proposed RS scheme is investigated. By numerically optimizing the power splitting coefficients, it is shown than in all scenarios the proposed RS scheme provides a significant gain over the schemes of Chapter 3, and this gain improves by increasing the number of antennas M. It is also observed that the impacts of increasing the number of users, the number of cells, cell radius, correlation magnitude as well as shadow fading on the performance of the proposed RS scheme are similar to those observed in Chapter 3. Lastly, the results of Chapter 4 reveal that by replacing the optimized values of the power splitting coefficients, found via exhaustive search, with their precomputed average values (over a large number of realizations), the performance loss is quite negligible. This means that in a practical implementation of a multi-cell massive MIMO system performing the proposed RS scheme, the optimization complexity can be reduced.

1.6 Organization of the thesis

The rest of this thesis is organized as follows.

- In Chapter 2, we present an extensive survey of various pilot contamination mitigation/reduction techniques proposed in the literature. We further classify different mitigation techniques based on the rational behind them and provide a comprehensive discussion of their advantages and limitations. Next, in order to provide some

background in connection to our work, we present the multi-cell massive MU-MIMO system model with imperfect CSI, and discuss some well-known existing results on uplink/downlink achievable throughput under pilot contamination effect. We further provide the derivations of TDD-based CSI estimation for both cases of a spatially correlated channel and an uncorrelated channel.

- In Chapter 3, the performance of interference decoding schemes based on *unique/non-unique* decoding of pilot contamination users is studied. Closed-form lower bounds for both uplink/downlink of a multi-cell massive MIMO system are also established, and the maximum symmetric rates of different schemes are evaluated via comprehensive simulation results.

- In Chapter 4, the performance of *partial* interference decoding schemes based on RS is investigated, where extensive simulation results are also provided for various scenarios.

- In Chapter 5, conclusions derived from the thesis are provided and several potential directions for future studies are also outlined.

1.7 Notation

We use the following notations in the thesis.

We use boldface upper and lower case symbols to represent matrices and vectors, respectively. The $N \times N$ identity matrix and the all-zero vector are denoted by $\boldsymbol{I}_N$ and $\boldsymbol{0}$, respectively.

The superscripts $(.)^T$, $(.)^\dagger$, and $(.)^{-1}$ denote the transpose, Hermitian transpose, and inverse operations. A diagonal matrix with vector $\boldsymbol{a}$ along its main diagonal is represented by $\text{diag}(\boldsymbol{a})$, and $\text{tr}(\boldsymbol{A})$ denotes the trace of a square matrix $\boldsymbol{A}$.

The expressions $\mathbb{E}[.]$ and $\text{var}[.]$ are used to denote the mean and variance of a random variable, respectively, and $\mathcal{CN}(\boldsymbol{m}, \boldsymbol{R})$ denotes the circular symmetric complex Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{R}$.

We use $h(X)$ to denote the differential entropy of a random variable X. Given the probability density function $f(x)$, it is defined by

$$h(X) := -\int f(x) \log f(x)\, dx.$$

The notation $h(X \mid Y)$ represents the conditional differential entropy of X given Y. Assuming that the joint and conditional probability density functions are given by $f(x, y)$ and $f(x \mid y)$, respectively, it is defined by

$$h(X \mid Y) := -\int\int f(x, y) \log f(x \mid y)\, dxdy.$$

The notation $I(X; Y)$ denotes the mutual information between X and Y defined by

$$I(X; Y) := h(X) - h(X \mid Y) = h(Y) - h(Y \mid X).$$

Chapter 2

Literature Survey and Background

2.1 Channel Estimation and Pilot Contamination

The focus of this thesis is on massive MIMO systems operating in TDD mode. In this chapter, we review various proposed techniques to mitigate or reduce the effects of pilot contamination in multi-cell massive MIMO systems, followed by some background and well-known results in the massive MIMO literature.

The availability of accurate channel state information (CSI) is deemed critical in wireless communication networks, and is considered a key component of massive MIMO communication systems [76]. In particular, the performance of MIMO systems relies heavily on the knowledge of CSI, which has been widely investigated in the literature [77–79]. In order to acquire CSI, several techniques have been proposed, among which the use of training pilots [80], semi-blind techniques [81] or blind techniques [82] are the most common approaches.

When uplink and downlink transmissions are separated in frequency, as in the FDD protocol, the uplink and downlink channels will be different. Hence, to estimate the downlink channels the BS needs to send downlink pilots to the users. After the estimation of the downlink channels, users send the channel estimates back to the BSs through a feedback link creating a backhaul overhead that grows with M, and that is significant when M is

large. On the other hand, with a TDD protocol, the uplink and downlink channels are the same due to channel reciprocity. This suggests the attractive solution of using one pilot sequence per user to estimate the uplink channel between the user and all antennas of the BS. As the uplink channels are the same as the downlink ones, the estimate of uplink channels can be used at the BSs to compute the precoding vectors during the downlink data transmission phase. Therefore, the significant overhead in FDD incurred due to feeding back the channel estimates from the users to the BSs is no longer an issue.

Elimination of inter-cell interference in multi-cell communication systems, under the assumption that full CSI is available at the BS, has been widely studied in the literature. For instance, the works of [83–85] have proposed coordinated beamforming to mitigate inter-cell interference in multi-cell multi-antenna systems assuming that the BS has full CSI. In practical implementations, however, CSI must be estimated. Particularly, in TDD based massive MIMO systems the channel estimate of an arbitrary user will be contaminated by the channel of users in other cells that are using the same pilot sequence. This phenomenon creates coherent inter-cell interference whose power grows at the same rate as that of the desired signal [86]. This results in the saturation of the achievable rates of users [87], which also adversely impacts spectral efficiency. This effect will be discussed in subsection 2.3.7 in more detail.

With respect to imperfect CSI, several approaches in the literature have been proposed to mitigate or reduce the effects of inter-cell interference caused by pilot contamination. Although sharing orthogonal pilots across different cells during the channel estimation phase is considered the main cause of pilot contamination, especially in the context of massive MIMO systems, pilot contamination can also be caused by hardware impairments. In a practical implementation, for instance, hardware impairments result in in-band and out-of-band distortions that interfere with pilots, and have been recognized as another source of pilot contamination in regular multi-antenna cellular systems [88] as well as massive MIMO communication systems [89, 90].

In Section 2.2 below, we present an extensive survey of proposed solutions to mitigate or reduce the effects of pilot contamination interference in multi-antenna communication systems, and more specifically in massive MIMO systems. Then, in Section 2.3, we present

some background followed by well-known results in massive MIMO literature.

2.2 Proposed Methods to Tackle Pilot Contamination

Based on the approach taken for channel estimation, we classify the proposed methods of coping with pilot contamination into two groups, i.e., pilot-based estimation approaches and data-aided estimation approaches. In the former, channel vectors are estimated using orthogonal training pilots, while the same set (or a rotated version thereof) of pilot sequences is shared either arbitrarily or in a smart manner among users across different cells. In contrast, in the data-aided approach, the users' channels are estimated with no or only a few pilot signals. Below, we discuss these two approaches and summarize the related works accordingly.

2.2.1 Pilot-Based Approaches

In [1], a time-shifting strategy during the pilot transmission phase is proposed that aims to mitigate inter-cell interference caused by pilot contamination. More specifically, the location of pilot signals is shifted in time frames within adjacent cells in such a way that the transmission of pilots is done at non-overlapping times (see Fig. 2.1). In addition, in [91] the performance of the time-shifting method of [1] was improved by including a power control algorithm. Even though this approach seems very attractive, a big challenge in practical implementations is how to adaptively synchronize the pilot transmissions over the entire network so that there is no overlap at all. As pointed out in [92], however, with the deployment of dynamic small cells over existing multi-layer HetNets, overlap in time and frequency seems inevitable. In [69,93], a channel estimation approach using the covariance matrix of the desired user in conjunction with the interfering users has been proposed. Under the condition that the covariance matrices of interfering users span a different subspace than that of the desired user, the authors have shown that when the number of BS antennas grows unbounded, the effects of pilot contamination asymptotically disappear.

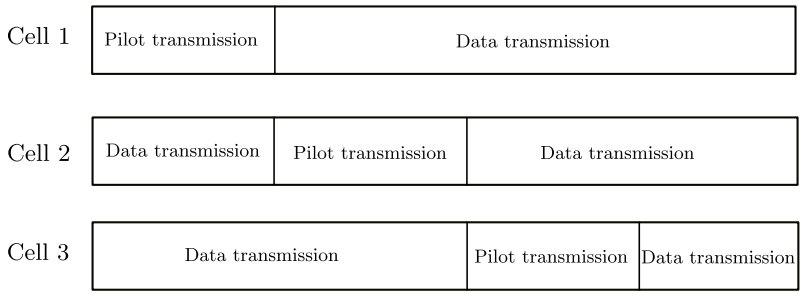


Figure 2.1: The time-shifted pilot scheme of [1].

In [94], a similar technique has been adopted for mitigating pilot contamination in a cognitive massive MIMO communication system. Therein, the authors reported that the effects of pilot contamination interference were greatly reduced and thus uplink/downlink achievable rates significantly increased. It should be pointed out that this approach requires the knowledge of the covariance matrix of all users' channels and hence is likely infeasible in practical implementations. In [95], the authors considered a semi-blind channel estimation technique to separate the subspace of the desired user channel from that of interfering users caused by pilot contamination. However, in order to completely eliminate pilot contamination, this method requires that the channel coherence time goes to infinity. Unfortunately, this assumption is not true in practice either.

In [96], a sophisticated two-stage processing approach is proposed to mitigate pilot contamination. Specifically, a downlink pilot transmission phase is followed by a scheduled uplink pilot transmission phase. During the downlink training phase, all BSs first send the pilot signals using only their first antenna to their users. Then, all BSs repeat this process using all of their antennas simultaneously. At the end of this two-step downlink training, users estimate the downlink channels by processing the received pilots. The estimation of downlink channels is then followed by an $(L+1)$-step (where L is the number of cells) scheduled uplink training phase. In particular, during the initial step, all users send their uplink pilots to their BSs. Then, during each of the next L steps of the uplink training phase, the users in cell l transmit a pre-distorted version of their pilots to BS l, while users in other cells transmit their uplink pilots to their BSs. More precisely, by transmitting the pre-distorted pilots, users encapsulate their non-contaminated downlink channels in the

uplink pilot signals so that each BS will be able to extract the channels of its own users from the received uplink pilots by removing pilots of users in other cells; thus, mitigating pilot contamination. A main disadvantage of this technique, however, is the significant overhead occurred during the sophisticated multi-step training phases. In [87], a multi-cell MMSE precoding technique is proposed so that the precoding matrix minimizes an objective function involving two parts: (i): the sum of the MSE of the downlink received signals at mobile terminals in the same cell, and (ii): the mean-square interference at users in other cells. Although, significant performance gains have been reported in [87] compared to conventional schemes, this approach requires full cooperation between BSs, the knowledge of second-order statistics of users' channels at BSs, and is not applicable to arbitrary pilot allocation schemes.

In another interesting line of work to tackle the rate saturation phenomena due to pilot contamination, several novel schemes have been proposed [38, 39, 64, 65, 68, 86, 95]. These solutions work well in the asymptotic regime provided that some assumptions and requirements are satisfied. In [68], a pilot contamination precoding (PCP) method has been proposed, assuming that the estimates of all large-scale fading coefficients as well as the data signals of all users across the network will be shared among all BSs. While the large-scale fading coefficients vary slowly (i.e., stay constant over many channel coherence intervals [38, 64, 74]) and can thus be estimated and tracked efficiently, the main drawback of this scheme is that the data signals of all users will need to be processed by all BSs. Also, the efficiency of this approach relies heavily on having a very large number of antennas, i.e., $M > 10^6$, as well as the accuracy of the computation of PCP matrices. This method has been further improved in [38, 65, 86] by adding an outer-cell processing called large-scale fading precoding (LSFP) and large-scale fading decoding (LSFD) creating additional computational complexities. More specifically, using matrix inversion the LSFD/LSFP matrices of all users across the network must be computed at a centralized processor (known as network controller), where the data signals of all users are shared. In addition, this method has been extended to the cases of spatially correlated channels [64], correlated channels with phase shift [70], maximizing the product of SINRs [41] as well as cell-free massive MIMO communication systems [40]. Different from these lines of work, the work

of [39] proposes a multi-cell MMSE precoding/combining technique and assumes that pilot-sharing users must have asymptotically linearly independent covariance matrices. This assumption, however, may not always be true and also requires the knowledge of channel covariance matrices at the BSs.

Smart pilot assignment approaches have also been proposed in the literature. By carefully assigning pilots to users, the effects of pilot contamination are reduced, but not fully eliminated [97–100]. However, properly assigning pilots to the users normally gives rise to solving combinatorial optimization problems that are typically dealt with heuristically. In particular, one approach is to let each BS assign and re-assign pilots to its users in an iterative fashion until a given objective function is optimized. For instance, if a cell edge user is experiencing substantial pilot contamination, it can switch its pilot sequence with a user close to the cell center with a stronger channel, thus reducing the adverse effects of pilot contamination on system performance. A location-based pilot assignment strategy was proposed in [101] to alleviate the effects of pilot contamination in a distributed cell-free MIMO communication system with a central control unit that has access to all BSs. In particular, the work of [101] proposes to group the users across the network into a number of clusters, i.e., sets of pilot-sharing users, based on their relative distance (with the help of a central unit) where non-orthogonal pilots are assigned to users that are far from each other.

2.2.2 Data-aided Estimation Approaches

The use of data-aided channel estimation method (also known as subspace based estimation) is deemed an attractive solution in the literature to mitigate pilot contamination and to increase spectral efficiency, as no pilot sequences are required. One of the early works in this area, which also predates the introduction of massive MIMO systems, is [102] which has proposed a subspace based blind channel estimation approach. An extension of this blind estimation method is reported in [103] and is based on the eigenvalue decomposition (EVD) of covariance matrices. In particular, the method is not impacted by pilot contamination and thus can potentially outperform estimation techniques based on

the transmission of pilot training signals. However, its accuracy and complete elimination of pilot contamination rely on having a very large number of BS antennas as well as unbounded coherence interval. In other words, if the length of the coherence interval does not grow unbounded, the channel estimates that result from this method will still suffer from residual pilot contamination. This approach has also been investigated in the context of multi-cell TDD massive MIMO systems, with the purpose of mitigating pilot contamination [66].

Further studies in the context of pilot decontamination reported in [67, 104, 105] have proposed a blind method for channel estimation in conjunction with power control using random matrix theory. More precisely, it is proposed to blindly estimate the system parameters in the subspace of the desired signal based on the singular value decomposition (SVD) of the received signal matrix. Nevertheless, a drawback of this technique is the condition that the channel of the desired user must be stronger than those of the interfering users, which is not always true in practice, especially in interference-limited scenarios with shadow fading. In [106], a similar approach using the maximum a-posteriori (MAP) principle for subspace based channel estimation in TDD massive MIMO systems is proposed. It is reported that this technique outperforms the results of the blind estimation methods in [67, 104, 105], but at the cost of increasing complexity. Moreover, in order to reduce the effects of pilot contamination, a fast EVD-based channel estimation method for massive MIMO systems using a diagonal Jacket matrix[1] along with iterative least-square with projection (ILSP) was developed in [108].

Recently, a technique based on message splitting has been adopted in the literature to cope with the interference due to pilot contamination [109], which does not fall into the above categories. In particular, the work of [109] has studied a *single-cell* massive MIMO system operating in TDD mode, where all users inside the cell share the *same* pilot sequence for channel estimation. Therein, assuming that each user applies single-user successive interference cancellation in downlink, it is shown that the spectral efficiency is increased compared to conventional schemes.

[1]The Jacket matrix is a generalization of the Hadamard matrix [107].

2.3 Massive MIMO System Model

In this section, we describe the system model for uplink and downlink of a multi-cell multi-user massive MIMO system and present some well known results associated with pilot contamination effect. We consider a multi-cell communication system with L cells, where each cell has a BS equipped with M antennas serving K ($M > K$) single antenna users. Assuming a spatially correlated channel model, the channel matrix between the M antennas of BS j and the users in cell l is denoted by $\boldsymbol{G}_{jl} = [\boldsymbol{g}_{j1l}, \boldsymbol{g}_{j2l}, ..., \boldsymbol{g}_{jKl}] \in \mathbb{C}^{M\times K}$. More precisely, the channel vector $\boldsymbol{g}_{jkl}$ associated with user k in cell l is described by

$$\boldsymbol{g}_{jkl} = \boldsymbol{R}_{jkl}^{1/2}\boldsymbol{h}_{jkl}, \tag{2.1}$$

where $\boldsymbol{h}_{jkl} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$, and $\boldsymbol{R}_{jkl}$ is the spatial correlation matrix of the channel, i.e., $\boldsymbol{R}_{jkl} = \mathbb{E}\left[\boldsymbol{g}_{jkl}\boldsymbol{g}_{jkl}^{\dagger}\right] \in \mathbb{C}^{M\times M}$ or $\boldsymbol{g}_{jkl} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_{jkl})$. Also, for the large-scale fading coefficients we have

$$\beta_{jkl} = \frac{\mathrm{tr}(\boldsymbol{R}_{jkl})}{M}, \tag{2.2}$$

which model shadowing and path loss effects. A standard block-fading model is considered here, where the channels are constant over one coherence interval with one independent realization in each block, whereas the large-scale fading coefficients are constant over many coherence time intervals. Furthermore, considering TDD operation, it is assumed that reciprocity holds between uplink and downlink channels. A frequency re-use factor of one is assumed, i.e., the whole frequency band is used in each cell. The following Lemma from [110] will be used later in the sequel.

Lemma 1. *Let* $\boldsymbol{x} := [x_1, x_2, ..., x_M]^T$ *and* $\boldsymbol{y} := [y_1, y_2, ..., y_M]^T$ *be mutually independent* $M \times 1$ *random vectors with distributions* $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_x^2\boldsymbol{I}_M)$ *and* $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_y^2\boldsymbol{I}_M)$, *respectively. Then, by the strong law of large numbers,*

$$\frac{\boldsymbol{x}^{\dagger}\boldsymbol{x}}{M} \xrightarrow{\text{a.s.}} \sigma_x^2, \quad \textit{and} \quad \frac{\boldsymbol{x}^{\dagger}\boldsymbol{y}}{M} \xrightarrow{\text{a.s.}} 0, \quad \textit{as} \quad M \to \infty, \tag{2.3}$$

where $\xrightarrow{\text{a.s.}}$ *denotes almost sure convergence.*

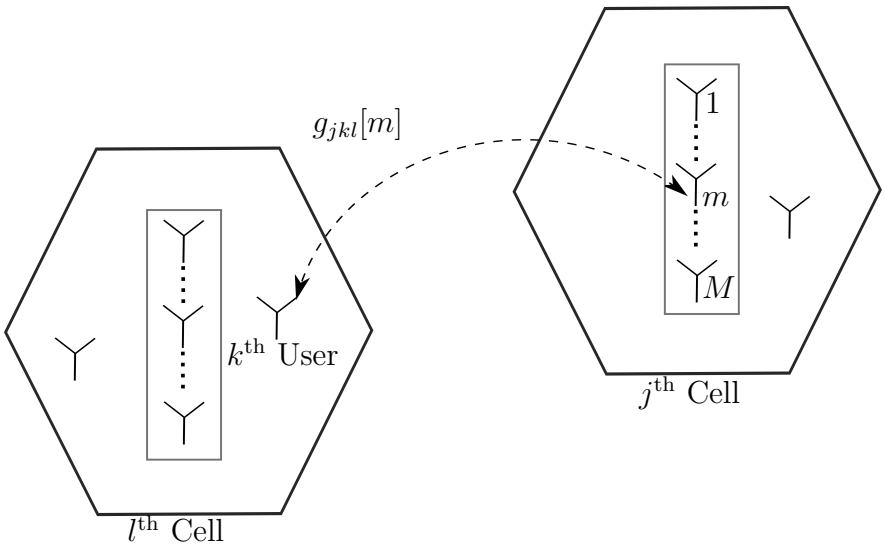


Figure 2.2: System model showing the channel gain between the m^{th} antenna of the BS in cell j and the k^{th} user in cell l.

2.3.1 Uplink Data Transmission

We point out that the model used for uplink/downlink data transmission in this thesis is similar to that of [32] with a slight change of notation. During the uplink data transmission phase, the BS in cell j receives the baseband signal $\boldsymbol{y}_j^{\text{ul}} \in \mathbb{C}^{M\times 1}$ where

$$\boldsymbol{y}_j^{\text{ul}} = \sum_{l=1}^{L}\sum_{k=1}^{K}\sqrt{\rho_{\text{ul}}}\boldsymbol{g}_{jkl}x_l^{\text{ul}}[k] + \boldsymbol{n}_j, \tag{2.4}$$

and $\boldsymbol{x}_l^{\text{ul}} = \left[x_l^{\text{ul}}[1], x_l^{\text{ul}}[2], ..., x_l^{\text{ul}}[K]\right]^T$ is the vector of transmit signals of the users in cell l, ρ_{ul} is the average uplink transmit power of the users, and $\boldsymbol{n}_j \sim \mathcal{CN}(\boldsymbol{0},\ \boldsymbol{I}_M)$ is the additive Gaussian noise vector at the BS in cell j. The assumption of unit noise variance is without loss of generality[2] and thus ρ_{ul} can be interpreted as the uplink transmit SNR of the users.

[2]Consider $\boldsymbol{y}_j^{\text{ul}'} = \sum_{l=1}^{L}\sum_{k=1}^{K}\sqrt{\rho'_{\text{ul}}}\boldsymbol{g}_{jkl}x_l^{\text{ul}}[k] + \boldsymbol{n}'_j$, where ρ'_{ul} is the transmit power and $\boldsymbol{n}'_j \sim \mathcal{CN}(\boldsymbol{0},\ \sigma^2\boldsymbol{I}_M)$. Then, let $\boldsymbol{y}_j^{\text{ul}} = \boldsymbol{y}_j^{\text{ul}'}/\sigma$ with $\rho_{\text{ul}} = \rho'_{\text{ul}}/\sigma^2$ being the transmit SNR.

2.3.2 Downlink Data Transmission

During the downlink data transmission phase, the i^{th} user in cell l receives the baseband signal

$$y_{il}^{\text{dl}} = \sum_{j=1}^{L} \sqrt{\rho_{\text{dl}}} \boldsymbol{g}_{jil}^{\dagger} \boldsymbol{x}_j^{\text{dl}} + z_{il}, \tag{2.5}$$

where $\boldsymbol{x}_j^{\text{dl}} = \left[x_j^{\text{dl}}[1], x_j^{\text{dl}}[2], ..., x_j^{\text{dl}}[M]\right]^T$ is the transmit signal of the BS in cell j, ρ_{dl} is the per-user transmit power of the BS, and $z_{il} \sim \mathcal{CN}(0,\ 1)$ is the receiver noise of the i^{th} user in cell l. Thus similar to ρ_{ul}, ρ_{dl} can be interpreted as the transmit SNR of the BSs. Also, defining $\boldsymbol{w}_{jkj} \in \mathbb{C}^{M \times 1}$ as the precoding vector of the BS in cell j, we have

$$\boldsymbol{x}_j^{\text{dl}} = \frac{1}{\sqrt{\lambda_j}} \sum_{k=1}^{K} \boldsymbol{w}_{jkj} s_j[k] \tag{2.6}$$

$$= \frac{\boldsymbol{W}_j \boldsymbol{s}_j}{\sqrt{\lambda_j}}, \tag{2.7}$$

where $\boldsymbol{s}_j = [s_j[1], s_j[2], ..., s_j[K]]^T$ is the vector of data symbols intended for the K users in cell j, $\boldsymbol{W}_j = [\boldsymbol{w}_{j1j}, \boldsymbol{w}_{j2j}, ..., \boldsymbol{w}_{jKj}] \in \mathbb{C}^{M \times K}$, and λ_j is a normalization factor used to make sure the following constraint is satisfied at BS j

$$\frac{\mathbb{E}\left[\boldsymbol{x}_j^{\text{dl}^{\dagger}} \boldsymbol{x}_j^{\text{dl}}\right]}{K} = 1, \tag{2.8}$$

i.e., the downlink SNR per user of the BS in cell j equals ρ_{dl}. Specific choices of precoding vectors will be discussed in subsection 2.3.6.

2.3.3 CSI estimation at BS

Following [38, 64, 65, 87, 91], it is assumed that the same set of pilot sequences $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_K \in \mathbb{C}^{\tau \times 1}$ of length τ (generally $\tau \geq K$, however here it is assumed that $\tau = K$) are used in all cells and thus the channel estimate will be corrupted by pilot contamination from the adjacent cells. Defining the pilot matrix $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ..., \boldsymbol{\psi}_K)^T \in \mathbb{C}^{K \times K}$, we

assume orthonormal pilots, i.e., $\boldsymbol{\Psi}\boldsymbol{\Psi}^{\dagger} = I_K$.

During the uplink training phase of the TDD protocol, user $k = 1, 2, ..., K$ in each cell transmits the pilot sequence $\boldsymbol{\psi}_k$ to its BSs. The BS in cell j then finds the estimate $\hat{\boldsymbol{G}}_{jj}$ of the local channels $\boldsymbol{G}_{jj}$. More specifically, the BS in cell j receives the matrix $\boldsymbol{Y}_j^{\mathrm{p}} \in \mathbb{C}^{M \times K}$, i.e.,

$$\boldsymbol{Y}_j^{\mathrm{p}} = \sum_{l=1}^{L} \sqrt{\rho_{\mathrm{p}}} \boldsymbol{G}_{jl} \boldsymbol{\Psi} + \boldsymbol{Z}_j, \tag{2.9}$$

where ρ_{p} is the average pilot transmission power, and $\boldsymbol{Z}_j$ is the additive white Gaussian noise (AWGN) at the BS with entries that are i.i.d $\mathcal{CN}(0, 1)$ random variables. Similar to uplink and downlink data transmission, ρ_{p} can be interpreted as the pilot SNR. Generally, ρ_{p} is a function of the average uplink transmit power of users ρ_{ul} and the length of pilot sequences τ. Multiplying $\boldsymbol{Y}_j^{\mathrm{p}}$ by $\boldsymbol{\Psi}^{\dagger}$, the k^{th} column of the resulting matrix is

$$\boldsymbol{r}_{jk} = \sum_{l=1}^{L} \sqrt{\rho_{\mathrm{p}}} \boldsymbol{g}_{jkl} + \tilde{\boldsymbol{z}}_{jk}, \tag{2.10}$$

where $\tilde{\boldsymbol{z}}_{jk} \sim \mathcal{CN}(\boldsymbol{0}, \ I_M)$. The MMSE estimate $\hat{\boldsymbol{g}}_{jkj}$ of $\boldsymbol{g}_{jkj}$, i.e., the vector $\hat{\boldsymbol{g}}_{jkj}$ that minimizes the mean squared error (MSE) $\mathbb{E}\left[\|\boldsymbol{g}_{jkj} - \hat{\boldsymbol{g}}_{jkj}\|^2\right]$, based on the observation $\boldsymbol{r}_{jk}$ is given by [111]

$$\hat{\boldsymbol{g}}_{jkj} = \sqrt{\rho_{\mathrm{p}}} \boldsymbol{R}_{jkj} \mathbb{E}\left[\boldsymbol{r}_{jk} \boldsymbol{r}_{jk}^{\dagger}\right]^{-1} \boldsymbol{r}_{jk} \tag{2.11}$$

$$= \sqrt{\rho_{\mathrm{p}}} \boldsymbol{R}_{jkj} \left(\sum_{l=1}^{L} \rho_{\mathrm{p}} \boldsymbol{R}_{jkl} + \boldsymbol{I}_M \right)^{-1} \boldsymbol{r}_{jk} \tag{2.12}$$

$$= \sqrt{\rho_{\mathrm{p}}} \boldsymbol{R}_{jkj} \left(\sum_{l=1}^{L} \rho_{\mathrm{p}} \boldsymbol{R}_{jkl} + \boldsymbol{I}_M \right)^{-1} \left(\underbrace{\sqrt{\rho_{\mathrm{p}}} \boldsymbol{g}_{jkj}}_{\text{Intended channel}} + \underbrace{\sum_{l=1, l \neq j}^{L} \sqrt{\rho_{\mathrm{p}}} \boldsymbol{g}_{jkl}}_{\text{Pilot contamination}} + \tilde{\boldsymbol{z}}_{jk} \right) \tag{2.13}$$

It can be seen from (2.13) that the estimate of $\boldsymbol{g}_{jkj}$ is contaminated by the channel of users in other cells that are using the same pilot sequence as user k in cell j. Due to

the orthogonality property of MMSE estimation, one can decompose the channel $\boldsymbol{g}_{jkj}$ as $\boldsymbol{g}_{jkj} = \hat{\boldsymbol{g}}_{jkj} + \boldsymbol{\epsilon}_{jkj}$, where $\boldsymbol{\epsilon}_{jkj}$ is an uncorrelated estimation error (thus statistically independent of the estimate $\hat{\boldsymbol{g}}_{jkj}$ as they are jointly Gaussian). The distribution of the channel estimate $\hat{\boldsymbol{g}}_{jkj}$ and the estimation error $\boldsymbol{\epsilon}_{jkj}$ are as follows [111]

$$\hat{\boldsymbol{g}}_{jkj} \sim \mathcal{CN}\left(\boldsymbol{0},\ \rho_{\mathrm{p}}\boldsymbol{R}_{jkj}\left(\sum_{l=1}^{L}\rho_{\mathrm{p}}\boldsymbol{R}_{jkl} + \boldsymbol{I}_M\right)^{-1}\boldsymbol{R}_{jkj}\right) \tag{2.14}$$

$$\boldsymbol{\epsilon}_{jkj} \sim \mathcal{CN}\left(\boldsymbol{0},\ \boldsymbol{R}_{jkj} - \rho_{\mathrm{p}}\boldsymbol{R}_{jkj}\left(\sum_{l=1}^{L}\rho_{\mathrm{p}}\boldsymbol{R}_{jkl} + \boldsymbol{I}_M\right)^{-1}\boldsymbol{R}_{jkj}\right). \tag{2.15}$$

Moreover, for the estimate of an inter-cell channel from users in other cells using the same pilot signal $\boldsymbol{\psi}_k$ to BS j, provided that $\boldsymbol{R}_{jkj}$ is invertible, using (2.11) we have

$$\hat{\boldsymbol{g}}_{jkl} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jkl}\left(\sum_{l=1}^{L}\rho_{\mathrm{p}}\boldsymbol{R}_{jkl} + \boldsymbol{I}_M\right)^{-1}\boldsymbol{r}_{jk} \tag{2.16}$$

$$= \boldsymbol{R}_{jkl}(\boldsymbol{R}_{jkj})^{-1}\hat{\boldsymbol{g}}_{jkj}. \tag{2.17}$$

This relation will be used in the proofs provided in the appendices. To simplify the notation, from now on we use the following definition

$$\boldsymbol{\Lambda}_{jk} := \sum_{l=1}^{L}\rho_{\mathrm{p}}\boldsymbol{R}_{jkl} + \boldsymbol{I}_M. \tag{2.18}$$

Uncorrelated Rayleigh Fading

For the special case of an uncorrelated Rayleigh fading channel, the spatial correlation matrix becomes diagonal, i.e., $\boldsymbol{R}_{jkj} = \beta_{jkj}\boldsymbol{I}_M$, and therefore (2.1) simplifies to

$$\boldsymbol{g}_{jkl} = \sqrt{\beta_{jkl}}\boldsymbol{h}_{jkl}. \tag{2.19}$$

Consequently, the MMSE estimate of (2.11) reduces to the following

$$\hat{\boldsymbol{g}}_{jkj} = \mathbb{E}\left[\boldsymbol{g}_{jkj}\boldsymbol{r}_{jk}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{r}_{jk}\boldsymbol{r}_{jk}^{\dagger}\right]^{-1}\boldsymbol{r}_{jk} \tag{2.20}$$

$$= \sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\left(1+\rho_{\mathrm{p}}\sum_{l=1}^{L}\beta_{jkl}\right)^{-1}\boldsymbol{r}_{jk} \tag{2.21}$$

$$= \alpha_{jkj}\left(\underbrace{\sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkj}}_{\text{Intended channel}} + \underbrace{\sum_{l=1,l\neq j}^{L}\sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkl}}_{\text{Pilot contamination}} + \tilde{\boldsymbol{z}}_{jk}\right), \tag{2.22}$$

where $\alpha_{jkj} := \frac{\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}}{1+\rho_{\mathrm{p}}\sum_{l=1}^{L}\beta_{jkl}}$. Substituting $\boldsymbol{R}_{jkj} = \beta_{jkj}\boldsymbol{I}_M$ into the distributions (2.14)-(2.15), one can obtain the following for an uncorrelated fading channel

$$\hat{\boldsymbol{g}}_{jkj} \sim \mathcal{CN}\left(\boldsymbol{0},\ \sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}I_M\right) \tag{2.23}$$

$$\boldsymbol{\epsilon}_{jkj} \sim \mathcal{CN}\left(\boldsymbol{0},\ \left(\beta_{jkj}-\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\right)I_M\right). \tag{2.24}$$

Lastly, using (2.17), the estimate of inter-cell channels from users sharing the same pilot signal $\boldsymbol{\psi}_k$ to BS j, is given by

$$\hat{\boldsymbol{g}}_{jkl} = \left(\frac{\beta_{jkl}}{\beta_{jkj}}\right)\hat{\boldsymbol{g}}_{jkj}, \tag{2.25}$$

which will be used in the proofs provided in the appendices.

2.3.4 Channel Hardening and Favorable Propagation

Below, we provide formal definitions for two important large-system effects in multi-antenna systems, namely, channel hardening and favorable propagation. We then present an example of these asymptotic effects based on the uncorrelated fading channel model.

Channel Hardening

By asymptotically eliminating the effects of small-scale fading components, channel hardening results in fast-fading channels asymptotically behaving like deterministic channels. In other words, the effective channel gains become "asymptotically constant" over many coherence intervals. Mathematically, this can be presented as follows:

$$\frac{\|\boldsymbol{g}_{jkl}\|^2}{\mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right]} \overset{\text{a.s.}}{\to} 1, \text{as} \quad M \to \infty. \tag{2.26}$$

This implies that as the number of BS antennas M grows to infinity, the gain of the fading channel $\|\boldsymbol{g}_{jkl}\|^2$ becomes close to the mean value $\mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right] = \text{tr}\left(\boldsymbol{R}_{jkl}\right)$.

Remark 1. *One should note that the asymptotic convergence in (2.26) does not result in* $\|\boldsymbol{g}_{jkl}\|^2 \overset{\text{a.s.}}{\to} \mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right]$*, as* $M \to \infty$*, as both the gain of the fading channel and its mean value can generally diverge. However, one can see that the condition of* $\|\boldsymbol{g}_{jkl}\|^2/M - \mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right]/M \overset{\text{a.s.}}{\to} 0$*, as* $M \to \infty$*, yields the asymptotic channel hardening in (2.26).*

Asymptotic Favorable Propagation

Asymptotic favorable propagation means that the direction of two distinct user channels becomes asymptotically orthogonal. More specifically, the channel is said to have asymptotic favorable propagation, if the following holds:

$$\frac{\boldsymbol{g}_{jkl}^{\dagger}\boldsymbol{g}_{nim}}{\sqrt{\mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right]\mathbb{E}\left[\|\boldsymbol{g}_{nim}\|^2\right]}} \overset{\text{a.s.}}{\to} 0, \text{ as} \quad M \to \infty, \ (j,k,l) \neq (n,i,m) \tag{2.27}$$

i.e., the normalized channel vectors $\boldsymbol{g}_{jkl}/\sqrt{\mathbb{E}\left[\|\boldsymbol{g}_{jkl}\|^2\right]}$ and $\boldsymbol{g}_{nim}/\sqrt{\mathbb{E}\left[\|\boldsymbol{g}_{nim}\|^2\right]}$ have asymptotically zero inner products.

Remark 2. *One should note that the asymptotic property in (2.27) does not mean that the channel vectors* $\boldsymbol{g}_{jkl}$ *and* $\boldsymbol{g}_{nim}$ *become asymptotically orthogonal. However, it implies that the direction of the two channel vectors becomes asymptotically orthogonal.*

Asymptotic favorable propagation helps the BS eliminate non-coherent inter-cell interference, and thus linear procoding/combining techniques become sufficient [111]. This property will be used later to present asymptotic limits of uplink/downlink achievable rates.

Example: Consider the uncorrelated Rayleigh fading channel model described by $\boldsymbol{g}_{jkl} = \sqrt{\beta_{jkl}}\boldsymbol{h}_{jkl}$, where $\boldsymbol{h}_{jkl} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$. It can be verified that this channel model satisfies both of these asymptotic properties with almost sure convergence [26], i.e.,

$$\frac{\boldsymbol{G}_{jj}^{\dagger}\boldsymbol{G}_{jj}}{M} \overset{\text{a.s.}}{\to} \text{diag}\left(\beta_{j1j}, \beta_{j2j}, ..., \beta_{jKj}\right), \text{ as } \quad M \to \infty. \tag{2.28}$$

2.3.5 A Worst-case Uncorrelated Noise Technique

In this part, a useful technique that is frequently used in massive MIMO literature to establish rate lower bounds for point-to-point channels is presented. Consider the following point-to-point channel model:

$$y = x^G + z, \tag{2.29}$$

where x^G is the zero mean input signal with complex Gaussian distribution, i.e., $x^G \sim \mathcal{CN}(0, P)$, and z is the zero mean additive noise having variance σ^2, and uncorrelated with the input x^G. It can be shown that the capacity of this channel is lower bounded by [112]

$$I(x^G; y) \geq I\left(x^G; y^G\right) \tag{2.30}$$

$$= C\left(\frac{P}{\sigma^2}\right), \tag{2.31}$$

where $C(x) := \log(1+x)$ is the Shannon rate function, $y^G = x^G + z^G$, and z^G is a zero mean complex Gaussian random variable having the same variance as z, i.e., $z^G \sim \mathcal{CN}(0, \sigma^2)$. In other words, (2.30) implies that the worst effect that the additive uncorrelated noise z can have on this channel is to behave as AWGN, hence the name *worst-case uncorrelated noise technique*. This technique will be used later in this chapter to obtain lower bounds for uplink/downlink achievable rates of massive MIMO systems.

2.3.6 Spectral Efficiency: Treating Interference as Noise (TIN)

In this part, we present well-known lower bounds on the uplink/downlink achievable rates of the multi-cell multi-user massive MIMO communication system discussed in subsections 2.3.1 and 2.3.2.

Uplink

Consider the baseband signal $\boldsymbol{y}_j^{\text{ul}}$ in (2.4) received by BS j during the uplink data transmission phase. Also, assume that the combining vector used by the BS to decode the signal transmitted by user i in cell j is denoted by $\boldsymbol{v}_{jij}$. By adding and subtracting a term associated with the mean of the effective channel $\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}$ in (2.4), the following is obtained over one coherence interval for the estimate of the signal transmitted by user i in cell j:

$$\hat{y}_{ji}^{\text{ul}} = \boldsymbol{v}_{jij}^{\dagger}\boldsymbol{y}_j^{\text{ul}} \tag{2.32}$$

$$= \underbrace{\sqrt{\rho_{\text{ul}}}\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}\right] x_j^{\text{ul}}[i]}_{\text{Desired signal}} + \underbrace{\sqrt{\rho_{\text{ul}}}\left(\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij} - \mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}\right]\right) x_j^{\text{ul}}[i]}_{\text{Interference due to beamforming gain uncertainty}}$$

$$+ \underbrace{\sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\sqrt{\rho_{\text{ul}}}\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}x_l^{\text{ul}}[k]}_{\text{Inter-cell interference caused by other users}} + \underbrace{\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_j}_{\text{Noise}} \tag{2.33}$$

$$= \underbrace{\eta_{ij}x_j^{\text{ul}}[i]}_{\text{Desired signal}} + \underbrace{z_{jij}}_{\text{Additive noise}}, \tag{2.34}$$

where the expectations are with respect to the channel realizations, $\eta_{ij} := \sqrt{\rho_{\text{ul}}}\mathbb{E}[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}]$ is the effective channel gain and z_{jij} is the additive noise term incorporating the last three terms in (2.33). Treating (2.34) as the output of a point-to-point channel similar to (2.29), it can be verified that both the input signal $\eta_{ij}x_j^{\text{ul}}[i]$ and the additive noise term z_{jij} are zero mean random variables, and also the additive noise is uncorrelated with the input. Defining R_{ij}^{ul} as the uplink rate of the i^{th} user in cell j, any rate tuple $\left(R_{i1}^{\text{ul}}, ..., R_{iL}^{\text{ul}}\right)$ is achievable if it satisfies the following set of inequalities

$$R_{ij}^{\text{ul}} \leq I\left(\hat{y}_{ji}^{\text{ul}}; x_j^{\text{ul}}[i]\right), \quad \text{for} \quad j = 1, ..., L,\ i = 1, ..., K. \tag{2.35}$$

Hence, assuming Gaussian signaling and applying the worst-case uncorrelated noise technique of (2.30), one obtains the following lower bound on the rate of decoding the signal transmitted by user i in cell j [111]

$$I\left(\hat{y}_{ji}^{\mathrm{ul}}; x_j^{\mathrm{ul}}[i]\right) \geq C\left(\frac{\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}\right]\right|^2}{\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right]-\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jij}\right]\right|^2+\mathbb{E}\left[\|\boldsymbol{v}_{jij}\|^2\right]}\right), \tag{2.36}$$

where the expectations are with respect to the channel realizations. Note that the numerator and the denominator in (2.36) are the variance of the desired signal $\eta_{ij}x_j^{\mathrm{ul}}[i]$ and the variance of the uncorrelated additive noise z_{jij}, respectively. It is worth mentioning that the rate lower bound of (2.36) can be applied with any combining vector $\boldsymbol{v}_{jij}$ and any channel estimator (including the MMSE estimation discussed in the previous section).

Downlink

Consider the baseband signal y_{il}^{dl} in (2.5) received by user i in cell l during the downlink data transmission phase. Furthermore, assume that the precoding vector used by BS l for transmission of the data stream intended for user i in cell l is denoted by $\boldsymbol{w}_{lil}$. Similar to the uplink, we add and subtract a term associated with the mean of the effective channel $\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}$, which results in the following

$$\begin{aligned} y_{il}^{\mathrm{dl}} = & \underbrace{\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_l}}\mathbb{E}\left[\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}\right]s_l[i]}_{\text{Desired signal}}+\underbrace{\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_l}}\left(\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}-\mathbb{E}\left[\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}\right]\right)s_l[i]}_{\text{Interference due to beamforming gain uncertainty}} \\ & +\underbrace{\sum_{j=1}^{L}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\sum_{k=1,k\neq i}^{K}\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}s_j[k]}_{\text{Inter-cell interference caused by other users}}+\underbrace{z_{il}}_{\text{Noise}} \end{aligned} \tag{2.37}$$

$$= \underbrace{\zeta_{il}s_l[i]}_{\text{Desired signal}}+\underbrace{z'_{il}}_{\text{Additive noise}}, \tag{2.38}$$

where the expectations are with respect to the channel realizations, $\zeta_{il} := \sqrt{\rho_{\mathrm{dl}}/\lambda_l}\mathbb{E}\left[\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}\right]$ is the effective channel gain and z'_{il} is the additive noise term incorporating the last three terms in (2.37). Treating (2.38) as the output of a point to point channel similar to (2.29), it can be verified that both the input signal $\zeta_{il}s_l[i]$ and the additive noise term z_{il} are zero mean random variables, and also the additive noise is uncorrelated with the input. Therefore, defining R_{il}^{dl} as the downlink rate associated with decoding $s_l[i]$, any rate tuple $\left(R_{i1}^{\mathrm{dl}},...,R_{iL}^{\mathrm{dl}}\right)$ is achievable if it satisfies the following set of inequalities

$$R_{il}^{\mathrm{dl}} \leq I\left(y_{il}^{\mathrm{dl}};s_l[i]\right), \quad \text{for} \quad l=1,...,L,\ i=1,...,K. \tag{2.39}$$

Thus, similar to the uplink case and assuming Gaussian signaling, by applying the worst-case uncorrelated noise technique of (2.30) one obtains the following lower bound on the downlink rate of decoding data stream $s_l[i]$ at user i in cell l [111]

$$I\left(y_{il}^{\mathrm{dl}};s_l[i]\right) \geq C\left(\frac{\frac{\rho_{\mathrm{dl}}}{\lambda_l}\left|\mathbb{E}\left[\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}\right]\right|^2}{\sum_{j=1}^{L}\sum_{k=1}^{K}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}\right|^2\right]-\frac{\rho_{\mathrm{dl}}}{\lambda_l}\left|\mathbb{E}\left[\boldsymbol{g}_{lil}^{\dagger}\boldsymbol{w}_{lil}\right]\right|^2+1}\right), \tag{2.40}$$

where the expectations are with respect to the channel realizations. Note that the numerator and the denominator in (2.40) are the variance of the desired signal $\zeta_{il}s_l[i]$ and the variance of the uncorrelated additive noise z_{il}, respectively. Similar to the uplink case, the rate lower bound of (2.40) can be applied with any precoding vector $\boldsymbol{w}_{lil}$ and any channel estimator (including the MMSE estimation discussed in the previous section). In the next section, we will present asymptotic limits of these rate lower bounds using the following conventional linear combining/precoding schemes.

- *Maximum ratio combining/transmission (MRC/MRT):* Under MRC (MRT), the BS tries to maximize the signal to noise ratio (SNR) for each data stream, while ignoring the effects of interference. The combining (precoding) vector for MRC (MRT) is given by $\boldsymbol{v}_{jij}=\hat{\boldsymbol{g}}_{jij}$ ($\boldsymbol{w}_{jij}=\hat{\boldsymbol{g}}_{jij}$). An important advantage of this scheme is that it is extremely easy to implement. However, as a disadvantage, one should bear in mind that as maximum ratio neglects the effects of interference, its performance is inferior

to other schemes in interference limited scenarios.

- *Zero forcing (ZF):* Unlike MRC/MRT, under ZF combining/precoding the effect of additive noise is neglected, while taking into account the effects of intra-cell interference. The ZF combining/precoding matrix at BS j is given by $\boldsymbol{V}_{jj} = [\boldsymbol{v}_{j1j}, ..., \boldsymbol{v}_{jKj}] = \hat{\boldsymbol{G}}_{jj} \left(\hat{\boldsymbol{G}}_{jj}^{\dagger} \hat{\boldsymbol{G}}_{jj} \right)^{-1}$. Noting that this is the pseudo inverse of the estimated channel matrix $\hat{\boldsymbol{G}}_{jj}$, it can be verified that $\boldsymbol{V}_{jj}^{\dagger} \hat{\boldsymbol{G}}_{jj} = \boldsymbol{I}_K$, and thereby

$$\boldsymbol{v}_{jkj}^{\dagger} \hat{\boldsymbol{g}}_{jmj} = \delta_{mk}, \tag{2.41}$$

  where δ_{mk} is the Kronecker delta function. In other words, the ZF scheme completely nulls out the interference by projecting each data stream onto the orthogonal complement of the intra-cell interference. Even though its computational complexity is higher than MRC/MRT due to the calculation of the pseudo inverse, it performs better in interference limited scenarios.

2.3.7 Large Antenna Regime

Using the uncorrelated Rayleigh fading channel model described in (2.19), we now present well-known asymptotic limits for uplink/downlink achievable rate lower bounds under the combining/precoding schemes discussed above.

Uplink

MRC: Assuming that MRC is used at BS j to estimate signal $x_j^{\mathrm{ul}}[i]$ received from user i in cell j, one can re-write (2.32) as

$$\hat{y}_{ji}^{\mathrm{ul}} = \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{y}_j^{\mathrm{ul}} \tag{2.42}$$

$$= \sum_{l=1}^{L} \sqrt{\rho_{\mathrm{ul}}} \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} x_l^{\mathrm{ul}}[i] + \sum_{l=1}^{L} \sum_{k=1, k \neq i}^{K} \sqrt{\rho_{\mathrm{ul}}} \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jkl} x_l^{\mathrm{ul}}[k] + \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{n}_j. \tag{2.43}$$

Using Lemma 1, the following is obtained

$$\frac{\hat{y}_{ji}^{\mathrm{ul}}}{M} \overset{\text{a.s.}}{\to} \sqrt{\rho_{\mathrm{ul}}\rho_{\mathrm{p}}}\alpha_{jij}\left(\underbrace{\beta_{jij}x_j^{\mathrm{ul}}[i]}_{\text{Desired signal}} + \underbrace{\sum_{l=1,l\neq j}^{L}\beta_{jil}x_l^{\mathrm{ul}}[i]}_{\text{Interference due to pilot contamination}}\right), \text{ as } \quad M\to\infty, \tag{2.44}$$

which confirms both the channel hardening effect and asymptotic favorable propagation. The decoding strategy which is widely used in the massive MIMO literature is called treating interference as noise (TIN). More precisely, assuming TIN in uplink, the BS in cell j only decodes the desired signal $x_j^{\mathrm{ul}}[i]$ while treating the remaining interfering signals $x_l^{\mathrm{ul}}[i],\ l\neq j$ as noise. Doing so, it has been shown in [26] that when MRC is used, the following lower bound is obtained on the uplink achievable rate:

$$I\left(\hat{y}_{ji}^{\mathrm{ul}}; x_j^{\mathrm{ul}}[i]\middle|\hat{\boldsymbol{g}}_{jij}\right) \overset{\text{a.s.}}{\to} C\left(\frac{\beta_{jij}^2}{\sum_{l=1,l\neq j}^{L}\beta_{jil}^2}\right), \text{ as } \quad M\to\infty. \tag{2.45}$$

ZF: Assuming that ZF combining is used at BS j to estimate signal $x_j^{\mathrm{ul}}[i]$ received from user i in cell j, (2.44) changes to the following:

$$\hat{y}_{ji}^{\mathrm{ul}} \overset{\text{a.s.}}{\to} \sqrt{\rho_{\mathrm{ul}}}\left(\underbrace{x_j^{\mathrm{ul}}[i]}_{\text{Desired signal}} + \underbrace{\sum_{l=1,l\neq j}^{L}\left(\frac{\beta_{jil}}{\beta_{jij}}\right)x_l^{\mathrm{ul}}[i]}_{\text{Interference due to pilot contamination}}\right), \text{ as } \quad M\to\infty, \tag{2.46}$$

i.e., both the channel hardening effect and asymptotic favorable propagation are observed here. Similar to MRC, assuming that BS j performs TIN by only decoding the desired signal $x_j^{\mathrm{ul}}[i]$ while treating the remaining interfering signals $x_l^{\mathrm{ul}}[i],\ l\neq j$ as noise, the following asymptotic lower bound is obtained

$$I\left(\hat{y}_{ji}^{\mathrm{ul}}; x_j^{\mathrm{ul}}[i]\middle|\hat{\boldsymbol{g}}_{jij}\right) \overset{\text{a.s.}}{\to} C\left(\frac{\beta_{jij}^2}{\sum_{l=1,l\neq j}^{L}\beta_{jil}^2}\right), \text{ as } \quad M\to\infty. \tag{2.47}$$

Interestingly, regardless of whether MRC or ZF is used in uplink, it is observed that the asymptotic limits on achievable rates are identical. In other words, as M grows to infinity, the performance of MRC and ZF converge in uplink [26, 111, 113].

Throughout this thesis, it is assumed that the noisy channel estimates $\hat{\boldsymbol{g}}_{jij}$ are known locally at the BSs. Thus, from now on, to simplify notation they will be omitted from the mutual information expressions.

Downlink

MRT: Analogously for downlink, now assume that MRT precoding is applied at BS j, i.e., $\boldsymbol{w}_{jkj} = \hat{\boldsymbol{g}}_{jkj}$. The normalization factor λ_j for this precoding vector can be found from (2.8) as follows

$$\begin{aligned}
\lambda_j^{\text{mrt}} &= \frac{\mathbb{E}\left[\text{tr}\left(\boldsymbol{W}_j^{\dagger}\boldsymbol{W}_j\right)\right]}{K} && (2.48)\\
&= \frac{\text{tr}\left(\mathbb{E}\left[\boldsymbol{W}_j^{\dagger}\boldsymbol{W}_j\right]\right)}{K}\\
&= \frac{\sum_{i=1}^{K}\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]}{K}\\
&= \frac{M\sum_{i=1}^{K}\sqrt{\rho_{\text{p}}}\beta_{jij}\alpha_{jij}}{K}. && (2.49)
\end{aligned}$$

A similar expression was derived in [32] using a slightly different notation for a spatially correlated Rayleigh fading channel. One can rewrite (2.5) for the i^{th} user in cell l as follows

$$y_{il}^{\text{dl}} = \sum_{j=1}^{L}\sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}}}\sum_{k=1}^{K}\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}s_j[k] + z_{il}. \tag{2.50}$$

Similar to (2.44), using Lemma 1 one can obtain

$$\frac{y_{il}^{\mathrm{dl}}}{\sqrt{M}} \overset{\text{a.s.}}{\to} \sqrt{K\rho_{\mathrm{dl}}\rho_{\mathrm{p}}} \left(\underbrace{\frac{\beta_{lil}\alpha_{lil}}{\sqrt{\sum_{k=1}^{K}\sqrt{\rho_{\mathrm{p}}}\beta_{lkl}\alpha_{lkl}}} s_l[i]}_{\text{Desired signal}} + \underbrace{\sum_{j=1,j\neq l}^{L} \frac{\beta_{jil}\alpha_{jij}}{\sqrt{\sum_{k=1}^{K}\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}}} s_j[i]}_{\text{Interference due to pilot contamination}} \right), \quad (2.51)$$

as $M \to \infty$, which again confirms both the channel hardening effect and asymptotic favorable propagation. Assuming TIN in downlink, user i in cell l only decodes the desired signal $s_l[i]$ and treats the remaining interfering signals $s_j[i]$, $j \neq l$, as noise. Hence, the following asymptotic lower bound on the downlink achievable rate is obtained

$$I\left(y_{il}^{\mathrm{dl}}; s_l[i]\right) \overset{\text{a.s.}}{\to} C\left(\frac{\beta_{lil}^2\alpha_{lil}^2 / \left(\sum_{k=1}^{K}\beta_{lkl}\alpha_{lkl}\right)}{\sum_{j=1,j\neq l}^{L}\beta_{jil}^2\alpha_{jij}^2 / \left(\sum_{k=1}^{K}\beta_{jkj}\alpha_{jkj}\right)} \right), \text{ as } \quad M \to \infty. \quad (2.52)$$

ZF: If ZF precoding is applied at BS j, i.e., $\boldsymbol{W}_j = \hat{\boldsymbol{G}}_{jj}\left(\hat{\boldsymbol{G}}_{jj}^{\dagger}\hat{\boldsymbol{G}}_{jj}\right)^{-1}$, the normalization factor λ_j from (2.8) is found as follows

$$\begin{aligned} \lambda_j^{\mathrm{zf}} &= \frac{\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{W}_j^{\dagger}\boldsymbol{W}_j\right]\right)}{K} \\ &= \frac{\sum_{i=1}^{K}\mathbb{E}\left[\boldsymbol{w}_{jij}^{\dagger}\boldsymbol{w}_{jij}\right]}{K} \\ &\overset{\text{(a)}}{=} \frac{1}{K(M-K)}\sum_{i=1}^{K}\frac{1}{\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}}, \end{aligned} \quad (2.53)$$

where (a) follows from a standard result in random matrix theory [114]. A similar expression was derived in [32] using a slightly different notation for a spatially correlated Rayleigh

fading channel. Using (2.41) and applying Lemma 1, it is obtained asymptotically

$$\frac{y_{il}^{\mathrm{dl}}}{\sqrt{M}} \overset{\text{a.s.}}{\to} \sqrt{K\rho_{\mathrm{dl}}} \left(\underbrace{\frac{1}{\sqrt{\sum_{k=1}^{K} \frac{1}{\sqrt{\rho_{\mathrm{p}}}\beta_{lkl}\alpha_{lkl}}}} s_l[i]}_{\text{Desired signal}} + \underbrace{\sum_{j=1,j\neq l} \frac{(\beta_{jil}/\beta_{jij})}{\sqrt{\sum_{k=1}^{K} \frac{1}{\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}}}} s_j[i]}_{\text{Interference due to pilot contamination}} \right), \quad (2.54)$$

as $M \to \infty$. Similar to MRT, by performing TIN, user i in cell l only decodes the desired signal $s_l[i]$ and treats the remaining interfering signals $s_j[i],\ j \neq l$, as noise. Doing so yields the following asymptotic lower bound on the downlink achievable rate

$$I\left(y_{il}^{\mathrm{dl}}; s_l[i]\right) \overset{\text{a.s.}}{\to} C\left(\frac{\frac{1}{\sum_{k=1}^{K}\left(1/\beta_{lkl}\alpha_{lkl}\right)}}{\sum_{j=1,j\neq l}^{L} \frac{\left(\beta_{jil}/\beta_{jij}\right)^2}{\sum_{k=1}^{K}\left(\beta_{jkj}\alpha_{jkj}\right)}} \right), \text{ as M} \to \infty. \quad (2.55)$$

Note that these asymptotic limits on the uplink/downlink achievable rates have been previously established in the literature with a slightly different notation (e.g., see the expressions of (7) and (11) in [91], (16) in [87], (4.51) and (4.52) in [111], (4.36) and (4.37) in [113]).

Remark 3. *One should note that as the number of BS antennas M becomes large enough in (2.44), (2.46), (2.51), and (2.54), the effects of all inter-user interference and noise terms vanish except the terms associated with the users in other cells using the same pilots. These terms are coherent inter-cell interference caused by pilot contamination. As explained in the previous chapter, in most works in the literature, a receiver only decodes the desired message coming from (or intended for) the main user inside its own cell while treating the interfering message of users sharing the same pilot in all other cells as noise. This approach, known as TIN, results in saturation of achievable rates even when M grows unbounded (see (2.45), (2.47), (2.52), and (2.55)).*

Remark 4. *It was shown in [111] that even with the assumption of a spatially correlated fading channel, applying TIN at the receiver yields the same conclusion and thus users' rate*

asymptotically saturate to a constant independent of M. It should, however, be pointed out that since the distributions of channel estimates and the estimation error are slightly more complicated in the case of a spatially correlated Rayleigh fading channel (see (2.14) *and* (2.15)*), more work is required to find the asymptotic limits of the achievable rates. For instance, the expressions of (4.49) and (4.50) in [111] show the asymptotically saturated rates after applying MRC and MRT, respectively.*

To avoid rate saturation, the decoding schemes proposed in the next two chapters of this thesis allow for the desired message from the current cell to be decoded along with either full decoding or partial decoding of the messages from users in other cells that are *sharing the same pilot as the desired user.*

Chapter 3

Performance of Full Interference Decoding

3.1 Interference Decoding

As discussed in the previous chapters, a common technique widely used in the literature to deal with the interference caused by pilot contamination in massive MIMO systems is to treat it as noise. More specifically, it was shown in Chapter 2 that, when the number of BS antennas grows unbounded the effects of additive noise, small-scale fading and non-coherent interference all vanish, except for the coherent interference due to pilot contamination from users of other cells sharing the same pilot as the desired user. When decoding the message of the desired user, by applying TIN at the receiver (i.e., treating the coherent pilot contamination interference as noise), the user's rate converges to a constant asymptotically and thus the benefits of increasing M saturate. For interference networks, while applying TIN is considered a better option when interference is weak [115, 116], it is known to be suboptimal in other interference regimes [71, 117]. In particular, when interference is strong, it is known that fully decoding interference along with the desired signal (also known as simultaneous or joint decoding) is optimal and achieves the sum-capacity of the interference channel (IC) [118]. While fully decoding interference or treating it as noise

are two extreme interference management strategies that are preferred in strong and weak interference regimes respectively, they are not always optimal. These two extreme policies will be bridged and reconciled by a partial interference decoding scheme presented in the next chapter.

Optimality and performance limits of schemes that work based on fully decoding interference have been extensively studied in the literature from an information theoretic perspective [71, 72, 117, 119–121]. In this chapter, we argue that when the number of BS antennas is sufficiently large, the power of inter-cell interference terms due to pilot contamination becomes strong enough that the interference can be decoded. More specifically, we show that when M is large enough, both simultaneous unique decoding (SD) and simultaneous non-unique decoding (SND) of pilot contamination interference outperform the conventional schemes based on TIN, which are commonly used in the massive MIMO literature. We further show that with a practical number of antennas M, SND outperforms both TIN and SD. Using a worst-case uncorrelated noise technique, we derive new closed-form expressions of achievable rates for the uplink and downlink of a multi-cell massive MIMO system with a spatially correlated/uncorrelated Rayleigh fading channel model. In addition, the performance of full interference decoding schemes (SD/SND) is investigated with MRC/MRT, ZF and RZF combining/precoding techniques, and it is shown that both RZF and ZF provide significantly higher rates than MRC/MRT, while RZF also outperforms ZF for a small number of antennas M. Structural results are also established in the case of a two-cell system with a symmetric setup as well as in the high SINR regime when $M \to \infty$. Moreover, the impact of increasing the number of cells, the number of users, cell radius, correlation magnitude as well as shadow fading on the performance of different schemes is numerically studied in this chapter.

3.2 Decoding Pilot Contamination

Following [38, 64, 65, 87, 91], it is assumed that the same set of orthogonal pilot sequences are used in all cells, resulting in pilot contamination interference. Hence, if one performs TIN, as the number of BS antennas M grows unbounded, the expressions of achievable

rates for uplink in (2.45) and (2.47), as well as those of the downlink in (2.52) and (2.55) saturate. In other words, treating interference due to pilot contamination as noise results in a fundamental limitation that constitutes a major bottleneck in the overall performance of massive MIMO systems with imperfect CSI [26].

In this chapter[1], as opposed to simply performing TIN, we propose a more advanced scheme based on full interference decoding. More specifically, we treat pilot contamination interference terms as individual users, similar to an interference channel (IC), and thus try to fully decode them. As will be seen in the subsequent part, this change of perspective results in new achievable rate expressions that grow without bound as $M \to \infty$.

3.2.1 Simultaneous unique Decoding (SD)

Note that in the expressions of the uplink received signal after performing MRC in (2.44) or ZF in (2.46), the first term is the desired signal and the remaining non-vanishing terms are all inter-cell interference caused by users in other cells that are sharing the same pilot sequence, $\boldsymbol{\psi}_i, i = 1, \ldots, K$, as the i^{th} user of cell j. Now, let us consider (2.43) which is the output of the j^{th} BS after performing MRC. If the baseband signals $\hat{y}_{ji}^{\text{ul}}$, for $j = 1, 2, ..., L,\ i = 1, ..., K$, are considered together, then these represent the outputs of K separate/non-interfering L-user ICs, one such L-user IC for each pilot sequence $\boldsymbol{\psi}_i, i = 1, \ldots, K$ as in Fig. 3.1: each L-user IC consists of L transmitters, i.e., the i^{th} user of each cell that is using the same pilot sequence $\boldsymbol{\psi}_i$, and L receivers, i.e., the BSs. One should also note that at each of the L receivers of each IC, an asymptotically noise-free L-user multiple access channel (MAC) is observed (see (2.44) for MRC and (2.46) for ZF). For instance, in the L-user MAC of (2.44), by uniquely jointly decoding the signals $\left[x_1^{\text{ul}}[i], x_2^{\text{ul}}[i], ..., x_L^{\text{ul}}[i]\right]^T$, unbounded rates are obtained as $M \to \infty$. A similar argument applies to ZF in uplink by considering the noise-free L-user MAC of (2.46).

Analogously for the downlink, by considering the received signals y_{il}^{dl}, for $l = 1, ..., L,\ i = 1, ..., K$, together we have the outputs of K separate/non-interfering L-user

[1]The results of this chapter were partially presented in [122].

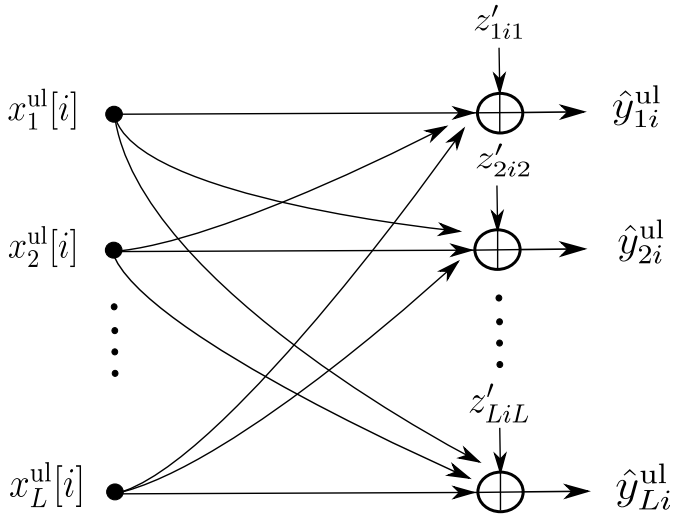


Figure 3.1: The L-user IC in uplink associated with the i^{th} users of each cell, sharing pilot sequence $\boldsymbol{\psi}_i$. There are a total of K separate/non-interfering such ICs in the network.

ICs, with input signals $[s_1[i], s_2[i], ..., s_L[i]]^T$ as in Fig. 3.2. Hence, the asymptotic expressions of (2.51) and (2.54) suggest that by performing MRT or ZF precoding at BSs, when M goes to infinity at a given user, say the i^{th} one in cell l, we have an asymptotically noise-free L-user MAC with infinite rate. Particularly, by uniquely jointly decoding the signals $[s_1[i], s_2[i], ..., s_L[i]]^T$, unbounded rates are obtained as $M \to \infty$.

Remark 5. *Note that since large-scale fading coefficients from contaminating users are unknown at the BS, and also the effective channel gains in the MAC of* (2.44) *are functions of these coefficients, this MAC can be regarded as a* compound *MAC [123], where the channel gains from transmitters to the receiver are unknown. It has been shown in [123] that the achievable rates of a compound MAC (i.e., a MAC with unknown channel gains) are the same as those of the standard MAC, where all channel gains are known. Therein, it has been shown that the lack of knowledge of channel gains at the receiver does not affect the achievable rates, i.e., the users' signal can still be successfully decoded. A similar argument is applied to the downlink if the large-scale fading coefficients (and thus the effective channel gains) are unknown at the users, and therefore the corresponding downlink L-user MACs can be treated as a* compound *MAC.*

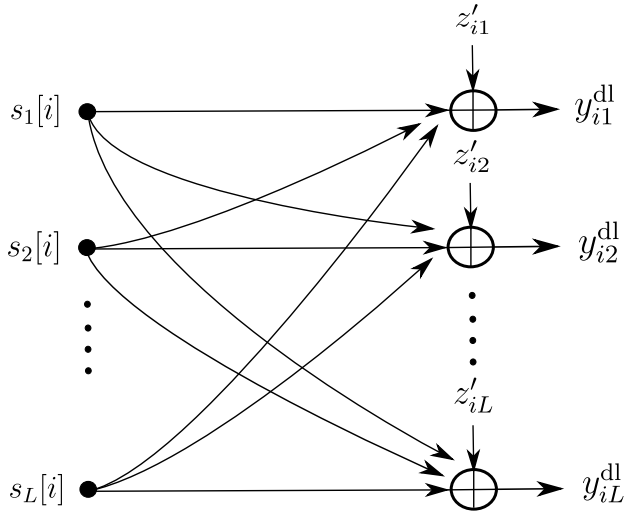


Figure 3.2: The L-user IC in downlink associated with the i^{th} user of each cell, sharing pilot sequence $\boldsymbol{\psi}_i$. There are a total of K separate/non-interfering such ICs in the network.

An intuitive way to understand why the result explained in Remark 5 is true is the following. Consider, for instance, the case of uplink where each user transmitting with codewords of length n can use $\log(n)$ training symbols interspersed across the frame to identify the channel gains at the receiver. As n grows, the resulting loss in rate due to $\log(n)$ can be made arbitrarily small. Moreover, as mentioned in the previous chapter, the large-scale fading coefficients remain fixed for many channel coherence intervals. Since we code over multiple coherence intervals, the BS essentially needs to estimate the large-scale fading coefficients only once for a number of coherence intervals. Therefore, if one allocates a small number of symbols across multiple coherence blocks, a negligible rate loss would be incurred.

Due to the finite coherence time of wireless channels resulting from user mobility, only a limited number of orthogonal pilot sequences are available. One way to address this issue is to re-use the *same* set of orthonormal pilots across all cells as described in Section 2.3.3. However, an alternative approach to that of Section 2.3.3 would be to use *different* sets of orthonormal pilots in different cells. To illustrate this alternative, assume that a single set of orthonormal pilots is picked for one cell, and different rotated versions of this set are

used in all other cells. In particular, user k in cell l transmits the pilot sequence $\boldsymbol{\psi}_{kl}$ to its BS, where the entire pilot matrix used in cell l is denoted by $\boldsymbol{\Psi}_l$. As the sequences of other pilot matrices, $\boldsymbol{\Psi}_j,\ j \neq l \in \{1, 2, ..., L\}$, are rotated versions of sequences in $\boldsymbol{\Psi}_l$, they have non-zero inner product.

After transmission of all pilot sequences, the BS in cell j receives the matrix $\boldsymbol{Y}_j^{\mathrm{p}} \in \mathbb{C}^{M\times K}$, given by

$$\boldsymbol{Y}_j^{\mathrm{p}} = \sum\nolimits_{l=1}^{L} \sqrt{\rho_{\mathrm{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi}_l + \boldsymbol{Z}_j. \tag{3.1}$$

Multiplying $\boldsymbol{Y}_j^{\mathrm{p}}$ by $\boldsymbol{\Psi}_j^{\dagger}$, the k^{th} column of the resulting matrix is

$$\boldsymbol{r}_{jk} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkj} + \sum\nolimits_{l=1,l\neq j}^{L} \sqrt{\rho_{\mathrm{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi}_l\boldsymbol{\psi}_{kj}^{\dagger} + \boldsymbol{q}_{jk}, \tag{3.2}$$

where $\boldsymbol{q}_{jk} \sim \mathcal{CN}(\boldsymbol{0},\ \boldsymbol{I}_M)$. Therefore, the MMSE estimate $\hat{\boldsymbol{g}}_{jkj}$ of $\boldsymbol{g}_{jkj}$ based on the observation $\boldsymbol{r}_{jk}$ is

$$\begin{aligned}\hat{\boldsymbol{g}}_{jkj} = {} & \mathbb{E}\left[\boldsymbol{g}_{jkj}\boldsymbol{r}_{jk}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{r}_{jk}\boldsymbol{r}_{jk}^{\dagger}\right]^{-1} \\ & \times\left(\sqrt{\rho_{\mathrm{p}}}\boldsymbol{g}_{jkj} + \sum\nolimits_{l=1,l\neq j}^{L}\sqrt{\rho_{\mathrm{p}}}\boldsymbol{G}_{jl}\boldsymbol{\Psi}_l\boldsymbol{\psi}_{kj}^{\dagger} + \boldsymbol{q}_{jk}\right).\end{aligned} \tag{3.3}$$

One can readily see from (3.3) that the channel estimate $\hat{\boldsymbol{g}}_{jkj}$ is now contaminated by the channel of all users in other cells. Thus, by letting $M \to \infty$, the non-vanishing terms in (2.44), (2.46), (2.51), and (2.54) will include the signal of every user in every other cell. In turn, when decoding pilot contamination interference, using the same set of pilots in different cells results in decoding L users, whereas using different sets of pilots in different cells, as explained above, results in decoding $K \times (L-1) + 1$ users. As will be explained in Remark 10 later in this chapter, this alternative approach that requires decoding $K\times(L-1)+1$ users (instead of L users) degrades the performance of interference decoding schemes, as compared to the approach of Section 2.3.3. Moreover, the complexity of jointly decoding $K \times (L-1) + 1$ users is larger than that of decoding L users. Hence, when decoding pilot contamination interference, using the same set of orthogonal pilots in different cells (as opposed to different sets of pilots) is preferable as it results in fewer interference terms to be decoded.

Below, we provide a detailed analysis of the uplink/downlink achievable rates by fully decoding pilot contamination interference terms along with the intended user in the finite M regime.

Uplink

In order to establish achievable rate lower bounds, we first need to generalize the worst-case uncorrelated noise technique of (2.30) to the case of a multi-user channel, i.e., MAC. This is formally presented in the following lemma.

Lemma 2. *Consider the L-user MAC given by* $y = \sum_{i=1}^{L} x_i^G + z$, *where the users' signals* $x_i^G,\ i = 1, ..., L$ *are independent with complex Gaussian distribution* $x_i^G \sim \mathcal{CN}(0,\ P_i)$, *and the additive noise* z *is a complex random variable with mean zero and variance* σ_z^2. *If* z *is uncorrelated from* $x_i^G,\ i = 1, ..., L$, *then*

$$I\left(\boldsymbol{x}_{\Omega}^{G};\ y^{G}\middle|\boldsymbol{x}_{\Omega^c}^{G}\right) \leq I\left(\boldsymbol{x}_{\Omega}^{G};\ y\middle|\boldsymbol{x}_{\Omega^c}^{G}\right), \tag{3.4}$$

where $\boldsymbol{x}_{\Omega}^{G}$ *is the vector with entries* $x_i^G, i \in \Omega \subseteq \{1, 2, ..., L\}, \Omega \neq \emptyset$, $\Omega^c := \{1, 2, ..., L\} \setminus \Omega$, $y^G = \sum_{i=1}^{L} x_i^G + z^G$, *and* $z^G \sim \mathcal{CN}(0,\ \sigma_z^2)$.

Proof. See Appendix A.1. □

Note that using Lemma 2, one can obtain an achievable lower bound on the capacity of a MAC with uncorrelated additive non-Gaussian noise by replacing the noise term with an independent zero mean Gaussian noise having the same variance. This is a natural extension of the worst-case uncorrelated noise result of [112], discussed in the previous chapter for a point-to-point channel, to the case of a multi-user channel. When the additive noise is independent of the users' signals, Gaussian noise has been proven to be the worst-case noise for point-to-point, MAC, degraded broadcast and MIMO channels [115]. However, the proof provided in Lemma 2 only requires the additive noise to be uncorrelated of the users' signals.

Let us revisit the received baseband signal of (2.33) in the uplink for a general combining vector $\boldsymbol{v}_{jij}$ applied at BS j. Following the above discussion, the i^{th} users of all cells, $\left\{x_l^{\text{ul}}[i]\right\}_{l=1}^{L}$, sharing pilot sequence $\boldsymbol{\psi}_i$, are now treated as the desired signals to be decoded jointly. Following an approach similar to that of (2.33), by adding and subtracting a term associated with the mean of the effective channel $\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}$ in (2.43), the expression below is obtained over one coherence interval

$$\hat{y}_{ji}^{\text{ul}} = \underbrace{\sqrt{\rho_{\text{ul}}}\sum\nolimits_{l=1}^{L}\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]x_l^{\text{ul}}[i]}_{\text{Desired signals}} + \underbrace{\sum\nolimits_{l=1}^{L}\sqrt{\rho_{\text{ul}}}\left(\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil} - \mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right)x_l^{\text{ul}}[i]}_{\text{Interference due to beamforming gain uncertainty}}$$
$$+ \underbrace{\sum\nolimits_{l=1}^{L}\sum\nolimits_{k=1,k\neq i}^{K}\sqrt{\rho_{\text{ul}}}\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}x_l^{\text{ul}}[k]}_{\text{Interference caused by other users}} + \underbrace{\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_j}_{\text{Noise}} \tag{3.5}$$
$$= \sum\nolimits_{l=1}^{L}\eta_{il}x_l^{\text{ul}}[i] + z'_{jij}, \tag{3.6}$$

where $\eta_{il} := \sqrt{\rho_{\text{ul}}}\mathbb{E}[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}]$ and z'_{jij} is the effective noise term incorporating the last three terms in (3.5).

Note that the power of the desired signals in (3.6) is proportional to $|\eta_{il}|^2$ and is thus proportional to M^2. Moreover, the power of the effective noise term z'_{jij} is proportional to M. Therefore, by uniquely jointly decoding the signals $\left\{x_l^{\text{ul}}[i]\right\}_{l=1}^{L}$ in (3.6), the achievable rates of the corresponding MAC grow unboundedly as $M \to \infty$.

We now consider the MAC of (3.6) at BS j. Using the usual definitions as in [124], each message $m_l \in [1 : 2^{nR_{il}^{\text{ul}}}]$, $l = 1, ..., L$ (distributed uniformly) is encoded into the codeword $\boldsymbol{x}_l^{\text{ul},n}[i](m_l)$ of length n which is generated iid $\mathcal{CN}(0,\ 1)$. By applying simultaneous unique decoding (SD) and the standard random coding analysis as in [124], it can be shown that the decoding error probability tends to zero as $n \to \infty$, i.e., the rate tuple $\left(R_{i1}^{\text{ul}}, ..., R_{iL}^{\text{ul}}\right)$ is achievable, if

$$\sum\nolimits_{l\in\Omega} R_{il}^{\text{ul}} \leq I\left(\hat{y}_{ji}^{\text{ul}};\ \boldsymbol{x}_{\Omega}^{\text{ul}} \,\middle|\, \boldsymbol{x}_{\Omega^c}^{\text{ul}}\right), \tag{3.7}$$

for all $\Omega \subseteq S = \{1, 2, ..., L\}$, $\Omega \neq \emptyset$, and $\boldsymbol{x}_{\Omega}^{\text{ul}}$ is the vector with entries $x_l^{\text{ul}}[i], l \in \Omega$. The set of rate vectors $\left(R_{i1}^{\text{ul}}, ..., R_{iL}^{\text{ul}}\right)$ that satisfies the inequalities of (3.7) defines the achievable region at BS j, denoted by $\mathcal{R}_{ij}^{\text{SD}}$, associated with uniquely jointly decoding the signals

$\left\{x_l^{\mathrm{ul}}[i]\right\}_{l=1}^{L}$ at this BS. Finally, to obtain the achievable region network-wide (at all BSs) for the i^{th} pilot-sharing users, one should take the intersection of achievable regions over all BSs, i.e., $\mathscr{R}_i^{\mathrm{SD}} = \bigcap_j \mathcal{R}_{ij}^{\mathrm{SD}}, j = 1, ..., L$.

Note that the effective noise term in (3.6), z'_{jij}, which contains the last three terms in (3.5) involves the inner product of random vectors such as $\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_j$, and hence is neither Gaussian nor independent of the users' signals. However, as will be shown below, it is uncorrelated from the users' signals.

The signals $x_l^{\mathrm{ul}}[i]$ and $x_j^{\mathrm{ul}}[k]$ are independent for $(l, i) \neq (j, k)$, and the interference term associated with channel estimation error is zero mean and thereby uncorrelated from the desired signals. It is easy to verify that the last two terms in (3.5) are also uncorrelated from the desired signals. Therefore, for transmission over multiple coherence intervals all interference and noise terms in (3.5) are uncorrelated from the desired signal components. Thus, applying Lemma 2, an achievable lower bound on $I(\hat{y}_{ji}^{\mathrm{ul}};\ \boldsymbol{x}_{\Omega}^{\mathrm{ul}} \mid \boldsymbol{x}_{\Omega^c}^{\mathrm{ul}})$ in (3.7), similar to (2.36) can be obtained. More specifically, replacing the uncorrelated noise z'_{jij} in (3.6) by an independent Gaussian noise with a variance equal to the sum of the variances of the interference and noise terms in (3.5), provides a lower bound on (3.7). This is formally presented in the following theorem.

Theorem 1. *Assuming Gaussian signaling, i.e.,* $\left[x_1^{\mathrm{ul}}[i], x_2^{\mathrm{ul}}[i], ..., x_L^{\mathrm{ul}}[i]\right]^T \sim \mathcal{CN}\left(\boldsymbol{0},\ \boldsymbol{I}_L\right)$ *for* $i \in \{1, 2, ..., K\}$*, the following set of achievable lower bounds is obtained for the MAC given in* (3.6) *at BS* j

$$I\left(\hat{y}_{ji}^{\mathrm{ul}};\ \boldsymbol{x}_{\Omega}^{\mathrm{ul}} \,\middle|\, \boldsymbol{x}_{\Omega^c}^{\mathrm{ul}}\right) \geq C\left(\frac{\sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2}{\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right] - \sum_{l=1}^{L}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2 + \mathbb{E}\left[\|\boldsymbol{v}_{jij}\|^2\right]}\right) \tag{3.8}$$

$$:= C_{\mathrm{LB}}^{\mathrm{ul}}(\Omega), \quad \forall \Omega \subseteq \{1, 2, ..., L\},$$

where the expectations are taken with respect to the channel realizations.

Proof. See Appendix A.2. □

Note that the numerator in (3.8) is the variance of the desired signal $\sum_{l\in\Omega}\eta_{il}x_l^{\mathrm{ul}}[i]$, and the denominator is the variance of the uncorrelated additive noise z'_{jij}.

Remark 6. *The lower bound of* (3.8) *is valid in general and does not depend on specific choices of the linear combining vector* $\boldsymbol{v}_{jij}$ *or channel distributions, and the expectations can be computed using the Monte Carlo simulation.*

Downlink

A similar analysis can be applied to the downlink in (2.37). Specifically, considering the MAC obtained on the downlink at user i in cell l, the intended symbols for the i^{th} users of all cells, i.e., $\{s_j[i]\}_{j=1}^L$ are now treated as the desired signals to be decoded jointly. Therefore, (2.37) changes to the following

$$y_{il}^{\mathrm{dl}} = \underbrace{\sum_{j=1}^{L}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]s_j[i]}_{\text{Desired signals}} + \underbrace{\sum_{j=1}^{L}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\left(\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij} - \mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]\right)s_j[i]}_{\text{Interference due to beamforming gain uncertainty}} \tag{3.9}$$

$$+ \underbrace{\sum\nolimits_{j=1}^{L}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\sum\nolimits_{k=1,k\neq i}^{K}\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}s_j[k]}_{\text{Inter-cell interference caused by other users}} + \underbrace{z_{il}}_{\text{Noise}}$$

$$= \underbrace{\sum_{j=1}^{L}\zeta_{ij}s_j[i]}_{\text{Desired signals}} + \underbrace{z'_{il}}_{\text{Additive noise}}, \tag{3.10}$$

where (3.10) is the output of the MAC obtained at user i in cell l, with input signals $\{s_j[i]\}_{j=1}^L$, and z'_{il} is the effective noise incorporating the last three terms of (3.9).

The power of the desired signals in (3.9) is proportional to $|\zeta_{ij}|^2$ and is thus proportional to M^2. Moreover, the power of the effective noise term z'_{il} is proportional to M. Hence, by unique joint decoding of the input signals $\{s_j[i]\}_{j=1}^L$ in (3.10), the achievable rates of the corresponding MAC grow without bound as $M\to\infty$.

Considering the MAC of (3.10) at user i in cell l, it is assumed that each message

$w_j \in \left[1 : 2^{nR_{ij}^{\mathrm{dl}}}\right]$, $j = 1, ..., L$ (distributed uniformly) is communicated over this MAC by encoding it into the codeword $\boldsymbol{s}_j^n[i](w_j)$ of length n, generated i.i.d. $\mathcal{CN}(0, 1)$. Therefore, by joint decoding of all desired messages uniquely as in SD, decoding error probability can be shown to approach zero for $n \to \infty$, if the set of rate tuples $\left(R_{i1}^{\mathrm{dl}}, ..., R_{iL}^{\mathrm{dl}}\right)$ is such that

$$\sum_{j \in \Omega} R_{ij}^{\mathrm{dl}} \leq I\left(y_{il}^{\mathrm{dl}};\ \boldsymbol{s}_{\Omega} \,\middle|\, \boldsymbol{s}_{\Omega^c}\right), \tag{3.11}$$

for all $\Omega \subseteq S = \{1, 2, ..., L\}$, and $\boldsymbol{s}_{\Omega}$ is the vector with entries $s_j[i]$, $j \in \Omega$. Finally, the network-wide achievable region is obtained by taking the intersection of the MAC achievable regions over all receivers.

Since $s_l^{\mathrm{dl}}[i]$ and $s_j^{\mathrm{dl}}[i]$ are independent for $l \neq j$, with a similar argument to that of uplink one can verify that for transmission over multiple coherence intervals the zero mean effective noise term z'_{il} is uncorrelated from the desired signal components. Hence, we are able to establish achievable lower bounds for the mutual information terms in (3.11), based on the worst-case uncorrelated noise technique of Lemma 2. In particular, as formally presented in the following theorem, the effective noise term z'_{il} is replaced by an independent Gaussian noise having the same variance.

Theorem 2. *Assuming Gaussian signaling, i.e.,* $[s_1[i], s_2[i], ..., s_L[i]]^T \sim \mathcal{CN}(\boldsymbol{0},\ \boldsymbol{I}_L)$ *for* $i \in \{1, 2, ..., K\}$*, the following set of achievable lower bounds is obtained for the MAC of* (3.10) *at user* i *in cell* l

$$I\left(y_{il}^{\mathrm{dl}};\ \boldsymbol{s}_{\Omega} \middle| \boldsymbol{s}_{\Omega^c}\right) \geq C\left(\frac{\sum_{j \in \Omega} \frac{\rho_{\mathrm{dl}}}{\lambda_j} \left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger} \boldsymbol{w}_{jij}\right]\right|^2}{\sum_{j=1}^{L} \sum_{k=1}^{K} \frac{\rho_{\mathrm{dl}}}{\lambda_j} \mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger} \boldsymbol{w}_{jkj}\right|^2\right] - \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j} \left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger} \boldsymbol{w}_{jij}\right]\right|^2 + 1}\right), \tag{3.12}$$

$$:= C_{\mathrm{LB}}^{\mathrm{dl}}(\Omega), \quad \forall \Omega \subseteq \{1, 2, ..., L\},$$

where the expectations are taken with respect to the channel realizations, and $\lambda_j, j = 1, ..., L$ *is found based on the choice of precoding vector* $\boldsymbol{w}_{jij}$*.*

Proof. See Appendix A.3. □

As the numerator in the lower bound of (3.12) is the variance of the desired signal and the denominator is the variance of the effective noise term z'_{il}, following the discussion below (3.10) this achievable lower now grows without bound as $M \to \infty$.

Remark 7. *The lower bound of* (3.12) *is valid in general and does not depend on specific choices of the linear precoding vector* $\boldsymbol{w}_{jij}$ *or channel distributions, and the expectations can be computed using the Monte Carlo simulation.*

In the following, by using a spatially correlated/uncorrelated Rayleigh fading channel model, we specialize the uplink/downlink lower bounds of (3.8) and (3.12) to specific choices of the linear combining/precoding schemes discussed in the previous chapter.

3.2.2 Spatially Correlated Rayleigh Fading

Uplink

Consider the uplink achievable rate of Theorem 1, and assume that the channel estimates $\hat{\boldsymbol{g}}_{jij}$ are obtained for a spatially correlated Rayleigh Fading model using the MMSE estimator described in (2.13), (2.14) and (2.15). When applying MRC at BS j, i.e., $\boldsymbol{v}_{jij} = \hat{\boldsymbol{g}}_{jij}$, the result presented in the following theorem computes the closed-form expression of this lower bound.

Theorem 3. *Using MRC combining based on the MMSE estimator for a spatially correlated Rayleigh fading channel model, the lower bound in* (3.8) *for finite* M *is found as below*

$$I\left(\hat{y}^{\mathrm{ul}}_{ji};\ \boldsymbol{x}^{\mathrm{ul}}_{\Omega} \,\middle|\, \boldsymbol{x}^{\mathrm{ul}}_{\Omega^c}\right) \geq C\left(\frac{\sum_{l\in\Omega}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\left|\operatorname{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right|^2}{\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}\operatorname{tr}\left(\boldsymbol{R}_{jkl}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)+\rho_{\mathrm{p}}\operatorname{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)}\right), \tag{3.13}$$

for all $\Omega \subseteq \{1,...,L\}$ *and* $\boldsymbol{\Lambda}_{ji}$ *is given in* (2.18).

Proof. See Appendix A.4. □

It is readily confirmed that the tr(.) operation in (3.13) scales proportional to M, and thus the numerator, which is the variance of the desired signal components, is proportional to M^2. On the other hand, the denominator, which is the variance of non-coherent interference plus the variance of noise, is proportional to M, and thereby the lower bound of (3.13) scales as $\mathcal{O}(\log M)$.

Downlink

Now, let us consider the downlink achievable rate of Theorem 2 for a spatially correlated Rayleigh fading channel model, and also assume that MRT precoding is applied at all BSs, i.e., $\boldsymbol{w}_{jij} = \hat{\boldsymbol{g}}_{jij}$. First, note that the normalization factor λ_j^{mrt} can be obtained as follows

$$\lambda_j^{\text{mrt}} = \frac{\sum_{i=1}^{K} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \hat{\boldsymbol{g}}_{jij}\right]}{K} \tag{3.14}$$

$$\overset{(a)}{=} \frac{\sum_{i=1}^{K} \text{tr}\left(\rho_{\text{p}} \boldsymbol{R}_{jij} \boldsymbol{\Lambda}_{ji}^{-1} \boldsymbol{R}_{jij}\right)}{K}, \tag{3.15}$$

where (a) follows from the distribution of $\hat{\boldsymbol{g}}_{jij}$ in (2.14). One can use the following theorem to find the closed-form expression of the lower bound in (3.12).

Theorem 4. *If MRT precoding based on the MMSE estimator is used at all BSs, the lower bound in (3.12) for finite M is found as below*

$$I\left(y_{il}^{\text{dl}};\ \boldsymbol{s}_{\Omega} \middle| \boldsymbol{s}_{\Omega^c}\right) \geq C\left(\frac{\sum_{j\in\Omega} \frac{\rho_{\text{dl}}\rho_{\text{p}}^2}{\lambda_j^{\text{mrt}}} \left| \text{tr}\left(\boldsymbol{R}_{jij} \boldsymbol{\Lambda}_{ji}^{-1} \boldsymbol{R}_{jil}\right) \right|^2}{\sum_{j=1}^{L} \sum_{k=1}^{K} \frac{\rho_{\text{dl}}\rho_{\text{p}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\boldsymbol{R}_{jil} \boldsymbol{R}_{jkj} \boldsymbol{\Lambda}_{jk}^{-1} \boldsymbol{R}_{jkj}\right) + 1}\right), \tag{3.16}$$

for all $\Omega \subseteq \{1, ..., L\}$ and λ_j^{mrt} is given in (3.15).

Proof. See Appendix A.5. □

Similar to the uplink, the numerator is the variance of the desired signal components, while the denominator is the variance of non-coherent interference plus the variance of noise. First, observe that λ_j^{mrt} in (3.14) grows linearly with M. Hence, it can be verified that as $M \to \infty$ the numerator in (3.16) scales proportional to M, whereas the denominator saturates; hence, the lower bound of (3.16) scales as $\mathcal{O}(\log M)$.

3.2.3 Uncorrelated Rayleigh Fading

In the following, we specialize the expressions of the uplink/downlink achievable lower bounds to the case of a spatially uncorrelated Rayleigh fading channel model, and further find the bounds of ZF combining/precoding for this special case.

3.2.4 Uplink

Let us consider the special case of an MMSE estimator for a spatially uncorrelated Rayleigh fading channel model given in (2.22) with corresponding distributions in (2.23) and (2.24). Using the substitution $\boldsymbol{R}_{jij} = \beta_{jij}\boldsymbol{I}_M$, the following is obtained

$$\boldsymbol{\Lambda}_{ji} = \left(1 + \sum_{l=1}^{L} \rho_{\text{p}} \beta_{jil}\right) \boldsymbol{I}_M. \tag{3.17}$$

MRC

Assuming MRC, one can use the following corollary to simplify the uplink achievable rate of (3.13).

Corollary 1. *Assuming that MMSE estimates are obtained for a spatially uncorrelated Rayleigh fading channel model, the lower bound in (3.8) (and therefore the one in (3.13)) simplifies to the following*

$$I\left(\hat{y}_{ji}^{\text{ul}};\ \boldsymbol{x}_{\Omega}^{\text{ul}} \,\middle|\, \boldsymbol{x}_{\Omega^c}^{\text{ul}}\right) \geq C\left(\frac{M^2 \sum_{l\in\Omega} \rho_{\text{ul}} \rho_{\text{p}} \beta_{jil}^2 \alpha_{jij}^2}{M\sqrt{\rho_{\text{p}}}\beta_{jij}\alpha_{jij}\left(\sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\text{ul}}\beta_{jkl} + 1\right)}\right), \tag{3.18}$$

for all $\Omega \subseteq \{1, ..., L\}$ *and* $\alpha_{jij} = \dfrac{\sqrt{\rho_{\mathrm{p}}}\beta_{jij}}{1 + \sum_{l=1}^{L} \rho_{\mathrm{p}}\beta_{jil}}$.

Proof. See Appendix A.6. □

As the array gains in the numerator and denominator of (3.18) are now apparent, it can be observed that the uplink achievable rate now grows as $\mathcal{O}(\log M)$.

ZF

Assuming that BS j applies ZF combining to estimate the signal of the i^{th} user in cell j, we obtain at BS j

$$\hat{y}_{ji}^{\mathrm{ul}} = \boldsymbol{v}_{jij}^{\dagger}\boldsymbol{y}_{j}^{\mathrm{ul}} \tag{3.19}$$

$$\overset{(a)}{=} \sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\left(\hat{\boldsymbol{g}}_{jkl} + \boldsymbol{\epsilon}_{jkl}\right) x_{l}^{\mathrm{ul}}[k] + \boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_{j}$$

$$= \underbrace{\sum_{l=1}^{L} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}x_{l}^{\mathrm{ul}}[i]}_{\text{Desired signals}}$$

$$+ \underbrace{\sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jkl}x_{l}^{\mathrm{ul}}[k]}_{\text{Interference caused by other users}} + \underbrace{\sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jkl}x_{l}^{\mathrm{ul}}[k]}_{\text{Interference due to estimation error}} + \underbrace{\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_{j}}_{\text{Noise}} \tag{3.20}$$

$$\overset{(b)}{=} \underbrace{\sum_{l=1}^{L} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\left(\frac{\beta_{jil}}{\beta_{jij}}\right)\hat{\boldsymbol{g}}_{jij}x_{l}^{\mathrm{ul}}[i]}_{\text{Desired signals}}$$

$$+ \underbrace{\sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\left(\frac{\beta_{jkl}}{\beta_{jkj}}\right)\hat{\boldsymbol{g}}_{jkj}x_{l}^{\mathrm{ul}}[k]}_{\text{Interference caused by other users}} + \underbrace{\sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{ul}}}\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jkl}x_{l}^{\mathrm{ul}}[k]}_{\text{Interference due to estimation error}} + \underbrace{\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{n}_{j}}_{\text{Noise}}, \tag{3.21}$$

where (a) follows from the decomposition $\boldsymbol{g}_{jij} = \hat{\boldsymbol{g}}_{jij} + \boldsymbol{\epsilon}_{jij}$, and (b) is due to (2.25). As already discussed in Chapter 2, by applying ZF the second term in (3.21) disappears as a

result of (2.41). Therefore, using (2.41), one can further simplify (3.21) as below

$$\hat{y}_{ji}^{\mathrm{ul}} = \underbrace{\sum_{l=1}^{L} \sqrt{\rho_{\mathrm{ul}}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right) x_l^{\mathrm{ul}}[i]}_{\text{Desired signals}} + \underbrace{\sum_{l=1}^{L}\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{ul}}} \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{\epsilon}_{jkl} x_l^{\mathrm{ul}}[k]}_{\text{Interference due to channel estimation error}} + \underbrace{\boldsymbol{v}_{jij}^{\dagger} \boldsymbol{n}_j}_{\text{Noise}} \tag{3.22}$$

$$= \underbrace{\sum_{l=1}^{L} \nu_{il} x_l^{\mathrm{ul}}[i]}_{\text{Desired signals}} + \underbrace{z''_{jij}}_{\text{Additive noise}}, \tag{3.23}$$

where $\nu_{il} = \sqrt{\rho_{\mathrm{ul}}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)$ and z''_{jij} is the effective noise term which is neither Gaussian nor independent of the desired signals. With a similar argument as before, however, it can be readily verified that the term associated with interference-plus-noise in (3.22) is zero mean and uncorrelated from the desired signal components. Consequently, one can apply the worst-case uncorrelated noise technique of Lemma 2 to obtain an achievable lower bound for ZF in uplink. The following theorem establishes this result.

Theorem 5. *Assuming ZF and also Gaussian signaling, i.e.,* $\left[x_1^{\mathrm{ul}}[i], x_2^{\mathrm{ul}}[i], ..., x_L^{\mathrm{ul}}[i]\right]^T \sim \mathcal{CN}\left(\boldsymbol{0},\, \boldsymbol{I}_L\right)$ *for* $i \in \{1, 2, ..., K\}$*, the following set of achievable rates is obtained for the MAC given in* (3.23) *at BS* j

$$I\left(\hat{y}_{ji}^{\mathrm{ul}};\, \boldsymbol{x}_{\Omega}^{\mathrm{ul}} \,\middle|\, \boldsymbol{x}_{\Omega^c}^{\mathrm{ul}}\right) \geq C\left(\frac{\sum_{l\in\Omega} \rho_{\mathrm{ul}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2}{\frac{1}{(M-K)\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}} \left(\sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\mathrm{ul}} \left(\beta_{jkl} - \sqrt{\rho_{\mathrm{p}}}\beta_{jkl}\alpha_{jkl}\right) + 1\right)}\right), \tag{3.24}$$

for all $\Omega \subseteq \{1, ..., L\}$.

Proof. See Appendix A.7. □

Using further algebraic simplifications, one can simplify (3.24) as follows

$$I\left(\hat{y}_{ji}^{\mathrm{ul}};\ \boldsymbol{x}_{\Omega}^{\mathrm{ul}}\ \middle|\ \boldsymbol{x}_{\Omega^c}^{\mathrm{ul}}\right) \geq C\left(\frac{(M-K)\sqrt{\rho_{\mathrm{p}}}\rho_{\mathrm{ul}}\sum_{l\in\Omega}\beta_{jil}\alpha_{jil}}{\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\left(\beta_{jkl}-\sqrt{\rho_{\mathrm{p}}}\beta_{jkl}\alpha_{jkl}\right)+1}\right). \tag{3.25}$$

Therefore, in the asymptotic regime where $M \to \infty$, the uplink achievable rates for ZF scale as $\mathcal{O}(\log M)$ and thus grow without bound.

3.2.5 Downlink

Considering the MRT/ZF precoding, we now specialize the downlink achievable rate of (3.12) to the case of a spatially uncorrelated Rayleigh fading channel model.

MRT

Assuming MRT, one can use the following corollary to simplify the downlink lower bound in (3.16).

Corollary 2. *Assuming that MMSE estimates are obtained for a spatially uncorrelated Rayleigh fading channel model, the lower bound in* (3.12) *(and therefore the one in* (3.16)*) simplifies to the following*

$$I\left(y_{il}^{\mathrm{dl}};\ \boldsymbol{s}_{\Omega}\middle|\boldsymbol{s}_{\Omega^c}\right) \geq C\left(\frac{M^2\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}}{\lambda_j^{\mathrm{mrt}}}\beta_{jil}^2\alpha_{jij}^2}{M\sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\sum_{k=1}^{K}\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\beta_{jil}+1}\right), \tag{3.26}$$

for all $\Omega \subseteq \{1,...,L\}$ *and* λ_j^{mrt} *is given in* (2.48).

Proof. See Appendix A.8. □

Noting the array gains in the numerator and denominator of (3.26), it is observed that the downlink rate of MRT scales as $\mathcal{O}(\log M)$.

ZF

Using ZF precoding, a similar analysis can be applied to the downlink at the i^{th} user in cell l. Thus, one can write

$$y_{il}^{\text{dl}} = \sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1}^{K} \left(\hat{\boldsymbol{g}}_{jil}^{\dagger} + \boldsymbol{\epsilon}_{jil}^{\dagger} \right) \boldsymbol{v}_{jkj} s_j[k] + w_{il} \tag{3.27}$$

$$= \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j}} \hat{\boldsymbol{g}}_{jil}^{\dagger} \boldsymbol{v}_{jij} s_j[i]}_{\text{Desired signals}}$$

$$+ \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1,k\neq i}^{K} \hat{\boldsymbol{g}}_{jil}^{\dagger} \boldsymbol{v}_{jkj} s_j[k]}_{\text{Non-coherent interference caused by other users}} + \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1}^{K} \boldsymbol{\epsilon}_{jil}^{\dagger} \boldsymbol{v}_{jkj} s_j[k]}_{\text{Iterference due to estimation error}} + \underbrace{w_{il}}_{\text{Noise}} \tag{3.28}$$

$$\overset{\text{(a)}}{=} \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \left(\frac{\beta_{jil}}{\beta_{jij}} \right) \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{v}_{jij} s_j[i]}_{\text{Desired signals}}$$

$$+ \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1,k\neq i}^{K} \left(\frac{\beta_{jil}}{\beta_{jij}} \right) \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{v}_{jkj} s_j[k]}_{\text{Non-coherent interference}} + \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1}^{K} \boldsymbol{\epsilon}_{jil}^{\dagger} \boldsymbol{v}_{jkj} s_j[k]}_{\text{Interference due to estimation error}} + \underbrace{w_{il}}_{\text{Noise}} \tag{3.29}$$

$$\overset{\text{(b)}}{=} \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \left(\frac{\beta_{jil}}{\beta_{jij}} \right) s_j[i]}_{\text{Desired signals}} + \underbrace{\sum_{j=1}^{L} \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \sum_{k=1}^{K} \boldsymbol{\epsilon}_{jil}^{\dagger} \boldsymbol{v}_{jkj} s_j[k] + w_{il}}_{\text{Interference + Noise}} \tag{3.30}$$

$$= \underbrace{\sum_{j=1}^{L} \theta_{ij} s_j[i]}_{\text{Desired signals}} + \underbrace{w_{il}''}_{\text{Additive noise}}, \tag{3.31}$$

where (a) follows from the decomposition $\boldsymbol{g}_{jij} = \hat{\boldsymbol{g}}_{jij} + \boldsymbol{\epsilon}_{jij}$, (b) is due to (2.25), $\theta_{ij} := \sqrt{\frac{\rho_{\text{dl}}}{\lambda_j^{\text{zf}}}} \left(\frac{\beta_{jil}}{\beta_{jij}} \right)$, w_{il}'' is the effective noise term which is neither Gaussian nor independent of the users' signals and λ_j^{zf} is given in (2.53). Using a similar argument as before, the

term associated with interference-plus-noise in (3.30) is zero mean and uncorrelated from the desired signal components. One can thus apply the worst-case uncorrelated noise technique of Lemma 2 to obtain an achievable lower bound for ZF in downlink. This is formally presented in the following theorem.

Theorem 6. *Assuming ZF and also Gaussian signaling, i.e.,* $[s_1[i], s_2[i], ..., s_L[i]]^T \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_L)$*, for* $i \in \{1, 2, ..., K\}$*, the following set of lower bounds can be achieved for the MAC of* (3.31)

$$I\left(y_{il}^{\mathrm{dl}};\ \boldsymbol{s}_{\Omega}\middle|\boldsymbol{s}_{\Omega^c}\right) \geq C\left(\frac{\sum_{j\in\Omega}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{zf}}}\right)\left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2}{\sum_{j=1}^{L}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{zf}}}\right)\sum_{k=1}^{K}\frac{\beta_{jil}-\sqrt{\rho_{\mathrm{p}}}\beta_{jil}\alpha_{jil}}{(M-K)\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}}+1}\right), \tag{3.32}$$

for all $\Omega \subseteq \{1, ..., L\}$.

Proof. See Appendix A.9. □

Noting the value calculated for λ_j^{zf} in (2.53), for fixed K and large M, $\frac{1}{\lambda_j^{\mathrm{zf}}} \propto M$, and thereby as M grows the downlink achievable rates of (3.32) for ZF precoding scale as $\mathcal{O}(\log M)$.

It is important to note that, in downlink, user i in cell l is only interested in correctly decoding $s_l[i]$. Thus, incorrectly decoding $s_j[i]$, $j \neq l$, should not penalize the rates achievable at this user. Furthermore, the power of the received signal for the users located in distant cells is very small, and thus trying to decode signals of such users can reduce achievable rates considerably.

Similarly for uplink, BS j is only interested in correctly decoding $x_j^{\mathrm{ul}}[i]$. Thus, incorrectly decoding $x_l^{\mathrm{ul}}[i]$, $l \neq j$, should not penalize the rates achievable at BS j. As will be illustrated later in this chapter, there exist scenarios where system performance is constrained by these distant cells. This motivates the need for more advanced decoding schemes such as simultaneous non-unique decoding (SND), which decodes pilot contamination interference signals non-uniquely. In particular, this decoding scheme achieves

higher rates than SD since the stringent conditions of *uniquely* decoding interference (i.e., applying SD) are now relaxed. Below, we discuss the SND scheme in more detail.

For the remainder of this chapter we only focus on the downlink. The analysis for uplink is similar.

3.2.6 Simultaneous Non-unique Decoding (SND)

In this part, we investigate the benefit of using SND and further show that it enlarges the region obtained by SD for finite M. The optimality of this decoding scheme for interference networks with point-to-point codes and time-sharing was shown in [72]. From the perspective of practical implementations, it is worth mentioning that low complexity techniques, known as sliding-window coded modulation (SWCM) and sliding-window superposition coding (SWSC) have recently been proposed in the literature and achieve performance close to that of the theoretical SND [125–131].

Let us recall the downlink IC associated with the i^{th} user of all cells which are sharing pilot sequence $\boldsymbol{\psi}_i$. More specifically, consider the downlink received signals y_{il}^{dl}, for $l = 1, ..., L$ in (3.10), which together constitute the L outputs of an L-user IC, with input signals $[s_1[i], s_2[i], ..., s_L[i]]^T$. Also, as mentioned before, note that there are in total K separate/non-interfering such L-user ICs in the downlink. The L BSs are the transmitters of this IC, while the i^{th} user of all cells are the receivers.

When performing the non-unique decoding scheme of SND, the i^{th} user in cell l simultaneously decodes the intended signal $s_l[i]$ and the interference signals $s_j[i]$, $j \neq l$, where incorrect decoding of the interference signals does not incur any penalty. More precisely, user i in cell l finds the unique message $\hat{w}_l$ such that $(\hat{\boldsymbol{s}}_l^n[i](\hat{w}_l), \hat{\boldsymbol{s}}_{S\setminus\{l\}}^n[i](w_{S\setminus\{l\}}), \boldsymbol{y}_{il}^{\text{dl},n})$ is jointly typical *for some* $w_{S\setminus\{l\}}$, where $\hat{\boldsymbol{s}}_{S\setminus\{l\}}^n[i](w_{S\setminus\{l\}})$ is the tuple of all codewords $\hat{\boldsymbol{s}}_j^n[i](w_j)$ for $j \in S \setminus \{l\}$.

In [72], the capacity region of an IC when point-to-point random codes are used was derived. Applying the results of [72] to the setting of the IC associated with the i^{th} users

across the L cells, the capacity region can be described by

$$\mathscr{R}_i^{\text{SND}} = \bigcap_{l \in S} \mathcal{R}_{il}^{\text{SND}}, \tag{3.33}$$

where $\mathcal{R}_{il}^{\text{SND}}$ is the rate region achievable at user i in cell l, given by

$$\mathcal{R}_{il}^{\text{SND}} = \bigcup_{\{l\} \subseteq \Omega \subseteq S} \mathcal{R}^i_{\text{MAC}(\Omega,l)}, \tag{3.34}$$

and $\mathcal{R}^i_{\text{MAC}(\Omega,l)}$ represents the achievable rate region obtained from uniquely jointly decoding signals $s_j[i],\ j \in \Omega$ at user i in cell l. Therefore, $\mathcal{R}^i_{\text{MAC}(\Omega,l)}$ can be described by the system of inequalities

$$\sum_{j \in \omega} R_{ij}^{\text{dl}} \leq I\left(y_{il}^{\text{dl}};\ \boldsymbol{s}_\omega \,\middle|\, \boldsymbol{s}_{\omega^c}\right), \quad \forall \omega \subseteq \Omega, \tag{3.35}$$

where $\boldsymbol{s}_\omega$ is the vector with entries $s_j[i],\ j \in \omega, \omega \neq \emptyset$. Note that Ω in (3.34) (the achievable region obtained by SND at user i in cell l) must contain the index of the intended signal $s_l[i]$.

The rate region $\mathcal{R}^i_{\text{MAC}(\Omega,l)}$ has the following properties:

[P1] The region does not include the rates $R_{ij}^{\text{dl}},\ j \in \Omega^c$, and is thus unbounded in these variables.

[P2] The signals $s_j[i],\ j \in \Omega^c$, are treated as noise in the rate expressions defining the region.

One can readily see that $\mathcal{R}_{il}^{\text{SND}}$ strictly contains the MAC region, obtained from SD, at user i in cell l. Therefore, the capacity region $\mathcal{C}_i$ in (3.33) (obtained from SND) is strictly larger than the intersection of the MAC regions at i^{th} users in cells $l = 1, ..., L$. Another important observation is that, due to [**P2**], $\mathcal{R}_{il}$ also contains the TIN region (a similar observation was also made in [72] and [71]). Hence, as it will also be confirmed later via simulation results, the performance of SND should always be at least as good as TIN and SD.

Recall that the SD scheme must decode all users, including the (possibly weak) pilot contamination interference terms. Hence, as will be seen later in this chapter, it achieves *worse rates* than SND.

Remark 8. *Note that there exists a complexity-performance trade-off between SD/SND and TIN. Specifically, while the proposed schemes of SD/SND have more complexity than TIN as they need to decode additional users, for large number of antennas M, the rates achieved by TIN saturate to a fixed value that does not increase with M. In contrast, the rates for SND/SD increase as $\mathcal{O}(\log M)$, and hence unbounded rates are obtained as $M \to \infty$.*

Remark 9. *Note that the successive interference cancellation (SIC) technique used in [37] is different from the SND/SD of this chapter in the following manner: the work of [37] considers a setting in downlink where each user is served by all BSs through the reception of L independent data symbols from the L BSs. In particular, each user applies SIC to sequentially decode the L intended data symbols transmitted by the BSs, while treating all interfering signals, including pilot-sharing interfering signals, as noise, thus resulting in the rate saturation problem. This is in contrast to the approach proposed in this chapter. As the receivers (i.e., BSs in uplink or users in downlink) try to jointly decode (either uniquely or non-uniquely) the intended signal along with the signal coming from the pilot-sharing users, there is no rate saturation as M increases.*

3.2.7 A Simplified Subset of SND (S-SND)

We now consider a simplified achievable region which is a subset of SND and also described in [115, Eq. (6.5)]. We refer to this region as S-SND, which is denoted by $\mathcal{R}_{il}^{\text{S-SND}}$ at user i in cell l, and described by the following set of inequalities

$$\sum\nolimits_{j\in\Omega} R_{ij}^{\text{dl}} \leq I\left(y_{il}^{\text{dl}};\ \boldsymbol{s}_{\Omega} \,\middle|\, \boldsymbol{s}_{\Omega^c}\right), \tag{3.36}$$

for all Ω such that $\{l\} \subseteq \Omega \subseteq \{1, 2, ..., L\}$. To obtain the network-wide achievable region across all cells (for the i^{th} pilot-sharing users), denoted by $\mathscr{R}_i^{\text{S-SND}}$, one should take the

intersection of $\mathcal{R}_{il}^{\text{S-SND}}$ over $l = 1, ..., L$, i.e., $\mathscr{R}_i^{\text{S-SND}} = \bigcap_l \mathcal{R}_{il}^{\text{S-SND}}$.

One can directly verify that $\mathcal{R}_{il}^{\text{S-SND}}$ can be obtained from $\mathcal{R}_{il}^{\text{SD}}$ by removing all $2^{L-1}-1$ inequalities in (3.7) that do not involve the rate R_{il}^{dl}. Hence, the region SD is strictly contained in S-SND. Furthermore, due to the results of (3.16) for a spatially correlated channel as well as (3.26) and (3.32) for an uncorrelated channel, it can be verified that the boundaries of $\mathcal{R}_{il}^{\text{S-SND}}$ and $\mathscr{R}_i^{\text{S-SND}}$ grow as $\mathcal{O}(\log M)$.

The motivation behind considering this region is as follows. It will be shown in the next section that, as opposed to SND, $\mathcal{R}_{il}^{\text{S-SND}}$ is in the form of a convex polytope which makes it easy for computing the maximum symmetric rate allocation. Therefore, even for large networks (e.g., large number of cells) the maximum symmetric rate of S-SND can be computed in a computationally efficient way. Furthermore, since S-SND is a subset of SND, it provides a lower bound to SND.

3.3 Maximum Symmetric Rate Allocation

Considering (3.10), it is evident that users with relatively small effective channel gains $\zeta_{ij},\ j = 1, ..., L,\ i = 1, ..., K$, suffer from smaller rates compared to users with stronger channels. Therefore, fairness among users when allocating resources should be considered. As a measure of fairness, we study the problem of maximum symmetric rate allocation (which is the same as maximizing the minimum achievable rate among all users) for various schemes. This measure of performance has been widely applied in the literature [111, 113, 132–137]. More specifically, we will compare the performance of the proposed full interference decoding schemes SD/SND as well as the sub-region of S-SND with that of TIN based on the maximum symmetric rate they can offer. In what follows, the analysis is shown only for the L-user IC associated with the i^{th} (i is arbitrary) users across all cells that are employing the same pilot sequence $\boldsymbol{\psi}_i$, since the same results hold for other sets of pilot-sharing users.

The maximum symmetric rate associated with the rate region achievable at user i in cell l is obtained by $R_{\text{Sym},l} = \max R$ such that the rate vector $\boldsymbol{R} = [R, R, ..., R]^T$ belongs to the

achievable region at user i in cell l. Therefore, the rate vector $[R_{\text{Sym},l}, R_{\text{Sym},l}, ..., R_{\text{Sym},l}]^T$ must lie at the intersection of the diagonal $\left(R_{i1}^{\text{dl}} = ... = R_{iL}^{\text{dl}}\right)$ with the boundary of the achievable region at user i in cell l.

One can verify that $\mathcal{R}_{il}^{\text{SD}}$ in (3.7) (achieved at user i in cell l) can be represented as the intersection of a finite number of closed half-spaces and is also bounded. Hence, it is a convex polytope, denoted by

$$\mathscr{R}_l = \left\{ \left[R_{i1}^{\text{dl}}, ..., R_{iL}^{\text{dl}}\right]^T : \sum\nolimits_{j\in\Omega} R_{ij}^{\text{dl}} \leq g_l(\Omega), \forall \Omega \subseteq \{1, 2, ..., L\} \right\}, \tag{3.37}$$

where the function $g_l(\Omega)$ is the r.h.s of the inequality in (3.7). Similarly, it can be verified that $\mathcal{R}_{il}^{\text{S-SND}}$ in (3.36) is of the form (3.37), except now $g_l(\Omega) = \infty$ if $l \notin \Omega$, and is also a convex polytope.

The following lemma can be used to find the maximum possible value for the minimum entry of a vector $\boldsymbol{R}$, where $\boldsymbol{R} \in \mathscr{R}_l$.

Lemma 3. *In the polytope $\mathscr{R}_l$, define*

$$\pi = \max \quad \min_{i\in S} \quad R_i \tag{3.38}$$

$$\text{subject to } \left[R_1, ..., R_L\right]^T \in \mathscr{R}_l, \tag{3.39}$$

where $S = \{1, 2, ..., L\}$. Then,

$$\pi = \min_{\Omega\subseteq\{1,2,...,L\}, \Omega\neq\emptyset} \frac{g_l(\Omega)}{|\Omega|}. \tag{3.40}$$

Proof. Following the steps of [138], consider an arbitrary vector $\boldsymbol{R} \in \mathscr{R}_l$, and define $\delta = \min_i \ R_i$. Hence, for all $\Omega \neq \emptyset$, we have $\delta \leq \sum_{i\in\Omega} R_i/|\Omega| \leq g_l(\Omega)/|\Omega|$. Therefore, $\min_{\Omega\neq\emptyset} g_l(\Omega)/|\Omega|$ is an upper bound on $\min_i \ R_i$. Choosing $\boldsymbol{R} = (\pi_0, ..., \pi_0) \in \mathscr{R}_l$, where $\pi_0 = \min_{\Omega\neq\emptyset} g_l(\Omega)/|\Omega|$, the upper bound is thus achieved. □

Therefore, the maximum symmetric rate (which also maximizes the minimum rate due

to Lemma 3) at user i in cell l is

$$R_{\text{Sym},l} = \min_{\Omega_l \subseteq \{1,2,...,L\}, \Omega_l \neq \emptyset} \frac{g_l(\Omega_l)}{|\Omega_l|}. \tag{3.41}$$

Finally, to find the maximum symmetric rate network-wide one needs to compute $\min R_{\text{Sym},l}$ for $l \in \{1, 2, ..., L\}$. In the following, we discuss how (3.41) can be solved over various regions.

SD: At user i in cell l, the minimization over $\mathcal{R}_{il}^{\text{SD}}$ in (3.7) can be carried out by solving

$$\min_{\Omega_l} \quad \frac{I(y_{il}^{\text{dl}};\ \boldsymbol{s}_{\Omega_l} \mid \boldsymbol{s}_{\Omega_l^c})}{|\Omega_l|} \tag{3.42}$$

$$\text{subject to} \quad \Omega_l \subseteq \{1, 2, ..., L\}. \tag{3.43}$$

The expressions of (3.16) for a spatially correlated channel or (3.26) and (3.32) for an uncorrelated channel then allow one to find an achievable lower bound to the above problem. For instance, when applying MRT and assuming that channels are spatially uncorrelated, one needs to solve the following

$$[\mathcal{P}1] \qquad \min_{\Omega_l} \quad \frac{1}{|\Omega_l|} \log \left(1 + \frac{MK \sum_{j \in \Omega_l} \left(\frac{\rho_{\text{dl}} \rho_{\text{p}} \beta_{jil}^2 \alpha_{ij}^2}{\sum_{k=1}^{K} \sqrt{\rho_{\text{p}}} \beta_{jkj} \alpha_{jkj}} \right)}{K \sum_{j=1}^{L} \rho_{\text{dl}} \beta_{jil} + 1} \right) \tag{3.44}$$

$$\text{subject to} \quad \Omega_l \subseteq \{1, 2, ..., L\}. \tag{3.45}$$

SND: It can be seen from (3.34) that the region $\mathcal{R}_{il}^{\text{SND}}$ achieved by SND at user i in cell l can not in general be represented by the intersection of a finite number of half-spaces and thus does not fall in the category of convex polytopes. However, $\mathcal{R}_{il}^{\text{SND}}$ in (3.34) is represented as the union of a finite number of convex polytopes. Hence, to find the maximum symmetric rate of SND at user i in cell l, one calculates the maximum symmetric rate over each of these convex polytopes using Lemma 3, and then picks the largest of these quantities. We will simulate the performance of SND in different scenarios

and for different numbers of cells (i.e., $L = 2, 3, 4$, and 7) at the end of this chapter. Moreover, the special cases of two-cell and three-cell systems, for which the SND region is easier to investigate, will be analytically studied and discussed at the end of this section. Below, we also investigate the maximum symmetric rate of S-SND which provides a lower bound to SND.

S-SND: Using $\mathcal{R}_{il}^{\mathrm{S-SND}}$, problem $[\mathcal{P}1]$ changes to the following

$$[\mathcal{P}2] \qquad \min_{\Omega_l} \quad \frac{1}{|\Omega_l|} \log \left(1 + \frac{MK \sum_{j \in \Omega_l} \left(\frac{\rho_{\mathrm{dl}} \rho_{\mathrm{p}} \beta_{jil}^2 \alpha_{jij}^2}{\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{p}}} \beta_{jkj} \alpha_{jkj}} \right)}{K \sum_{j=1}^{L} \rho_{\mathrm{dl}} \beta_{jil} + 1} \right) \tag{3.46}$$

$$\text{subject to} \quad \{l\} \subseteq \Omega_l \subseteq \{1, 2, ..., L\}. \tag{3.47}$$

Note that even though $[\mathcal{P}1]$ and $[\mathcal{P}2]$ have the same objective function, following the discussion below (3.36) the solution Ω_l of $[\mathcal{P}2]$ must include the index l associated with the rate R_{il}^{dl}, and is thus not necessarily identical to that of $[\mathcal{P}1]$. For an uncorrelated channel with ZF precoding, (3.26) in $[\mathcal{P}1]$ and $[\mathcal{P}2]$ is replaced with (3.32), while for a spatially correlated channel with MRT it is replaced with (3.16).

To tackle $[\mathcal{P}1]$ (or $[\mathcal{P}2]$), we first consider the extreme regime of high SINR.

3.3.1 High SINR regime

Assume that ZF precoding is applied at the BSs. In this regime, the values of M and L are such that

$$\log \left(1 + \frac{MK \sum_{j \in \Omega_l} \left(\frac{\rho_{\mathrm{dl}} \rho_{\mathrm{p}} \beta_{jil}^2 \alpha_{jij}^2}{\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{p}}} \beta_{jkj} \alpha_{jkj}} \right)}{K \sum_{j=1}^{L} \rho_{\mathrm{dl}} \beta_{jil} + 1} \right) \approx \log(M). \tag{3.48}$$

For instance, this approximation holds when the number of BS antennas M is truly large but finite while the number of cells L and users K are fixed. A similar approximation can be obtained assuming that MRT is used.

Thus, in this regime the minimization in both $[\mathcal{P}1]$ and $[\mathcal{P}2]$ is achieved by $\Omega_l^* = \{1, 2, ..., L\}$, and thereby the maximum symmetric rate in cell l is given by

$$R_{\text{Sym},l}^{\text{SD}} = R_{\text{Sym},l}^{\text{S-SND}} = \frac{I\left(y_{il}^{\text{dl}};\ s_1[i], s_2[i], ..., s_L[i]\ \right)}{L}, \tag{3.49}$$

which scales as $\mathcal{O}(\log M)$. As discussed before, the performance of SND is at least as good as SD and S-SND, i.e., $R_{\text{Sym},l}^{\text{SND}} \geq R_{\text{Sym},l}^{\text{SD}} = R_{\text{Sym},l}^{\text{S-SND}}$. Thus, in the high SINR regime the maximum symmetric rate of SND occurs on one of its region boundaries that scales as $\mathcal{O}(\log M)$. In other words, from (3.34) the maximum symmetric rate achieved by SND in the high SINR regime belongs to the full MAC, i.e., $R_{\text{Sym},l}^{\text{SND}} \in \mathcal{R}^i_{\text{MAC}(\{1,...,L\},l)}$. Therefore, in the high SINR regime one can upper bound $R_{\text{Sym},l}^{\text{SND}}$ by $R_{\text{Sym},l}^{\text{SND}} \leq \frac{1}{L} I\left(y_{il}^{\text{dl}};\ s_1[i], s_2[i], ..., s_L[i]\ \right)$. Consequently, we obtain for the high SINR regime $R_{\text{Sym},l}^{\text{SND}} = R_{\text{Sym},l}^{\text{SD}} = R_{\text{Sym},l}^{\text{S-SND}}$. To find the allocation network-wide, denoted by R_{Sym}, one needs to calculate the smallest value of (3.49) across all cells, i.e.,

$$R_{\text{Sym}} = \min_j \frac{I\left(y_{ij}^{\text{dl}};\ s_1[i], s_2[i], ..., s_L[i]\ \right)}{L}, \tag{3.50}$$

which is the same for all interference decoding schemes. Therefore, compared to TIN we obtain

$$R_{\text{Sym}} > \min_j I\left(y_{ij}^{\text{dl}};\ s_j[i]\ \right) = R_{\text{Sym}}^{\text{TIN}}, \tag{3.51}$$

i.e., joint decoding of all signals $\{s_j[i]\}_{j=1}^L$ performs strictly better than decoding only the desired signal (e.g., $s_l[i]$ in cell l), while treating the interference signals (e.g., $\{s_j[i]\}_{j=1,l\neq j}^L$ in cell l) as noise (TIN).

Remark 10. *Consider the alternative approach of using different pilots in different cells, as explained before* (3.1)*. One should note that, for the regime of large but finite M,*

decoding all $K(L-1)+1$ number of interfering users at the current cell will generally produce a smaller symmetric rate than the approach of Section 2.3.3 which only decodes L users, due to the much smaller pre-log factor in the former case. For instance, in the regime of high SINR, using (3.48)-(3.50), the achieved maximum symmetric rate of the former case is $\approx 1/(K(L-1)+1)\log(M)$, whereas that of the latter case is $\approx 1/L\log(M)$. Hence, when decoding pilot contamination interference, re-using orthonormal pilots cross all cells is preferred as, for finite M, it results in larger symmetric rate across the network.

3.3.2 General case

Now, consider the problem of determining the maximum symmetric rate over the regions TIN/SD/SND/S-SND in general, where the approximation of high SINR is no longer assumed. Since it is difficult to comment on the performance of maximum symmetric rate for SND in general due to the structure of the SND region, we next study the two special cases of two-cell and three-cell systems which are analytically tractable. For a symmetric two-cell system, we will find that either TIN is optimal (SND is always optimal, and thus the performance of TIN equals that of SND) or the interference decoding scheme of SD is optimal (and hence the performance of SD equals that of SND). Whereas, for the three-cell system we will briefly illustrate examples where SND outperforms *all the other schemes*. Cases with more cells (i.e., $L=4$ and $L=7$) will be evaluated at the end of this chapter via comprehensive simulation results in different scenarios.

Two-cell system

We now consider a cellular system consisting of only two cells, and denote the indices of the cells by $j=1,2$. For the downlink IC associated with the i^{th} pilot-sharing users, $i=1,2,...,K$, the rate regions achieved by different schemes in cell 1 are given as below.

SD: From (3.7), we obtain for $\mathcal{R}_{il}^{\text{SD}}$

$$R_{i1}^{\text{dl}} \leq I\left(y_{i1}^{\text{dl}};\ s_1[i] \,\middle|\, s_2[i]\right) \tag{3.52}$$

$$R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_2[i] \,\middle|\, s_1[i]\right) \tag{3.53}$$

$$R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ \right). \tag{3.54}$$

SND: From (3.34), we obtain for $\mathcal{R}_{il}^{\mathrm{SND}}$

$$R_{i1}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_1[i] \,\middle|\, s_2[i]\right) \tag{3.55}$$

$$R_{i1}^{\mathrm{dl}} + \min\left\{R_{i2}^{\mathrm{dl}},\ I\left(y_{i1}^{\mathrm{dl}};\ s_2[i] \,\middle|\, s_1[i]\right)\right\} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ \right). \tag{3.56}$$

S-SND: From (3.36), we obtain for $\mathcal{R}_{il}^{\mathrm{S-SND}}$

$$R_{i1}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_1[i] \,\middle|\, s_2[i]\right) \tag{3.57}$$

$$R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ \right). \tag{3.58}$$

Note that, depending on the choice of precoding vector and whether the channel is correlated or not, the $I\left(.;.\middle|.\right)$ terms above can be readily replaced by the bounds of (3.16), (3.26) or (3.32), which scale as $\mathcal{O}(\log M)$.

Remark 11. *One can similarly obtain the rate regions in cell* 2 *by replacing* y_{i1}^{dl} *with* y_{i2}^{dl} *and swapping appropriate indices in* (3.52)-(3.58).

Note that in a two-cell system, $\mathcal{R}_{il}^{\mathrm{SND}}$ is the union of $\mathcal{R}_{il}^{\mathrm{SD}}$ and the TIN region (or alternatively the union of $\mathcal{R}_{il}^{\mathrm{S-SND}}$ and TIN). We now aim to investigate the performance of different schemes with maximum symmetric rate allocation. For the two-cell system, we first define the following cases:

Case (i): In this case, we have

$$I\left(y_{i1}^{\mathrm{dl}};\ s_2[i] \,\middle|s_1[i]\ \right) < I\left(y_{i1}^{\mathrm{dl}};\ s_1[i]\ \right). \tag{3.59}$$

Case (ii): In this case, we have

$$\frac{1}{2} I\left(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ \right) \leq \min\left\{I\left(y_{i1}^{\mathrm{dl}};\ s_1[i] \,\middle|\, s_2[i]\right),\ I\left(y_{i1}^{\mathrm{dl}};\ s_2[i] \,\middle|\, s_1[i]\right)\right\}. \tag{3.60}$$

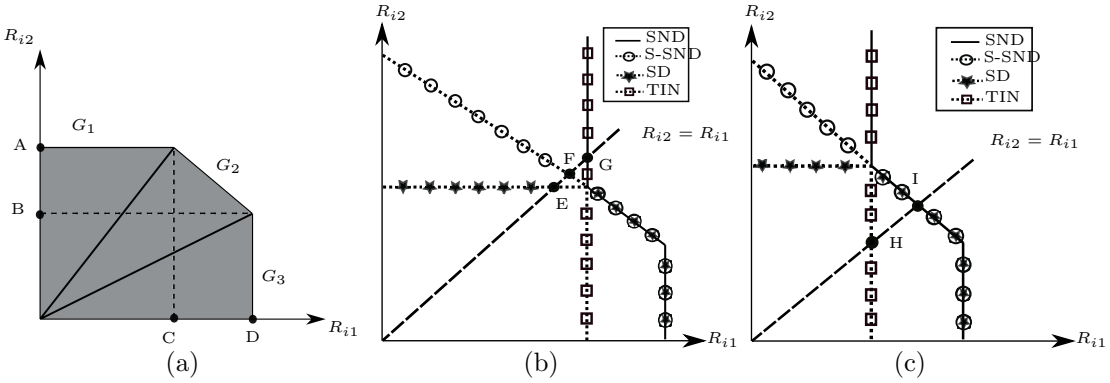


Figure 3.3: (a) Illustration of $\mathcal{R}_{i1}^{\text{SD}}$ in cell 1 representing the 3 sub-regions G_1, G_2 and G_3 over which the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ will intersect a particular facet of the rate region, (b) Illustration of the rate regions achieved under TIN/SND/S-SND/SD in cell 1 for case (i): the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ intersects SD at point E, S-SND at point F, and SND/TIN at point G, resulting in (3.66), (c) Illustration of the rate regions achieved under TIN/SND/S-SND/SD in cell 1 for case (ii): the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ intersects TIN at point H, and SND/S-SND/SD at point I, resulting in (3.67).

Case (iii): In this case, we have

$$I\left(y_{i1}^{\text{dl}};\ s_1[i]\ \middle|s_2[i]\ \right) < I\left(y_{i1}^{\text{dl}};\ s_2[i]\ \right). \tag{3.61}$$

From the perspective of the maximum symmetric rate, cases (i)-(iii) refer to conditions (in terms of mutual information) under which the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ intersects one of the three facets of the rate region.

More specifically, consider the rate region $\mathcal{R}_{i1}^{\text{SD}}$ achieved by SD in cell 1, depicted in Fig. 3.3a, where the entire region is divided into 3 sub-regions G_1, G_2 and G_3. Also, from (3.52)-(3.54), note that the corner points are given by $(C,\ A) = \left(I\left(y_{i1}^{\text{dl}};\ s_1[i]\ \right),\ I\left(y_{i1}^{\text{dl}};\ s_2[i]\ \middle|s_1[i]\ \right)\right)$ and $(D,\ B) = \left(I\left(y_{i1}^{\text{dl}};\ s_1[i]\ \middle|s_2[i]\ \right),\ I\left(y_{i1}^{\text{dl}};\ s_2[i]\ \right)\right)$. Now, the conditions under which the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ lies in sub-regions G_1, G_2 or G_3, are equivalent to the conditions of the three cases of (3.59)-(3.61) as follows: the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ lies in G_1, i.e., case (i) is true, iff $C > A$; the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ lies in G_2, i.e., case (ii) is true, iff $C \leq A$ and $B \leq D$; the diagonal $R_{i2}^{\text{dl}} = R_{i1}^{\text{dl}}$ lies in G_3, i.e., case (iii) is true, iff $B > D$. Specifically, the conditions for case (i) in (3.59) and case (iii) in (3.61)

are exactly those given by $C > A$ and $B > D$, respectively.

For case (ii), note that one can also write

$$I(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ ) = I(y_{i1}^{\mathrm{dl}};\ s_1[i]\ ) + I(y_{i1}^{\mathrm{dl}};\ s_2[i] \mid s_1[i]\ ) \tag{3.62}$$

$$= I(y_{i1}^{\mathrm{dl}};\ s_2[i]\ ) + I(y_{i1}^{\mathrm{dl}};\ s_1[i] \mid s_2[i]\ ). \tag{3.63}$$

Hence, in case (ii) where we have $C \leq A$ and $B \leq D$, by replacing C and B with their respective identity from (3.62) and (3.63), we reach the following conditions

$$\frac{1}{2} I(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ ) \leq I(y_{i1}^{\mathrm{dl}};\ s_2[i] \mid s_1[i]\ ) \tag{3.64}$$

$$\frac{1}{2} I(y_{i1}^{\mathrm{dl}};\ s_1[i],\ s_2[i]\ ) \leq I(y_{i1}^{\mathrm{dl}};\ s_1[i] \mid s_2[i]\ ), \tag{3.65}$$

resulting in (3.60).

Remark 12. *If the worst-case uncorrelated noise bounds of* (3.16), (3.26) *or* (3.32) *are substituted for the mutual information expressions in* (3.59)-(3.61)*, case (iii) should not happen as the effects of small-scale fading vanish in these three bounds and thus the received power of* $s_2[i]$ *in cell* 1 *can not be larger than that of* $s_1[i]$ *in cell* 1*. Hence, case (i) and case (ii) can be viewed as two complimentary and exhaustive conditions for a two-cell system in cell* 1*.*

Further note that the bounds of (3.16), (3.26) *or* (3.32) *differ from the mutual expressions in* (3.61) *due to two factors: (a) the expressions in* (3.61) *depend on the specific fading gains, and (b) the effective noise is not necessarily Gaussian. However, as* M *increases the channel hardening of* (2.51) *and* (2.54) *minimize the effects of (a). Moreover, due to the channel hardening and favorable propagation effects in* (2.51) *and* (2.54) *as well as the assumption of Gaussian signaling in the results of* (3.16), (3.26) *and* (3.32)*, the interference terms (effective noise) in* (2.43) *are asymptotically Gaussian.*

The performance comparison of various schemes in cell 1 is summarized in the following corollary.

Corollary 3. *If the condition of case (i) in* (3.59) *holds in cell* 1*, then*

$$R_{Sym,1}^{SD} < R_{Sym,1}^{S\text{-}SND} < R_{Sym,1}^{SND} = R_{Sym,1}^{TIN}, \tag{3.66}$$

otherwise, if the condition of case (ii) in (3.60) *holds in cell* 1*, then*

$$R_{Sym,1}^{TIN} \leq R_{Sym,1}^{SD} = R_{Sym,1}^{SND} = R_{Sym,1}^{S\text{-}SND}, \tag{3.67}$$

with strict equality in (3.67) *if and only if* (3.60) *holds with strict equality.*

Proof. See Appendix A.10. □

Fig. 3.3 illustrates an example of this corollary. Sub-figure (b) represents case (i) and its consequence in (3.66), whereas sub-figure (c) represents case (ii) and its consequence in (3.67). It is also verified from Fig. 3.3 that if the condition of case (iii) in (3.61) is active in cell 1, its consequence is identical to that of case (ii) in (3.67), however it is not discussed here due to Remark 12.

To further comment on the performance of various schemes over both cells, we consider a symmetric setting which is easy to analyze, and provides insights into the benefits of employing interference decoding schemes.

We define the symmetric setting as a scenario where the MACs seen in both cells 1 and 2 are identical, i.e., a symmetric two-user IC. Therefore, if case (i) is active in cell 1, it is also active in cell 2, and the resulting rates are equal in both cells. Following Remark 11 it is thus obtained network-wide that

$$R_{\text{Sym}}^{\text{SD}} < R_{\text{Sym}}^{\text{S-SND}} < R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{TIN}}. \tag{3.68}$$

Observation: Both SND and TIN achieve the same performance and strictly outperform SD and S-SND. Thus, TIN may be the better choice of strategy in practice due to its simplicity.

Similarly, if case (ii) is active with strict inequality in cell 1, it is also active with strict inequality in cell 2, and the resulting rates are equal in both cells. Following Remark 11 it

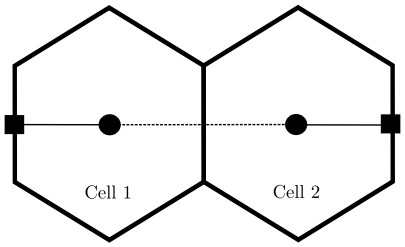


Figure 3.4: An example of symmetric geometry in a two-cell system, where users are located at the same location on the cell edge, denoted by solid squares, and BSs are located at the center of the cells.

is thus obtained network-wide that

$$R_{\text{Sym}}^{\text{TIN}} < R_{\text{Sym}}^{\text{SD}} = R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{S-SND}}. \tag{3.69}$$

Observation: The interference decoding schemes SD/SND achieve the same performance and strictly outperform TIN. Thus, SD may be the simpler one to implement in practice. Examples of these cases will be demonstrated in the simulation results section.

Consider, for instance, a setup where all users are located at the same location on the cell edge, denoted by solid squares, as in Fig. 3.4. With respect to the bounds of (3.16), (3.26) and (3.32), this setup is symmetric as the effects of small-scale fading vanish in these bounds.

Note that in a two-cell system, as discussed in Fig. 3.3, all faces of $\mathcal{R}_{il}^{\text{SND}}, l = 1, ..., L$, are achieved by either TIN or SD. Next, we will briefly illustrate scenarios for a three-cell system, where SND can strictly outperform *all the other schemes*.

Three-cell system

Now, consider a cellular system consisting of only three cells. In this case, the rate regions under SD/S-SND can be obtained by a straightforward extension of (3.52)-(3.54) and (3.57)-(3.58) to the three-cell system, thus omitted for brevity. Moreover for SND, the rate

region $\mathcal{R}_{i1}^{\text{SND}}$ achieved in cell 1 can be found using (3.34) as follows:

$$R_{i1}^{\text{dl}} \leq I\left(y_{i1}^{\text{dl}};\ s_1[i] \,\middle|\, s_2[i], s_3[i]\ \right) \tag{3.70}$$

$$\begin{aligned} R_{i1}^{\text{dl}} + \min\left\{I\left(y_{i1}^{\text{dl}};\ s_2[i] \,\middle|\, s_1[i], s_3[i]\ \right),\ R_{i2}^{\text{dl}}\right\} \\ \leq I\left(y_{i1}^{\text{dl}};\ s_1[i], s_2[i] \,\middle|\, s_3[i]\ \right) \end{aligned} \tag{3.71}$$

$$\begin{aligned} R_{i1}^{\text{dl}} + \min\left\{I\left(y_{i1}^{\text{dl}};\ s_3[i] \,\middle|\, s_1[i], s_2[i]\ \right),\ R_{i3}^{\text{dl}}\right\} \\ \leq I\left(y_{i1}^{\text{dl}};\ s_1[i], s_3[i] \,\middle|\, s_2[i]\ \right) \end{aligned} \tag{3.72}$$

$$\begin{aligned} R_{i1}^{\text{dl}} + \min\Big\{&I\left(y_{i1}^{\text{dl}};\ s_2[i], s_3[i] \,\middle|\, s_1[i]\ \right), \\ &R_{i2}^{\text{dl}} + I\left(y_{i1}^{\text{dl}};\ s_3[i] \,\middle|\, s_1[i], s_2[i]\ \right), \\ &R_{i3}^{\text{dl}} + I\left(y_{i1}^{\text{dl}};\ s_2[i] \,\middle|\, s_1[i], s_3[i]\ \right), \\ &R_{i2}^{\text{dl}} + R_{i3}^{\text{dl}}\Big\} \\ \leq I\left(y_{i1}^{\text{dl}};\ s_1[i], s_2[i], s_3[i]\ \right). \end{aligned} \tag{3.73}$$

Similar to the two-cell system, depending on the choice of precoding vector and whether the channel is correlated or not, the $I\left(.;.|.\right)$ terms above can be readily replaced by the bounds of (3.16), (3.26) or (3.32), which scale as $\mathcal{O}(\log M)$. An example of this region is plotted in Fig. 3.5, where the dashed lines indicate that the region in cell 1 is unbounded in variables R_{i2}^{dl} and R_{i3}^{dl}, which is in agreement with property [**P1**] of the achievable region. Also, note that following Remark 11 the regions corresponding to cells 2 and 3 can be similarly found.

By comparing (3.70)-(3.73) with the achievable regions of SD, $\mathcal{R}_{i1}^{\text{SD}}$, and S-SND, $\mathcal{R}_{i1}^{\text{S-SND}}$, it is noted that there are four faces in Fig. 3.5 that are only achieved by SND and not by any other schemes. More precisely in cell 1, it is possible for $R_{\text{Sym},1}$ to achieve one of the rates, $I(y_{i1}^{\text{dl}};\ s_1[i] \,|\, s_3[i]\ )$, $I(y_{i1}^{\text{dl}};\ s_1[i] \,|\, s_2[i]\ )$, $\frac{1}{2}I(y_{i1}^{\text{dl}};\ s_1[i], s_2[i]\ )$ or $\frac{1}{2}I(y_{i1}^{\text{dl}};\ s_1[i], s_3[i]\ )$. Note that the first rate $I(y_{i1}^{\text{dl}};\ s_1[i] \,|\, s_3[i]\ )$ can be interpreted as the maximum rate of the i^{th} user of cell 1, while treating the i^{th} user in cell 2 as noise. The second rate $I(y_{i1}^{\text{dl}};\ s_1[i] \,|\, s_2[i]\ )$ can be interpreted similarly. Moreover, the rate $\frac{1}{2}I(y_{i1}^{\text{dl}};\ s_1[i], s_2[i]\ )$ can be interpreted as the maximum symmetric rate achieved by joint decoding of the i^{th} users of cells 1 and 2, while treating the i^{th} user of cell 3 as noise. The fourth rate

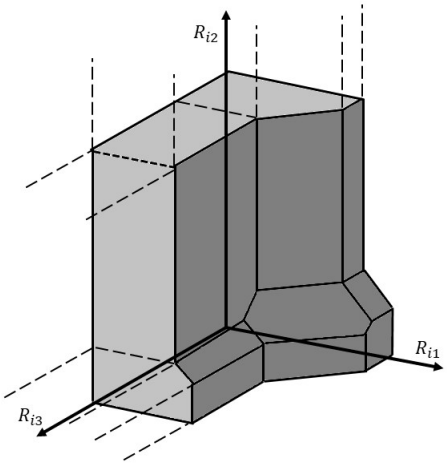


Figure 3.5: An example of the rate region, R_{i1}^{SND}, obtained by SND in cell 1.

$\frac{1}{2}I(y_{i1}^{\mathrm{dl}};\ s_1[i], s_3[i]\ )$ can be interpreted similarly. Therefore, neither SD/S-SND nor TIN can provide these rates, in which case SND can strictly outperform *all the other schemes*. More discussion will be provided in the simulation results section.

3.4 Simulation Results

To illustrate the performance of the different interference management schemes, TIN/SD/SND, with maximum symmetric spectral efficiency (SE in units of bits/sec/Hz) allocation, we simulate the downlink of a multi-cell massive MIMO system experiencing pilot contamination. In particular, we consider hexagonal cells with a radius of $r = 400$ m where BSs are located at the center of the cell and $K = 15$ users are uniformly distributed at random within the area of each cell. To evaluate the performance, the average of maximum symmetric SEs is calculated over 150 random realizations of user locations. The downlink transmit power of each BS is taken to be 40 W, and to model large-scale

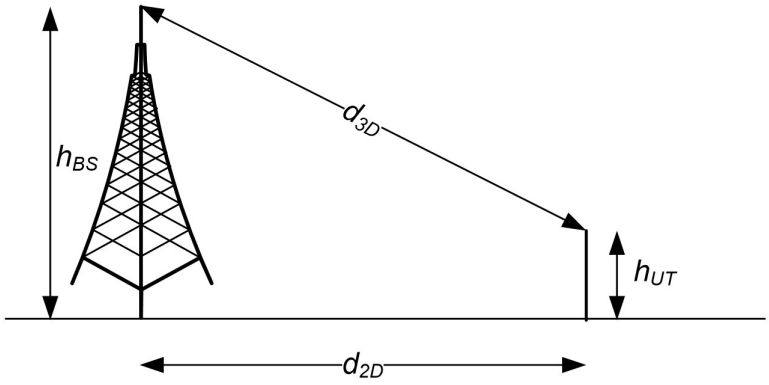
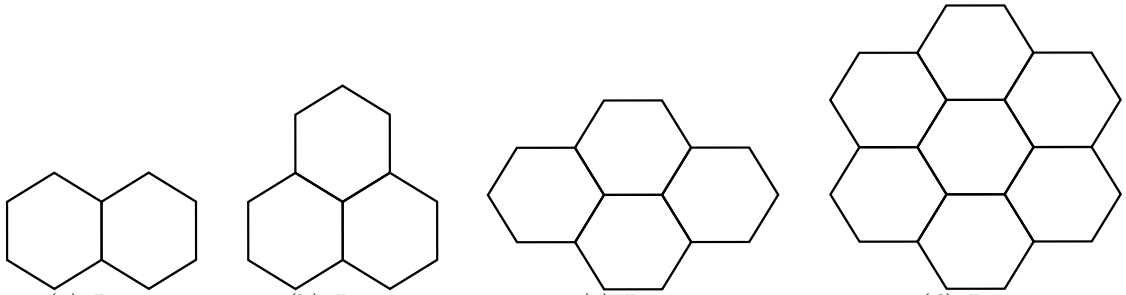


Figure 3.6: The 3D distance model of [2], where h_{BS} is the BS height taken to be 25 m, and d_{2D} is the 2D distance from the user to the BS.



(a) $L = 2$ (b) $L = 3$ (c) $L = 4$ (d) $L = 7$

Figure 3.7: Cell configurations for different scenarios depending on the value of L, (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, (d) $L = 7$.

fading coefficients, β_{jkl}, a path-loss model adopted from [2] is considered:

$$[\beta_{jkl}]_{\text{dB}} = -13.54 - 39.08 \log_{10}\left(d_{jkl}^{3D}\right) - 20 \log_{10}\left(f_c\right) + 0.6\left(h_{UT} - 1.5\right), \tag{3.74}$$

where d_{jkl}^{3D} is the 3D distance (in meters) from user k in cell l to BS j (see Fig. 3.6), the carrier frequency is $f_c = 3.5$ GHz, h_{UT} is the user height which is taken to be 1.5 m. Also, with a system bandwidth of 20 MHz, the noise variance is assumed to be -101 dBm. Note that while the effects of shadowing are neglected in (3.74), we will investigate its impact on system performance separately at the end of this section. It is further assumed that

users are located at least 35 m away from their BSs, i.e., $d_{2D} \geq 35$ m. To study the impact of changing the number of cells, L, on system performance we also consider four different cell configurations when $L \in \{2, 3, 4, 7\}$, as shown in Fig. 3.7. It should also be mentioned that wrap around topology is not used for the simulation results corresponding to the cell configurations of Fig. 3.7. However, the effect of wrap around topology for $L = 7$ will be studied separately at the end of this section.

For the sake of completeness, we also plot the performance of the sub-region S-SND in all scenarios. Note that S-SND is not a communication scheme, but only a simple sub-region of SND that serves as a good lower bound (clearly tighter than SD) to SND, and thus is only used for performance comparison purposes. Below, we investigate the two cases of a spatially correlated Rayleigh fading channel and an uncorrelated Rayleigh fading channel separately.

3.4.1 Spatially correlated

We now study the downlink performance of MRT and ZF precoding when a spatially correlated channel model is used. We adopt the exponential correlation model of [139], i.e.,

$$\boldsymbol{R}_{jkl} = \beta_{jkl} \begin{pmatrix} 1 & r_{jkl}^{*} & \cdots & \left(r_{jkl}^{*}\right)^{M-1} \\ r_{jkl} & 1 & \cdots & \left(r_{jkl}^{*}\right)^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \left(r_{jkl}\right)^{M-1} & \left(r_{jkl}\right)^{M-2} & \cdots & 1 \end{pmatrix}, \tag{3.75}$$

which is widely used in the literature [39, 64, 140–143], where $\boldsymbol{R}_{jkl}$ is the correlation matrix from user k in cell l to BS j. In particular, in this model $r_{jkl} = \kappa e^{j\phi_{jkl}}$ is the correlation coefficient, $\kappa \in [0,\ 1]$ is the correlation magnitude and ϕ_{jkl} is the user angle to the antenna array boresight. Unless otherwise specified, we assume $\kappa = 0.4$, i.e., moderate spatial correlation.

Figs. 3.8 and 3.9 show the performance of ZF and MRT for $L = 7$ versus M, respectively. While in both cases the achievable symmetric SEs increase with M, as discussed below (2.41) ZF achieves significantly higher SEs compared to MRT. It can be seen from Fig. 3.8

that the gain offered by SND compared to TIN is about 19% and 33% for $M = 128$ and $M = 256$, respectively, and this gain increases to about 52% when $M = 1024$. On the other hand, Fig. 3.9 shows that when using MRT the gain offered by SND is about 7% and 12% for $M = 128$ and $M = 256$, respectively, and this gain increases to about 43% when $M = 1024$. Lastly, the two figures confirm that SD performs poorly in both scenarios, as it tries to blindly decode pilot contamination interference terms regardless of their strength. This is as opposed to the non-unique decoding scheme of SND, which automatically determines which pilot contamination interference terms should be decoded along with the message of interest while the remaining interference terms will be treated as noise. Since MRT performs poorly compared to ZF, as observed in Fig. 3.9, we only focus on the performance of ZF for the rest of the considered scenarios.

Figs. 3.10-3.12 show the performance of ZF for the cases of $L = 4$, $L = 3$ and $L = 2$, respectively. While a pattern similar to that of Fig. 3.8 is apparent, one can notice that the proposed schemes achieve higher SEs as the number of cells decreases. This is due to the fact that by reducing L, while keeping the other system parameters fixed, the amount of interference decreases, thus increasing the spectral efficiency. As a consequence, when performing ZF for instance, the gain offered by SND decreases as the number of cells reduces. In particular, for $L = 4$ this gain is about 18% when $M = 256$ (down from 33% in the case of $L = 7$) and it decreases to about 7% when $L = 2$ and $M = 256$.

Next, we study the performance of regularized zero-forcing (RZF) precoding. Here, the precoding matrix at BS j is given by

$$\boldsymbol{W}_j^{\text{rzf}} = \hat{\boldsymbol{G}}_{jj} \left(\hat{\boldsymbol{G}}_{jj}^{\dagger} \hat{\boldsymbol{G}}_{jj} + \delta \boldsymbol{I}_K \right)^{-1}, \quad \forall j, \tag{3.76}$$

where δ is a regularization factor, which improves the numerical stability of the inverse operation. Note that the choice of δ is arbitrary and could be further optimized (see for example [87, Theorem 6] and [144]). The two choices of $\delta = K/\rho_{\text{dl}}$ (suggested by [145]) and $\delta = M/\rho_{\text{dl}}$ (suggested by [32]) were investigated by simulation. The former provided better performance for the setup and system parameters considered in this work. Therefore, in this thesis we take $\delta = K/\rho_{\text{dl}}$. One can verify that when M is large, the diagonal entries

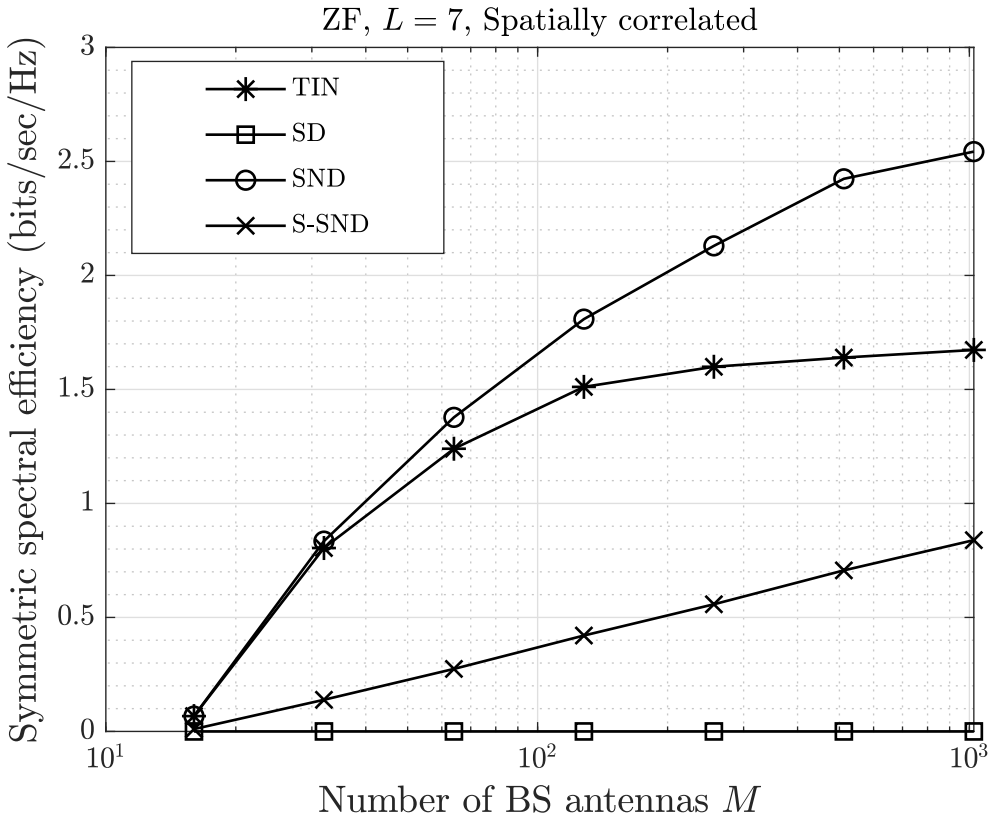


Figure 3.8: Performance of maximum symmetric SE for seven cells when ZF precoding and a spatially correlated channel model are used.

of $\hat{\boldsymbol{G}}_{jj}^{\dagger}\hat{\boldsymbol{G}}_{jj}$ increase with M and therefore the approximation below can be used

$$\left(\hat{\boldsymbol{G}}_{jj}^{\dagger}\hat{\boldsymbol{G}}_{jj} + \delta \boldsymbol{I}_K\right)^{-1} \approx \left(\hat{\boldsymbol{G}}_{jj}^{\dagger}\hat{\boldsymbol{G}}_{jj}\right)^{-1}. \tag{3.77}$$

Hence, when M is large, one expects the performance of RZF to resemble that of ZF. On the other hand, when M is small, with a proper choice of the regularization factor RZF can outperform ZF [145]. These results are confirmed in the next figures. Specifically, Fig. 3.13 shows the performance of TIN/SD/SND with maximum symmetric SE for $L = 7$ and when RZF precoding is applied at the BSs. Similar to the cases of ZF and MRT, as M increases the gain offered by SND improves, while unsurprisingly SD is always outperformed by the

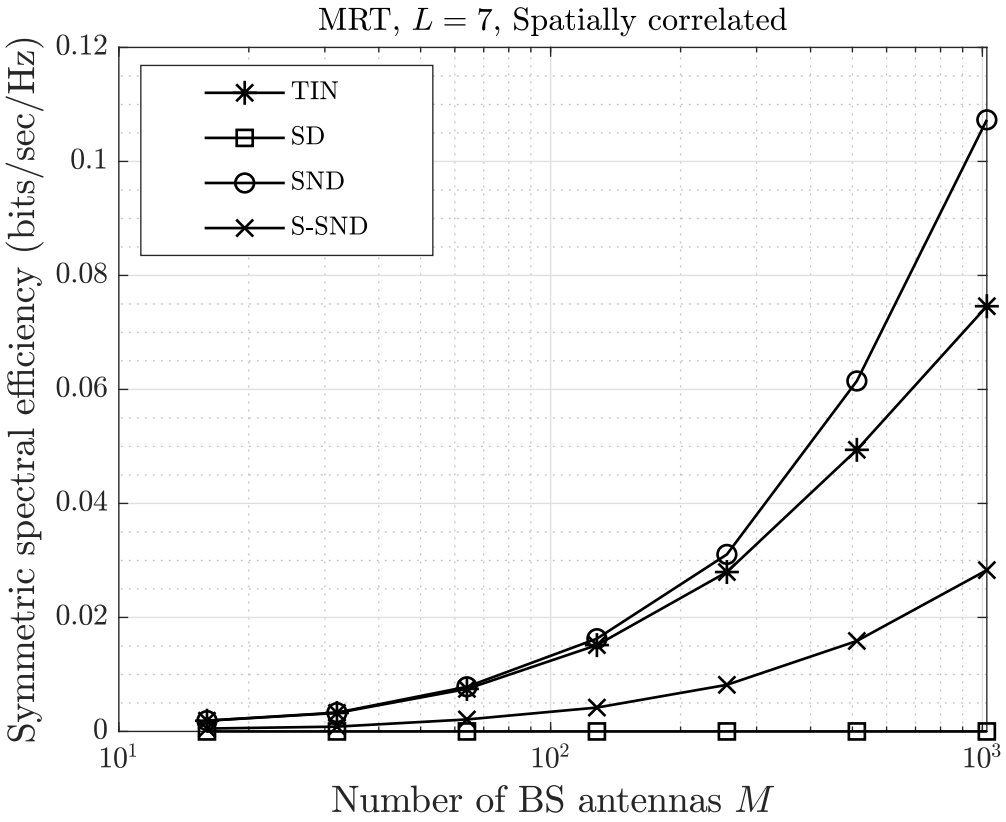


Figure 3.9: Performance of maximum symmetric SE for seven cells when MRT precoding and a spatially correlated channel model are used.

other schemes for this range of antennas. Particularly, the gain offered by SND is about 19% and 32% for $M = 128$ and $M = 256$, respectively, and this gain increases to about 51% when $M = 1024$. This is similar to ZF, and is indeed in agreement with the previous discussion, as M is relatively large in these scenarios. This comparison becomes more clear by looking at Fig. 3.14, which shows the performance of SND for RZF and ZF in the same plot. One can observe that when M is small (i.e., $M \leq 64$), there is a visible gain offered by RZF, while for large M the performance of RZF converges to that of ZF.

Next, we study the impact of changing the number of users K and the correlation magnitude κ on the performance of the different schemes. To this end, we assume that ZF precoding is applied at the BSs. Fig. 3.15 shows the performance of TIN/SD/SND with

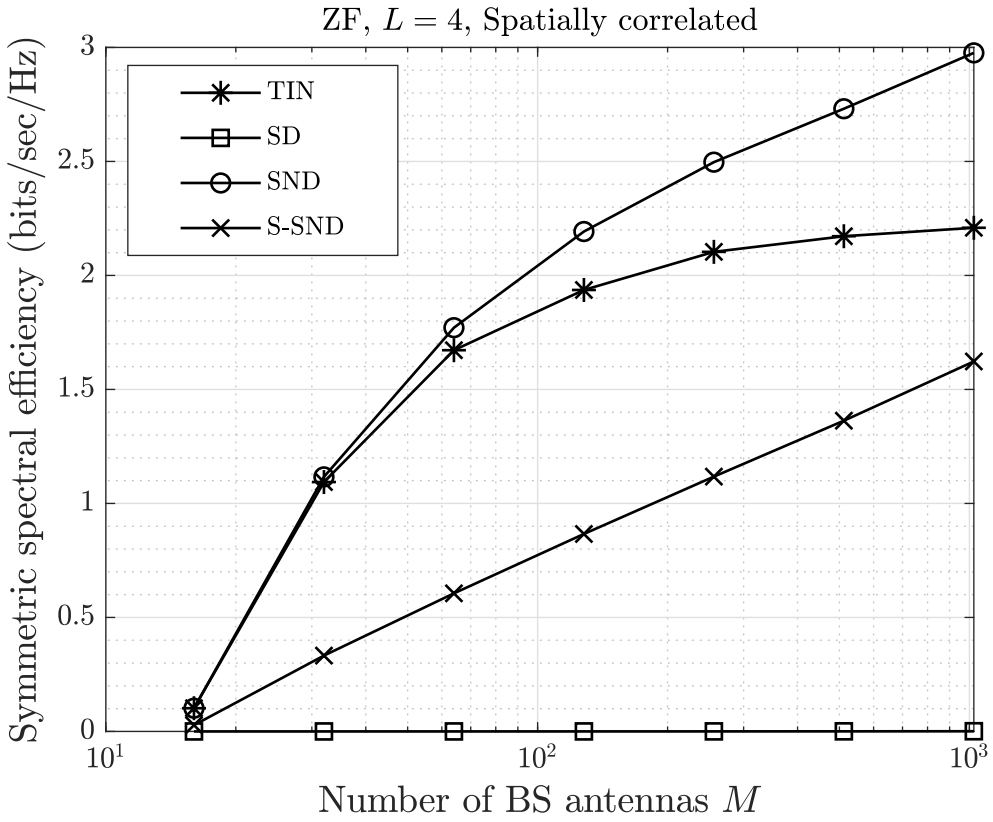


Figure 3.10: Performance of maximum symmetric SE for four cells when ZF precoding and a spatially correlated channel model are used.

maximum symmetric SE versus the number of users K, when $L = 7$ and $M = 256$. It can be observed that increasing K results in smaller achievable symmetric SEs. This is similar to the case of increasing L, as serving a larger number of users leads to smaller symmetric SEs. Consequently, the gain offered by SND improves when K increases. More specifically, Fig. 3.15 shows a gain of about 11% when $K = 2$, which increases up to about 33% when $K = 15$.

The impact of changing the correlation magnitude κ on system performance is shown in Fig. 3.16, where $L = 7$, $M = 256$ and $K = 15$. It can be seen from the figure that increasing the correlation magnitude from 0 (equivalent to the case of uncorrelated fading) to 0.8 (equivalent to strong spatial correlation) results in improving the performance of

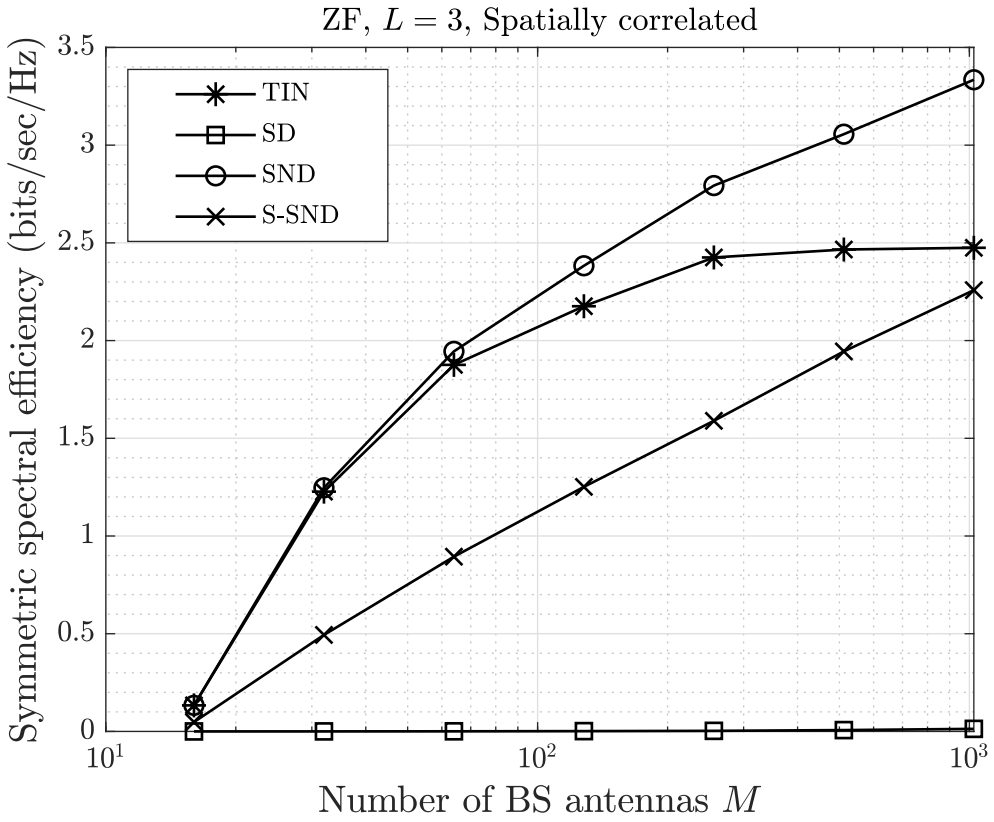


Figure 3.11: Performance of maximum symmetric SE for three cells when ZF precoding and a spatially correlated channel model are used.

both TIN and SND, while the performance of SD does not change much. Consequently, as the performance of both schemes improves by transitioning towards the strong spatial correlation regime, the gain offered by SND reduces. More precisely, SND provides a gain of about 56% in the spatially uncorrelated case ($\kappa = 0$), while this gain gradually reduces to about 6% in the strong spatial correlation regime ($\kappa = 0.8$). Also, in the regime of moderate spatial correlation ($\kappa = 0.4$) the gain of SND is about 33%.

The improvement of the performance obtained by TIN is in agreement with the results reported in [111] and [64]. Particularly, it is known that spatial correlation can improve the quality of MMSE channel estimates resulting in reduced pilot contamination effects in massive MIMO systems, provided that users have different spatial correlation charac-

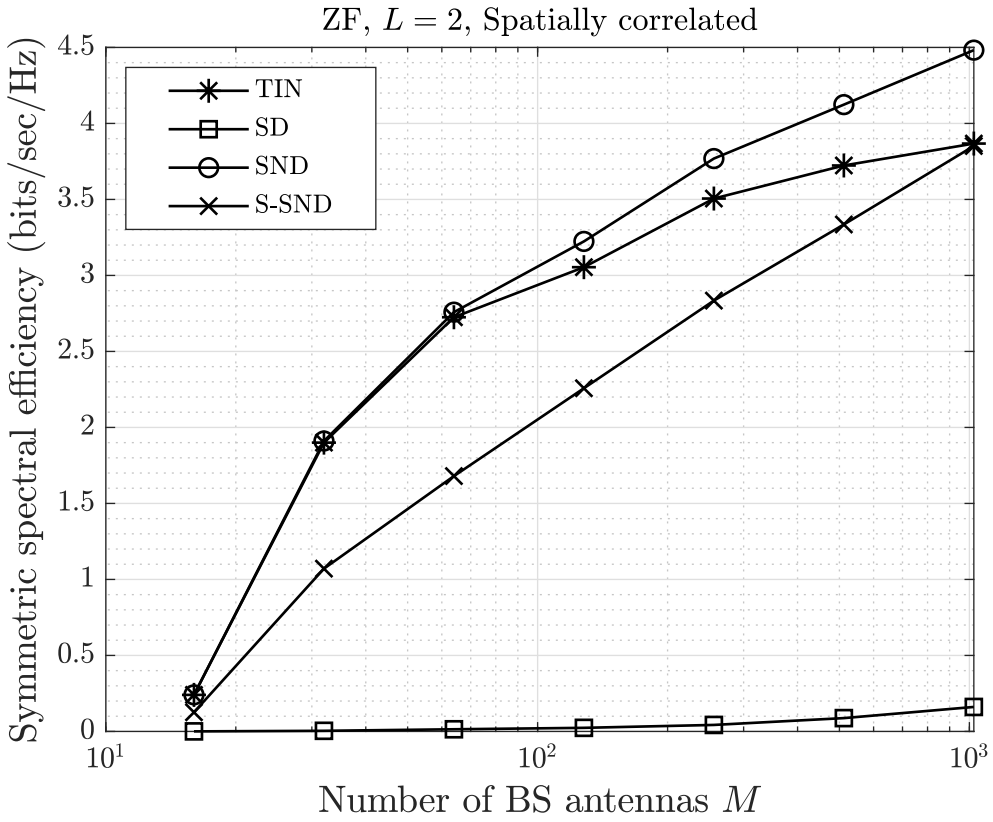


Figure 3.12: Performance of maximum symmetric SE for two cells when ZF precoding and a spatially correlated channel model are used.

teristics [111]. In other words, when pilot sharing users have different spatial correlation characteristics, pilot contamination interference becomes negligible in the strong spatial correlation regime. One should note that while the channel estimates of pilot sharing users are correlated, these users can have very different correlation matrices. For instance, this can happen when pilot sharing users have completely different angles to the antenna boresight in (3.75). On the other hand, when these users have similar spatial correlation characteristics, this interference becomes strong. An example of this scenario is the special case of an uncorrelated channel where the correlation matrices are very similar (recall that in this case $\boldsymbol{R}_{jkl} = \beta_{jkl}\boldsymbol{I}_M$). This is also confirmed by the fact that TIN has the minimum symmetric SE (i.e., maximum pilot contamination interference) in Fig. 3.16 when $\kappa = 0$.

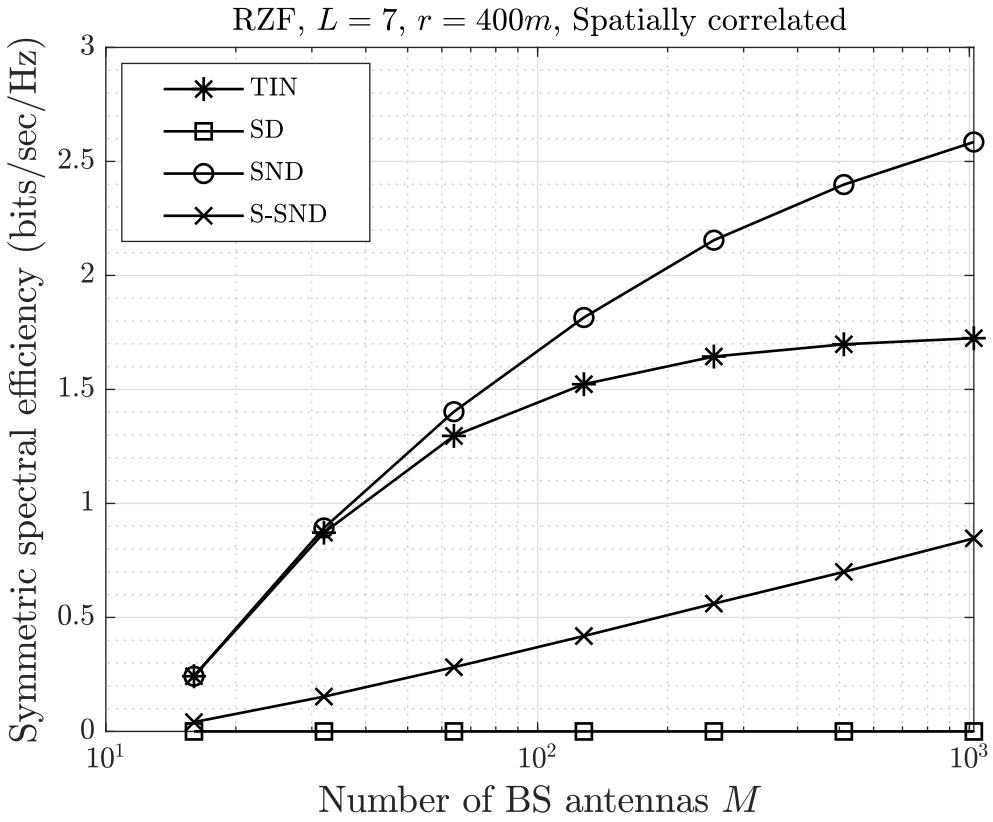


Figure 3.13: Performance of maximum symmetric SE for seven cells when RZF precoding and a spatially correlated channel model are used.

The performance improvement obtained by SND in the spatial correlation regime can also be explained as follows. When performing SND, a receiver (an arbitrary user in the downlink) automatically decides to decode pilot contamination interference non-uniquely, only if its signal strength is good enough, otherwise it will be treated as noise. In other words, when dealing with a spatially correlated channel, if a pilot sharing user is creating strong enough interference at the receiver (e.g., when spatial correlation characteristics of pilot sharing users are similar), it will automatically be decoded under SND. On the other hand, if this user is creating weak interference (e.g., when its spatial correlation matrix is different than that of the user of interest), it will effectively be treated as noise under SND. Thus, it is expected that spatial correlation will improve the performance of SND as well.

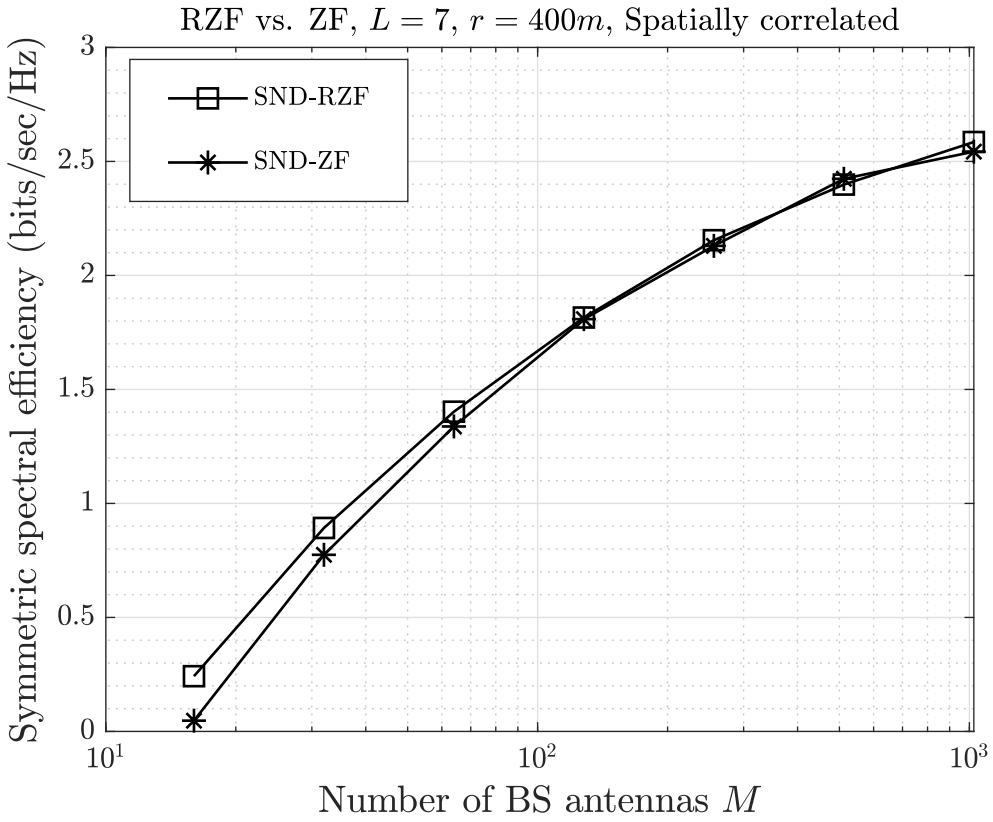


Figure 3.14: Performance comparison between RZF and ZF for $L = 7$, when SND is used.

Next, we study the impact of shadow fading on the performance of the proposed schemes. In particular, we assume that a term associated with shadow fading is now added to the large-scale fading model of (3.74) with a standard deviation of σ_{shadow} in dB. Fig. 3.17 shows the achieved symmetric SEs of the different schemes where the standard deviation of shadow fading, σ_{shadow}, varies in the range from 0 dB to 5 dB. The parameters for this figure are the same as those in Fig. 3.16 except that correlation magnitude is now fixed at $\kappa = 0.4$. It can be observed that, as expected, by increasing the shadow fading the SEs achieved by all schemes reduce. Nevertheless, as σ_{shadow} becomes larger the gain provided by SND over TIN increases, which shows the importance of the proposed scheme in practical scenarios. In particular, when there is no shadowing in the path-loss model of (3.74), SND provides 33% improvement over TIN, whereas when the shadowing increases

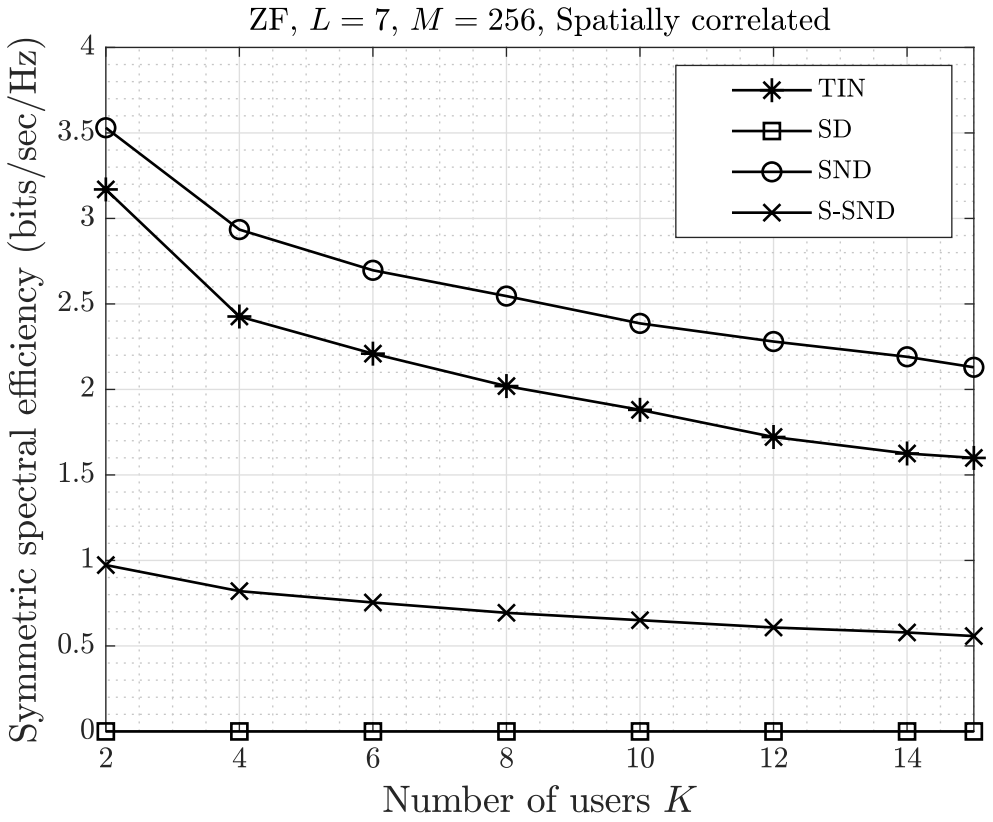


Figure 3.15: Performance of maximum symmetric SE versus the number of users K, where $L = 7$, $M = 256$, and ZF precoding with a spatially correlated channel model are used.

to $\sigma_{\text{shadow}} = 3$ dB the gain provided by SND over TIN improves to 186%, and it continues to grow for larger values of σ_{shadow}.

Next, we consider the seven-cell configuration illustrated in Fig. 3.7d with wrap around topology. In this case, the interference environment seen by the center cell will not change; nevertheless, the six surrounding cells will be subject to greater interference. More specifically, wrap around will change the effective user-BS distances and therefore lead to greater interference for these six cells. Hence, it is expected that this scenario will degrade the achieved SEs. Fig. 3.18 shows the performance of the different schemes TIN/SD/SND versus M when a wrap around topology is used with seven cells. It can be seen that while the SEs have slightly reduced, due to the increased interference in the six surrounding

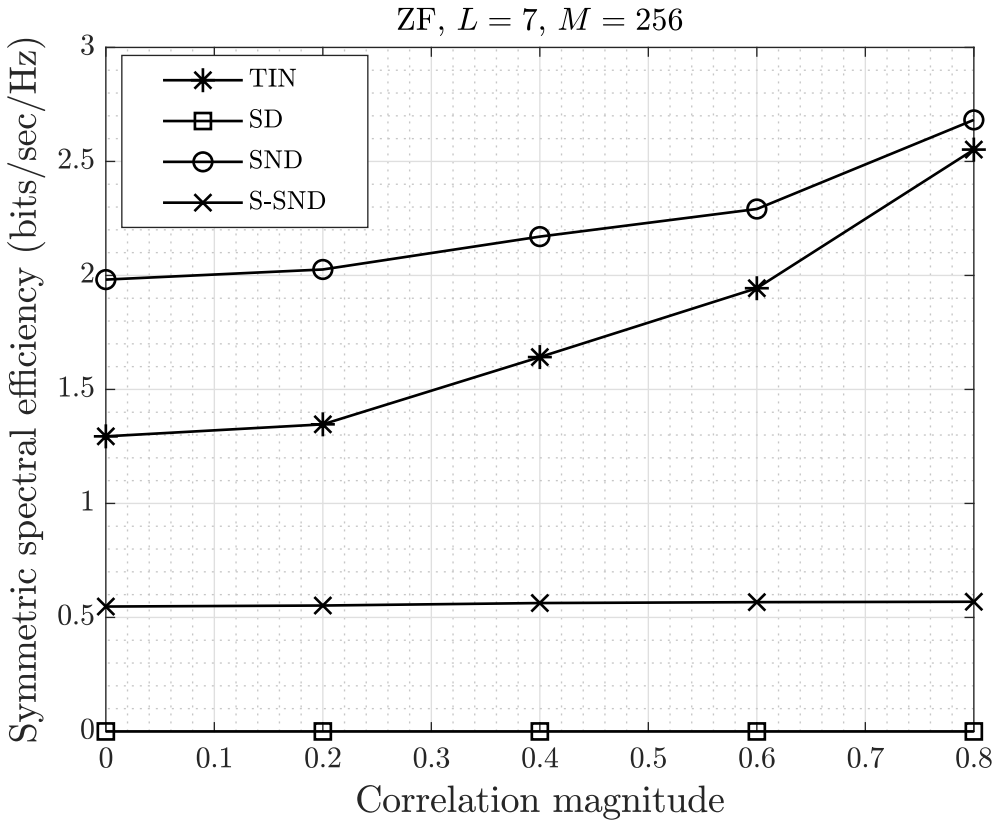


Figure 3.16: Performance of maximum symmetric SE versus the correlation magnitude κ, where $L = 7$, $M = 256$ and ZF precoding is used.

cells, SND still provides a notable gain over TIN, which again confirms the significance of SND in realistic settings. In particular, this gain is about 30% and 40% for $M = 128$ and $M = 256$, respectively, and improves to about 60% when $M = 1024$.

3.4.2 Spatially uncorrelated

We now consider the case of an uncorrelated Rayleigh fading channel model discussed in Section 3.2.3, which is a special case of a spatially correlated channel model. Also, using the closed-form expressions of the rate lower bounds for an uncorrelated channel in Section 3.2.3, we are able to compute the performance for a significantly wider range

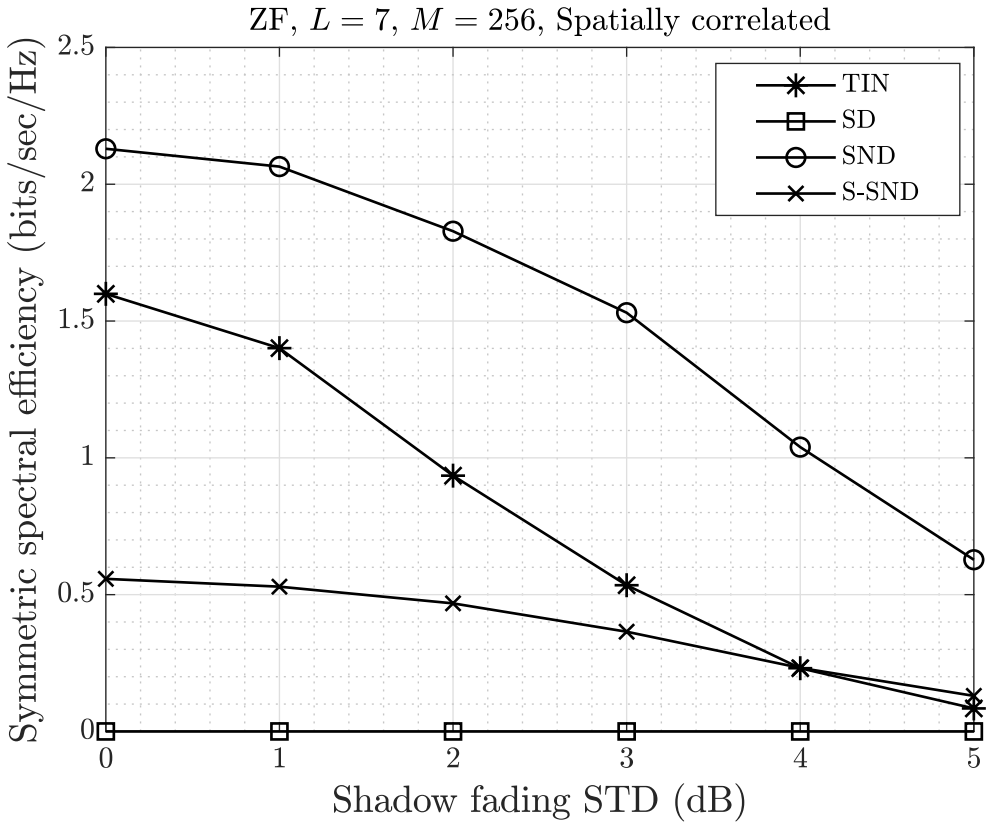


Figure 3.17: Performance of maximum symmetric SE versus the standard deviation of shadow fading σ_{shadow}, where $L = 7$, $M = 256$ and ZF precoding is used.

of M, thus providing insights into the asymptotic performance limits. Specifically, we investigate the impact of changing the number of antennas M, number of cells L and cell radius r on system performance. To do so, we simulate a multi-cell massive MIMO system with spatial correlation matrices given by $\boldsymbol{R}_{jkl} = \beta_{jkl}\boldsymbol{I}_M$, and parameters similar to the correlated case. In particular, the performance of the different schemes TIN/SD/SND with maximum symmetric SE and ZF is studied for $L = 2,\ 3,\ 4$ and 7 and for two choices of cell radius $r = 400$ m, 800 m. To evaluate the performance, the average of the maximum symmetric SEs is calculated over 200 random realizations of user locations.

Fig. 3.19 shows these results for a range of moderately large M, while Fig. 3.20 shows the same for a range of extremely large M, when $L = 7$. While the latter covers a range

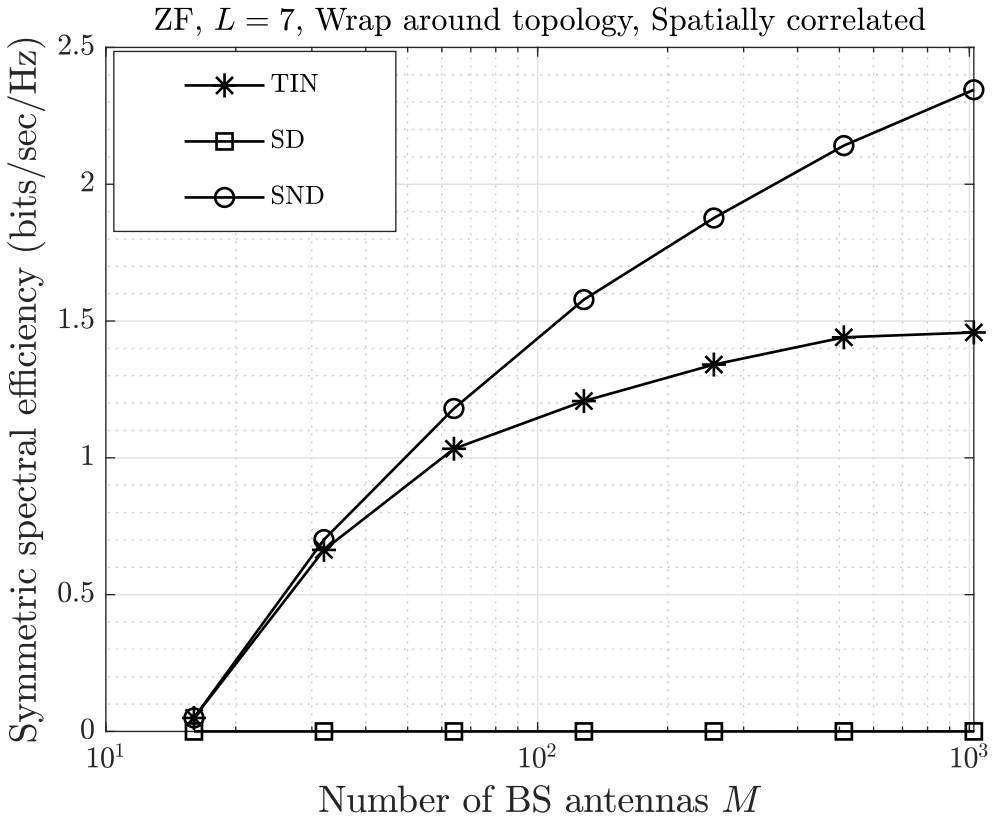
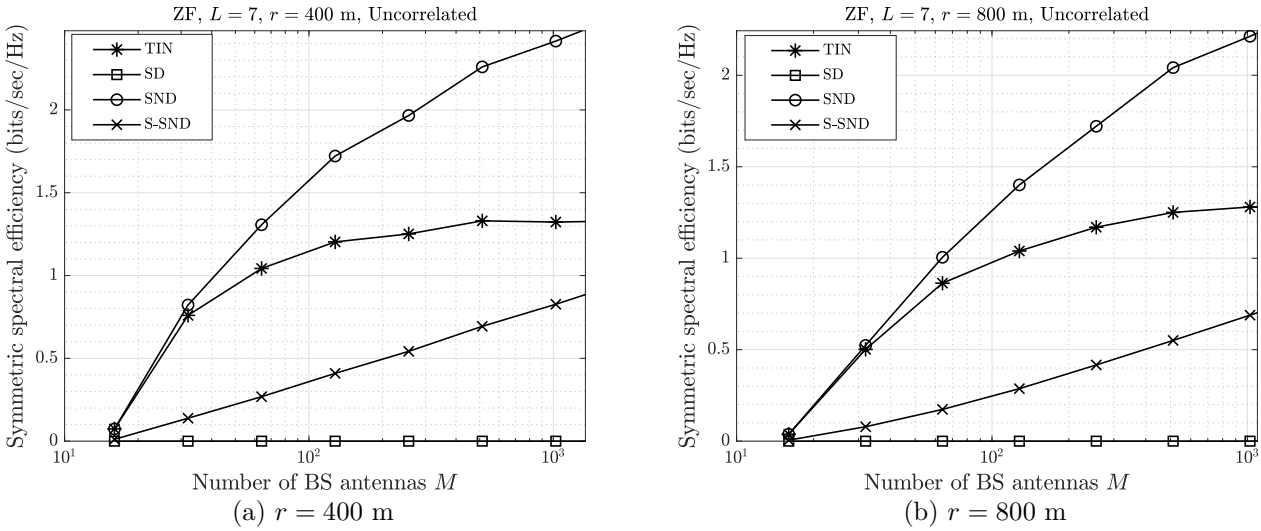


Figure 3.18: Performance of maximum symmetric SE for seven cells with wrap around topology, when ZF precoding and a spatially correlated channel model are used.

of M that is beyond practical, the results of Fig. 3.20 can be used to confirm asymptotic performance limits as $M \to \infty$. By comparing Figs. 3.19a and 3.19b, it is concluded that increasing the cell radius will slightly decrease the gain provided by SND. More specifically, while the gain of SND for $M = 128$ and $M = 256$ is 43% and 56% when $r = 400$ m, respectively, these gains reduce to 34% and 47% when $r = 800$ m. Also, when $M = 1024$ this gain increases to about 82% for $r = 400$ m, whereas it reaches approximately 72% for $r = 800$ m. It is also confirmed via Fig. 3.20 that, when M grows unbounded, the performance of TIN saturates to a constant value while the performance of both SD and SND continue to improve and thus asymptotically converge. In other words, as discussed before, this result confirms that when M becomes truly large, the system enters the high

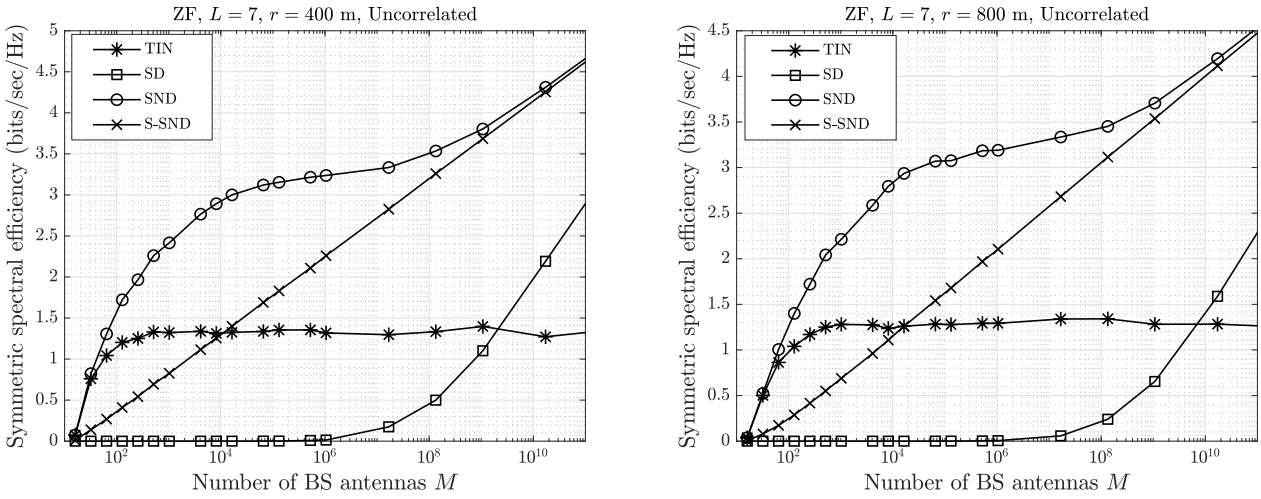


(a) $r = 400$ m (b) $r = 800$ m

Figure 3.19: Performance of maximum symmetric SE for seven cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

SINR regime so that the optimal performance is obtained by jointly decoding signals of all pilot sharing users; hence almost identical performance is obtained under both SND and SD. In addition, by comparing these results with the case of a spatially correlated channel, one can clearly see that spatial correlation improves the performance of both TIN and SND (as discussed before), which leads to a slightly smaller SND gain. Furthermore, as expected, by increasing the cell radius more antennas will be required to achieve the same symmetric SEs.

A similar pattern is observed in Figs. 3.21 to 3.26. Note that Figs. 3.21, 3.23 and 3.25 show the results for a range of moderately large M, whereas Figs. 3.22, 3.24 and 3.26 show the same results for a range of extremely large M. One can notice that for $L = 4, 3$ and 2, the performance improves with increasing M, while achieved symmetric SEs slightly reduce when cell radius increases from $r = 400$ m to $r = 800$ m, which results in smaller SND gains. Moreover, similar to the case of $L = 7$, increasing the cell radius in these scenarios gives rise to requiring more antennas to achieve the same SEs. Lastly,



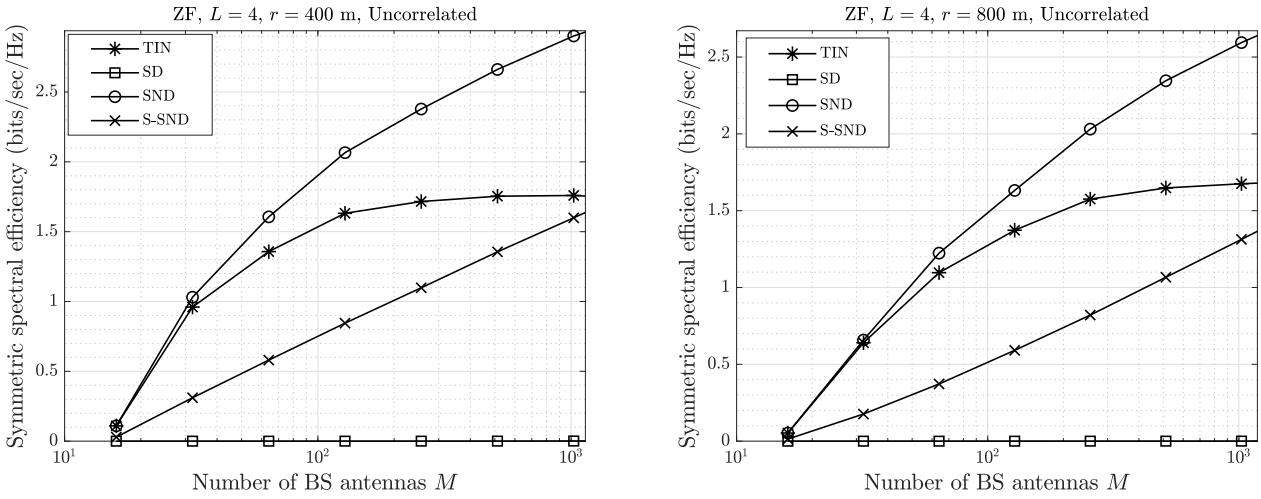
(a) $r = 400$ m (b) $r = 800$ m

Figure 3.20: Performance of maximum symmetric SE for seven cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

one can also conclude from these figures that reducing the number of cells yields larger symmetric SEs and thereby smaller SND gains; a pattern that was also observed for a spatially correlated channel. For instance, in the case of $L = 2$, $M = 256$ and $r = 400$ m, Fig. 3.25 shows a gain of 17% provided by SND (down from 56% when $L = 7$ with the same parameters).

Next, we consider a somewhat pessimistic scenario where all users are located on the cell edge at the farthest distance from their BSs. Specifically, consider the scenario depicted in Fig. 3.27 for $L = 2$, where users' positions are denoted by solid squares. While the position of users in cell 1 is fixed, the position of users in cell 2 varies on an inscribed circle inside the cell based on the angle $\theta \in [0°,\ 360°]$. Note that the users' location in this setup is somewhat in favor of TIN, especially when θ is not close to 180°. This is due to the fact that this assumption makes the power of the signal of interest as well as that of the interfering user from another cell at each receiver very small.

Fig. 3.28 shows the performance of different schemes for a two-cell system with a sym-

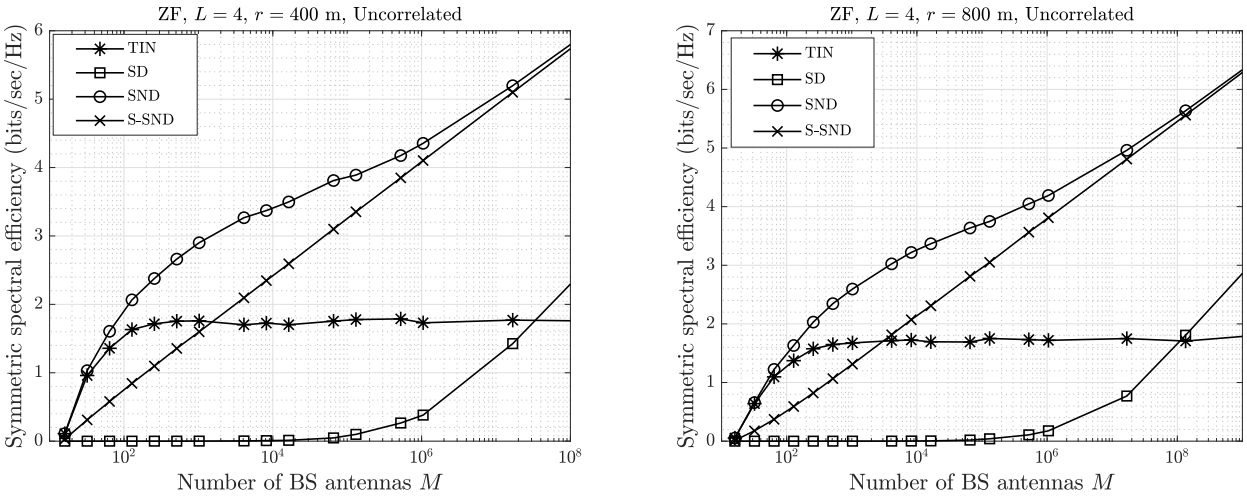


(a) $r = 400$ m (b) $r = 800$ m

Figure 3.21: Performance of maximum symmetric SE for four cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

metric setup, i.e., $\theta = 0°$ in Fig. 3.27, and thus the MACs seen at both cells are identical. The results of Fig. 3.28 are used to validate the analytical findings of Corollary 3 as well as that of the high SINR regime.

It can be observed from Fig. 3.28 that approximately for $M < 10^6$, condition of case (i) in (3.59) is active; thus, SND and TIN have the same performance and strictly outperform SD, i.e., $R_{\text{Sym}}^{\text{SD}} < R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{TIN}}$. In other words, for $M < 10^6$, to achieve the optimum performance in each cell, one should only decode the signal of its own user while treating the signal of pilot contamination interference as noise. On the other hand, when approximately $M > 10^6$, the condition of case (ii) in (3.60) is active; thus, both interference decoding schemes SD/SND are optimal, i.e., $R_{\text{Sym}}^{\text{TIN}} < R_{\text{Sym}}^{\text{SD}} = R_{\text{Sym}}^{\text{SND}}$. Consequently, for significantly large values of M, to achieve the optimum performance in each cell, one should jointly decode both the signal of its own user as well as that of pilot contamination interference. This observation also matches with the consequence of the high SINR regime for truly large M in (3.51). These observations are all in agreement with the analysis performed in



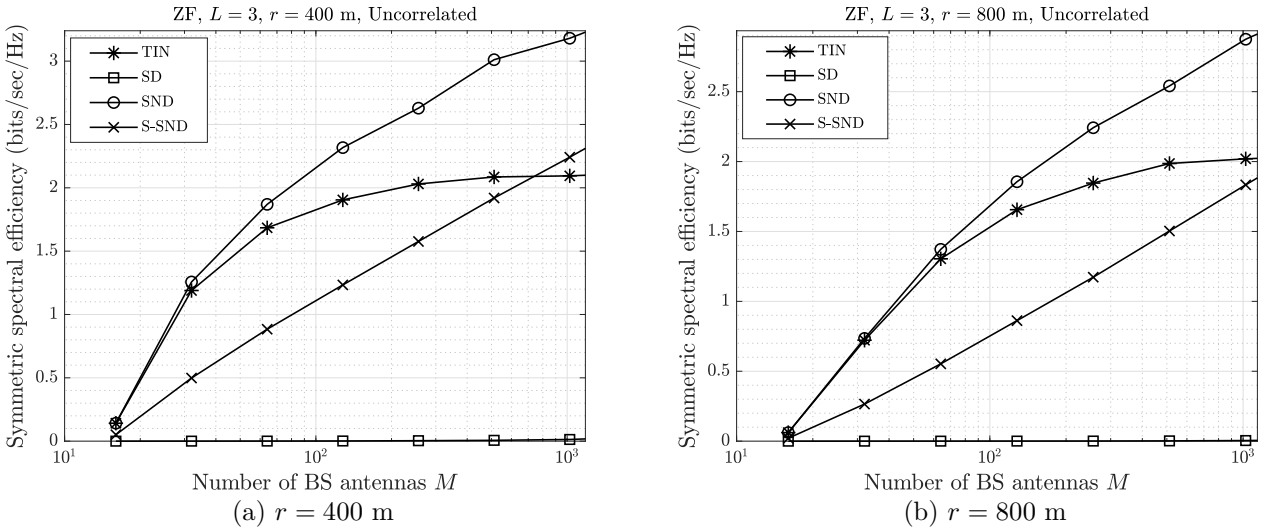
(a) $r = 400$ m (b) $r = 800$ m

Figure 3.22: Performance of maximum symmetric SE for four cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

Section 3.3.2.

Lastly, the results of maximum symmetric SE with ZF versus θ are shown in Fig. 3.29, for four different values of $M = 128,\ 256,\ 512,\ 1024$. First, notice that since this is a somewhat pessimistic scenario and in favor of TIN (as discussed above), for all values of M there exists only a small range of θ where SND outperforms TIN. Second, it can be verified that as M is increased and the performance of SND is thus improved, the range of θ over which SND outperforms TIN expands. Specifically, for $M = 128$ this range is about $152^\circ \leq \theta \leq 204^\circ$, while for $M = 1024$ it increases to about $136^\circ \leq \theta \leq 220^\circ$. In addition, one can notice that as θ increases and approaches 180°, unsurprisingly its performance constantly degrades. This is due to the fact that when θ approaches 180° the strength of pilot contamination interference from users of cell 2 at BS 1 increases, as explained in more detail below.

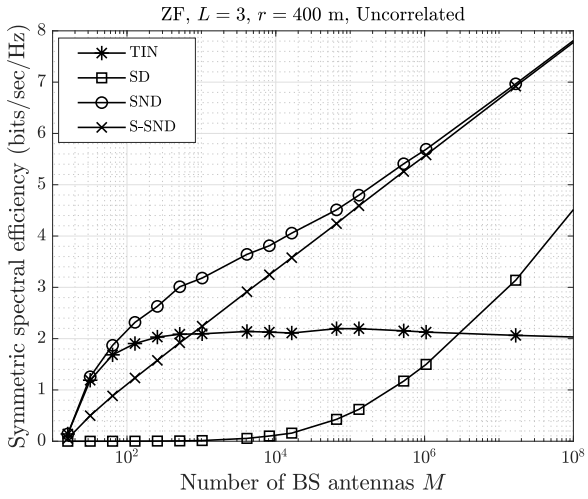
The characteristics of SEs can be classified into 2 regimes of θ: regime-1 where θ is close to 180° and SND outperforms TIN (e.g., $136^\circ \leq \theta \leq 220^\circ$ when $M = 1024$), and



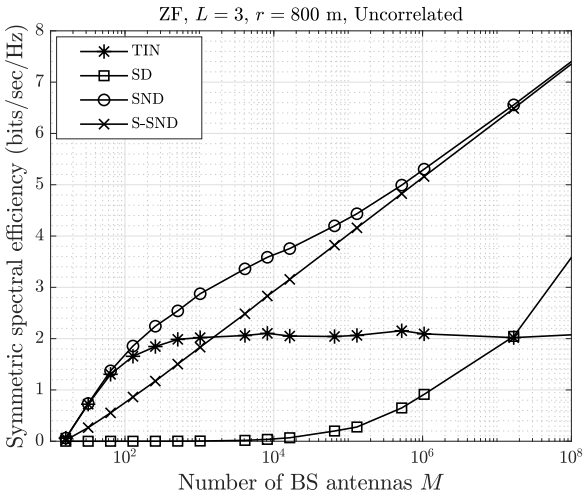
(a) $r = 400$ m (b) $r = 800$ m

Figure 3.23: Performance of maximum symmetric SE for three cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

regime-2 where θ is far from 180° and SND and TIN have the same performance (i.e., $\theta \leq 136°$ and $\theta \geq 220°$ when $M = 1024$). In regime-1, we have $R_{\text{Sym}}^{\text{SND}} > R_{\text{Sym}}^{\text{TIN}} > R_{\text{Sym}}^{\text{SD}}$, as explained in the following. Note that θ captures the distance between users and the BSs in different cells, and when θ enters regime-1, users in cell 2 become relatively closer to BS 1. Therefore, users in cell 2 create "strong" pilot contamination interference at BS 1; hence, TIN performs poorly whereas SND outperforms all other schemes. In contrast, for θ in regime-2, we have $R_{\text{Sym}}^{\text{SND}} = R_{\text{Sym}}^{\text{TIN}} > R_{\text{Sym}}^{\text{SD}}$. Here, users in cell 2 are somewhat far from BS 1, and thus the resulting pilot contamination interference becomes "weak" at BS 1. Hence, performing TIN at both cells is optimal and provides identical performance to that of SND. In the next chapter, we will propose a more advanced decoding scheme based on partial interference decoding, and will observe that the proposed scheme outperforms both TIN and SND for a much wider range of θ.

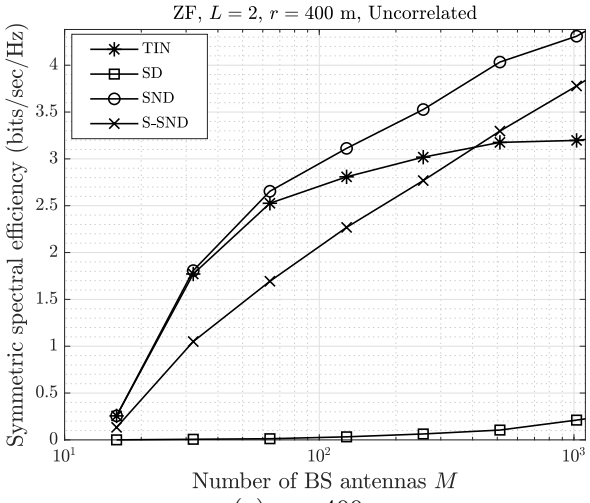


(a) $r = 400$ m

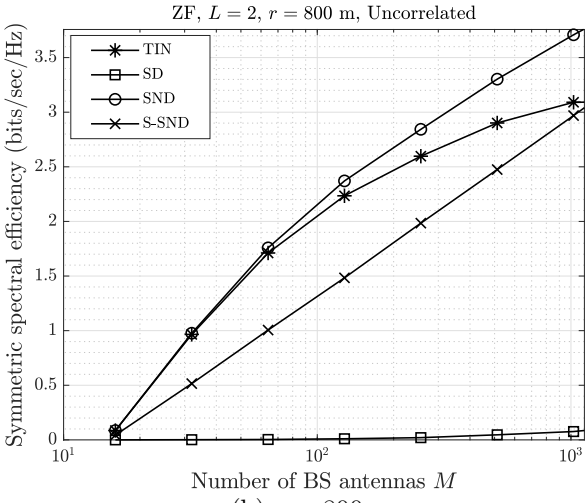


(b) $r = 800$ m

Figure 3.24: Performance of maximum symmetric SE for three cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

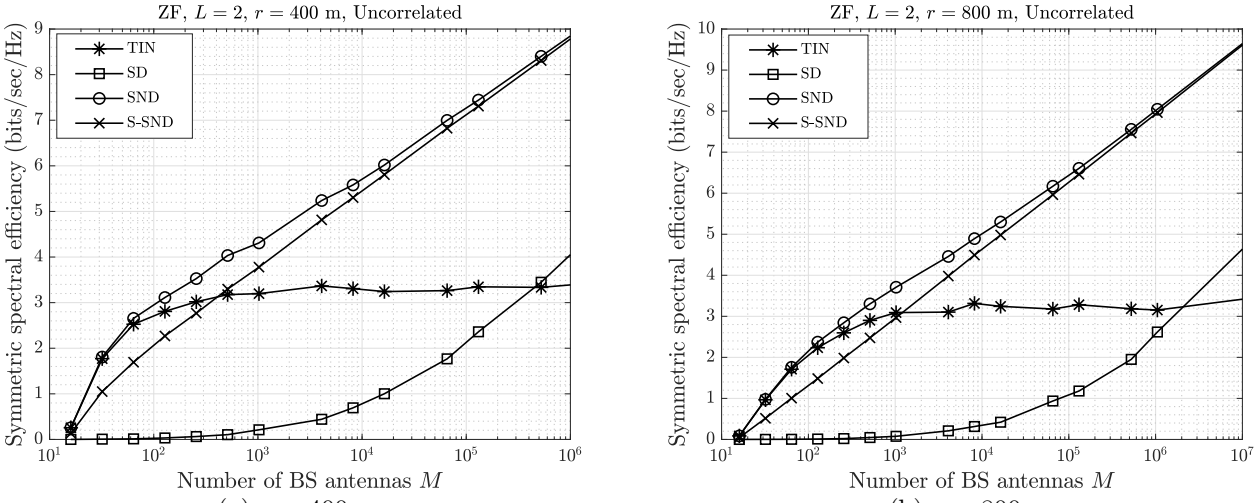


(a) $r = 400$ m



(b) $r = 800$ m

Figure 3.25: Performance of maximum symmetric SE for two cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.



(a) $r = 400$ m (b) $r = 800$ m

Figure 3.26: Performance of maximum symmetric SE for two cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

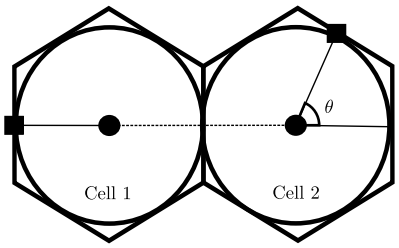


Figure 3.27: Illustration of a two-cell system, where all users of the left cell are located on the cell edge at the farthest distance from the BSs located at the center of the cells, whereas the position of users in the right cell is changing on a circle inside the cell over $0° \leq \theta \leq 360°$. The position of users is denoted by solid squares.

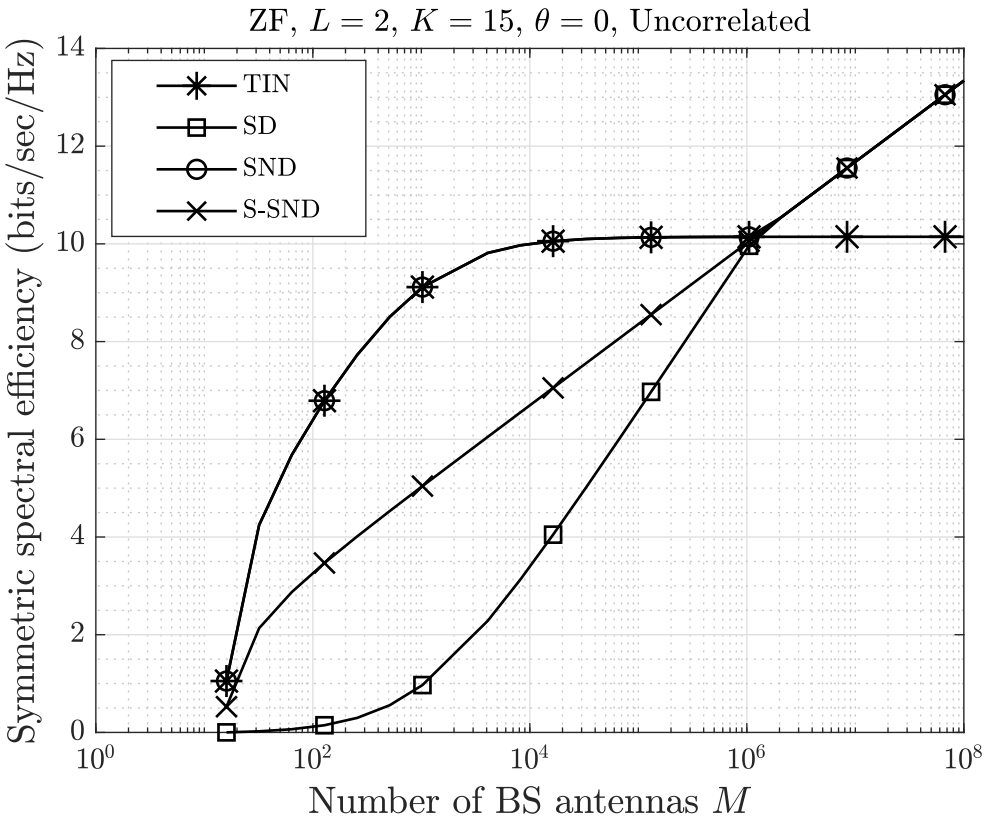
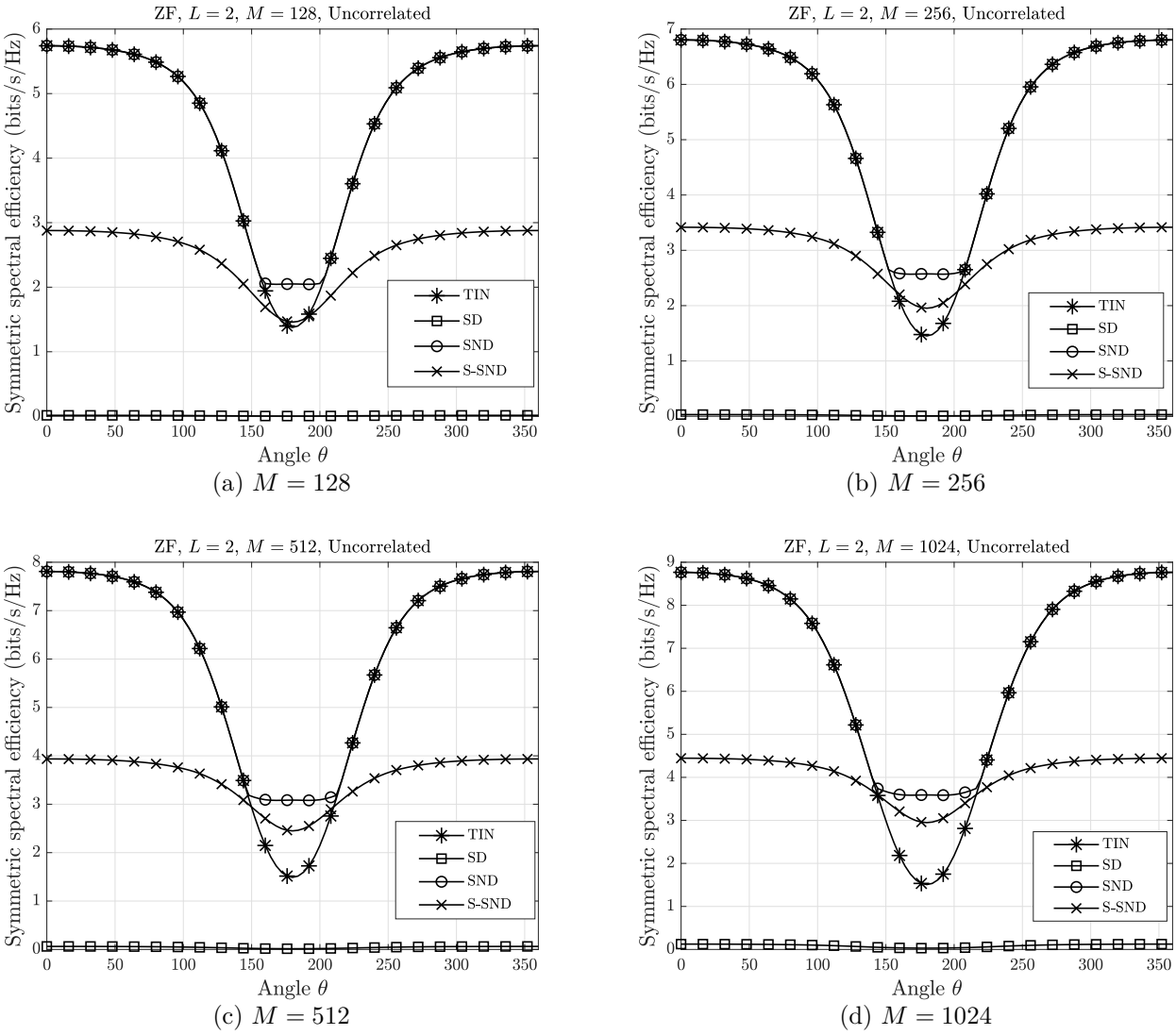


Figure 3.28: Performance of TIN/SD/SND in a two-cell system with a symmetric setup, i.e., $\theta = 0^\circ$ in Fig. 2.1, as a function of M, when ZF precoding and an uncorrelated channel model are used.



(a) $M = 128$ (b) $M = 256$

(c) $M = 512$ (d) $M = 1024$

Figure 3.29: Performance of TIN/SD/SND in a two-cell system with maximum symmetric SE versus θ, when ZF precoding and an uncorrelated channel model are used, (a) $M = 128$, (b) $M = 256$, (c) $M = 512$, (d) $M = 1024$.

Chapter 4

Performance of Partial Interference Decoding

4.1 Partial Interference Decoding

While fully decoding interference or fully treating it as noise are two common strategies for managing interference, it is known that these extreme strategies are not always optimal. For instance, as discussed in Chapter 3, TIN is preferred when interference is weak, whereas SD is only preferred when interference is strong (i.e., when M is truly large). Furthermore, while the results of the previous chapter revealed that SND outperforms both of these schemes, the SND decoder is still *effectively* faced with only two options: treating the interfering signal as noise or fully decoding it[1]; although which one to choose can now be adapted to the strength of the interference. Therefore, the proposed interference decoding schemes of Chapter 3 do not have the flexibility to decode only part of the interference while treating the remaining part as noise. For instance, such flexible decoding can be obtained by the celebrated Han-Kobayashi (HK) scheme [75], which provides the best

[1]Recall from Chapter 3 that each point inside the SND region is equivalent to the rate of decoding a subset of interfering users fully, while treating the remaining ones as noise. However, under SD, an interfering signal is *always* fully decoded; hence performing poorly in the weak and moderate interference regimes.

known achievable performance for the IC.

Motivated in part by the recent introduction of practical sliding-widow codes that can achieve performance close to the HK inner bound for the two-user IC [146] as well as their extension to practical 5G settings [128], in this chapter[2], we depart from the viewpoints of these two extremes (i.e., SD/SND at one extreme end and TIN at the other) and propose a *partial* interference decoding scheme based on rate splitting (RS) and superposition coding techniques. In this proposed partial interference decoding strategy, all users' messages will be partitioned into two independent layers so that each pilot contamination interference term is split into two parts, an inner and an outer layer, based on a power splitting coefficient. By varying the power splitting coefficients, this scheme enables each receiver to partially decode one interference layer while treating the remaining layer as noise; hence achieving a wider range of achievable rates. Therefore, by bridging the extreme strategies of fully decoding the interference or fully treating it as noise, we show that the partial interference decoding scheme of this chapter achieves higher spectral efficiencies for the same number of BS antennas and thereby outperforms the proposed schemes of Chapter 3. However, this performance improvement is achieved at the cost of additional computational complexity due to centrally calculating the optimized rates of individual layers for all BSs.

4.2 Rate Splitting (RS)

When performing conventional RS, the message of each user is split into two layers (each treated as a virtual user) that are superimposed to generate a single codeword at the transmitter: one layer is intended to be decoded only at the targeted receiver (commonly known as the private or outer part) while the other layer can be decoded at a subset/all of the receivers (commonly known as the public or inner part). The total available transmit power is divided into the transmit power of these two layers according to a power splitting coefficient. One can also consider a more general RS scheme where the message of each user is divided into more than two parts such that each part is decoded by a subset of receivers [148, 149].

[2]The results of this chapter are partially presented in [147].

The RS technique was first introduced by Carleial [150] for an IC and was later used in the seminal work of [75] to establish the best known achievable rates for a two-user IC, which contain all other known schemes as special cases (e.g., joint decoding or TIN). Since the introduction of the HK scheme, much effort has been made to simplify this region in closed-form for various special cases of the IC. For instance, the work of [151] has proposed a simplified HK-based RS scheme that can achieve the capacity of a two-user Gaussian IC to within one bit in the weak interference regime. The idea of splitting users' messages in conjunction with superposition coding has been considered in the literature for the purpose of interference mitigation in cellular networks [152–155]. The work of [152] proposed a RS-based scheme in the downlink of a multi-cell network with perfect CSI to jointly design beamforming vectors for public and private parts. Therein, it was shown that by doing single-user successive decoding with a fixed decoding order, higher rates are achieved by this RS scheme compared to conventional TIN. Motivated by the HK scheme, [153] proposed an interference cancellation technique via message splitting at the transmitter along with the SIC decoder at the receiver, that maximizes the sum-rate in heterogeneous networks. A similar technique has been adopted in [154] to mitigate inter-cell interference in a multi-cell multi-user MIMO interference network. In [155], a single RS-based approach has been developed to mitigate interference in the downlink of a MISO BC while minimizing the total transmit power. Specifically, the approach of [155] splits the message of one user only (i.e., the one whose channel is most aligned with the other channels) at the BS, where single-user SIC is used by this user to recover its private message. In another line of work, the idea of message splitting has been used to enhance the efficiency of medium access techniques. For instance, the work of [156] has proposed a RS multiple access (RSMA) technique that improves the performance of schemes such as SDMA and Non-Orthogonal Multiple Access (NOMA). The energy efficiency improvement provided by the RSMA technique has also been investigated in [157].

Recently, RS has also been utilized in the context of massive MIMO communications with imperfect CSI [109, 158, 159]. Specifically, a novel hierarchical RS scheme has been proposed in [158] for the downlink of a single-cell massive MIMO system operating in FDD mode. Therein, the precoding vectors of each public part is designed so as to maximize

the minimum rate of the public part achieved by each user. In [159], the benefits of RS scheme have been investigated to tackle adverse effects of hardware impairments in the downlink of a TDD-based massive MISO broadcast channel. Compared to the case with no RS, the authors in [159] have shown significant improvement from the view point of sum-rate, especially in the high SNR regime. Lastly, the work of [109] has addressed the pilot contamination problem in a *single-cell* massive MIMO system operating in TDD mode, where all users inside the cell share the *same* pilot sequence. While the authors have shown that the decoding scheme of [109] achieves higher sum SE compared to the case without RS, by applying a single-user SIC decoder the intra-cell interference is still treated as noise. This is different than the decoding scheme proposed in this chapter, which tries to partially decode pilot contamination interference (jointly with the intended signal), as will be discussed in the sequel.

In the following, we first discuss how RS can be applied in a two-cell massive MIMO system. Then, motivated by this scheme, we propose a generalization of this scheme to more than two cells (i.e., an L-user IC with $L \geq 3$). Lastly, we study the performance of this RS scheme with the maximum symmetric rate allocation, and elaborate on how the corresponding optimization problem can be solved for each of these cases.

4.2.1 Two-cell system

Consider the two-user IC in Fig. 4.1, associated with the i^{th} user of each cell in the downlink of a two-cell massive MIMO system. To achieve the HK inner bound for this IC, we follow the (simple) scheme of [115, Section 6.5]. As shown in [160], this scheme achieves the same inner bound as the original HK scheme [75].

Recall that in SND, a message at BS l, $m_{il} \in \left[1 : 2^{nR_{il}^{\text{dl}}}\right], l = 1, 2$, is encoded into a single codeword. In contrast, encoding in the RS scheme proceeds as follows. Adopting the scheme of [115] for the Gaussian case, at BS 1, message m_{i1} is first partitioned into two independent parts $m_{i1}^{(a)} \in \left[1 : 2^{nR_{i1}^{\text{dl,(a)}}}\right]$ and $m_{i1}^{(b)} \in \left[1 : 2^{nR_{i1}^{\text{dl,(b)}}}\right]$ such that $R_{i1}^{\text{dl}} = R_{i1}^{\text{dl,(a)}} + R_{i1}^{\text{dl,(b)}}$. Then, part $m_{i1}^{(b)}$ is encoded into codeword $\boldsymbol{s}_1^{(b)}[i](m_{i1}^{(b)})$ of length n (known as the "cloud center" which carries "coarse information"), while part $m_{i1}^{(a)}$ is encoded into another code-

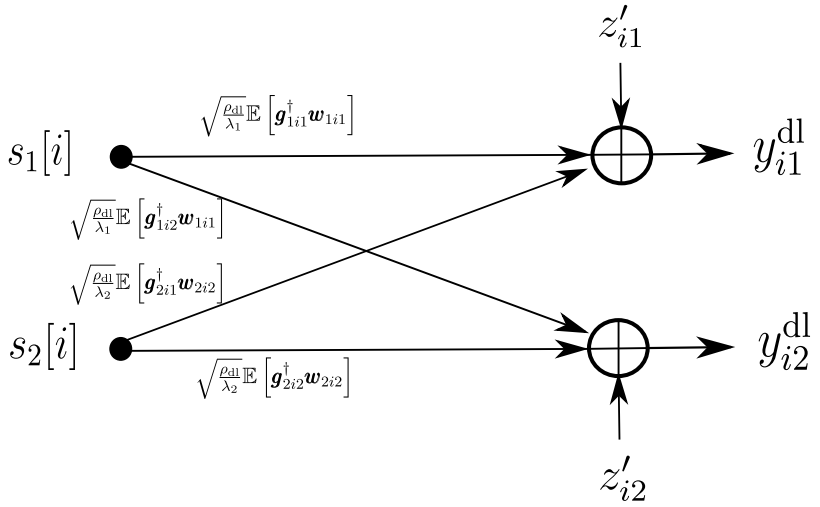


Figure 4.1: The 2-user IC in downlink associated with the $i^{\rm th}$ user of each cell, sharing pilot sequence $\boldsymbol{\psi}_i$. There are a total of K separate/non-interfering such ICs in this two-cell network.

word $\boldsymbol{s}_1^{(a)}[i](m_{i1}^{(a)}, m_{i1}^{(b)})$ of length n; finally the latter codeword is superimposed (or layered) on the former to produce a single codeword for transmission $\boldsymbol{s}_1[i](m_{i1}^{(a)}, m_{i1}^{(b)}) = \boldsymbol{s}_1^{(b)}[i](m_{i1}^{(b)}) + \boldsymbol{s}_1^{(a)}[i](m_{i1}^{(a)}, m_{i1}^{(b)})$ (known as the "satellite codeword" which carries the full information). Similarly, at BS 2, message m_{i2} is first partitioned into two independent parts $m_{i2}^{(a)} \in \left[1 : 2^{nR_{i2}^{\rm dl,(a)}}\right]$ and $m_{i2}^{(b)} \in \left[1 : 2^{nR_{i2}^{\rm dl,(b)}}\right]$ such that $R_{i2}^{\rm dl} = R_{i2}^{\rm dl,(a)} + R_{i2}^{\rm dl,(b)}$. Part $m_{i2}^{(b)}$ is then encoded into codeword $\boldsymbol{s}_2^{(b)}[i](m_{i2}^{(b)})$ (i.e., the "cloud center"), while part $m_{i2}^{(a)}$ is encoded into another codeword $\boldsymbol{s}_2^{(a)}[i](m_{i2}^{(a)}, m_{i2}^{(b)})$. Finally, similar to the superposition encoding at BS 1, the latter codeword is superimposed on the former to produce a single codeword for transmission $\boldsymbol{s}_2[i](m_{i2}^{(a)}, m_{i2}^{(b)}) = \boldsymbol{s}_2^{(b)}[i](m_{i2}^{(b)}) + \boldsymbol{s}_2^{(a)}[i](m_{i2}^{(a)}, m_{i2}^{(b)})$ (i.e., the "satellite codeword"). The total transmit power budget for the $i^{\rm th}$ user at each BS is split into two fixed parts according to the power splitting coefficient $\mu_{il} \in [0, 1],\ l = 1, 2$: the fraction μ_{il} of the budget is allocated to the "outer" layer $\boldsymbol{s}_l^{(a)}[i],\ l = 1, 2$, while the fraction $(1 - \mu_{il})$ of the budget is allotted to the "inner" layer $\boldsymbol{s}_l^{(b)}[i],\ l = 1, 2$. Finally, $\boldsymbol{s}_l^{(a)}[i]$ and $\boldsymbol{s}_l^{(b)}[i]$ are chosen to be i.i.d., zero-mean circularly symmetric complex Gaussian, with powers determined by μ_{il} for $l = 1, 2$. In other words, if the total available power budget for $s_l[i]$ is given by P_{il}, then we have $s_l^{(a)}[i] \sim \mathcal{CN}(0,\ \mu_{il}P_{il})$ and $s_l^{(b)}[i] \sim \mathcal{CN}(0,\ (1 - \mu_{il})P_{il})$.

Note that while full information $(m_{il}^{(a)}, m_{il}^{(b)})$, $l = 1, 2$, is carried in the satellite codeword $\boldsymbol{s}_l[i]$, $l = 1, 2$, the inner layer $\boldsymbol{s}_l^{(b)}[i]$, $l = 1, 2$, only carries coarse information $m_{il}^{(b)}$. Therefore, due to the code construction, the inner layer can be decoded without decoding $m_{il}^{(a)}$ in the outer layer (while treating the outer layer as noise), whereas the outer layer can be decoded either jointly with the inner layer, i.e., $(m_{il}^{(a)}, m_{il}^{(b)})$, or only after $m_{il}^{(b)}$ is decoded first in the inner layer. Hence, when the channel condition is poor (e.g., when M is small), a user may only decode $m_{il}^{(b)}$, while for strong channels (e.g., when M is large) it may choose to decode $(m_{il}^{(a)}, m_{il}^{(b)})$ from both layers. Such adaptability helps the RS scheme outperform the SND scheme of Chapter 3, where messages are not layered, and thus full interference is either non-uniquely decoded or treated as noise.

The decoding at user i in cell 1 proceeds as follows: it decodes both the inner and outer layers of the intended message $(m_{i1}^{(a)}, m_{i1}^{(b)})$ uniquely, and only tries to *non-uniquely* decode the inner layer of the interfering message $m_{i2}^{(b)}$, while treating the outer layer $m_{i2}^{(a)}$ as noise. Similarly at user i in cell 2, both the inner and outer layers of the intended message $(m_{i2}^{(a)}, m_{i2}^{(b)})$ are decoded uniquely, and the inner layer of the interfering message $m_{i1}^{(b)}$ is decoded *non-uniquely*. Using the standard random coding analysis as in [115, Section 6.5.1], such a decoding procedure at user i in cell 1 is successful as $n \to \infty$, when the rates of the inner and outer layers, $R_{i1}^{\mathrm{dl,(a)}}$, $R_{i1}^{\mathrm{dl,(b)}}$ and $R_{i2}^{\mathrm{dl,(b)}}$, satisfy the following constraints

$$R_{i1}^{\mathrm{dl,(a)}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) \tag{4.1}$$

$$R_{i1}^{\mathrm{dl,(a)}} + R_{i1}^{\mathrm{dl,(b)}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i] \middle| s_2^{(b)}[i]\right) \tag{4.2}$$

$$R_{i1}^{\mathrm{dl,(a)}} + R_{i2}^{\mathrm{dl,(b)}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i] \middle| s_1^{(b)}[i]\right) \tag{4.3}$$

$$R_{i1}^{\mathrm{dl,(a)}} + R_{i1}^{\mathrm{dl,(b)}} + R_{i2}^{\mathrm{dl,(b)}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i]\right). \tag{4.4}$$

Similarly, as $n \to \infty$, the error probability of this RS scheme at user i in cell 2 goes to zero, when the rates of the inner and outer layers, $R_{i2}^{\mathrm{dl,(a)}}$, $R_{i2}^{\mathrm{dl,(b)}}$ and $R_{i1}^{\mathrm{dl,(b)}}$, satisfy the following

$$R_{i2}^{\mathrm{dl,(a)}} \leq I\left(y_{i2}^{\mathrm{dl}}; s_2[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) \tag{4.5}$$

$$R_{i2}^{\mathrm{dl,(a)}} + R_{i2}^{\mathrm{dl,(b)}} \leq I\left(y_{i2}^{\mathrm{dl}}; s_2[i] \middle| s_1^{(b)}[i]\right) \tag{4.6}$$

	$m_{i1}^{(\mathrm{a})}$	$m_{i1}^{(\mathrm{b})}$	$m_{i2}^{(\mathrm{b})}$	Joint pdf
E_1	1	1	1	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_1[i], \boldsymbol{s}_2^{(\mathrm{b})}[i])$
E_2	$\times$	1	1	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_2^{(\mathrm{b})}[i])$
E_3	$\times$	$\times$	1	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_2^{(\mathrm{b})}[i])$
E_4	1	$\times$	1	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_2^{(\mathrm{b})}[i])$
E_5	$\times$	1	$\times$	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_1^{(\mathrm{b})}[i])$
E_6	1	$\times$	$\times$	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}})$
E_7	$\times$	$\times$	$\times$	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}})$
E_8	1	1	$\times$	$p(\boldsymbol{s}_1^{(\mathrm{b})}[i], \boldsymbol{s}_1[i]) p(\boldsymbol{s}_2^{(\mathrm{b})}[i]) p(\boldsymbol{y}_{i1}^{\mathrm{dl}} \vert \boldsymbol{s}_1[i])$

Table 4.1: The joint pdfs induced by various $\left(m_{i1}^{(\mathrm{a})}, m_{i1}^{(\mathrm{b})}, m_{i2}^{(\mathrm{b})}\right)$ triples, where '$\times$' denotes an incorrectly decoded message.

$$R_{i2}^{\mathrm{dl},(\mathrm{a})} + R_{i1}^{\mathrm{dl},(\mathrm{b})} \leq I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}[i] \middle| s_2^{(b)}[i]\right) \tag{4.7}$$

$$R_{i2}^{\mathrm{dl},(\mathrm{a})} + R_{i2}^{\mathrm{dl},(\mathrm{b})} + R_{i1}^{\mathrm{dl},(\mathrm{b})} \leq I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}[i]\right). \tag{4.8}$$

The network-wide region is obtained by taking the intersection of the achievable region in cell 1, i.e., (4.1)-(4.4) and the achievable region in cell 2, i.e., (4.5)-(4.8).

In the following, we discuss a brief proof sketch of the achievability, as it provides insights that are particularly important when considering more than two cells, for which we will use similar arguments to describe the achievable region. Assume, without loss of generality, that the message pairs $(m_{i1}^{(\mathrm{a})}, m_{i1}^{(\mathrm{b})}) = (1,1)$ and $(m_{i2}^{(\mathrm{a})}, m_{i2}^{(\mathrm{b})}) = (1,1)$ are sent. As for the error events that result from different combinations of the triples $(m_{i1}^{(\mathrm{a})}, m_{i1}^{(\mathrm{b})}, m_{i2}^{(\mathrm{b})})$ at receiver 1 and $(m_{i2}^{(\mathrm{a})}, m_{i2}^{(\mathrm{b})}, m_{i1}^{(\mathrm{b})})$ at receiver 2, there are 8 cases to be considered at each receiver. We only consider the 8 cases at receiver 1 (listed in Table 4.1 adopted from [115, Section 6.5.1]) giving rise to (4.1)-(4.4). Similar arguments can be applied at receiver 2 leading to (4.5)-(4.8). By the law of large numbers, the probability of the error event that results from E_1 tends to zero, as $n \to \infty$. Also, since $m_{i2}^{(\mathrm{b})}$ is decoded non-uniquely at receiver 1, E_8 does not cause an error. In addition, as $n \to \infty$, the probabilities of the error events that result from E_2 and E_5 tend to zero if (4.1) and (4.3)

are satisfied, respectively. Moreover, as can be seen in the last column of Table 4.1, due to the structure of the satellite codewords, E_3 and E_4 have the same joint pdf. Therefore, as $n \to \infty$, the probability of the error event that results from each of these cases tends to zero if (4.2) is satisfied. Similarly, since E_6 and E_7 have the same joint pdf (due to the code construction), the probability of the error events that result from these two cases tends to zero if (4.4) is satisfied. Hence, as $n \to \infty$, by the union bound the average probability of error at receiver 1 tends to zero if (4.1)-(4.4) are satisfied. For a more detailed proof refer to [160, Appendix B]. One can notice that, due to the special structure of the satellite codewords, as explained, the constraints induced by E_4 and E_6 on $R_{i1}^{\mathrm{dl},(\mathrm{b})}$ and $R_{i1}^{\mathrm{dl},(\mathrm{b})} + R_{i2}^{\mathrm{dl},(\mathrm{b})}$, respectively, are redundant and can be removed. In other words, it is concluded that due to the code construction, if both messages $(m_{il}^{(a)}, m_{il}^{(b)})$ are decoded at receiver l, then the constraints that involve $R_{il}^{\mathrm{dl},(\mathrm{b})}$ but not $R_{il}^{\mathrm{dl},(\mathrm{a})}$ are not needed and can thus be omitted from the rate region. This observation will be used in the next section to establish the rate region in the case of more than two cells.

One can also apply Fourier-Motzkin elimination procedure, as in [115, Appendix D], to present the rate region only in terms of R_{i1}^{dl} and R_{i2}^{dl}. Particularly, by substituting $R_{i1}^{\mathrm{dl},(\mathrm{a})} = R_{i1}^{\mathrm{dl}} - R_{i1}^{\mathrm{dl},(\mathrm{b})}$ and $R_{i2}^{\mathrm{dl},(\mathrm{a})} = R_{i2}^{\mathrm{dl}} - R_{i2}^{\mathrm{dl},(\mathrm{b})}$ into (4.1)-(4.4) and (4.5)-(4.8) and performing Fourier-Motzkin elimination with the additional constraints $0 \leq R_{i1}^{\mathrm{dl},(\mathrm{b})} \leq R_{i1}^{\mathrm{dl}}$ and $0 \leq R_{i2}^{\mathrm{dl},(\mathrm{b})} \leq R_{i2}^{\mathrm{dl}}$, the following seven inequalities are obtained

$$R_{i1}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i] \middle| s_2^{(b)}[i]\right) \tag{4.9}$$

$$R_{i2}^{\mathrm{dl}} \leq I\left(y_{i2}^{\mathrm{dl}}; s_2[i] \middle| s_1^{(b)}[i]\right) \tag{4.10}$$

$$R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}\right) \tag{4.11}$$

$$R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i] \middle| s_1^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}[i] \middle| s_2^{(b)}[i]\right) \tag{4.12}$$

$$R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) \tag{4.13}$$

$$2R_{i1}^{\mathrm{dl}} + R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) + I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}[i] \middle| s_2^{(b)}[i]\right) \tag{4.14}$$

$$R_{i1}^{\mathrm{dl}} + 2R_{i2}^{\mathrm{dl}} \leq I\left(y_{i1}^{\mathrm{dl}}; s_1[i], s_2^{(b)}[i] \middle| s_1^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i] \middle| s_1^{(b)}[i], s_2^{(b)}[i]\right) + I\left(y_{i2}^{\mathrm{dl}}; s_2[i], s_1^{(b)}[i]\right) \tag{4.15}$$

Note that the RS region in (4.9)-(4.15) is obtained for a fixed power splitting strategy μ_{i1} and μ_{i2}. To obtain the overall achievable region, one needs to take the union over all possible such strategies satisfying $(\mu_{i1}, \mu_{i2}) \in [0, 1]$. Some special choices of μ_{i1} and μ_{i2} have interesting interpretations as given in the remark below.

Remark 13. *On the one hand, setting $\mu_{i1} = \mu_{i2} = 1$ (equivalent to $s_1^{(b)}[i] = s_2^{(b)}[i] = 0$) means that user i of each cell only decodes its own message while treating the interfering signal of the other cell entirely as noise; hence, achieving the TIN region of Chapter 3. On the other hand, setting $\mu_{i1} = \mu_{i2} = 0$ (equivalent to $s_1^{(b)}[i] = s_1[i]$ and $s_2^{(b)}[i] = s_2[i]$) means that user i of each cell jointly decodes its own message along with the entire part of the interfering signal; hence, achieving the SD region of Chapter 3. Moreover, by taking the union over the four different possibilities of $\mu_{i1} \in \{0, 1\}$ and $\mu_{i2} \in \{0, 1\}$, one achieves the SND region of Chapter 3, which is the union of TIN and SD (as discussed before).*

The above remark confirms that by taking the union of all possible power splitting strategies, the RS scheme achieves all other schemes of Chapter 3 (TIN/SD/SND) in addition to providing the flexibility of partially decoding pilot contamination interference while treating the remaining part as noise.

Using the assumption of Gaussian signaling and a fixed power splitting strategy of μ_{i1} at BS 1 and μ_{i2} at BS 2, one can use the lower bounds established in Chapter 3 to evaluate $I\left(.; .\middle|.\right)$ terms in the region described by (4.1)-(4.4) and (4.5)-(4.8). For instance, assuming that ZF precoding is applied at the BSs and using the bound of (3.32) from the previous chapter, define the following

$$N_l := K \sum_{j=1}^{L} \rho_{\mathrm{dl}} \beta_{jil} + 1 \tag{4.16}$$

$$P_{jl} := \frac{MK \rho_{\mathrm{dl}} \rho_{\mathrm{p}} \beta_{jil}^2 \alpha_{iij}^2}{\sum_{k=1}^{K} \sqrt{\rho_{\mathrm{p}}} \beta_{jkj} \alpha_{jkj}}, \tag{4.17}$$

where $N_l, l = 1, 2$, is the noise power at receiver l and $P_{jl}, j = 1, 2, l = 1, 2$, is the received power of the i^{th} user from cell j at receiver l (refer to the 2-user IC in Fig. 4.1). Note that when $j = l$, P_{jl} is interpreted as the received power of the intended user, whereas when $j \neq l$, it is interpreted as the received power of pilot contamination interference. Therefore, $I\left(.;.\middle|.\right)$ terms in the expressions of (4.1)-(4.4) and (4.5)-(4.8) are lower bounded by

$$I\left(y_{i1}^{\text{dl}}; s_1[i]\middle|s_1^{(b)}[i], s_2^{(b)}[i]\right) \geq C\left(\frac{\mu_{i1}P_{11}}{\mu_{i2}P_{21}+N_1}\right) \tag{4.18}$$

$$I\left(y_{i1}^{\text{dl}}; s_1[i]\middle|s_2^{(b)}[i]\right) \geq C\left(\frac{P_{11}}{\mu_{i2}P_{21}+N_1}\right) \tag{4.19}$$

$$I\left(y_{i1}^{\text{dl}}; s_1[i], s_2^{(b)}[i]\middle|s_1^{(b)}[i]\right) \geq C\left(\frac{\mu_{i1}P_{11}+(1-\mu_{i2})P_{21}}{\mu_{i2}P_{21}+N_1}\right) \tag{4.20}$$

$$I\left(y_{i1}^{\text{dl}}; s_1[i], s_2^{(b)}[i]\right) \geq C\left(\frac{P_{11}+(1-\mu_{i2})P_{21}}{\mu_{i2}P_{21}+N_1}\right) \tag{4.21}$$

$$I\left(y_{i2}^{\text{dl}}; s_2[i]\middle|s_1^{(b)}[i], s_2^{(b)}[i]\right) \geq C\left(\frac{\mu_{i2}P_{22}}{\mu_{i1}P_{12}+N_2}\right) \tag{4.22}$$

$$I\left(y_{i2}^{\text{dl}}; s_2[i]\middle|s_1^{(b)}[i]\right) \geq C\left(\frac{P_{22}}{\mu_{i1}P_{12}+N_2}\right) \tag{4.23}$$

$$I\left(y_{i2}^{\text{dl}}; s_2[i], s_1^{(b)}[i]\middle|s_2^{(b)}[i]\right) \geq C\left(\frac{\mu_{i2}P_{22}+(1-\mu_{i1})P_{12}}{\mu_{i1}P_{12}+N_2}\right) \tag{4.24}$$

$$I\left(y_{i2}^{\text{dl}}; s_2[i], s_1^{(b)}[i]\right) \geq C\left(\frac{P_{22}+(1-\mu_{i1})P_{12}}{\mu_{i1}P_{12}+N_2}\right). \tag{4.25}$$

4.2.2 Beyond two cells

When going beyond two cells (i.e., $L \geq 3$), one possible generalization of the RS scheme can be obtained by considering one power splitting coefficient for each user in the corresponding IC, i.e., L different coefficients $\mu_{il} \in [0, 1], l = 1, ..., L$, for the L-user IC of Fig. 4.2. However, taking the union over the combination of all such power splitting strategies seems infeasible, especially for networks with large number of cells. This motivates the need for a more feasible generalization of the RS scheme which enables the use of a much simpler power splitting strategy. In the following, we propose one possible application of RS to more than two cells that uses only one power splitting coefficient per IC, and show

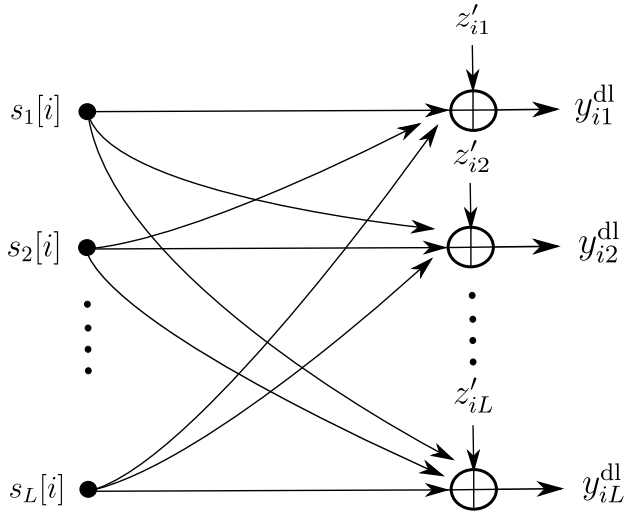


Figure 4.2: The L-user IC in downlink associated with the i^{th} user of each cell, sharing pilot sequence $\boldsymbol{\psi}_i$.

that it can outperform all other schemes of Chapter 3 (TIN/SD/SND).

Consider the downlink L-user IC of Fig. 4.2, which corresponds to the i^{th} user in each of the L cells across the network. Below, we discuss the encoding and decoding stages of the proposed RS scheme for this IC, and further characterize the rate region achieved by this scheme. Recall that there are K non-interfering such ICs in the network and the analysis is thus the same with respect to the index $i \in \{1, ..., K\}$ of the users sharing the same pilot sequence $\boldsymbol{\psi}_i$. As such, for the rest of this chapter, to simplify notation, the index i will be removed from equations.

Encoding: Encoding is similar to the case of a two-cell system in Section 4.2.1 that uses superposition coding, except that now only one power splitting coefficient μ is utilized by all users of the IC in Fig. 4.2. In particular, message $m_l, l = 1, ..., L$, is first partitioned into two independent parts $m_l^{(a)} \in \left[1 : 2^{nR_l^{\text{dl,(a)}}}\right]$ and $m_l^{(b)} \in \left[1 : 2^{nR_l^{\text{dl,(b)}}}\right]$ such that $R_l^{\text{dl}} = R_l^{\text{dl,(a)}} + R_l^{\text{dl,(b)}}, l = 1, ..., L$. Next, part $m_l^{(b)}$ is encoded into codeword $\boldsymbol{s}_l^{(b)}(m_l^{(b)}), l = 1, ..., L$, of length n (i.e., the "cloud center" carrying only "coarse information"), while part $m_l^{(a)}$ is encoded into another codeword $\boldsymbol{s}_l^{(a)}(m_l^{(a)}, m_l^{(b)}), l = 1, ..., L$, of length n. Finally, the latter codeword is superimposed on the former to produce a single

codeword for transmission $\boldsymbol{s}_l(m_l^{(a)}, m_l^{(b)}) = \boldsymbol{s}_l^{(b)}(m_l^{(b)}) + \boldsymbol{s}_l^{(a)}(m_l^{(a)}, m_l^{(b)}), l = 1, ..., L$, (i.e., the "satellite codeword" carrying the full information).

The total transmit power budget at all BSs is split into two fixed parts according to the power splitting coefficient $\mu \in [0, 1]$: the fraction μ of the budget is allocated to the "outer" layer $\boldsymbol{s}_l^{(a)}$, $l = 1, ..., L$, while the fraction $(1 - \mu)$ of the budget is allotted to the "inner" layer $\boldsymbol{s}_l^{(b)}$, $l = 1, ..., L$. Lastly, $\boldsymbol{s}_l^{(a)}$ and $\boldsymbol{s}_l^{(b)}$ are chosen to be i.i.d., zero-mean circularly symmetric complex Gaussian, with powers determined by μ.

Decoding: In the decoding stage, the SND scheme of the previous chapter is applied to *non-uniquely* decode each layer of all pilot contamination interference terms. Specifically, the decoder at receiver l (i.e., user of cell l) uniquely decodes both the inner and outer layers of its own message $(m_l^{(a)}, m_l^{(b)})$, and *non-uniquely* decodes each layer of all interfering messages $\left\{m_j^{(a)}, m_j^{(b)}\right\}, j \in \{1, ..., L\} \setminus \{l\}$. In Appendix A.11, a detailed derivation of the achievable region for a two-cell system is provided. Below, the general achievable region for $L \geq 2$ is presented. The achievability proof follows the same steps as that in Appendix A.11, but is significantly more tedious.

To characterize the rate region achieved by this RS scheme, we first need to define the following sets:

$$\mathcal{S}_l := \left\{\left\{m_l^{(a)}, m_l^{(b)}\right\}\right\} \times \mathcal{A}_1 \times ... \times \mathcal{A}_{l-1} \times \mathcal{A}_{l+1} \times ... \times \mathcal{A}_L, \tag{4.26}$$

where $\times$ denotes the Cartesian product and $\mathcal{A}_j$ is given by

$$\mathcal{A}_j := \left\{\emptyset, \left\{m_j^{(b)}\right\}, \left\{m_j^{(a)}, m_j^{(b)}\right\}\right\}. \tag{4.27}$$

Furthermore, denote the achievable region for the rate vector $\left[R_1^{\mathrm{dl,(a)}}, R_1^{\mathrm{dl,(b)}}, ..., R_L^{\mathrm{dl,(a)}}, R_L^{\mathrm{dl,(b)}}\right]^T$ obtained by the proposed RS scheme at receiver l and the network-wide achievable region by $\mathcal{R}_l^{\mathrm{RS}}$ and $\mathscr{R}^{\mathrm{RS}}$, respectively. Then, following the discussion in Appendix A.11, we have

$$\mathscr{R}^{\mathrm{RS}} = \bigcap_l \mathcal{R}_l^{\mathrm{RS}}, \quad l = 1, ..., L, \tag{4.28}$$

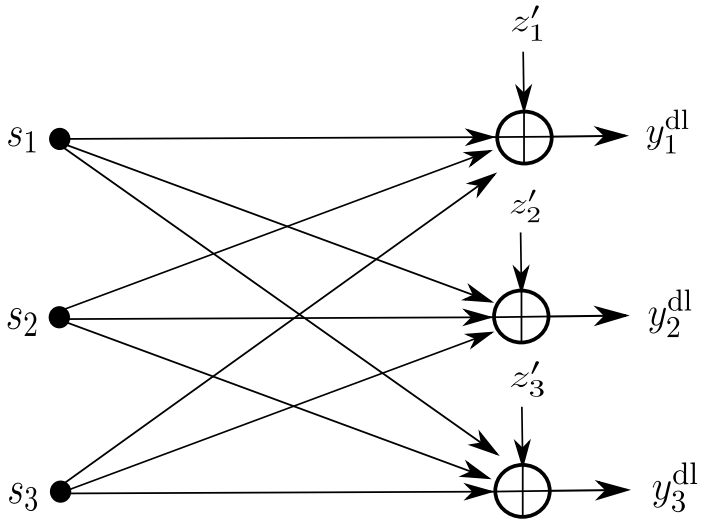


Figure 4.3: The 3-user IC in downlink with one user from each cell, sharing the same pilot sequence.

where

$$\mathcal{R}_l^{\mathrm{RS}} = \bigcup_{\Omega_l \in \mathcal{S}_l} \mathcal{R}_{\mathrm{MAC}(\Omega_l,l)}^{\mathrm{RS}}, \tag{4.29}$$

where $\mathcal{R}_{\mathrm{MAC}(\Omega_l,l)}^{\mathrm{RS}}$ is a modified MAC region (as will be explained in the following) obtained from decoding of the messages included in the set Ω_l, where Ω_l is an element of $\mathcal{S}_l$ defined in (4.26) and thus $\left\{m_l^{(a)}, m_l^{(b)}\right\} \subseteq \Omega_l$.

One should note that the $|\Omega_l|$-user MAC of $\mathcal{R}_{\mathrm{MAC}(\Omega_l,l)}^{\mathrm{RS}}$ has less than $2^{|\Omega_l|}-1$ constraints ($|\Omega_l|$ is the cardinality of the set Ω_l), as some of the constraints will be relaxed because of the following. As shown in Appendix A.11, if Ω_l contains messages of both layers $(m_j^{(a)}, m_j^{(b)})$, for some j, then those constraints that contain $R_j^{\mathrm{dl,(b)}}$ but not $R_j^{\mathrm{dl,(a)}}$ will be removed from the rate region. In Appendix A.11, for the case of a two-cell system we have explicitly identified these constraints at each receiver. Below, we provide the example of a three-cell system and discuss its achievable rate region under the proposed RS scheme.

Dropping the index i, consider a massive MIMO system consisting of only three cells modeled by a 3-user IC as in Fig. 4.3. To apply the proposed RS scheme to this IC, the message of each transmitter is first partitioned into two independent parts: $m_1^{(a)}, m_1^{(b)}$ at

BS 1, $m_2^{(a)}, m_2^{(b)}$ at BS 2, and $m_3^{(a)}, m_3^{(b)}$ at BS 3. Then, by applying superposition coding and non-unique decoding as explained above, the set $\mathcal{S}_l, l = 1, 2, 3$, at each BS is given by:

$$\mathcal{S}_1 = \left\{\left\{m_1^{(a)}, m_1^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_2^{(b)}\right\}, \left\{m_2^{(a)}, m_2^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_3^{(b)}\right\}, \left\{m_3^{(a)}, m_3^{(b)}\right\}\right\} \quad (4.30)$$

$$\mathcal{S}_2 = \left\{\left\{m_2^{(a)}, m_2^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_1^{(b)}\right\}, \left\{m_1^{(a)}, m_1^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_3^{(b)}\right\}, \left\{m_3^{(a)}, m_3^{(b)}\right\}\right\} \quad (4.31)$$

$$\mathcal{S}_3 = \left\{\left\{m_3^{(a)}, m_3^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_1^{(b)}\right\}, \left\{m_1^{(a)}, m_1^{(b)}\right\}\right\} \times \left\{\emptyset, \left\{m_2^{(b)}\right\}, \left\{m_2^{(a)}, m_2^{(b)}\right\}\right\}, \quad (4.32)$$

where $\times$ denotes the Cartesian product. Therefore, the achievable region at each BS using the RS scheme with non-unique decoding is obtained by taking the union of 9 modified MAC regions as discussed before. For instance, at BS 1, one needs to take the union of regions $\mathcal{R}^{\text{RS}}_{\text{MAC}(\Omega_1,1)}$ over the following 9 elements of $\mathcal{S}_1$, denoted by $\Omega_1^{(j)}, j = 1, ..., 9$,

$$\Omega_1^{(1)} = \left\{m_1^{(a)}, m_1^{(b)}\right\} \quad (4.33)$$

$$\Omega_1^{(2)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(b)}\right\} \quad (4.34)$$

$$\Omega_1^{(3)} = \left\{m_1^{(a)}, m_1^{(b)}, m_3^{(b)}\right\} \quad (4.35)$$

$$\Omega_1^{(4)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(a)}, m_2^{(b)}\right\} \quad (4.36)$$

$$\Omega_1^{(5)} = \left\{m_1^{(a)}, m_1^{(b)}, m_3^{(a)}, m_3^{(b)}\right\} \quad (4.37)$$

$$\Omega_1^{(6)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(b)}, m_3^{(b)}\right\} \quad (4.38)$$

$$\Omega_1^{(7)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(a)}, m_2^{(b)}, m_3^{(b)}\right\} \quad (4.39)$$

$$\Omega_1^{(8)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(b)}, m_3^{(a)}, m_3^{(b)}\right\} \quad (4.40)$$

$$\Omega_1^{(9)} = \left\{m_1^{(a)}, m_1^{(b)}, m_2^{(a)}, m_2^{(b)}, m_3^{(a)}, m_3^{(b)}\right\}. \quad (4.41)$$

One can similarly obtain 9 possible choices of Ω_2 and Ω_3, at BSs 2 and 3, respectively, by swapping the appropriate indices. Moreover, as explained before, for each choice of Ω_1 some of the rate constraints in the corresponding MAC region are removed, i.e.,

- for $\Omega_1^{(1)}$, $\Omega_1^{(2)}$, $\Omega_1^{(3)}$ and $\Omega_1^{(6)}$: constraints involving $R_1^{\mathrm{dl,(b)}}$ but not $R_1^{\mathrm{dl,(a)}}$ are relaxed,
- for $\Omega_1^{(4)}$, $\Omega_1^{(7)}$: constraints involving $R_1^{\mathrm{dl,(b)}}$ but not $R_1^{\mathrm{dl,(a)}}$ or $R_2^{\mathrm{dl,(b)}}$ but not $R_2^{\mathrm{dl,(a)}}$ are relaxed,
- for $\Omega_1^{(5)}$, $\Omega_1^{(8)}$: constraints involving $R_1^{\mathrm{dl,(b)}}$ but not $R_1^{\mathrm{dl,(a)}}$ or $R_3^{\mathrm{dl,(b)}}$ but not $R_3^{\mathrm{dl,(a)}}$ are relaxed,
- for $\Omega_1^{(9)}$: constraints involving $R_1^{\mathrm{dl,(b)}}$ but not $R_1^{\mathrm{dl,(a)}}$ or $R_2^{\mathrm{dl,(b)}}$ but not $R_2^{\mathrm{dl,(a)}}$ or $R_3^{\mathrm{dl,(b)}}$ but not $R_3^{\mathrm{dl,(a)}}$ are relaxed.

Regardless of what μ is, some choices of Ω_1 have special interpretations that are discussed in the following remark.

Remark 14. *Note that $\Omega_1^{(1)}$ is equivalent to decoding $\boldsymbol{s}_1$ while treating $(\boldsymbol{s}_2, \boldsymbol{s}_3)$ as noise, i.e., performing TIN at BS 1. Also, $\Omega_1^{(9)}$ is equivalent to jointly decoding $(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$, i.e., performing SD at BS 1. Moreover, $\Omega_1^{(4)}$ is equivalent to jointly decoding $(\boldsymbol{s}_1, \boldsymbol{s}_2)$ while treating $\boldsymbol{s}_3$ as noise, whereas $\Omega_1^{(5)}$ is equivalent to jointly decoding $(\boldsymbol{s}_1, \boldsymbol{s}_3)$ while treating $\boldsymbol{s}_2$ as noise. Consequently, by taking the union of regions $\mathcal{R}^{\mathrm{RS}}_{MAC(\Omega_1,1)}$ over these 4 choices of Ω_1, the SND region is obtained at BS 1 (see (3.34)). Similarly, it can be verified that there are 4 choices of Ω_2 at BS 2 and 4 choices of Ω_3 at BS 3 that give rise to SND region at BSs 1 and 2, respectively. Therefore, notwithstanding the value of μ, the SND region is strictly contained in the region obtained by the proposed RS scheme.*

Remark 15. *If we choose $R_l^{\mathrm{dl,(a)}} = 0, l = 1, ..., L$, in the code construction, then $m_l^{(a)} = 1$, and the codewords are $\boldsymbol{s}_l(1, m_l^{(b)}) = \boldsymbol{s}_l^{(b)}(m_l^{(b)}) + \boldsymbol{s}_l^{(a)}(1, m_l^{(b)}) := \boldsymbol{s}_l(m_l^{(b)}), l = 1, ..., L$, and these are all i.i.d Gaussian. Since in cell l, the messages $m_j^{(b)}, j \neq l$ (i.e., messages from the other cells) are decoded non-uniquely, and also $m_l^{(a)}$ has only one possible value, then for all $\mu \in [0, 1]$, $\mathcal{S}_l$ in (4.26) becomes the set of feasible message combinations of the form $\left\{\boldsymbol{m}_\Omega^{(b)}\right\}_{\{l\}\subseteq\Omega\subseteq\{1,...,L\}}$, where $\boldsymbol{m}_\Omega^{(b)}$ is the vector with entries $m_j^{(b)}, j \in \Omega$. Also, as pointed out in [72, Section 2], the region given by the resulting set of constraints is equivalent to the SND region, i.e., $\left[R_1^{\mathrm{dl}}, ..., R_L^{\mathrm{dl}}\right] \in \mathscr{R}^{\mathrm{SND}} \Longleftrightarrow \left[0, R_1^{\mathrm{dl}}, ..., 0, R_L^{\mathrm{dl}}\right] \in \mathscr{R}^{\mathrm{RS}}$, $\forall \mu \in [0, 1]$.*

Remark 14 confirms that for the case of $L = 3$, considering 4 of the 9 choices is sufficient to reproduce the SND region. Thus, considering the 5 additional choices of $\Omega_l, l = 1, 2, 3$, at each receiver leads to additional flexibility due to partially decoding pilot contamination interference while treating the remaining part as noise; hence, providing room for improving the performance of SND. More specifically, taking the union over all possible choices of μ enables the proposed RS scheme to enlarge the region achieved by SND; thus, outperforming the schemes of Chapter 3, TIN/SD/SND.

Note that this generalization of the RS scheme is different from the HK scheme applied to the two-cell case [115], for which the coarse information is non-uniquely decoded with the outer layer treated as noise and we thus have $\mathcal{A}_j^{\text{HK}} := \left\{\left\{m_j^{(b)}\right\}\right\}$.

4.3 Maximum Symmetric Rate Allocation

Similar to the previous chapter, here the maximum symmetric rate allocation is taken as the measure of fairness among users, and thus the performance of the proposed RS schemes of this chapter will be compared with those of the previous chapter based on the maximum symmetric rate they can offer. Below, we separately discuss how the corresponding optimization problem is solved for the case of a two-cell system and that of more than two cells.

4.3.1 Two-cell system

For a fixed μ_1 and μ_2 in the case of a two-cell system, the maximum symmetric rate is obtained by solving the following problem

$$[\mathcal{P}1] \qquad \max \quad \min \quad \left\{R_1^{\text{dl,(a)}} + R_1^{\text{dl,(b)}}, R_2^{\text{dl,(a)}} + R_2^{\text{dl,(b)}}\right\} \tag{4.42}$$

$$\text{subject to} \quad (4.1) - (4.4) \quad \text{and} \quad (4.5) - (4.8) \tag{4.43}$$

$$R_1^{\text{dl,(a)}}, R_1^{\text{dl,(b)}}, R_2^{\text{dl,(a)}}, R_2^{\text{dl,(b)}} \geq 0, \tag{4.44}$$

where the constraints of (4.1)-(4.4) and (4.5)-(4.8) describe the achievable region obtained by the HK scheme in cell 1 and 2, respectively. $[\mathcal{P}1]$ can be solved by introducing an auxiliary variable t as follows

$$
\begin{aligned}
[\mathcal{P}1'] \qquad & \max \quad t && (4.45)\\
& \text{subject to} \quad (4.1)-(4.4) \quad \text{and} \quad (4.5)-(4.8) && (4.46)\\
& \qquad\qquad R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} \geq t && (4.47)\\
& \qquad\qquad R_2^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \geq t && (4.48)\\
& \qquad\qquad R_1^{\mathrm{dl,(a)}}, R_1^{\mathrm{dl,(b)}}, R_2^{\mathrm{dl,(a)}}, R_2^{\mathrm{dl,(b)}}, t \geq 0. && (4.49)
\end{aligned}
$$

It is readily verified that $[\mathcal{P}1']$ is a linear programming (LP) problem. Specifically, define $\boldsymbol{x} := \left[R_1^{\mathrm{dl,(a)}}, R_1^{\mathrm{dl,(b)}}, R_2^{\mathrm{dl,(a)}}, R_2^{\mathrm{dl,(b)}}, t\right]^T$, $\boldsymbol{c} := [0,0,0,0,1]^T$, and further rewrite (4.47) and (4.48) as

$$
\begin{aligned}
-R_1^{\mathrm{dl,(a)}} - R_1^{\mathrm{dl,(b)}} + t \leq 0 && (4.50)\\
-R_2^{\mathrm{dl,(a)}} - R_2^{\mathrm{dl,(b)}} + t \leq 0. && (4.51)
\end{aligned}
$$

As the constraints of the optimization problem $[\mathcal{P}1']$ are a finite number of closed half-spaces, one can write them in the form of a matrix inequality $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}(\mu_1, \mu_2)$, where $\boldsymbol{A}$ is an 10×5 matrix and $\boldsymbol{b}(\mu_1, \mu_2)$ is an 10×1 column vector (the first 8 rows correspond to (4.46) and the last two rows correspond to (4.50)-(4.51)) which is also a function of power splitting coefficients μ_1 and μ_2. Hence, the following LP is obtained

$$
\begin{aligned}
[\mathcal{P}1'] \qquad & \max \quad \boldsymbol{c}^T\boldsymbol{x} && (4.52)\\
& \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}(\mu_1, \mu_2) && (4.53)\\
& \qquad\qquad \boldsymbol{x} \geq 0, && (4.54)
\end{aligned}
$$

which can be solved efficiently, with the optimal solution denoted by $t^*(\mu_1, \mu_2)$. Note that this LP is solved for a fixed choice of μ_1 and μ_2. However, the overall achievable region is obtained by taking the union over the combination of all power splitting strategies. As

such, the optimal solution to the maximum symmetric rate problem is obtained by taking the maximum over all possible power splitting coefficients, i.e.,

$$\max_{0\leq\mu_1,\mu_2\leq 1} \quad t^*(\mu_1,\mu_2), \tag{4.55}$$

which is solved by numerically searching over the interval $\mu_1,\mu_2\in[0,1]$ to find the best symmetric rate.

4.3.2 Beyond two cells

In the case of more than two cells, we first fix μ and solve the following

$$[\mathcal{P}2] \qquad \max \quad \min_{l} \quad R_l^{\mathrm{dl,(a)}}+R_l^{\mathrm{dl,(b)}} \tag{4.56}$$

$$\text{subject to} \quad \left[R_1^{\mathrm{dl,(a)}},R_1^{\mathrm{dl,(b)}},...,R_L^{\mathrm{dl,(a)}},R_L^{\mathrm{dl,(b)}}\right]\in\mathscr{R}^{\mathrm{RS}}, \tag{4.57}$$

$$R_l^{\mathrm{dl,(a)}},R_l^{\mathrm{dl,(b)}}\geq 0, \quad \forall l\in\{1,...,L\}, \tag{4.58}$$

where $\mathscr{R}^{\mathrm{RS}}$ is given in (4.28). Note that the region $\mathcal{R}^{\mathrm{RS}}_{\mathrm{MAC}(\Omega_l,l)}$ in (4.29) is in the form of a convex polytope and the intersection of a finite number of these convex polytopes yields another convex polytope. Therefore, by distributing the intersection in (4.28) over the union in (4.29) (using the distributive law) the network-wide region $\mathscr{R}^{\mathrm{RS}}$ can be re-written as the union of a finite number of convex polytopes, i.e., $\mathscr{R}^{\mathrm{RS}}=\bigcup_n\tilde{\mathscr{R}}^{\mathrm{RS}}_n$, $n\in\mathcal{I}^{\mathrm{RS}}:=\{1,2,...,N^{\mathrm{RS}}\}$, where N^{RS} is the total number of these convex polytopes, and each n corresponds to one choice of $(\Omega_1,\Omega_2,...,\Omega_L)\in\mathcal{S}_1\times...\times\mathcal{S}_L$,. Similar to $[\mathcal{P}1']$ in the case of a two-cell system, solving the maximum symmetric rate problem over one of these convex polytopes can be formulated as an LP. Specifically, we first define the two $(2L+1)\times 1$ (where $L\geq 3$) column vectors $\boldsymbol{x}:=\left[R_1^{\mathrm{dl,(a)}},R_1^{\mathrm{dl,(b)}},...,R_L^{\mathrm{dl,(a)}},R_L^{\mathrm{dl,(b)}},t\right]^T$ and $\boldsymbol{c}:=[0,...,0,1]^T$. Then, by writing $\tilde{\mathscr{R}}^{\mathrm{RS}}_n$ in matrix form as $\tilde{\boldsymbol{A}}_n\boldsymbol{x}\leq\tilde{\boldsymbol{b}}_n(\mu)$ and the constraints $-R_l^{\mathrm{dl,(a)}}-R_l^{\mathrm{dl,(b)}}+t\leq 0, l=1,...,L$, in matrix form as $\boldsymbol{D}\boldsymbol{x}\leq\boldsymbol{0}$, the following equivalent

LP is thus obtained

$$[\mathcal{P}2'] \quad \max_{n,\boldsymbol{x}} \quad \boldsymbol{c}^T\boldsymbol{x} \tag{4.59}$$

$$\text{subject to} \quad \tilde{\boldsymbol{A}}_n\boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(\mu) \tag{4.60}$$

$$\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{0} \tag{4.61}$$

$$\boldsymbol{x} \geq 0 \tag{4.62}$$

$$n \in \mathcal{I}^{\text{RS}}. \tag{4.63}$$

Denote the optimal value of $[\mathcal{P}2']$ over one of these convex polytopes (i.e., solving $[\mathcal{P}2']$ for a fixed n) by $\tilde{t_n}$. Then, as the network-wide region $\mathscr{R}^{\text{RS}}$ is the union of the convex polytopes $\tilde{\boldsymbol{A}}_n\boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(\mu)$, the optimal solution to $[\mathcal{P}2]$ is found by taking the maximum value of $\tilde{t_n}$, denoted by $t^*(\mu)$, over all sub-regions $\tilde{\mathscr{R}}_n^{\text{RS}}$, i.e., $t^*(\mu) := \max_{n\in\mathcal{I}^{\text{RS}}} \tilde{t_n}$. Lastly, noting that $[\mathcal{P}2]$ (or $[\mathcal{P}2']$) is solved for a fixed μ and also recalling that the overall region is obtained by taking the union over the combination of all possible power splitting strategies, the optimal solution to the symmetric rate problem is found as below

$$\max_{0\leq\mu\leq 1} \quad t^*(\mu), \tag{4.64}$$

which is solved by numerically searching over the interval $\mu \in [0, 1]$ to find the best symmetric rate. It is verified from (4.33)-(4.41) that for $L = 3$, one needs to solve (4.59) over 729 sub-regions, i.e., when $L = 3$ we have $N^{\text{RS}} = 729$ and thereby $\mathcal{I}^{\text{RS}} = \{1, 2, ..., 729\}$. However, as will be discussed in the next section, for the cases of $L = 4$ and $L = 7$ we will only consider a subset of those sub-regions (i.e., a subset of convex polytopes defining the network-wide region) and show that this subset still provides a significant gain over the schemes of Chapter 3. Also, we will see in the next section that the optimization problems in the last step of each case, i.e., (4.55) for two cells and (4.64) for more than two cells, are not necessarily needed, and one can skip numerically optimizing μ and instead use pre-computed average values of splitting coefficients with negligible performance loss.

4.4 Simulation Results

In this section, the performance of rate splitting techniques for the cases of two cells and more than two cells is illustrated using the maximum symmetric SE criteria. Specifically, similar to the previous chapter, the downlink of a multi-cell massive MIMO system is simulated, where the simple HK scheme of Section 4.2.1 for two cells and the RS scheme of Section 4.2.2 for more than two cells are used to partially decode pilot contamination interference while treating the remaining part as noise. For the sake of comparison, in all scenarios we also show the performance of TIN and SND from the previous chapter. The network configuration, path-loss model and parameters of the setup are identical to those of Chapter 3. In particular, unless otherwise specified, we assume $K = 15$ users are uniformly distributed at random within the area of each cell (at least 35 m away from the BS) with a radius of $r = 400$ m and the downlink transmit power of each BS is 40 W. The noise variance is also taken to be -101 dBm, and the large-scale fading coefficients are modeled using (3.74) with the same parameters. Moreover, we only focus on the use of ZF and RZF precoding, since, as observed in Chapter 3, MRT performs poorly compared to these two precoding schemes.

Below, we study the two cases of a spatially correlated Rayleigh fading channel and an uncorrelated Rayleigh fading channel separately.

4.4.1 Spatially Correlated

The exponential correlation model of (3.75) is used here with a correlation magnitude of $\kappa = 0.4$. Below, we first investigate the performance of the HK scheme for the case of $L = 2$ followed by the performance of the RS scheme for $L = 3, 4$ and 7, where similar to the previous chapter, the average of the maximum symmetric SEs is calculated over 150 random realizations of user locations. Also, we separately study the effect of wrap around topology with seven cells at the end of this section.

To compute the maximum symmetric SE of the HK scheme, two different approaches are utilized. In the first approach, the maximum symmetric SE is found by solving (4.55)

and numerically optimizing $0 \leq \mu_1, \mu_2 \leq 1$ (with a step size of 0.01) to find the optimum values of the power splitting coefficients. Then, the average of the optimum choices of μ_1 and μ_2 over 150 realizations is calculated and stored for each value of M. In the second approach, rather than numerically optimizing μ_1 and μ_2, (4.55) is solved for 150 new random realizations of user locations by searching over the combinations of $\mu_1 \in \{0, \mu_{\text{avg}}, 1\}$ and $\mu_2 \in \{0, \mu_{\text{avg}}, 1\}$ in each realization, where μ_{avg} is the pre-computed average value of optimum choices of μ_1 and μ_2 obtained in different random realizations of the first approach. As such, the computational cost of numerically optimizing μ_1 and μ_2 in the first approach is now reduced in the second approach, highlighting its importance in practical implementations. The justification behind the inclusion of the 4 additional combinations $\mu_1 \in \{0, 1\}$ and $\mu_2 \in \{0, 1\}$ is to make sure that the performance obtained in the second approach is always at least as good as SND (cf. Remark 13).

Fig. 4.4 shows the performance of ZF with the HK scheme, when $L = 2$, using the achieved maximum symmetric SE obtained from the two approaches explained above. Interestingly, this figure shows that for each M it is sufficient to use only the pre-computed values μ_{avg} as in the second approach. Specifically, it is revealed in Fig. 4.4 that calculating the maximum symmetric SE using the second approach yields almost the same performance as that obtained from the first approach, and thus using the pre-computed average values of the splitting coefficients can reduce the optimization complexity. In addition, it can be seen that as M is increased the symmetric SE obtained by the HK scheme increases, an observation that was also made in Chapter 3. This figure also shows that the performance gain offered by the HK scheme compared to TIN/SND improves, as M is increased. In particular, it is verified that the HK scheme provides gains of 78% and 80% over TIN, for $M = 128$ and $M = 256$, respectively, and this gain improves to about 100% for $M = 1024$. Also, compared to SND, the HK scheme offers 68% and 70% gains when $M = 128$ and $M = 256$, respectively, while this gain increases to about 75% for $M = 1024$.

Figs. 4.5-4.7 show the performance of ZF for the cases of $L = 3, 4$ and 7, respectively. As discussed in Section 4.3.2, when the number of cells increases beyond $L = 2$, an alternative RS scheme is proposed that uses only one power splitting coefficient μ for the entire IC representing a set of pilot sharing users. The maximum symmetric SEs achieved by this

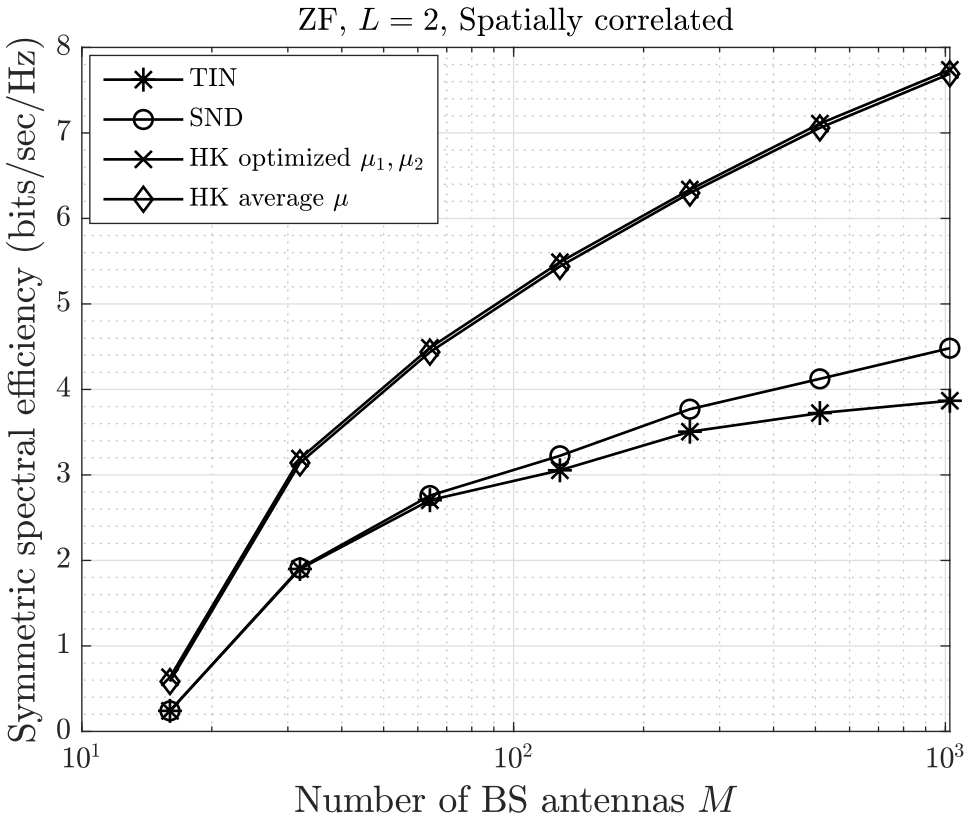


Figure 4.4: Performance of message splitting strategy using maximum symmetric SE for two cells when ZF precoding and a spatially correlated channel model are used.

RS scheme are illustrated in Figs. 4.5-4.7. Again, two different approaches are taken to compute the symmetric SE in each scenario. Specifically, in the first approach, the maximum symmetric SE is found by solving (4.64) and numerically optimizing $0 \leq \mu \leq 1$ (with a step size of 0.02) to find the optimum value of the power splitting coefficient. Then, for each value of M, the average of the optimum choices of μ over 150 realizations is calculated and stored. In the second approach, rather than numerically optimizing μ in (4.64), [$\mathcal{P}2$] is solved using the pre-computed average value of μ and validated on 150 new random realizations of user locations. It is revealed from these three figures that, similar to the case of $L = 2$, the performance obtained using the second approach is almost the same as the one obtained from the first approach, showing the advantage of using pre-computed

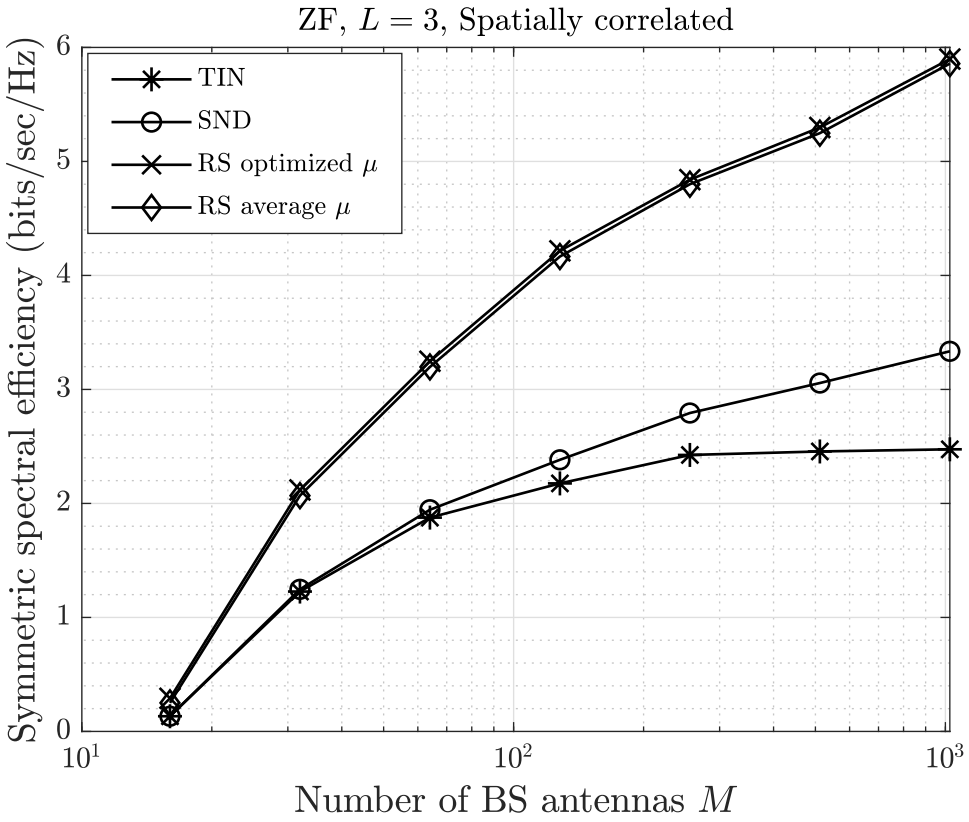


Figure 4.5: Performance of message splitting strategy using maximum symmetric SE for three cells when ZF precoding and a spatially correlated channel model are used.

average values of μ in practical implementations.

Fig. 4.5 shows the performance of ZF for $L = 3$, as a function of M. It can be observed that, compared to the SND scheme of Chapter 3, for all values of M the proposed RS scheme provides significantly higher gains over TIN. Specifically, this gain is about 92% and 99% for $M = 128$ and $M = 256$, respectively, and it increases to about 138% when $M = 1024$. As explained before, this is due to the fact that the proposed RS scheme provides the additional flexibility of decoding part of the interference (depending on the strength of the signal which is a function of M) while treating the remaining part as noise, resulting in extra degrees of freedom in the decoding stage that are not offered by SND.

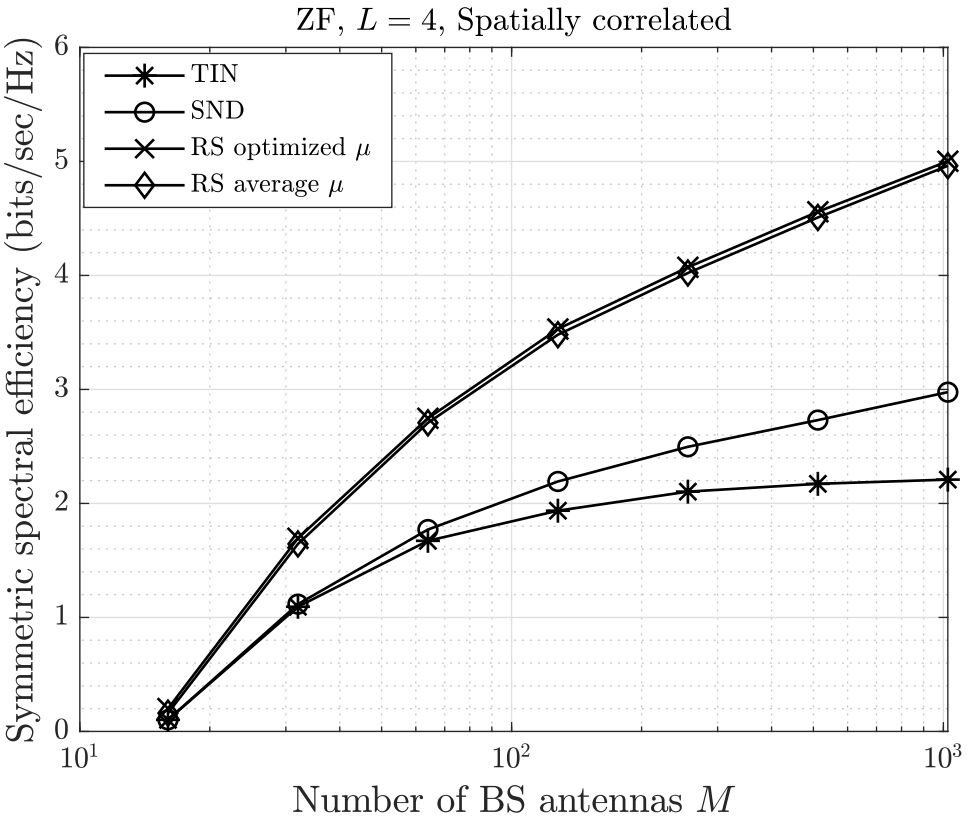


Figure 4.6: Performance of message splitting strategy using maximum symmetric SE for four cells when ZF precoding and a spatially correlated channel model are used.

Note that following the discussion provided in Section 4.3.2, when $L = 4$ or 7, to obtain the maximum symmetric SE over the entire region $\mathscr{R}^{\mathrm{RS}}$ one needs to consider the union of a large number of convex polytopes, which may not be feasible. Hence, to provide some insights into the benefits of using the proposed RS scheme in networks with large number of cells (e.g., $L \geq 4$) we consider an achievable sub-region of $\mathscr{R}^{\mathrm{RS}}$ which provides a lower bound to the true performance of the RS scheme, while offering a significant gain over TIN and SND.

First, define a subset of $\mathcal{S}_l$ as follows

$$\mathcal{S}_l^{\mathrm{Sub}} := \left\{ \left\{ m_l^{(\mathrm{a})}, m_l^{(\mathrm{b})} \right\} \right\} \times \left\{ \left\{ m_1^{(\mathrm{b})} \right\}, ..., \left\{ m_{l-1}^{(\mathrm{b})} \right\}, \left\{ m_{l+1}^{(\mathrm{b})} \right\}, ..., \left\{ m_L^{(\mathrm{b})} \right\} \right., \tag{4.65}$$

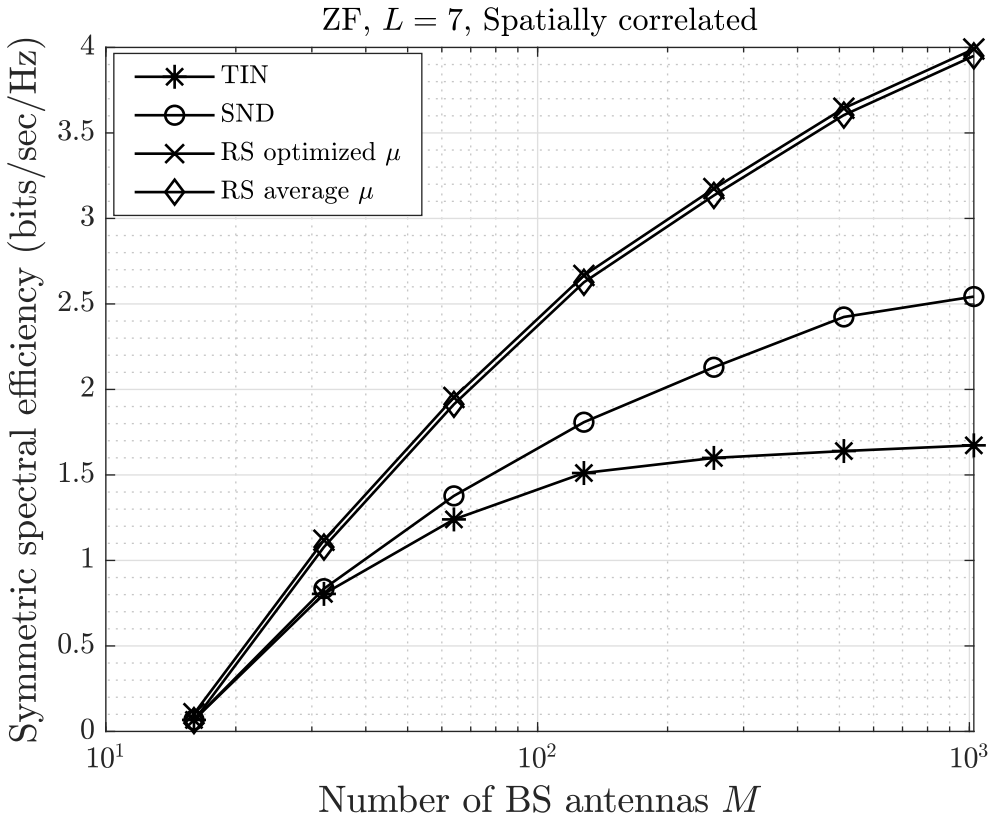


Figure 4.7: Performance of message splitting strategy using maximum symmetric SE for seven cells when ZF precoding and a spatially correlated channel model are used.

$$\left\{m_1^{(\mathrm{b})}, ..., m_{l-1}^{(\mathrm{b})}, m_{l+1}^{(\mathrm{b})}, ..., m_L^{(\mathrm{b})}\right\}\Big\},$$

which gives rise to $\mathscr{R}^{\mathrm{Sub}} \subset \mathscr{R}^{\mathrm{RS}}$ as follows

$$\mathscr{R}^{\mathrm{Sub}} = \bigcap_{l} \mathcal{R}_l^{\mathrm{Sub}}, \quad l = 1, ..., L, \tag{4.66}$$

where

$$\mathcal{R}_l^{\mathrm{Sub}} = \bigcup_{\Omega_l \in \mathcal{S}_l^{\mathrm{Sub}}} \mathcal{R}_{\mathrm{MAC}(\Omega_l, l)}^{\mathrm{RS}}. \tag{4.67}$$

This sub-region $\mathscr{R}^{\mathrm{Sub}}$ of $\mathscr{R}^{\mathrm{RS}}$ can be represented as the union of a finite number of convex

polytopes, i.e., $\mathscr{R}^{\text{Sub}} = \bigcup_n \tilde{\mathscr{R}}_n^{\text{RS}},\ n \in \mathcal{I}^{\text{Sub}}$, where $\mathcal{I}^{\text{Sub}} \subset \mathcal{I}^{\text{RS}}$ since $\mathcal{S}_l^{\text{Sub}} \subset \mathcal{S}_l$.

Note that one can also reproduce the SND region, $\mathscr{R}^{\text{SND}} \subseteq \mathbb{R}^L$, as a "sub-region" of RS, $\mathscr{R}^{\text{RS}} \subseteq \mathbb{R}^{2L}$, as follows. Assuming $\mu = 0$, from the description of the encoding stage in Section 4.2.2, we have $R_l^{\text{dl,(a)}} = 0$ and thus $R_l^{\text{dl}} = R_l^{\text{dl,(b)}}$. Define the projection operator, $\pi:\ \mathbb{R}^{2L} \longrightarrow \mathbb{R}^L$, as follows

$$\pi\left(R_1^{\text{dl,(a)}}, R_1^{\text{dl,(b)}}, ..., R_L^{\text{dl,(a)}}, R_L^{\text{dl,(b)}}\right) = \left[R_1^{\text{dl,(a)}} + R_1^{\text{dl,(b)}}, ..., R_L^{\text{dl,(a)}} + R_L^{\text{dl,(b)}}\right]. \tag{4.68}$$

Therefore, for $\mu = 0$, we can write

$$\mathscr{R}^{\text{SND}} = \pi\left(\bigcap_{l\in\{1,...,L\}} \mathcal{R}_l^{\text{SND}}\right), \tag{4.69}$$

where

$$\mathcal{R}_l^{\text{SND}} = \bigcup_{\Omega_l \in \mathcal{S}_l^{\text{SND}}} \mathcal{R}^{\text{RS}}_{\text{MAC}(\Omega_l, l)}, \tag{4.70}$$

and the set $\mathcal{S}_l^{\text{SND}}$ at receiver l is given by

$$\mathcal{S}_l^{\text{SND}} := \left\{\left\{m_l^{(a)}, m_l^{(b)}\right\}\right\} \times \mathcal{A}'_1 \times ... \times \mathcal{A}'_{l-1} \times \mathcal{A}'_{l+1} \times ... \times \mathcal{A}'_L, \tag{4.71}$$

where $\mathcal{A}'_j := \left\{\emptyset, \left\{m_j^{(a)}, m_j^{(b)}\right\}\right\},\ j \neq l$. Also, from (4.71), it is readily verified that $\mathcal{S}_l^{\text{SND}} \subset \mathcal{S}_l$.

One can similarly re-write the region, $\bigcap_{l\in\{1,...,L\}} \mathcal{R}_l^{\text{SND}}$, given in (4.69)-(4.70) as the union of a finite number of convex polytopes, i.e., $\bigcap_{l\in\{1,...,L\}} \mathcal{R}_l^{\text{SND}} = \bigcup_n \tilde{\mathscr{R}}_n^{\text{RS}},\ n \in \mathcal{I}^{\text{SND}}$, where $\mathcal{I}^{\text{SND}} \subset \mathcal{I}^{\text{RS}}$. Since $\mathcal{I}^{\text{SND}} \subset \mathcal{I}^{\text{RS}}$ and $\mathcal{I}^{\text{Sub}} \subset \mathcal{I}^{\text{RS}}$, it is readily verified that the following problem provides a lower bound to (4.64)

$$[\mathcal{P}3] \qquad \max_{\mu,n,\boldsymbol{x}} \quad \boldsymbol{c}^T\boldsymbol{x} \tag{4.72}$$

$$\text{subject to} \quad \tilde{\boldsymbol{A}}_n\boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(\mu) \tag{4.73}$$

$$\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{0} \tag{4.74}$$

$$0 \leq \mu \leq 1 \tag{4.75}$$

$$\boldsymbol{x} \geq 0 \tag{4.76}$$

$$n \in \mathcal{I}^{\text{SND}} \cup \mathcal{I}^{\text{Sub}}. \tag{4.77}$$

Further define function $f(\mu, \mathcal{I})$ as follows

$$f(\mu, \mathcal{I}) := \quad \max_{n, \boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x} \tag{4.78}$$

$$\text{subject to} \quad \tilde{\boldsymbol{A}}_n \boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(\mu) \tag{4.79}$$

$$\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{0} \tag{4.80}$$

$$\boldsymbol{x} \geq 0 \tag{4.81}$$

$$n \in \mathcal{I}. \tag{4.82}$$

i.e., the optimum value of [$\mathcal{P}3$] for a fixed μ and subset $\mathcal{I} \subseteq \mathcal{I}^{\text{RS}}$. Also, define $g(\mathcal{I}) := \max_{0 \leq \mu \leq 1} \; f(\mu, \mathcal{I})$. For instance, [$\mathcal{P}3$] can be written as $g\left(\mathcal{I}^{\text{SND}} \cup \mathcal{I}^{\text{Sub}}\right)$. Now, it can be verified that

$$g\left(\mathcal{I}^{\text{SND}} \cup \mathcal{I}^{\text{Sub}}\right) = \max\left\{g\left(\mathcal{I}^{\text{SND}}\right),\; g\left(\mathcal{I}^{\text{Sub}}\right)\right\} \tag{4.83}$$

$$= \max\left\{\max_{0 \leq \mu \leq 1} \; f\left(\mu, \mathcal{I}^{\text{SND}}\right),\; \max_{0 \leq \mu \leq 1} \; f\left(\mu, \mathcal{I}^{\text{Sub}}\right)\right\} \tag{4.84}$$

$$\geq \max\left\{\; f\left(0, \mathcal{I}^{\text{SND}}\right),\; \max_{0 \leq \mu \leq 1} \; f\left(\mu, \mathcal{I}^{\text{Sub}}\right)\right\} \tag{4.85}$$

$$= \max\left\{t^{\text{SND},*},\; g\left(\mathcal{I}^{\text{Sub}}\right)\right\}, \tag{4.86}$$

where $t^{\text{SND},*}$ is the optimal value of the objective function in the problem below

$$\max_{n, \boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x} \tag{4.87}$$

$$\text{subject to} \quad \tilde{\boldsymbol{A}}_n \boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(0) \tag{4.88}$$

$$\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{0} \tag{4.89}$$

$$\boldsymbol{x} \geq 0 \tag{4.90}$$

$$n \in \mathcal{I}^{\text{SND}}, \tag{4.91}$$

and $g\left(\mathcal{I}^{\text{Sub}}\right)$ is the optimal value of the objective function given below

$$\max_{\mu,n,\boldsymbol{x}} \quad \boldsymbol{c}^T\boldsymbol{x} \tag{4.92}$$

$$\text{subject to} \quad \tilde{\boldsymbol{A}}_n\boldsymbol{x} \leq \tilde{\boldsymbol{b}}_n(\mu) \tag{4.93}$$

$$\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{0} \tag{4.94}$$

$$0 \leq \mu \leq 1 \tag{4.95}$$

$$\boldsymbol{x} \geq 0 \tag{4.96}$$

$$n \in \mathcal{I}^{\text{Sub}}. \tag{4.97}$$

Therefore, by calculating $\max\left\{t^{\text{SND},*},\ g\left(\mathcal{I}^{\text{Sub}}\right)\right\}$, a lower bound to (4.72) is obtained. Also, note that $t^{\text{SND},*}$ is the maximum symmetric rate of the SND region from the previous chapter, i.e, $t^{\text{SND},*} = R_{\text{Sym}}^{\text{SND}}$. As such, for $L = 4$ and $L = 7$, $\max\left\{R_{\text{Sym}}^{\text{SND}}, g\left(\mathcal{I}^{\text{Sub}}\right)\right\}$ is found, thus providing a lower bound to the true performance of the RS scheme.

Figs. 4.6 and 4.7 show the performance of ZF for the cases of $L = 4$ and $L = 7$, respectively. Notice that a pattern similar to those of Figs. 4.4 and 4.5 is apparent in Figs. 4.6 and 4.7. Specifically, it can be seen from Fig. 4.6 that the gain offered by the RS scheme over TIN is at least 79% and 92% for $M = 128$ and $M = 256$, respectively, and it increases to more than 125% for $M = 1024$. Recall from Chapter 3 that increasing the number of cells results in improving the gains provided by SND over TIN. While a similar observation can be made here when moving from $L = 2$ to $L = 3$ (i.e., the performance gains for $L = 3$ are larger than those for $L = 2$), these gains slightly drop when moving to $L = 4$. This can be explained by the fact that for $L = 3$ the true performance obtained by the RS scheme is calculated, whereas for $L = 4$ and $L = 7$ a lower bound to the true performance of the RS scheme is calculated. Nevertheless, as the case of $L = 7$ is treated in a similar manner, the gains provided by the RS scheme over TIN for seven cells are larger than those in the case of four cells. In particular, when $L = 7$, Fig. 4.7 shows gains of 79% and 98% for $M = 128$ and $M = 256$, respectively, and this gain increases to about 138% for $M = 1024$.

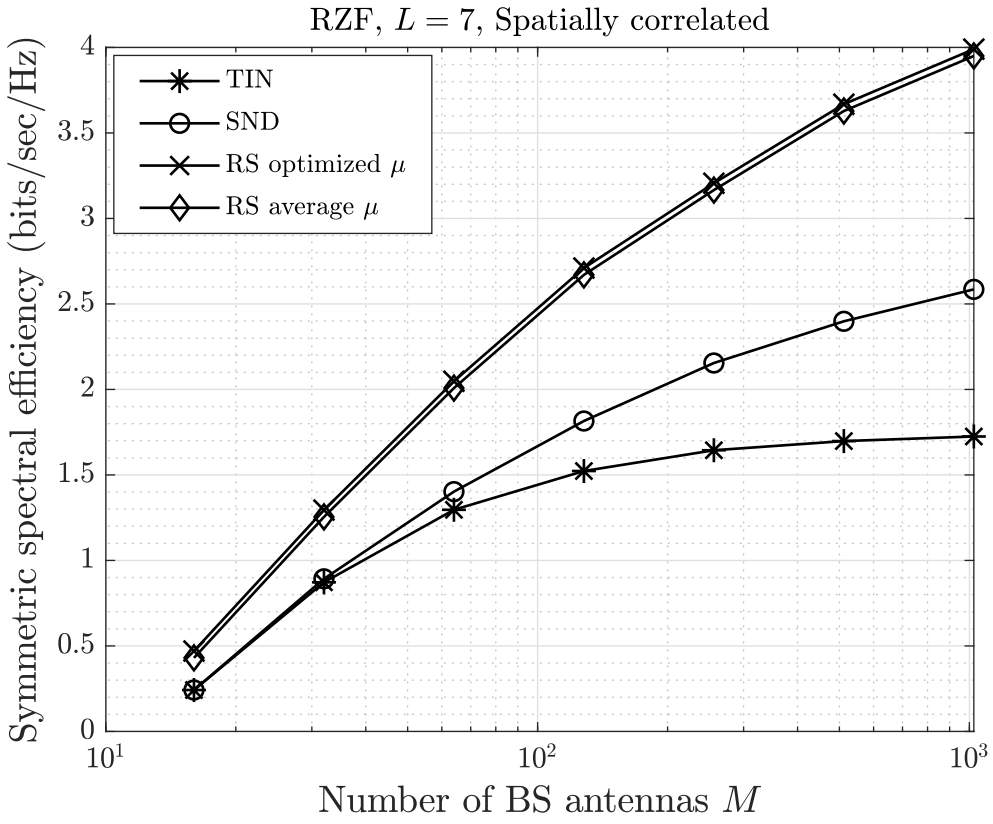


Figure 4.8: Performance of message splitting strategy using maximum symmetric SE for seven cells when RZF precoding and a spatially correlated channel model are used.

The performance of RZF precoding is illustrated in Fig. 4.8 which shows a similar pattern to those of Figs. 4.4-4.7, and the precoding matrix is given in (3.76) where $\delta = K/\rho_{\text{dl}}$. As discussed in the previous chapter, it is expected that when M is small RZF outperforms ZF, whereas for large values of M, the performance gap between RZF and ZF vanishes. This is confirmed in Fig. 4.9. More specifically, Fig. 4.9 reveals that for approximately $M \leq 64$ a notable gain is provided by RZF compared to ZF, while this gain gradually disappears for larger values of M.

The impact of increasing the number of users K on system performance is shown in Fig. 4.10, when $L = 7$ and $M = 256$. Similar to the results of Chapter 3, it is observed that while increasing the number of users degrades the performance of the RS scheme, the gain

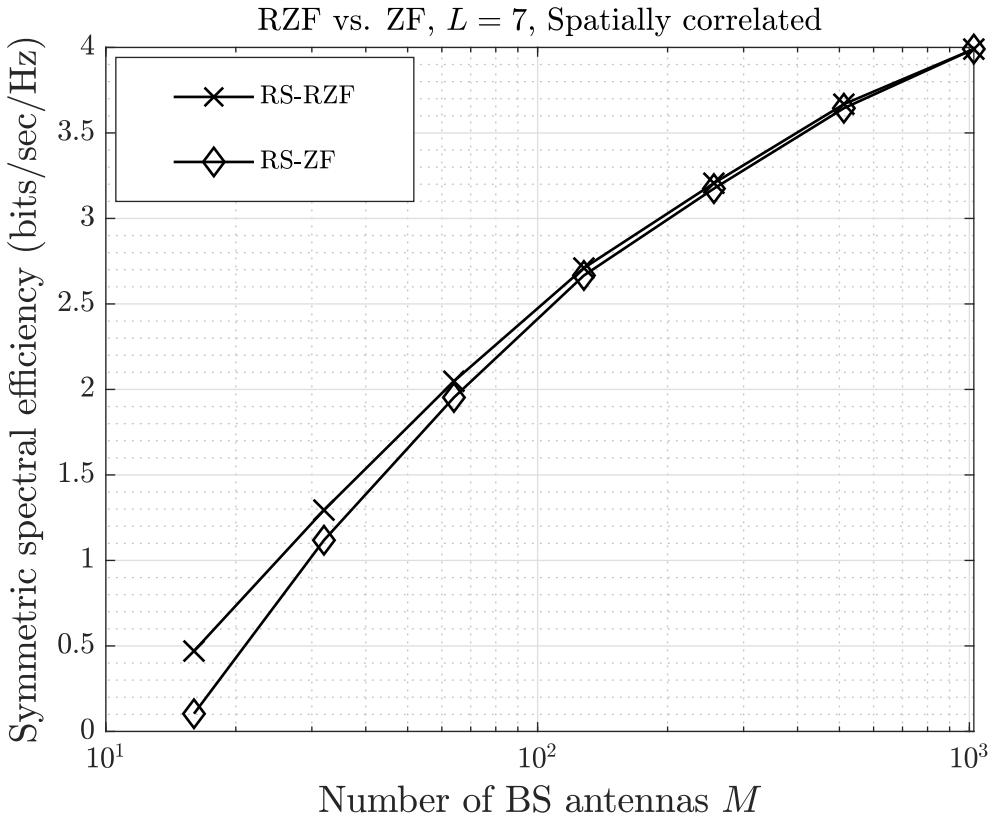


Figure 4.9: Performance comparison between RZF and ZF for $L = 7$, when the RS scheme is used.

provided by the partial interference decoding scheme over TIN improves. In particular, Fig. 4.10 shows that this gain is at least 30% when $K = 2$, and it increases to more than 98% when $K = 15$.

The impact of increasing the correlation magnitude on system performance is shown in Fig. 4.11, when $L = 7$ and $M = 256$. The results presented in this figure are in agreement with the discussions provided in Chapter 3. Specifically, as the spatial channel correlation improves the performance of all schemes, the gain offered by the RS scheme over TIN is at least 138% when $\kappa = 0$ (i.e., uncorrelated channel) and it reduces to about 34% when $\kappa = 0.8$ (i.e., strong spatial correlation), however this gain is still significantly higher than the one provided by SND.

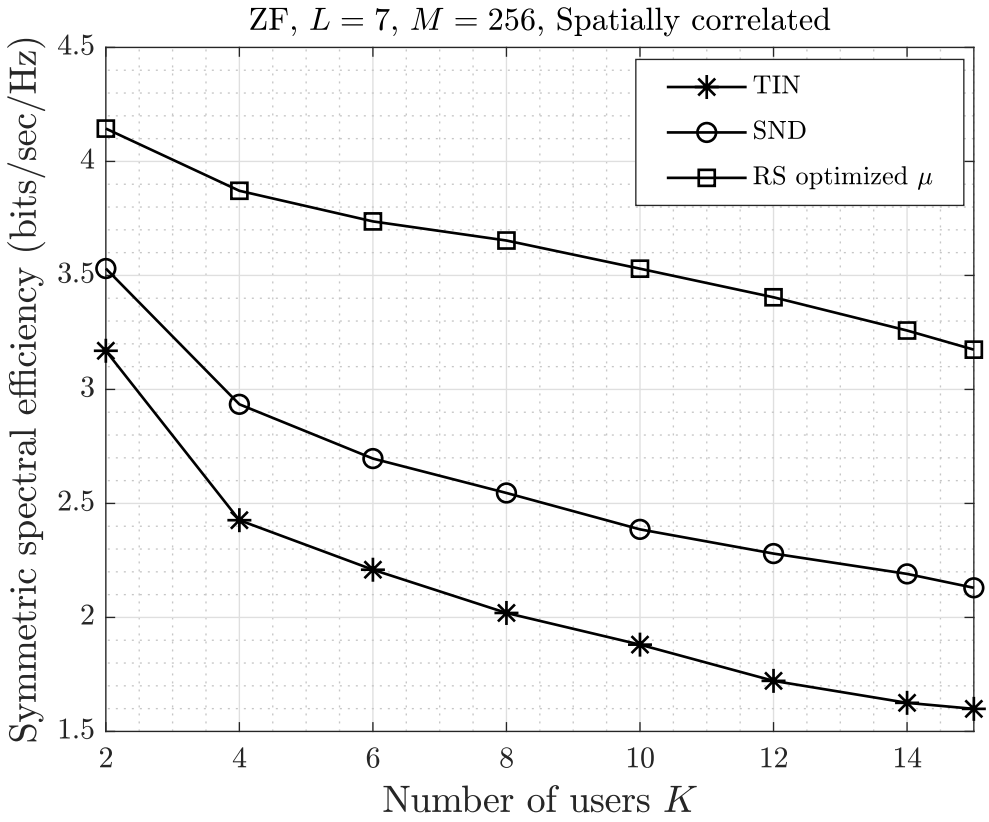


Figure 4.10: Performance of maximum symmetric SE versus the number of users K, where $L = 7$, $M = 256$, and ZF precoding with a spatially correlated channel model are used.

Lastly, we study the impact of shadow fading on the performance of the proposed RS scheme. Similar to the previous chapter, we assume that a term associated with shadow fading is now added to the large-scale fading model with a standard deviation of σ_{shadow} in dB. Fig. 4.12 shows this impact, where the standard deviation of shadow fading σ_{shadow} increases from 0 dB to 5 dB. The parameters for this figure are the same as those in Fig. 4.11 except that correlation magnitude is now fixed at $\kappa = 0.4$. It is observed that, similar to TIN and SND, by increasing the shadow fading the symmetric SE of the RS scheme reduces, yet for all values of σ_{shadow} it offers larger gains over TIN compared to SND. Nevertheless, as σ_{shadow} becomes larger the gain provided by the RS scheme over TIN increases, which shows the importance of the proposed scheme in practical scenarios.

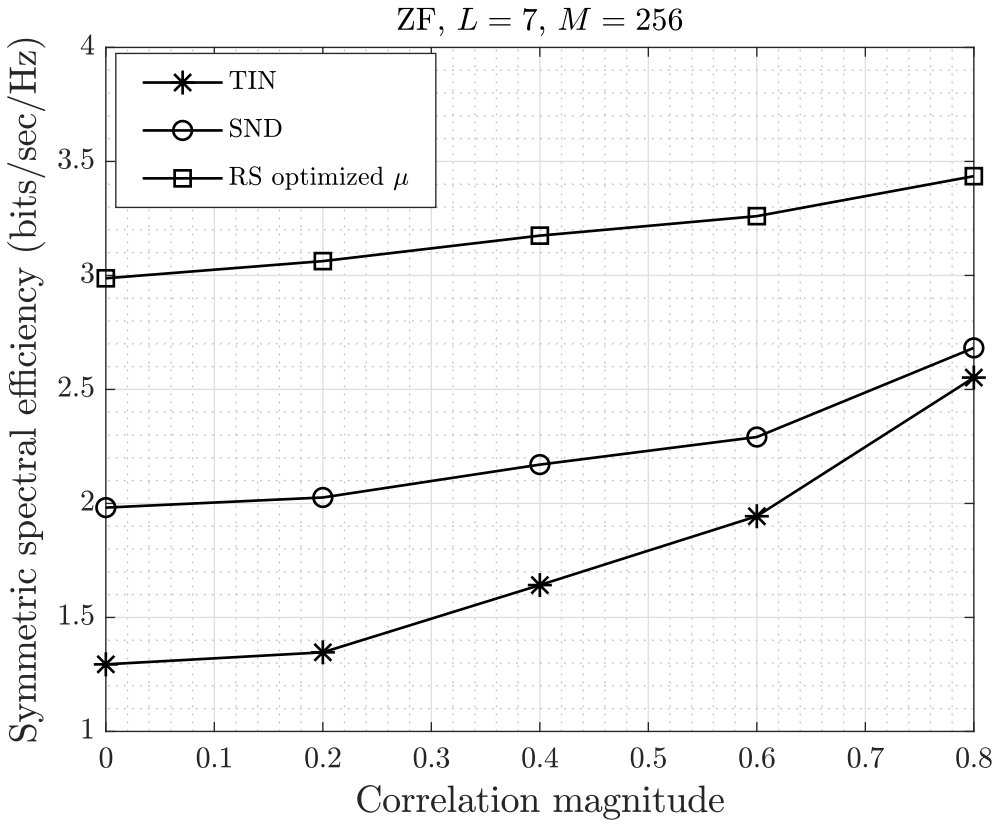


Figure 4.11: Performance of maximum symmetric SE versus the correlation magnitude κ, where $L = 7$, $M = 256$ and ZF precoding is used.

In particular, there is a gain of at least 98% for $\sigma_{\text{shadow}} = 0$ dB, and it improves to more than a factor of 4 when $\sigma_{\text{shadow}} = 3$ dB. This gain continues to improve for larger values of σ_{shadow}.

Next, we assume that a wrap around topology is considered for the seven-cell configuration illustrated in Fig. 3.7d. Similar to the observations made in Chapter 3, it is expected that the SEs achieved by RS will be slightly degraded; nevertheless, since RS can better manage the interference compared to TIN, it should still provide a significant gain over TIN while outperforming SND. The results are shown in Fig. 4.13. It can be seen that the gain offered by RS over TIN is at least 86% and 106% for $M = 128$ and $M = 256$, respectively, and reaches about 136% when $M = 1024$. Hence, the use of RS or SND with

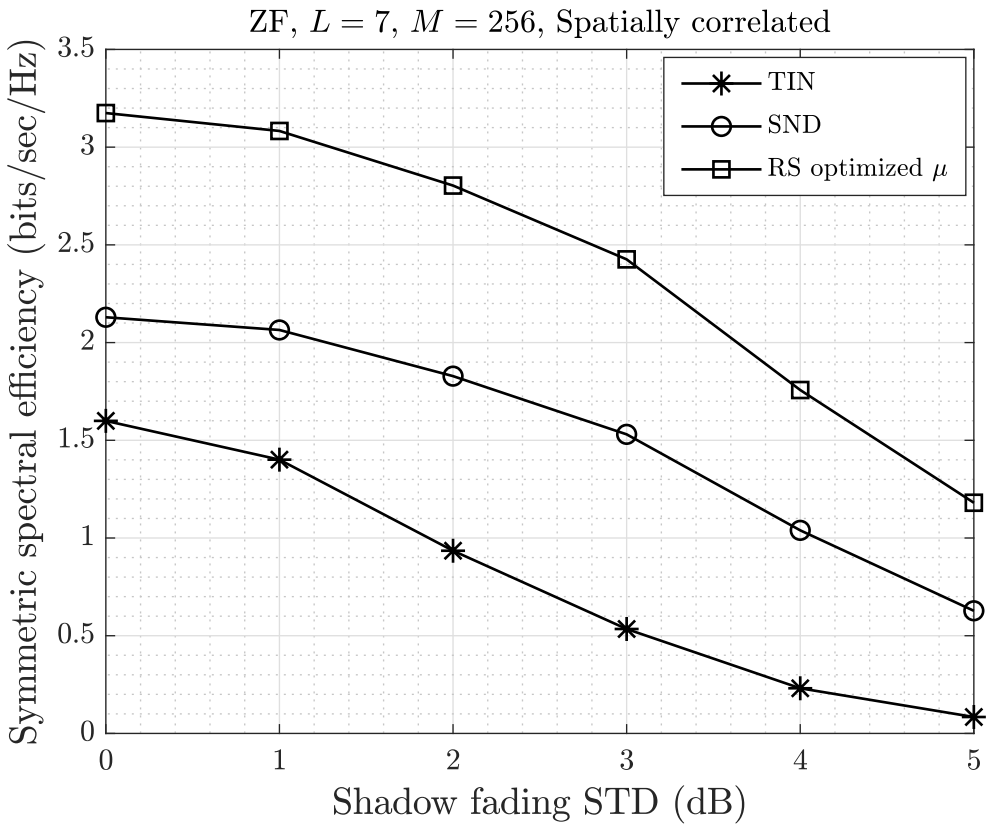


Figure 4.12: Performance of maximum symmetric SE versus the standard deviation of shadow fading σ_{shadow}, where $L = 7$, $M = 256$ and ZF precoding is used.

ZF can be considered a viable solution to combat the rate saturation problem due to pilot contamination in practical implementations.

4.4.2 Spatially Uncorrelated

We now consider the special case of an uncorrelated Rayleigh fading channel, and evaluate the performance of the message splitting techniques by simulating scenarios identical to those of Section 3.4.2. Similar to the previous chapter, we study the impact of changing the number of antennas M, number of cells L and cell radius r on the performance of the proposed RS scheme. To this end, the downlink of a multi-cell massive MIMO system

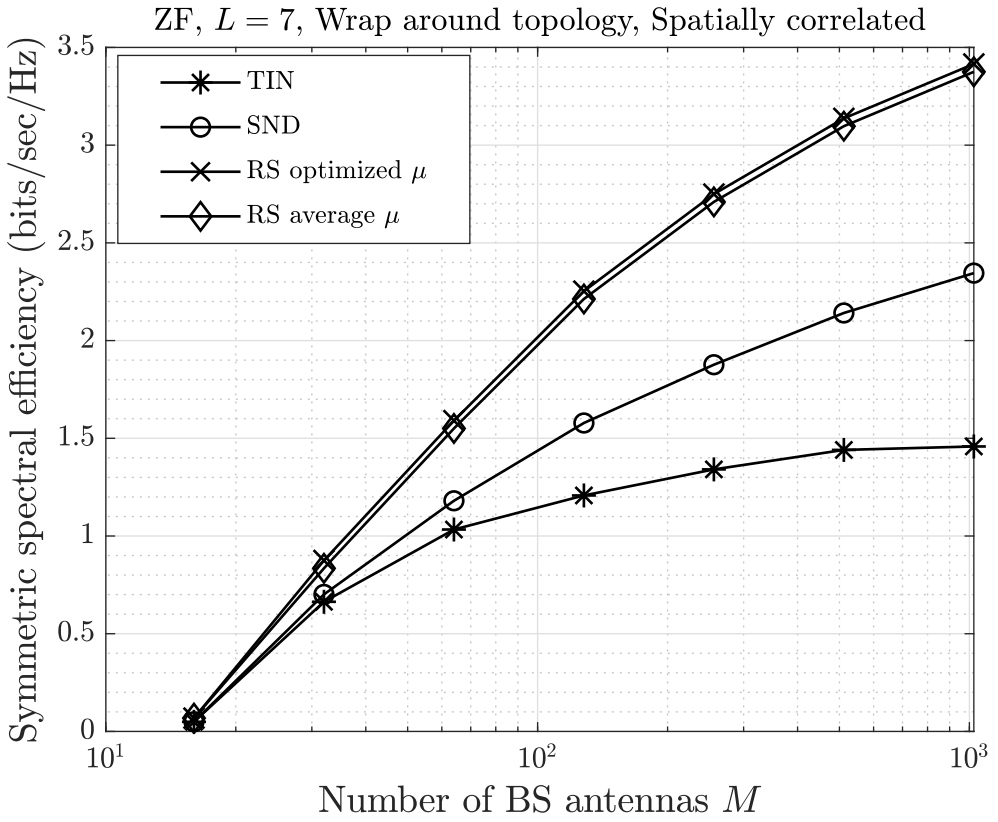
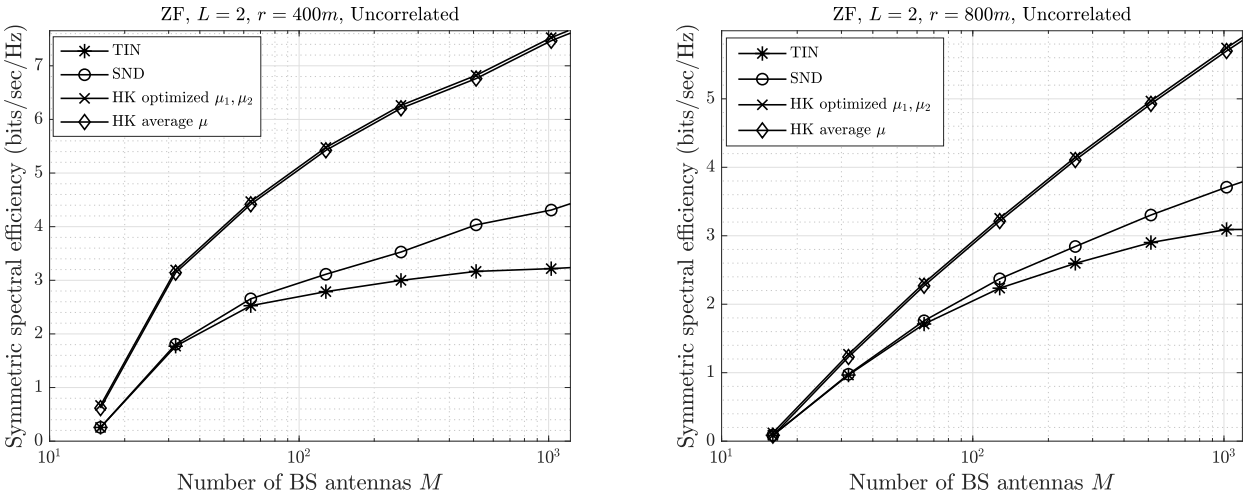


Figure 4.13: Performance of maximum symmetric SE for seven cells with wrap around topology, when ZF precoding and a spatially correlated channel model are used.

with spatial correlation matrices given by $\boldsymbol{R}_{jkl} = \beta_{jkl}\boldsymbol{I}_M$ is simulated with parameters identical to the correlated case. In particular, the performance of the proposed message splitting techniques along with TIN and SND from the previous chapter (for the sake of comparison) are illustrated with ZF precoding and for $L = 2,\ 3,\ 4,\ 7$, with two choices of cell radius: $r = 400$ m, 800 m. To evaluate the performance, the average of the maximum symmetric SEs is calculated over 200 random realizations of user locations. Also, for each scenario, the symmetric SE of the proposed partial decoding scheme is calculated using the two approaches discussed in the previous section, namely, using the optimized value of the splitting coefficient and using the pre-computed average value of the splitting coefficient. Moreover, using the closed-form expressions of the rate lower bounds for an uncorrelated

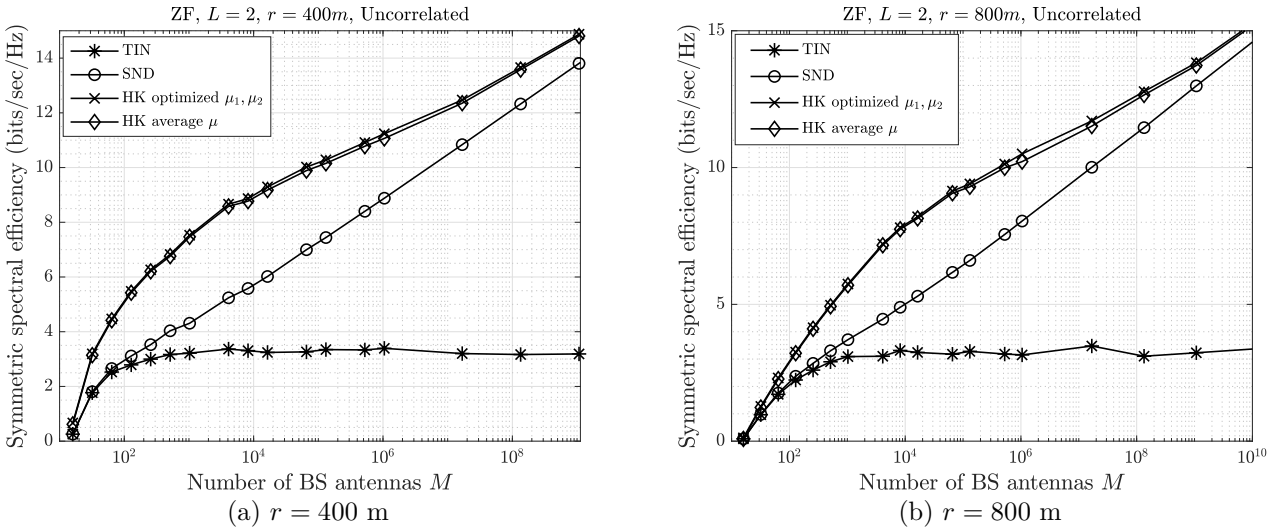


(a) $r = 400$ m (b) $r = 800$ m

Figure 4.14: Performance of maximum symmetric SE for two cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

channel in Chapter 3, we are able to compute the performance for a significantly larger range of M.

Fig. 4.14 shows the performance of the HK scheme for a range of moderately large M, while Fig. 4.15 shows the same for a range of extremely large M, when $L = 2$. While the latter covers a range of M that is beyond practical, the results of Fig. 4.15 can be used to confirm asymptotic performance limits as $M \to \infty$. The first observation to make here is that, similar to the case of a spatially correlated channel, the symmetric SEs obtained using the optimized values of splitting coefficients (i.e., first approach) have almost the same values as the ones obtained using the pre-computed average values of the splitting coefficients (i.e., second approach). Moreover, the results of this figure reveal that the gains provided by the HK scheme over TIN here are larger than those observed in the spatially correlated channel. A similar observation was already confirmed in Fig. 4.11. It can also be seen that, for both choices of the cell radius, this performance gain improves as M is increased. Specifically, when $r = 400$ m, this gain is about 96% and 108% for $M = 128$ and

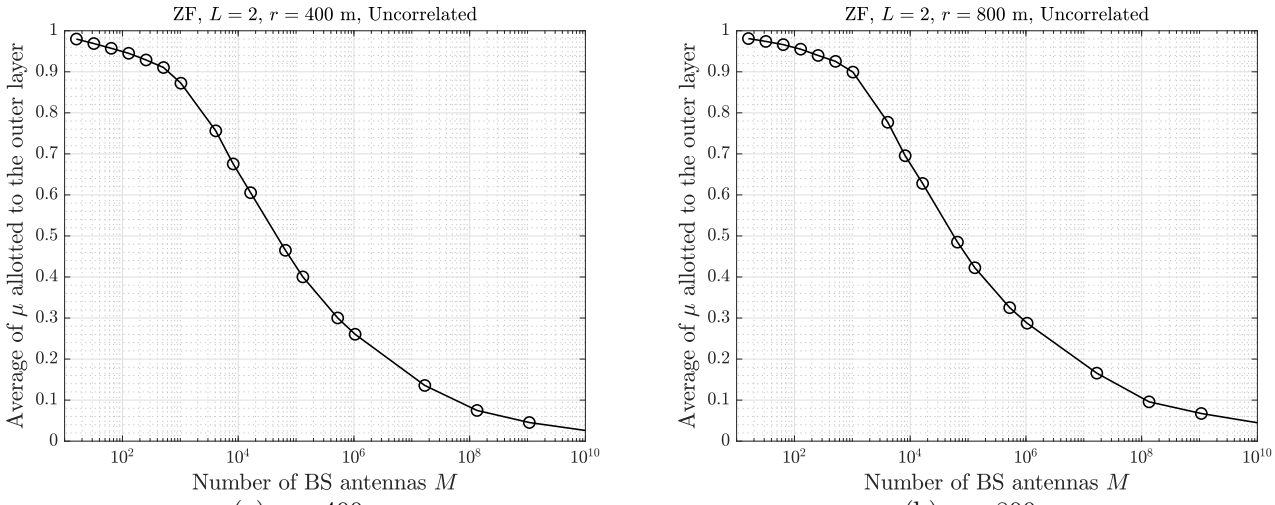


(a) $r = 400$ m (b) $r = 800$ m

Figure 4.15: Performance of maximum symmetric SE for two cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

$M = 256$, respectively, and it increases to about 133% for $M = 1024$. It is also observed that increasing the cell radius reduces the achieved symmetric SEs and results in smaller performance gains; nonetheless, these gains are still larger than those provided by SND. Particularly, when $r = 800$ m, the HK scheme offers gains of about 45% and 60% for $M = 128$ and $M = 256$, respectively, and the gain improves to about 85% for $M = 1024$.

Interestingly, in Fig. 4.15 one can notice that as M increases, the performance gain provided by the HK scheme relative to SND gradually increases up to a point (i.e., approximately somewhere in the range $10^4 < M < 10^5$ in this case), beyond which the gap to SND gradually reduces. This means that as M grows and thus the strength of the interfering signal increases, the power of the inner layer can be increased. This result is better illustrated in Fig. 4.16, which demonstrates the average of the optimized value of splitting coefficients over 200 random realizations of user locations. Specifically, this figure shows that increasing M yields a smaller value of μ_{avg} (i.e., larger value of $1 - \mu_{\text{avg}}$ which is the fraction of power allotted to the inner layer), and as M grows unbounded μ_{avg} approaches zero (i.e., $1 - \mu_{\text{avg}}$ approaches one) and thereby a larger fraction of the interference can



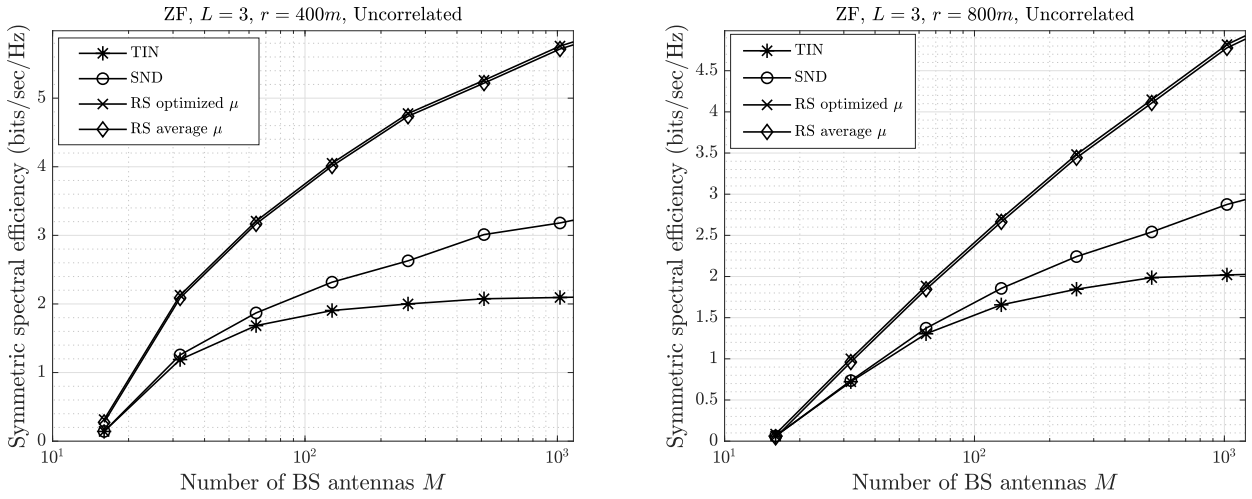
(a) $r = 400$ m (b) $r = 800$ m

Figure 4.16: Average of power splitting coefficients μ_1, μ_2 (i.e., fraction of power allotted to the outer layer) for two cells as a function of M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

be decoded at each receiver. Another observation to make here is that, as expected, when the cell radius increases the curve in Fig. 4.16 shifts slightly up, indicating that on average more antennas will be required to be able to decode the same fraction of the interference signal.

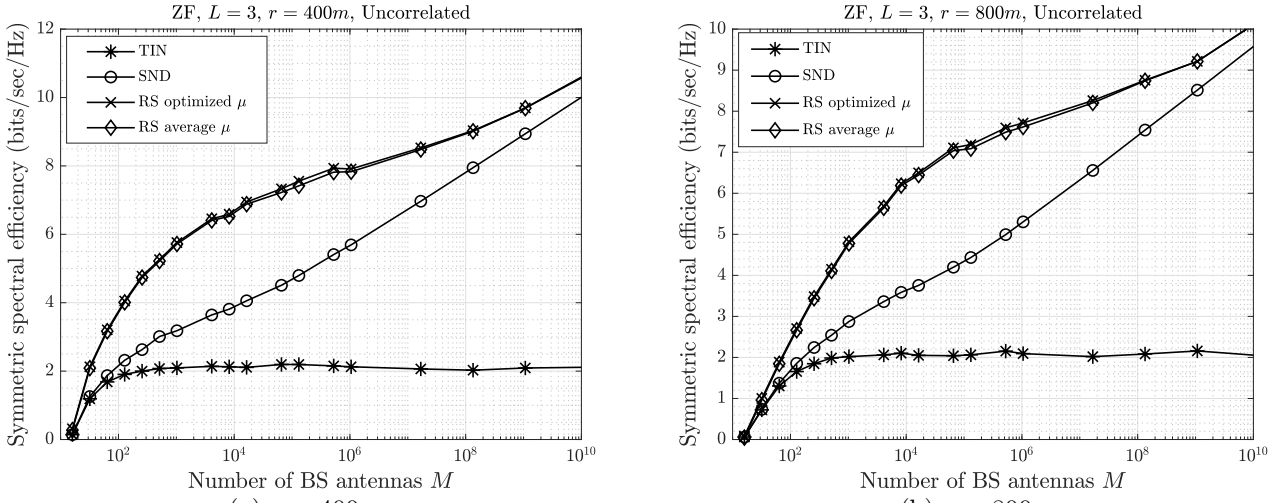
The performance of the proposed RS scheme with ZF precoding is shown in Figs. 4.17 to 4.25, for $L = 3, 4$ and 7. Note that Figs. 4.17, 4.20 and 4.23 illustrate the performance for a range of moderately large values of M, while Figs. 4.18, 4.21 and 4.24 demonstrate the same plots for a range of truly large values of M, and Figs. 4.19, 4.22 and 4.25 show the average of the optimized splitting coefficients as a function of M. In addition, it should be pointed out that the performance of the RS scheme for the cases of $L = 4$ and $L = 7$ is evaluated using the lower bound obtained by the achievable sub-region discussed in Section 4.4.1.

First, notice that patterns similar to those of the two-cell case are observed in all these figures. Also, as expected, it can be seen that increasing the number of cells gives rise to



(a) $r = 400$ m (b) $r = 800$ m

Figure 4.17: Performance of maximum symmetric SE for three cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.



(a) $r = 400$ m (b) $r = 800$ m

Figure 4.18: Performance of maximum symmetric SE for two cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

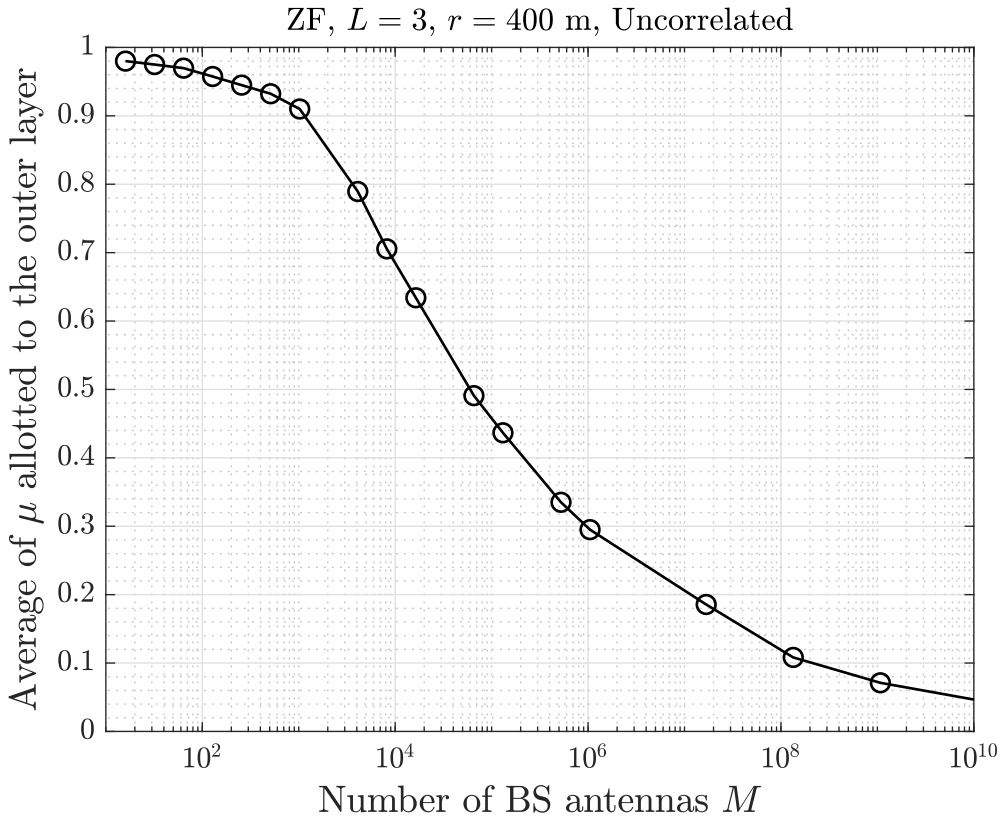
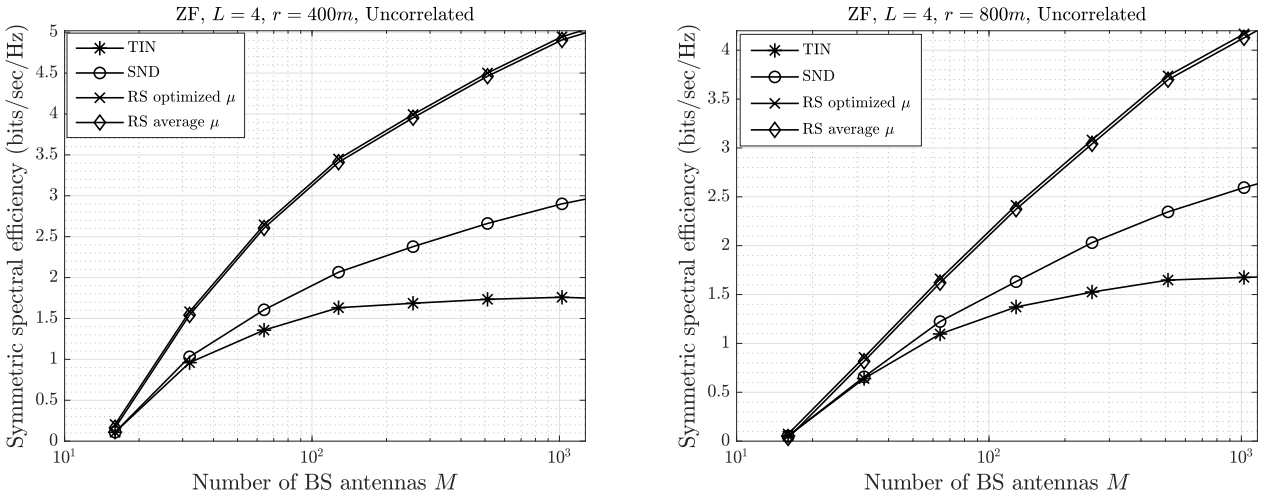


Figure 4.19: Average of power splitting coefficients μ (i.e., fraction of power allotted to the outer layer) for three cells as a function of M, when $r = 400$ m ZF precoding and an uncorrelated channel model are used

reducing the symmetric SEs of the RS scheme; nevertheless it still provides a larger gain over TIN compared to SND. Also, as the number of cells increases the performance gain offered by the RS scheme improves, except for the small drop in gain when moving from $L = 3$ to $L = 4$ due to considering the sub-region rather than the entire region of $\mathscr{R}^{\mathrm{RS}}$.

More specifically, it is observed from Fig. 4.17a that when $L = 3$ and $r = 400$ m, the gain provided by the RS scheme over TIN is about 111% and 137% for $M = 128$ and $M = 256$, respectively, and it improves to about 173% for $M = 1024$. Also, Fig. 4.17b shows that when the cell radius increases to $r = 800$ m, this gain reduces to about 63% and 88% for $M = 128$ and $M = 256$, respectively, and it increases to about 138% for $M = 1024$.

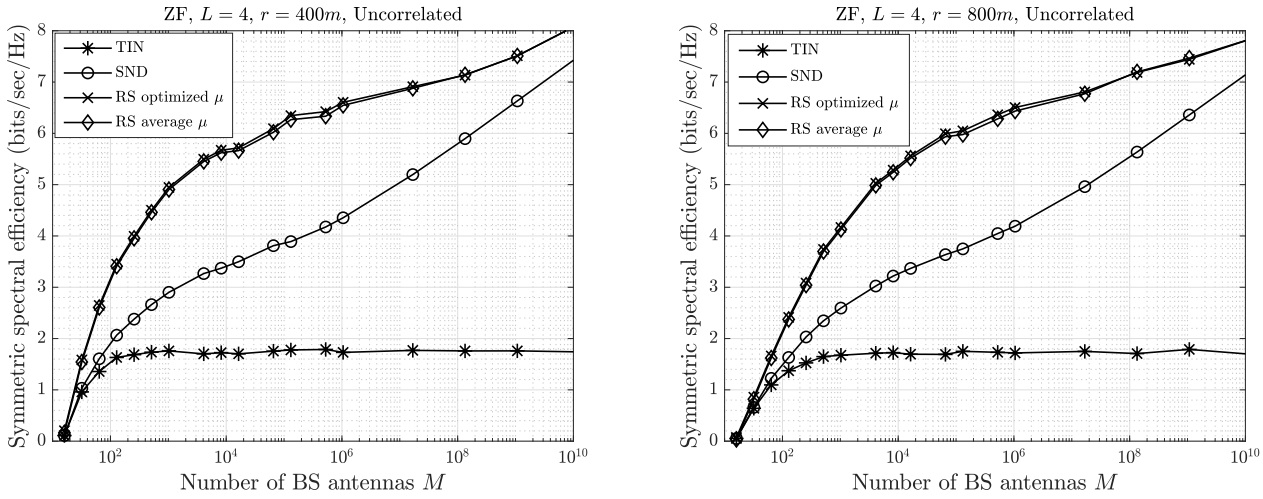


(a) $r = 400$ m (b) $r = 800$ m

Figure 4.20: Performance of maximum symmetric SE for four cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

Further note that by increasing the number of cells to $L = 7$ in Fig. 4.23a, when $r = 400$ m the gain of the RS scheme over TIN improves to at least 112% and 144% for $M = 128$ and $M = 256$, respectively, and it increases to more than 184% for $M = 1024$. One can also observe, as expected, the slight degradation of the performance gains in Fig. 4.23b when $r = 800$ m. More precisely, Fig. 4.23b shows a gain of at least 87% and 116% for $M = 128$ and $M = 256$, respectively, whereas this gain is more than 164% for $M = 1024$.

Lastly, Fig. 4.26 shows a scenario identical to the one depicted in Fig. 3.27 which is somewhat pessimistic and also in favor of TIN, and the results are similar to those of Fig. 3.29. In particular, it is assumed that $L = 2$ and all users of the left cell are located on the cell edge at the farthest distance from the BSs, whereas the position of users in the right cell is on the boundary of an inscribed circle inside the cell, determined by $0° \leq \theta \leq 360°$. The achieved maximum symmetric SEs with ZF as a function of θ are shown in Fig. 4.26, for four different values of $M = 128,\ 256,\ 512,\ 1024$, where the performance of TIN and SND from the previous chapter are plotted again for the sake of comparison.



(a) $r = 400$ m (b) $r = 800$ m

Figure 4.21: Performance of maximum symmetric SE for four cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

It can be observed that the symmetric SEs achieved by the RS scheme follow a pattern similar to those of TIN and SND. Specifically, Fig. 4.26 shows that, for a fixed value of M, as θ increases and approaches 180°, users of the right cell become closer to BS 1 and the interference thus becomes stronger (see the performance of TIN). Therefore, the achieved symmetric SEs of the RS scheme reduce, nevertheless its performance gain over TIN improves due to the severe degradation of the performance obtained by TIN. Finally, one can notice that the RS scheme outperforms both TIN and SND for a significantly wider range of θ, as it benefits from the additional flexibility of decoding part of the interference. In other words, it is only for a small range where θ is close to either 0° or 360° (i.e., users of the right cell are located at the farthest distance from BS 1), that the RS scheme achieves a performance identical to those of TIN and SND. It is worth noting that, for a fixed value of M, TIN offers its maximum possible performance in this range of θ.

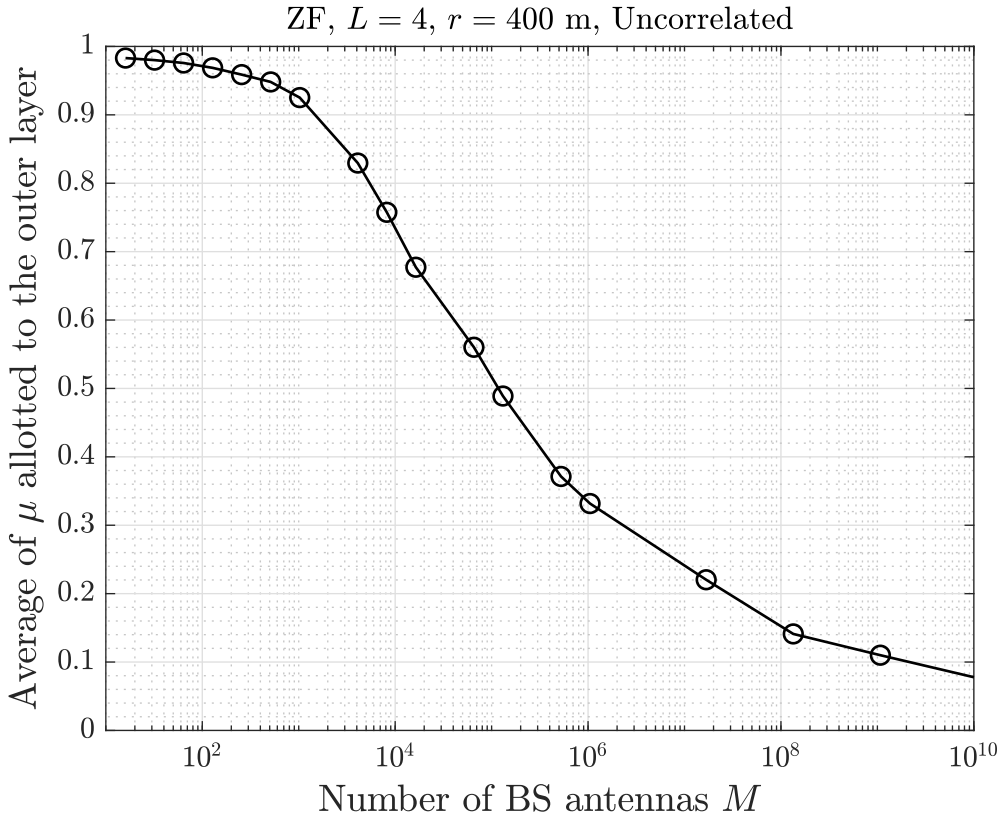
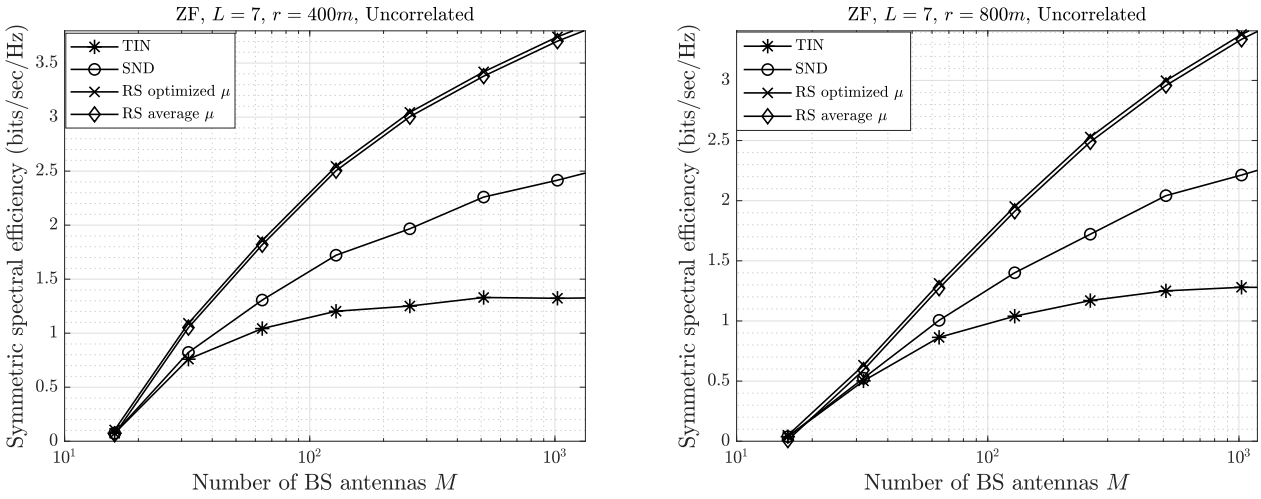
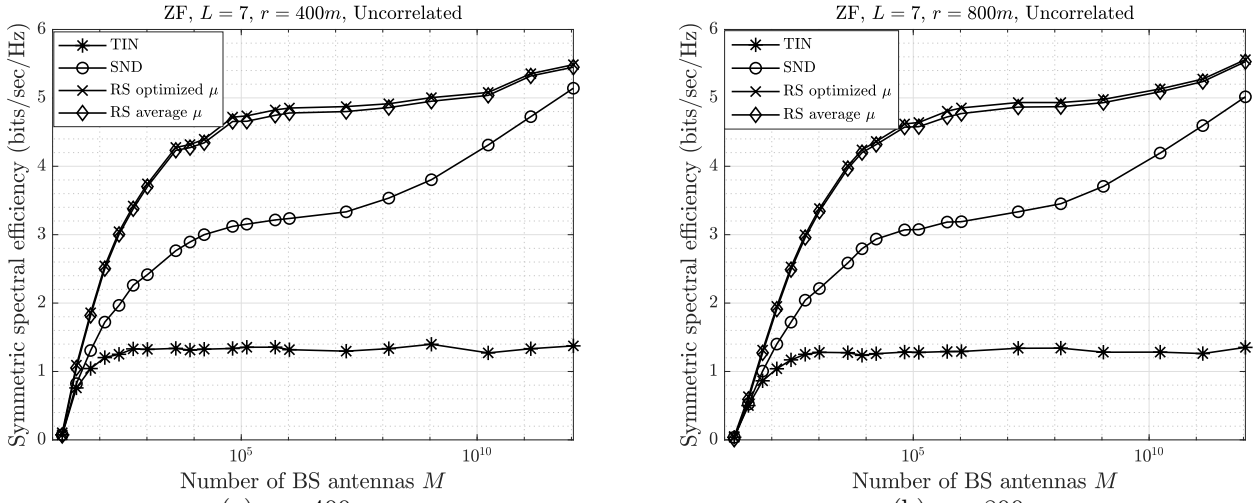


Figure 4.22: Average of power splitting coefficients μ (i.e., fraction of power allotted to the outer layer) for four cells as a function of M, when $r = 400$ m ZF precoding and an uncorrelated channel model are used



(a) $r = 400$ m (b) $r = 800$ m

Figure 4.23: Performance of maximum symmetric SE for seven cells with moderately large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.



(a) $r = 400$ m (b) $r = 800$ m

Figure 4.24: Performance of maximum symmetric SE for seven cells with truly large M, when ZF precoding and an uncorrelated channel model are used, (a) $r = 400$ m, (b) $r = 800$ m.

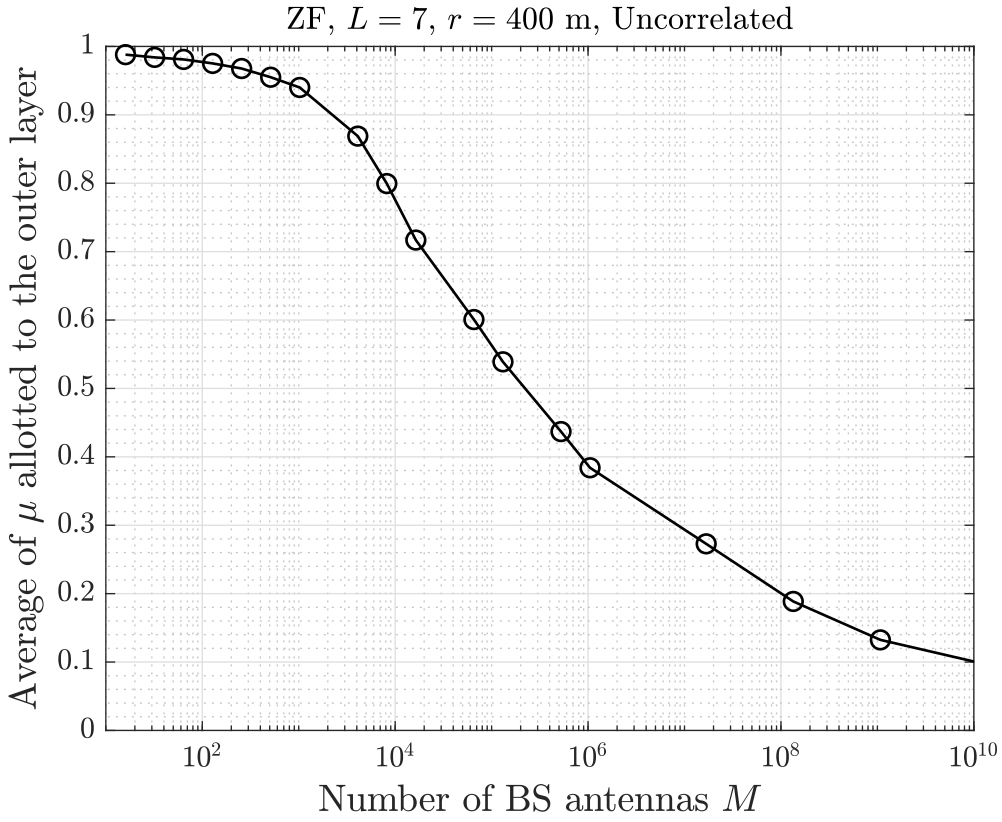
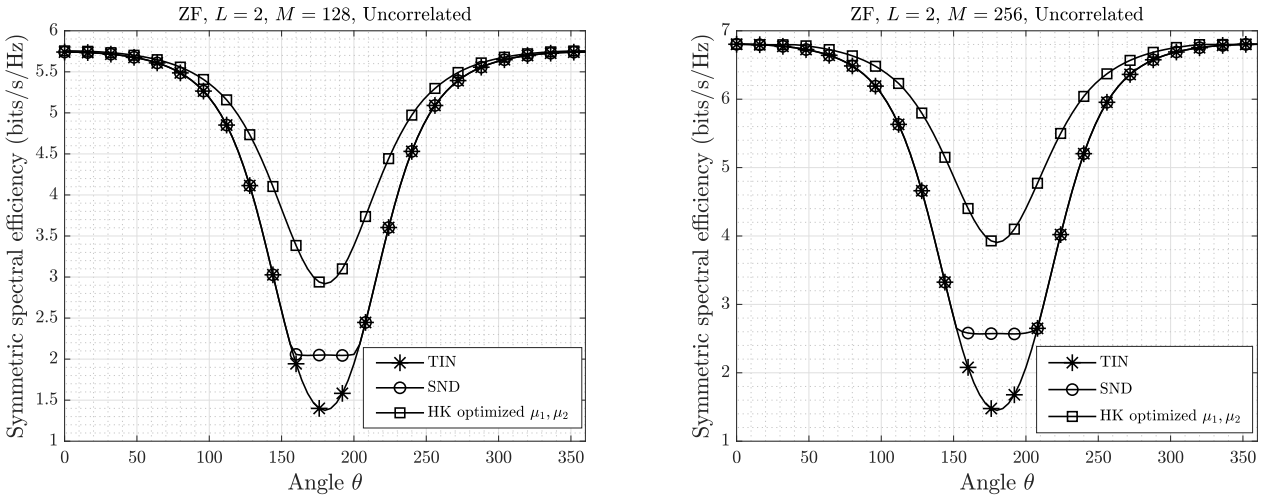
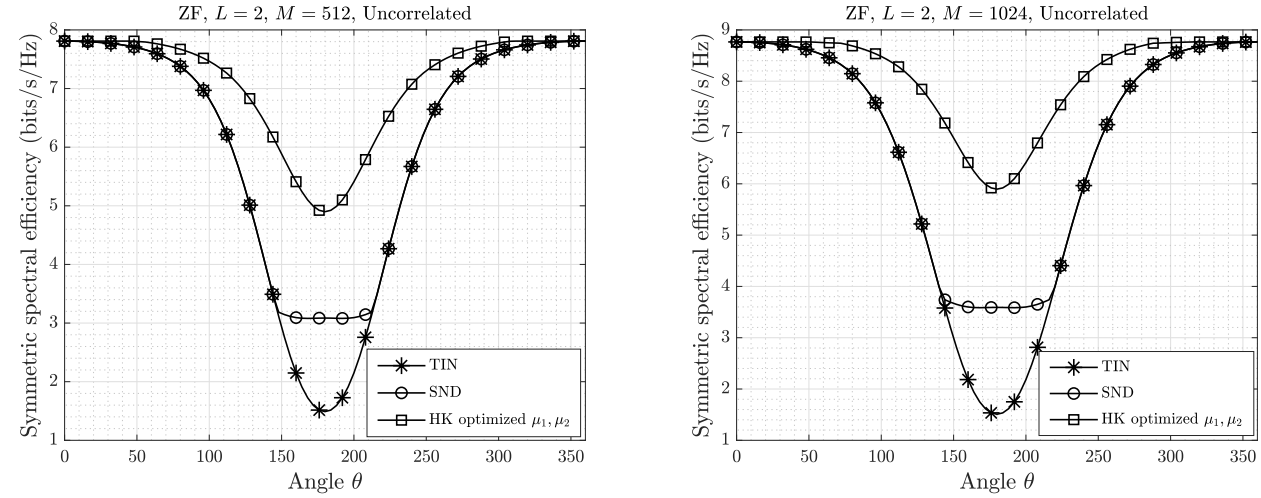


Figure 4.25: Average of power splitting coefficients μ (i.e., fraction of power allotted to the outer layer) for seven cells as a function of M, when $r = 400$ m ZF precoding and an uncorrelated channel model are used



(a) $M = 128$

(b) $M = 256$



(c) $M = 512$

(d) $M = 1024$

Figure 4.26: Performance of TIN/SND/HK in a two-cell system with maximum symmetric SE versus θ, when ZF precoding and an uncorrelated channel model are used, (a) $M = 128$, (b) $M = 256$, (c) $M = 512$, (d) $M = 1024$.

Chapter 5

Concluding Remarks

5.1 Summary of Contributions and Conclusions

The performance gains offered by massive MIMO systems rely heavily on the availability of accurate CSI, which needs to be estimated. A popular technique to acquire CSI in TDD-based massive MIMO systems that benefit from channel reciprocity is to transmit uplink training pilots, which will be used at the BSs to estimate the channels. In a multi-cell system, however, since the length of the channel coherence interval is small and finite due to users mobility, the number of available orthogonal pilot sequences is limited. Therefore, a natural approach to cope with this limitation is to re-use the orthogonal pilots (used in one cell to obtain an estimation of the users' channel in that cell) in other cells across the network. This causes pilot contamination interference, also known as the coherent interference, that scales at the same rate as the desired signal, and therefore does not vanish asymptotically when the number of BS antennas, M, grows to infinity. Thus, by treating this interference term as noise (TIN) which is a common technique used in the massive MIMO literature, one obtains achievable rate expressions that, in the asymptotic limit of $M \to \infty$, converge to a constant independent of M. Hence, as a result of performing TIN at the BSs in uplink or at the users' side in downlink, the benefits of increasing the number of BS antennas saturate, which constitutes a bottleneck in practical implementations of

massive MIMO systems.

In this thesis, we have addressed pilot contamination problem (due to the re-use of orthogonal pilots in different cells) for a multi-user multi-cell massive MIMO system. We analytically studied both uplink and downlink with two well-known linear combining and precoding techniques, i.e., MRC/MRT and ZF. The case of RZF was also numerically studied. We showed that when the number of BS antennas increases, pilot contamination interference is not detrimental to the system performance, provided that it is carefully decoded jointly along with the desired signal. As a result, we proposed to jointly decode the main signal along with pilot contamination interference terms (either fully or partially) as opposed to performing TIN, and showed that by doing so the rate saturation phenomena is resolved and one can therefore achieve unbounded rates as $M \to \infty$.

The major contributions of the thesis are as follows:

In Chapter 3, for the uplink and downlink of a multi-cell massive MIMO system with spatially correlated Rayleigh fading as well as uncorrelated Rayleigh fading channel models, we studied the performance of interference decoding schemes based on SD and SND. Specifically, rather than treating pilot contamination interference as noise, we proposed to jointly decode it (either *uniquely* as in SD or *non-uniquely* as in SND) along with the useful signal, resulting in achievable rate expressions that grow with M.

The major findings from this study are as follows:

1. We intuitively showed that after performing linear processing techniques and letting $M \to \infty$, the resulting signal can be treated as a noise-free L-user MAC where the capacity region grows unbounded. It was also shown that when decoding the interference caused by pilot contamination, the re-use of orthogonal pilots across cells (as opposed to using different pilots in different cells) is preferable, as it results in decoding significantly fewer terms.

2. We extended the previously established result on the worst-case uncorrelated noise of a point-to-point channel to the multi-user case. Using this technique, we derived general expressions of achievable rate lower bounds for both uplink and downlink of

a multi-cell massive MIMO system that applies joint decoding to each set of pilot-sharing users. These lower bounds are independent of the choice of linear processing technique, and can thus be used in conjunction with Monte Carlo simulations to evaluate the system performance under various types of combining/precoding schemes.

3. Using the derivations of MMSE channel estimate in the case of a spatially correlated Rayleigh fading channel, we specialized the achievable lower bounds to the case of MRC/MRT in uplink/downlink, and obtained new closed-form expressions for these linear processing techniques that scale as $\mathcal{O}(\log M)$. Moreover, assuming that an uncorrelated Rayleigh fading channel model is used (as a special case of the correlated Rayleigh fading model), we simplified these rate lower bounds, and further obtained new closed-form expressions for ZF in both uplink and downlink, which again scale as $\mathcal{O}(\log M)$.

4. As a measure of fairness, we studied the performance of maximum symmetric rate allocation and compared the performance of the two proposed full interference decoding schemes, SD/SND, with that of TIN based on this fairness criteria. We also introduced a simplified achievable sub-region of SND, known as S-SND, and showed that it strictly contains SD and also provides a lower bound to the performance of SND. First, we considered the extreme regime of high SINR (i.e., when M is truly large) and observed that both interference decoding schemes, SD/SND, have identical performance in this regime and also outperform TIN. Also, it was shown that in the high SINR regime pilot contamination interference is strong enough that the optimum performance is obtained by uniquely decoding all signals of pilot-sharing users together.

5. We obtained structural results for a symmetric two-cell system, where the MACs seen in both cells are identical. For this symmetric setting, it was found that when M is small and pilot contamination interference is thus *weak*, SND and TIN have identical performance while outperforming SD. On the other hand, when M is sufficiently large and pilot contamination interference is therefore *strong*, SND and SD have the same performance while outperforming TIN. We also found conditions in terms of

mutual information expressions under which these results are valid.

6. We provided extensive numerical results by simulating the downlink of a multi-cell massive MIMO system with correlated/uncorrelated Rayleigh fading channel models in different scenarios. The first finding was that in all cases, unless M is truly large and beyond practical limit, SD provides poor performance and is outperformed by the other schemes. We also observed that, for a practical range of M (e.g., $M \geq 128$) SND offers a notable gain over TIN, and this gain improves by either increasing M or increasing the number of cells. It was further shown that both RZF and ZF produce significantly larger rates compared to MRC/MRT, due to the better mitigation of multi-user interference, while RZF also outperforms ZF for small values of M (e.g., $M < 64$) only.

7. In addition, we numerically studied the impact of changing the number of users, the cell radius, the correlation magnitude and the standard deviation of shadow fading on system performance. It was observed that while increasing the cell radius (leading to reduced received power) or the number of users (leading to increased interference) results in reducing the achievable rates of all schemes, it increases the performance gain provided by SND over TIN. A similar observation was also made when increasing the standard deviation of the shadow fading, i.e., increasing the shadow fading degrades the performance of all schemes (as expected) while improving the gain of SND over TIN. Lastly, we observed that spatial correlation improves the performance of all schemes, resulting in reducing the gap between SND and TIN. Nevertheless, it was seen that under moderate spatial correlation and with a reasonable number of antennas SND still provides a significant gain over TIN. This means that the gains offered by SND in the case of an uncorrelated Rayleigh fading channel are larger compared to the correlated case.

In Chapter 4, we studied the performance of partial interference decoding schemes based on RS and *non-unique* decoding in the downlink of a multi-cell massive MIMO system. Specifically, in this chapter, we proposed to *non-uniquely* decode part of pilot contamination interference, while treating the remaining part as noise. We showed that

these additional decoding flexibilities lead to achieving significantly larger rates compared to the schemes of Chapter 3, but at the cost of higher complexity.

We obtain the following findings from this study:

1. In the case of a two-cell system, we investigated the performance of an RS technique based on the well-known HK scheme (for an IC) and non-unique decoding. In particular, the message of both users (i.e., the intended user and pilot contamination interference) were partitioned into two independent layers, known as the "inner layer" and the "outer layer", where at the receiver's side both parts of the intended message were decoded along with non-uniquely decoding only the inner layer of the interfering signal. The rates of the individual layers at each cell are also adjusted based on two power splitting coefficients that are numerically optimized.

2. This partial interference decoding scheme was also carefully generalized to the case of more than two cells, while using only one power splitting coefficient per IC. Specifically, unlike the two-cell case, it was proposed to non-uniquely decode each layer of the interfering signals along with uniquely decoding both parts of the intended message at each receiver. By doing so, it was shown that even though all pilot-sharing users are applying the same numerically-optimized power splitting coefficient, additional decoding flexibilities (compared to SND) will be provided at each receiver; hence outperforming the schemes of Chapter 3.

3. To evaluate the performance of the proposed RS schemes, we numerically studied the performance of maximum symmetric rate allocation for both cases of spatially correlated Rayleigh fading and uncorrelated Rayleigh fading, when the cellular network comprises two, three, four or seven cells. We showed that the maximum symmetric rate allocation problem can be equivalently formulated as an LP that can be efficiently solved. While for the case of two and three cells the true performance of the RS scheme was investigated, in the case of four and seven cells an achievable sub-region of the proposed RS scheme was considered. In all cases, we observed that by numerically optimizing the power splitting coefficients, the proposed RS schemes produce significantly larger rates compared to SND and TIN, and the performance

improves by increasing the number of antennas M. Furthermore, it was observed that increasing the number of cells results in improving the gain provided by the RS scheme such that the maximum gains in all scenarios were obtained in the case of seven cells; thus showing the importance of the proposed scheme in practical implementations.

4. Similar to Chapter 3, the impacts of increasing the number of users, the cell radius, the correlation magnitude and the shadow fading on the achieved maximum symmetric rate of the proposed RS scheme were also investigated. It was observed that increasing the number of users, the cell radius and the shadow fading degrade the performance of the RS scheme, while improving the gain it can offer. Besides, increasing the correlation magnitude leads to improving the performance of the proposed RS scheme (similar to the results of Chapter 3), thereby reducing the gain provided by the RS scheme. Nonetheless, we observed that with a moderate spatial correlation and a practical value of M (e.g., $M = 128$), the RS scheme still provides a significant gain over the schemes of Chapter 3.

5. Lastly, we showed that by replacing the numerically-optimized values of the power splitting coefficients with their mean value over a large number of random realizations of users' locations, the performance loss is quite negligible; thus reducing the complexity of the optimization problem in practical settings.

5.2 Directions for Future Work

In the following, we mention a few noteworthy future extensions of the current work. While some of these are only consolidations or extensions of the present work, new approaches/techniques are possibly needed for the other problems.

5.2.1 Other measures of fairness

In this work, we compared the performance of the different interference management schemes, TIN/SD/SND/RS with the maximum symmetric rate they can offer. However, other measures of fairness among users such as proportional fairness (i.e., maximizing the geometric mean of the users' rate) can also be considered and is expected to provide larger rates (compared to the maximum symmetric rate allocation) to users with stronger channels.

5.2.2 Cell-free massive MIMO systems

Cell-free massive MIMO refers to a special network configuration, where the BS antennas (or access points) are distributed over a large area, rather than being co-located in a compact array, and users in the network can potentially be served by all BSs [161]. It is thus clear that the large-scale fading coefficients in a cell-free massive MIMO system depend on both the BS antenna index as well as the user index. A number of papers have studied the performance of cell-free massive MIMO under the assumption that perfect CSI is available everywhere [162–166], though massive MIMO gains rely heavily on the accurate estimation of CSI. Moreover, the work of [167] has investigated the performance of MRC in uplink assuming that channel estimates are obtained using orthogonal pilots across the entire network, thus eliminating pilot contamination effect. Recently, cell-free massive MIMO systems with imperfect CSI have been also studied in the literature, where techniques such as LSFD [40] have been proposed to cope with pilot contamination problem. One possible direction for future work could be the investigation of the proposed interference decoding schemes of this thesis (as opposed to TIN) in cell-free configurations to improve the system performance. While studying the performance of the interference decoding schemes, it is expected that user scheduling (i.e., which users should be served by which BSs) and proper pilot assignment among users can also be challenges that need to be addressed [101, 168].

5.2.3 Multi-cell systems with a central unit connected to all BSs via fronthaul links

In a cloud radio access network (C-RAN), it is assumed that several remote radio heads (RRH) (or BSs) are connected to a central unit (CU) to perform digital baseband processing tasks [169]. A large amount of research has been reported in the literature addressing the viability of massive MIMO implementations with C-RANs [170–172]. Hence, another direction for future work would be to consider a C-RAN with imperfect CSI (thus experiencing pilot contamination) consisting of distributed RRHs that have a massive number of antenna arrays. In turn, when sending all the BSs' signals to the CU, it has access to the signals of both strong and possibly weak users (pilot contamination interference terms). Thus, by performing an additional layer of processing based on decoding pilot contamination interference terms at the CU, one can potentially achieve an improved throughput while requiring possibly smaller number of antennas at each RRH compared to regular massive MIMO systems. In the simplest model, we can assume that RRHs are connected to the CU via infinite-capacity fronthaul links and study the performance of interference decoding schemes at the CU. In the next step, one can consider finite-capacity fronthaul links, thus the need for proper signal quantization at RRHs while exploring the benefits of interference decoding schemes.

5.2.4 Multi-antenna users

We have observed that the performance gains of massive MIMO systems performing full/partial interference decoding schemes scale with the number of BS antennas. However, nothing in our analysis precludes the use of multi-antenna users. In particular, a multi-antenna user can potentially benefit from a throughput increase that is proportional to the number of antennas it possesses. Furthermore, multi-antenna users are capable of performing more advanced interference suppression techniques. In the simplest case, one could model each multi-antenna terminal as a multiplicity of single-antenna terminals. Investigating the benefits of multi-antenna terminals in massive MIMO systems is still under an active area of research, in both regular multi-cell configurations [173,174] as well as the

cell-free deployments [175, 176]. Therefore, another interesting direction for future work is the study of the performance of pilot contamination interference decoding schemes in massive MIMO systems, where each BS serves several multi-antenna users.

References

[1] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *Int. Conf. Commun. (ICC)*, pp. 1–5, IEEE, 2010.

[2] "3GPP; technical specification group radio access network; study on channel model for frequencies from 0.5 to 100 GHz (release 16)," *[Online]. Available: https://portal.3gpp.org/*, 2020.

[3] Cisco, "Cisco annual internet report (2018–2023) white paper," tech. rep., [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, 2020.

[4] A. Hoglund, D. P. Van, T. Tirronen, O. Liberg, Y. Sui, and E. A. Yavuz, "3GPP release 15 early data transmission," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 90–96, 2018.

[5] *[Online]. Available: https://www.3gpp.org/release-16*, 2020.

[6] *[Online]. Available: https://www.itu.int/md/R15-SG05-C-0040/en*, 2020.

[7] Ericsson, "Ericsson mobility report june 2020," tech. rep., [Online]. Available: https://www.ericsson.com/en/mobility-report/reports/june-2020, 2020.

[8] I. Vision, "Framework and overall objectives of the future development of IMT for 2020 and beyond," tech. rep., ITU-R Std. M.2083-0, 2015.

[9] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[10] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 119–125, 2018.

[11] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, *et al.*, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, 2017.

[12] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, Č. Stefanović, *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE access*, vol. 6, pp. 28969–28992, 2018.

[13] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality-what is next? five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[14] M. Amine, A. Walid, A. Kobbane, and J. Ben-Othman, "New user association scheme based on multi-objective optimization for 5G ultra-dense multi-RAT HetNets," in *IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2018.

[15] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, 2019.

[16] D. Gesbert, M. Kountouris, R. W. Heath Jr, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE signal processing magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[17] S. Anderson, M. Millnert, M. Viberg, and B. Wahlberg, "An adaptive array for mobile communication systems," *IEEE transactions on Vehicular technology*, vol. 40, no. 1, pp. 230–236, 1991.

[18] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 1, pp. 56–67, 1990.

[19] J. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Transactions on Communications*, vol. 35, no. 11, pp. 1222–1230, 1987.

[20] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array system for spatially selective transmission," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 651–660, 1995.

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

[22] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on selected areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[23] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[24] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE transactions on information theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[25] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 361–374, 2006.

[26] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[27] 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, 2014.

[28] 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, 2013.

[29] E. Björnson, E. G. Larsson, and M. Debbah, "Optimizing multi-cell massive MIMO for spectral efficiency: How many users should be scheduled?," in *IEEE Global Conf. Signal and Inf. Process. (GlobalSIP)*, pp. 612–616, IEEE, 2014.

[30] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Communication, Control, and Computing (Allerton), 49th Annual Allerton Conference on*, pp. 1272–1279, IEEE, 2011.

[31] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE signal processing magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[32] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[33] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving massive MIMO spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[34] C. Guthy, W. Utschick, and M. L. Honig, "Large system analysis of sum capacity in the Gaussian MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 149–159, 2013.

[35] E. Bjornson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Telecommunications (ICT), 20th International Conference on*, pp. 1–5, IEEE, 2013.

[36] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, 2013.

[37] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, 2016.

[38] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, "Uplink interference reduction in large-scale antenna systems," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2194–2206, 2017.

[39] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.

[40] . Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive MIMO with Rician fading and phase shifts," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5299–5315, 2019.

[41] Ö. T. Demir and E. Björnson, "Large-scale fading precoding for maximizing the product of SINRs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5150–5154, IEEE, 2020.

[42] Z. Gülgün, E. Björnson, and E. G. Larsson, "Is massive MIMO robust against distributed jammers?," *IEEE Transactions on Communications*, 2020.

[43] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, 2015.

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

[45] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014.

[46] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Öwall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *Globecom Workshops (GC Wkshps)*, pp. 287–293, IEEE, 2014.

[47] Ericsson, "Ericsson first to deliver 5G NR radio," tech. rep., [Online]. Available: https://www.ericsson.com/en/press-releases/2016/8/ericsson-firstto-deliver-5g-nr-radio, 2016.

[48] E. Björnson, "A look at an LTE-TDD massive MIMO product," tech. rep., [Online]. Available: http://ma-mimo.ellintech.se/2018/08/27/a-look-at-an-lte-tdd-massive-mimo-product/, 2019.

[49] Huawei, "Huawei launches 5G simplified solution," tech. rep., [Online]. Available: https://www.huawei.com/en/press-events/news/2019/2/huawei-5g-simplified-solution, 2019.

[50] S. K. Mohammed, "Impact of transceiver power consumption on the energy efficiency of zero-forcing detector in massive MIMO systems," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3874–3890, 2014.

[51] S. Mumtaz, J. Rodriguez, and L. Dai, *MmWave massive MIMO: a paradigm for 5G*. Academic Press, 2016.

[52] T. E. Bogale and L. B. Le, "Massive MIMO and mmwave for 5G wireless HetNet: Potential benefits and challenges," *IEEE Vehicular Technology Magazine*, vol. 11, no. 1, pp. 64–75, 2016.

[53] N. Tawa, T. Kuwabara, Y. Maruta, M. Tanio, and T. Kaneko, "28 GHz downlink multi-user MIMO experimental verification using 360 element digital AAS for 5G massive MIMO," in *48th European Microwave Conference (EuMC)*, pp. 934–937, IEEE, 2018.

[54] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3205–3219, 2019.

[55] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE transactions on information theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[56] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," in *Information Sciences and Systems (CISS), 46th Annual Conference on*, pp. 1–6, IEEE, 2012.

[57] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-advanced to FDD massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 2007–2016, 2015.

[58] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Wireless Communications and Networking Conference (WCNC), IEEE*, pp. 2890–2895, IEEE, 2013.

[59] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 1465–1470, IEEE, 2019.

[60] A. Abdallah and M. M. Mansour, "Efficient angle-domain processing forFDD-based cell-free massive MIMO systems," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2188–2203, 2020.

[61] F. Rottenberg, T. Choi, P. Luo, C. J. Zhang, and A. F. Molisch, "Performance analysis of channel extrapolation in FDD massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2728–2741, 2020.

[62] C. Tian, A. Liu, M. B. Khalilsarai, G. Caire, W. Luo, and M.-J. Zhao, "Randomized channel sparsifying hybrid precoding forFDD massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5447–5460, 2020.

[63] J. Guo, C.-K. Wen, and S. Jin, "CAnet: Uplink-aided downlink channel acquisition in FDD massive MIMO using deep learning," *arXiv preprint arXiv:2101.04377*, 2021.

[64] T. Van Chien, C. Mollén, and E. Björnson, "Large-scale-fading decoding in cellular massive MIMO systems with spatially correlated channels," *IEEE Trans. Commun.*, 2019.

[65] A. Ashikhmin, L. Li, and T. L. Marzetta, "Interference reduction in multi-cell massive MIMO systems with large-scale fading precoding," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6340–6361, 2018.

[66] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, pp. 3249–3252, IEEE, 2012.

[67] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, 2014.

[68] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Information Theory Proceedings (ISIT), IEEE International Symposium on*, pp. 1137–1141, IEEE, 2012.

[69] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on selected areas in communications*, vol. 31, no. 2, pp. 264–273, 2013.

[70] Ö. T. Demir and E. Björnson, "Large-scale fading precoding for spatially correlated Rician fading with phase shifts," *arXiv preprint arXiv:2006.14267*, 2020.

[71] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2966–2975, 2011.

[72] B. Bandemer, A. El Gamal, and Y.-H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, 2015.

[73] H. Yoon, S. Park, and S. Choi, "Efficient feedback mechanism and rate adaptation for LTE-based D2D communication," in *IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1–9, IEEE, 2017.

[74] A. Adhikary and A. Ashikhmin, "Uplink massive MIMO for channels with spatial correlation," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2018.

[75] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE transactions on information theory*, vol. 27, no. 1, pp. 49–60, 1981.

[76] T. L. Marzetta, G. Caire, M. Debbah, I. Chih-Lin, and S. K. Mohammed, "Special issue on massive MIMO," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 333–337, 2013.

[77] L. Berriche, K. Abed-Meraim, and J.-C. Belfiore, "Investigation of the channel estimation error on MIMO system performance," in *Signal Processing Conference, 13th European*, pp. 1–4, IEEE, 2005.

[78] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," in *Communications, IEEE International Conference on*, vol. 2, pp. 808–813, IEEE, 2004.

[79] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 3055–3069, 2002.

[80] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE transactions on signal processing*, vol. 54, no. 3, pp. 884–893, 2006.

[81] T. Wo and P. A. Hoeher, "Semi-blind channel estimation for frequency-selective MIMO systems," *Proc. 14-th IST Mobile & Wireless Communications Summit, Dresden, Germany*, 2005.

[82] H. Bolcskei, R. W. Heath, and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Transactions on signal Processing*, vol. 50, no. 1, pp. 96–109, 2002.

[83] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.

[84] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7771–7786, 2011.

[85] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE transactions on wireless communications*, vol. 9, no. 5, 2010.

[86] A. Ashikhmin, T. L. Marzetta, and L. Li, "Interference reduction in multi-cell massive MIMO systems i: Large-scale fading precoding and decoding," *arXiv preprint arXiv:1411.4182*, 2014.

[87] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[88] E. Björnson, P. Zetterberg, and M. Bengtsson, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *Global Communications Conference (GLOBECOM), IEEE*, pp. 4775–4780, IEEE, 2012.

[89] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[90] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 711–723, 2015.

[91] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, 2013.

[92] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO 5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2016.

[93] H. Yin, D. Gesbert, M. C. Filippou, and Y. Liu, "Decontaminating pilots in massive MIMO systems," in *Communications (ICC), IEEE International Conference on*, pp. 3170–3175, IEEE, 2013.

[94] M. Filippou, D. Gesbert, and H. Yin, "Decontaminating pilots in cognitive massive MIMO networks," in *Wireless Communication Systems (ISWCS), International Symposium on*, pp. 816–820, IEEE, 2012.

[95] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, 2016.

[96] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 759–772, 2014.

[97] S. Jin, M. Li, Y. Huang, Y. Du, and X. Gao, "Pilot scheduling schemes for multi-cell massive multiple-input–multiple-output transmission," *IET Communications*, vol. 9, no. 5, pp. 689–700, 2015.

[98] M. Li, S. Jin, and X. Gao, "Spatial orthogonality-based pilot reuse for multi-cell massive MIMO transmission," in *International Conference on Wireless Communications and Signal Processing*, pp. 1–6, IEEE, 2013.

[99] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on selected areas in communications*, vol. 31, no. 2, pp. 264–273, 2013.

[100] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1644–1647, 2015.

[101] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. Srinivas, "Resource allocation and scheduling in non-coherent user-centric cell-free MIMO," *arXiv preprint arXiv:2102.02968*, 2021.

[102] A.-J. Van Der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 173–190, 1997.

[103] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. i. algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, 1996.

[104] L. Cottatellucci, R. R. Muller, and M. Vehkapera, "Analysis of pilot decontamination based on power control," in *Vehicular Technology Conference (VTC Spring), IEEE 77th*, pp. 1–5, IEEE, 2013.

[105] R. R. Muller, M. Vehkapera, and L. Cottatellucci, "Analysis of blind pilot decontamination," in *Signals, Systems and Computers, Asilomar Conference on*, pp. 1016–1020, IEEE, 2013.

[106] D. Neumann, A. Gruendinger, M. Joham, and W. Utschick, "Pilot coordination for large-scale multi-cell TDD systems," in *Smart Antennas (WSA), 18th International ITG Workshop on*, pp. 1–6, VDE, 2014.

[107] M. Lee, "Jacket matrices: Constructions and its applications for fast cooperative wireless signal processing," *LAP LAMBERT Publishing, Germany*, 2012.

[108] M. A. L. Sarker and M. H. Lee, "A fast channel estimation and the reduction of pilot contamination problem for massive MIMO based on a diagonal Jacket matrix," in *Fiber Optics in Access Network (FOAN), 4th International Workshop on*, pp. 26–30, IEEE, 2013.

[109] C. K. Thomas, B. Clerckx, L. Sanguinetti, and D. Slock, "A rate splitting strategy for mitigating intra-cell pilot contamination in massive MIMO," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2020.

[110] H. Cramér, *Random variables and probability distributions*, vol. 36. Cambridge University Press, 2004.

[111] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[112] 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, 2003.

[113] T. L. Marzetta and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[114] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004.

[115] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[116] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1753–1767, 2015.

[117] F. Baccelli, A. El Gamal, and N. David, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, 2011.

[118] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE transactions on information theory*, vol. 27, no. 6, pp. 786–788, 1981.

[119] A. S. Motahari and A. K. Khandani, "To decode the interference or to consider it as noise," *IEEE transactions on information theory*, vol. 57, no. 3, pp. 1274–1283, 2011.

[120] G. Zhou, W. Xu, and G. Bauch, "Is MAC joint decoding optimal for interference channels?," in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2015.

[121] R. Chaluvadi, M. Bolli, and S. Bhashyam, "On the optimality of interference decoding schemes for k-user Gaussian interference channels," *Entropy*, vol. 21, no. 11, p. 1053, 2019.

[122] M. S. Motlagh, S. Majhi, and P. Mitran, "Performance of multi-cell massive MIMO systems with interference decoding," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 930–946, 2019.

[123] A. Sabharwal, D. Dash, and S. Diggavi, "Compound Gaussian multiple access channels with noisy feedback," in *Annu. Allerton Conf. Commun., Control, and Comput.*, pp. 887–894, IEEE, 2008.

[124] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[125] L. Wang, E. Şaşoğlu, and Y.-H. Kim, "Sliding-window superposition coding for interference networks," in *IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2749–2753, IEEE, 2014.

[126] H. Park, Y.-H. Kim, and L. Wang, "Interference management via sliding-window superposition coding," in *IEEE Globecom Workshops (GC Wkshps)*, pp. 972–976, IEEE, 2014.

[127] K. T. Kim, S.-K. Ahn, Y.-H. Kim, H. Park, L. Wang, C.-Y. Chen, and J. Park, "Adaptive sliding-window coded modulation in cellular networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–7, IEEE, 2015.

[128] K. T. Kim, S.-K. Ahn, Y.-S. Kim, J. Park, C.-Y. Chen, and Y.-H. Kim, "Interference management via sliding-window coded modulation for 5G cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 82–89, 2016.

[129] S. Ahn, K. Kwangtaik, and Y.-H. Kim, "Apparatus and method for scheduling in wireless communication system using sliding window superposition coding scheme," Apr. 8 2021. US Patent App. 15/733,203.

[130] Y.-h. Kim, W. Lele, H. Park, and E. Sasoglu, "Methods and apparatus for coding for interference network," Oct. 12 2017. US Patent App. 15/514,092.

[131] K. Kwangtaik, S. Ahn, Y.-h. Kim, H. Park, W. Lele, C. Chiao-Yi, and J. Park, "Method and apparatus for transmitting signal using sliding-window superposition coding in wireless network," July 4 2017. US Patent 9,699,008.

[132] S. Zarei, J. Aulin, W. Gerstacker, and R. Schober, "Max–min multicell-aware precoding and power allocation for downlink massive MIMO systems," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1433–1437, 2017.

[133] T. Van Chien, T. N. Canh, E. Björnson, and E. G. Larsson, "Power control in cellular massive MIMO with varying user activity: A deep learning solution," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5732–5748, 2020.

[134] A. Ghazanfari, H. V. Cheng, E. Björnson, and E. G. Larsson, "A fair and scalable power control scheme in multi-cell massive MIMO," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4499–4503, IEEE, 2019.

[135] R. S. Chaves, E. Cetin, M. V. Lima, and W. A. Martins, "On the convergence of max-min fairness power allocation in massive MIMO systems," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2873–2877, 2020.

[136] G. Interdonato, M. Karlsson, E. Bjornson, and E. G. Larsson, "Downlink spectral efficiency of cell-free massive MIMO with full-pilot zero-forcing," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1003–1007, IEEE, 2018.

[137] S. Shahsavari, M. Nosrati, P. Hassanzadeh, A. Ashikhmin, T. L. Marzetta, and E. Erkip, "Multi-point coordination in massive MIMO systems with sectorized antennas," *arXiv preprint arXiv:2104.10805*, 2021.

[138] S. Fujishige, *Submodular functions and optimization*, vol. 58. Elsevier, 2005.

[139] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Communications letters*, vol. 5, no. 9, pp. 369–371, 2001.

[140] E. Björnson, E. De Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random access protocol for pilot allocation in crowded massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2220–2234, 2017.

[141] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2000–2015, 2018.

[142] B. Sun, Y. Zhou, J. Yuan, and J. Shi, "Interference cancellation based channel estimation for massive MIMO systems with time shifted pilots," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6826–6843, 2020.

[143] I. Boukhedimi, A. Kammoun, and M.-S. Alouini, "Multi-cell MMSE combining over correlated Rician channels in massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 12–16, 2019.

[144] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero-forcing beamforming," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 2839–2853, 2019.

[145] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[146] L. Wang, Y.-H. Kim, C.-Y. Chen, H. Park, and E. Şaşoğlu, "Sliding-window superposition coding: Two-user interference channels," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3293–3316, 2020.

[147] M. Shahrbaf Motlagh, S. Majhi, P. Mitran, and H. Ochiai, "On downlink interference decoding in multi-cell massive MIMO systems," in *Biennial Symposium on Communications (BSC)*, June 2021.

[148] G. Zhou, W. Xu, and G. Bauch, "Interference mitigation with rate splitting in multi-cell wireless networks," in *IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 219–224, IEEE, 2012.

[149] G. Zhou, G. Bauch, J. Berkmann, and W. Xu, "Multi-layer rate splitting scheme for interference mitigation in tri-sectored wireless networks," in *IEEE International Conference on Communications (ICC)*, pp. 4202–4206, IEEE, 2012.

[150] A. Carleial, "Interference channels," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 60–70, 1978.

[151] R. H. Etkin, N. David, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on information theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[152] H. Dahrouj and W. Yu, "Multicell interference mitigation with joint beamforming and common message decoding," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2264–2273, 2011.

[153] O. Sahin, J. Li, Y. Li, and P. J. Pietraski, "Interference mitigation via successive cancellation in heterogeneous networks," in *International Symposium on Wireless Communication Systems*, pp. 720–724, IEEE, 2011.

[154] E. Che, H. D. Tuan, H. H. M. Tam, and H. H. Nguyen, "Successive interference mitigation in multiuser MIMO channels," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2185–2199, 2015.

[155] M. Medra and T. N. Davidson, "Robust downlink transmission: An offset-based single-rate-splitting approach," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2018.

[156] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP journal on wireless communications and networking*, vol. 2018, no. 1, pp. 1–54, 2018.

[157] Y. Mao, B. Clerckx, and V. O. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–5, IEEE, 2018.

[158] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, 2016.

[159] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8196–8211, 2017.

[160] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3188–3195, 2008.

[161] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, 2019.

[162] M. Matthaiou, C. Zhong, M. R. McKay, and T. Ratnarajah, "Sum rate analysis of ZF receivers in distributed MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 180–191, 2013.

[163] A. Yang, Y. Jing, C. Xing, Z. Fei, and J. Kuang, "Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5659–5671, 2015.

[164] M. Sadeghi, C. Yuen, and Y. H. Chew, "Sum rate maximization for uplink distributed massive MIMO systems with limited backhaul capacity," in *Globecom Workshops (GC Wkshps)*, pp. 308–313, IEEE, 2014.

[165] Y. Huang, C. W. Tan, and B. D. Rao, "Efficient SINR fairness algorithm for large distributed multiple-antenna networks," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, pp. 5026–5030, IEEE, 2013.

[166] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 954–965, 2014.

[167] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive MIMO systems," in *Signals, Systems and Computers, Asilomar Conference on*, pp. 1318–1323, IEEE, 2013.

[168] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2018.

[169] D. Pliatsios, P. Sarigiannidis, S. Goudos, and G. K. Karagiannidis, "Realizing 5G vision through Cloud RAN: technologies, challenges, and trends," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–15, 2018.

[170] A. Burr, M. Bashar, and D. Maryopi, "Cooperative access networks: Optimum fronthaul quantization in distributed massive MIMO and cloud ran," in *IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2018.

[171] G. Femenias and F. Riera-Palou, "Multi-layer downlink precoding for cloud-RAN systems using full-dimensional massive MIMO," *IEEE Access*, vol. 6, pp. 61583–61599, 2018.

[172] H. Peng, W. Tärneberg, E. Fitzgerald, and M. Kihl, "Massive MIMO pilot scheduling over cloud RAN for industry 4.0," in *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 1–6, IEEE, 2020.

[173] F. Rezaei and A. Tadaion, "Multi-layer beamforming in uplink/downlink massive MIMO systems with multi-antenna users," *Signal Processing*, vol. 164, pp. 58–66, 2019.

[174] J. Singh and D. Kedia, "Performance improvement in large-scale MU-MIMO system with multiple antennas on user side in a single-cell downlink system," *Arabian Journal for Science and Engineering*, vol. 45, no. 8, pp. 6769–6789, 2020.

[175] T. C. Mai, H. Q. Ngo, and T. Q. Duong, "Cell-free massive MIMO systems with multi-antenna users," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 828–832, IEEE, 2018.

[176] T. C. Mai, H. Q. Ngo, and T. Q. Duong, "Downlink spectral efficiency of cell-free massive MIMO systems with multi-antenna users," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4803–4815, 2020.

[177] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4353–4368, 2015.

Appendices

Appendix A

A.1 Proof of Lemma 2

Without loss of generality assume that $\Omega = \{1, 2, ..., l\}$ and thus $\Omega^c = \{l+1, ..., L\}$. We start by expanding the r.h.s of (3.4) as follows

$$I\left(\boldsymbol{x}_{\Omega}^{G};\ y\middle|\boldsymbol{x}_{\Omega^c}^{G}\right) = h\left(\boldsymbol{x}_{\Omega}^{G}\middle|\boldsymbol{x}_{\Omega^c}^{G}\right) - h\left(\boldsymbol{x}_{\Omega}^{G}\middle|y, \boldsymbol{x}_{\Omega^c}^{G}\right) \tag{A.1}$$

$$\overset{(a)}{=} h\left(\boldsymbol{x}_{\Omega}^{G}\right) - h\left(\boldsymbol{x}_{\Omega}^{G}\middle|y, \boldsymbol{x}_{\Omega^c}^{G}\right) \tag{A.2}$$

$$\overset{(b)}{=} \log\left((\pi e)^l \Pi_{i=1}^{l} P_i\right) - h\left(\boldsymbol{x}_{\Omega}^{G}\middle|y, \boldsymbol{x}_{\Omega^c}^{G}\right), \tag{A.3}$$

where (a) is because the entries of $\boldsymbol{x}_{\Omega}^{G}$ and $\boldsymbol{x}_{\Omega^c}^{G}$ are independent, and (b) follows from the entropy of a complex Gaussian vector with independent entries. Also, using the chain rule one can write

$$\begin{aligned} & h\left(\boldsymbol{x}_{\Omega}^{G}\middle|y, \boldsymbol{x}_{\Omega^c}^{G}\right) \qquad (A.4) \\ &= \sum_{i\in\Omega} h\left(x_i^G \middle| x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^{G}\right) \\ &= \sum_{i\in\Omega} h\left(x_i^G - \alpha_i\left(y - \sum_{j=1}^{i-1} x_j^G - \sum_{k\in\Omega^c} x_k^G\right) \middle| x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^{G}\right), \end{aligned}$$

where α_i is any constant. Defining $\tilde{y}_i = y - \sum_{j=1}^{i-1} x_j^G - \sum_{k\in\Omega^c} x_k^G$, we obtain

$$\sum_{i\in\Omega} h\left(x_i^G - \alpha_i\tilde{y}_i \middle| x_1^G, ..., x_{i-1}^G, y, \boldsymbol{x}_{\Omega^c}^G\right) \tag{A.5}$$

$$\overset{(c)}{\leq} \sum_{i\in\Omega} h\left(x_i^G - \alpha_i\tilde{y}_i\right)$$

$$\overset{(d)}{\leq} \sum_{i\in\Omega} \log\left((\pi e)\text{var}\left[x_i^G - \alpha_i\tilde{y}_i\right]\right), \tag{A.6}$$

where (c) is due to the fact that conditioning reduces the entropy and (d) follows as Gaussian distributions maximize entropy. To obtain the tightest upper bound, one should minimize $\text{var}[x_i^G - \alpha_i\tilde{y}_i]$, i.e., $\alpha_i\tilde{y}_i$ must be the LMMSE estimate of x_i^G. More precisely, one can choose $\alpha_i = \mathbb{E}[\tilde{y}_i^*\tilde{y}_i]^{-1}\mathbb{E}[x_i^G\tilde{y}_i^*] = P_i/(\sum_{j=i}^{l} P_j + \sigma_z^2)$, where the second equality follows since z is uncorrelated from the users' signals. Thus

$$\text{var}\left[x_i^G - \alpha_i\tilde{y}_i\right] = \frac{P_i\left(\sum_{j=i+1}^{l} P_j + \sigma_z^2\right)}{\sum_{j=i}^{l} P_j + \sigma_z^2}, \tag{A.7}$$

and therefore we obtain

$$h\left(\boldsymbol{x}_\Omega^G \middle| y, \boldsymbol{x}_{\Omega^c}^G\right) \leq \sum_{i\in\Omega} \log\left((\pi e)\frac{P_i\left(\sum_{j=i+1}^{l} P_j + \sigma_z^2\right)}{\sum_{j=i}^{l} P_j + \sigma_z^2}\right) \tag{A.8}$$

$$= \log\left((\pi e)^l \frac{\left(\Pi_{i\in\Omega} P_i\right)\sigma_z^2}{\sum_{i\in\Omega} P_i + \sigma_z^2}\right). \tag{A.9}$$

Hence, from (A.3) the following lower bound is obtained

$$I\left(\boldsymbol{x}_\Omega^G;\ y \middle| \boldsymbol{x}_{\Omega^c}^G\right)$$

$$\geq \log\left((\pi e)^l \Pi_{i\in\Omega} P_i\right) - \log\left((\pi e)^l \frac{\left(\Pi_{i\in\Omega} P_i\right)\sigma_z^2}{\sum_{i\in\Omega} P_i + \sigma_z^2}\right) \tag{A.10}$$

$$= \log\left(1 + \frac{\sum_{i\in\Omega} P_i}{\sigma_z^2}\right) = I\left(\boldsymbol{x}_\Omega^G;\ y^G \middle| \boldsymbol{x}_{\Omega^c}^G\right). \tag{A.11}$$

A.2 Proof of Theorem 1

The expression of $\hat{y}_{ji}^{\text{ul}}$ in (3.5) includes four different sums where each term in each sum is zero-mean and also uncorrelated with every other term in the four sums. As the last three sums in (3.5) are uncorrelated with $x_l^{\text{ul}}[i], l = 1, ..., L$, the variance of these can be treated as the variance of an uncorrelated noise. Therefore, to prove the desired result, one needs to calculate the variance of the interference due to beamforming gain uncertainty (denoted by P_2^{ul}), the variance of the interference caused by other uses (denoted by P_3^{ul}) and the variance of noise (denoted by P_4^{ul}). It can be verified that the variance of the desired signal components is

$$P_1^{\text{ul}} = \sum_{l=1}^{L} \rho_{\text{ul}} \left| \mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right] \right|^2. \tag{A.12}$$

Also, since these four sums are uncorrelated, one can write

$$P_2^{\text{ul}} + P_3^{\text{ul}} + P_4^{\text{ul}} = \text{var}\left[\hat{y}_{ji}^{\text{ul}} \right] - P_1^{\text{ul}}. \tag{A.13}$$

Hence, we need to compute $\text{var}\left[\hat{y}_{ji}^{\text{ul}} \right]$. First, observe that the uplink baseband signal $\hat{y}_{ji}^{\text{ul}}$ can be equivalently written as

$$\hat{y}_{ji}^{\text{ul}} = \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\rho_{\text{ul}}} \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{g}_{jkl} x_l^{\text{ul}}[k] + \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{n}_j. \tag{A.14}$$

Therefore, $\text{var}\left[\hat{y}_{ji}^{\text{ul}} \right]$ can be computed as follows

$$\text{var}\left[\hat{y}_{ji}^{\text{ul}} \right] = \sum_{l=1}^{L} \sum_{k=1}^{K} \rho_{\text{ul}} \mathbb{E}\left[\left| \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{g}_{jkl} \right|^2 \right] + \mathbb{E}\left[\left| \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{n}_j \right|^2 \right] \tag{A.15}$$

$$= \sum_{l=1}^{L} \sum_{k=1}^{K} \rho_{\text{ul}} \mathbb{E}\left[\left| \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{g}_{jkl} \right|^2 \right] + \mathbb{E}\left[\left\| \boldsymbol{v}_{jij} \right\|^2 \right], \tag{A.16}$$

where we have used the fact that $\mathbb{E}\left[\left|x_l^{\mathrm{ul}}[k]\right|^2\right] = 1, \forall l, k$, and the last step follows by noting that the elements of the noise vector $\boldsymbol{n}_j$ have unit variance. We thus have

$$P_2^{\mathrm{ul}} + P_3^{\mathrm{ul}} + P_4^{\mathrm{ul}} = \sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right] + \mathbb{E}\left[\|\boldsymbol{v}_{jij}\|^2\right] - \sum_{l=1}^{L} \rho_{\mathrm{ul}} \left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2. \quad \text{(A.17)}$$

Furthermore, note that for the numerator in the r.h.s of (3.8) we have

$$\mathrm{var}\left[\sqrt{\rho_{\mathrm{ul}}}\sum_{l\in\Omega}\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right] x_l^{\mathrm{ul}}[i]\right] = \sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2. \quad \text{(A.18)}$$

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the lower bound in (3.8).

A.3 Proof of Theorem 2

The proof follows steps similar to those of Theorem 1. Specifically, note that the expression of y_{il}^{dl} in (3.9) includes four zero-mean sums that are also uncorrelated. Denote the variance of the desired signals by P_1^{dl}, the variance of the interference due to beamforming gain uncertainty by P_2^{dl}, the variance of inter-cell interference caused by other users by P_3^{dl} and the variance of noise by P_4^{dl}. Hence, since these four sums are uncorrelated with each other, we have $\mathrm{var}\left[y_{il}^{\mathrm{dl}}\right] = P_1^{\mathrm{dl}} + P_2^{\mathrm{dl}} + P_3^{\mathrm{dl}} + P_4^{\mathrm{dl}}$. It is easy to verify that the variance of the desired signals is

$$P_1^{\mathrm{dl}} = \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]\right|^2, \quad \text{(A.19)}$$

which is followed from $\mathbb{E}\left[|s_j[k]|^2\right] = 1, \forall j, k$. In addition, one can equivalently write y_{il}^{dl} as

$$y_{il}^{\mathrm{dl}} = \sum_{j=1}^{L}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\sum_{k=1}^{K}\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}s_j[k] + z_{il}. \quad \text{(A.20)}$$

Therefore, we obtain for var $\left[y_{il}^{\mathrm{dl}}\right]$

$$\mathrm{var}\left[y_{il}^{\mathrm{dl}}\right] = \sum_{j=1}^{L}\sum_{k=1}^{K}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}\right|^2\right] + 1, \tag{A.21}$$

where we have used the fact that the noise term has unit variance, i.e., $\mathbb{E}\left[|z_{il}|^2\right] = 1$. As such, $P_2^{\mathrm{dl}} + P_3^{\mathrm{dl}} + P_4^{\mathrm{dl}}$ is found as

$$P_2^{\mathrm{dl}} + P_3^{\mathrm{dl}} + P_4^{\mathrm{dl}} = \sum_{j=1}^{L}\sum_{k=1}^{K}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jkj}\right|^2\right] + 1 - \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]\right|^2. \tag{A.22}$$

Lastly, for the numerator in the r.h.s of (3.12) we have

$$\mathrm{var}\left[\sum_{j\in\Omega}\sqrt{\frac{\rho_{\mathrm{dl}}}{\lambda_j}}\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]s_j[i]\right] = \sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]\right|^2. \tag{A.23}$$

Therefore, since the effective noise is zero-mean and uncorrelated with the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the lower bound in (3.12).

A.4 Proof of Theorem 3

Following a similar approach to that of [64, Appendix C], for the numerator of the lower bound in (3.13), we calculate $\sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2$ with the combining vector given by $\boldsymbol{v}_{jij} = \hat{\boldsymbol{g}}_{jij}$, i.e.,

$$\sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2 = \sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\left(\hat{\boldsymbol{g}}_{jil} + \boldsymbol{\epsilon}_{jil}\right)\right]\right|^2 \tag{A.24}$$

$$\overset{(a)}{=} \sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right]\right|^2 \tag{A.25}$$

$$= \sum_{l\in\Omega}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\mathrm{tr}\left(\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right)\right]\right|^2 \tag{A.26}$$

$$= \sum_{l\in\Omega} \rho_{\text{ul}} \left| \text{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil} \hat{\boldsymbol{g}}_{jij}^{\dagger} \right] \right) \right|^2, \tag{A.27}$$

where (a) follows by the fact that the estimate and the estimation error are uncorrelated and have zero mean. Recall from Chapter 2 that $\hat{\boldsymbol{g}}_{jil} = \boldsymbol{R}_{jil}\boldsymbol{R}_{jij}^{-1}\hat{\boldsymbol{g}}_{jij}$. Therefore,

$$\sum_{l\in\Omega} \rho_{\text{ul}} \left| \text{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil} \hat{\boldsymbol{g}}_{jij}^{\dagger} \right] \right) \right|^2 = \sum_{l\in\Omega} \rho_{\text{ul}} \left| \text{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}^{-1} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij} \hat{\boldsymbol{g}}_{jij}^{\dagger} \right] \right) \right|^2 \tag{A.28}$$

$$= \sum_{l\in\Omega} \rho_{\text{ul}} \left| \text{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}^{-1} \rho_{\text{p}} \boldsymbol{R}_{jij} \boldsymbol{\Lambda}_{ji}^{-1} \boldsymbol{R}_{jij} \right) \right|^2 \tag{A.29}$$

$$= \sum_{l\in\Omega} \rho_{\text{ul}} \rho_{\text{p}}^2 \left| \text{tr}\left(\boldsymbol{R}_{jil} \boldsymbol{\Lambda}_{ji}^{-1} \boldsymbol{R}_{jij} \right) \right|^2. \tag{A.30}$$

The first term in the denominator of the lower bound in (3.13) is computed as follows.

$$\sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\text{ul}} \mathbb{E}\left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jkl} \right|^2 \right] = \sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right|^2 \right] + \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K} \rho_{\text{ul}} \mathbb{E}\left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jkl} \right|^2 \right], \tag{A.31}$$

i.e., we separately compute the summation over the users that share the same pilot sequence as user i in cell j (the first summation in (A.31)) and the summation over users that are using a different pilot sequence (the second summation in (A.31)). For the first summation, we have

$$\sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right|^2 \right] = \sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\left(\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right) \left(\boldsymbol{g}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right) \right] \tag{A.32}$$

$$= \sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \left(\hat{\boldsymbol{g}}_{jil} + \boldsymbol{\epsilon}_{jil} \right) \left(\hat{\boldsymbol{g}}_{jil}^{\dagger} + \boldsymbol{\epsilon}_{jil}^{\dagger} \right) \hat{\boldsymbol{g}}_{jij} \right] \tag{A.33}$$

$$\overset{(b)}{=} \sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \hat{\boldsymbol{g}}_{jil} \hat{\boldsymbol{g}}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right] + \sum_{l=1}^{L} \rho_{\text{ul}} \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{\epsilon}_{jil} \boldsymbol{\epsilon}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right], \tag{A.34}$$

where (b) is due to the fact that the estimate and the estimation error are uncorrelated

and have zero mean. For the second summation in (A.34), we obtain

$$\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]=\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right)\right] \tag{A.35}$$

$$\overset{(c)}{=}\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right]\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\right) \tag{A.36}$$

$$=\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathrm{tr}\left(\left(\boldsymbol{R}_{jil}-\rho_{\mathrm{p}}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\left(\rho_{\mathrm{p}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right) \tag{A.37}$$

$$=\sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)$$

$$-\sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^{2}\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right), \tag{A.38}$$

where (c) follows by the fact that the estimate and the estimation error are uncorrelated. To compute the first summation in (A.34), we first need to present the following lemma from [177, Lemma 2].

Lemma 4. *Consider a zero-mean random vector* $\boldsymbol{q}$ *with distribution* $\boldsymbol{q}\sim\mathcal{CN}(\boldsymbol{0},\,\boldsymbol{\Sigma})$*, where* $\boldsymbol{\Sigma}\in\mathbb{C}^{M\times M}$ *is the covariance matrix, and an arbitrary deterministic matrix* $\boldsymbol{W}\in\mathbb{C}^{M\times M}$*. Then, the following identity holds*

$$\mathbb{E}\left[\left|\boldsymbol{q}^{\dagger}\boldsymbol{W}\boldsymbol{q}\right|^{2}\right]=\left|tr\left(\boldsymbol{\Sigma}\boldsymbol{W}\right)\right|^{2}+tr\left(\boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}^{\dagger}\right). \tag{A.39}$$

Next, we re-write the first summation in (A.34) as follows

$$\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right|^{2}\right]\overset{(d)}{=}\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\sqrt{\rho_{\mathrm{p}}}\boldsymbol{r}_{ji}^{\dagger}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}\right|^{2}\right] \tag{A.40}$$

$$=\sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^{2}\mathbb{E}\left[\left|\boldsymbol{r}_{ji}^{\dagger}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}\right|^{2}\right], \tag{A.41}$$

where (d) follows by the derivations of the MMSE channel estimate from Chapter 2, i.e.,

$\hat{\boldsymbol{g}}_{jij} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}$ and $\hat{\boldsymbol{g}}_{jil} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}$. Noting that $\boldsymbol{r}_{ji} \sim \mathcal{CN}(\boldsymbol{0},\ \boldsymbol{\Lambda}_{ji})$, and assuming that $\boldsymbol{q} = \boldsymbol{r}_{ji}$ and $\boldsymbol{W} = \boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}$, we can apply Lemma 4 to obtain the following

$$\sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\mathbb{E}\left[\left|\boldsymbol{r}_{ji}^{\dagger}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}\right|^2\right] = \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\left|\mathrm{tr}\left(\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\right)\right|^2 + \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\mathrm{tr}\left(\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\right) \tag{A.42}$$

$$= \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\left|\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\right)\right|^2 + \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\right) \tag{A.43}$$

$$= \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\left|\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right|^2 + \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right). \tag{A.44}$$

Now, (A.32) which is the summation of (A.38) and (A.44) is found as

$$\sum_{l=1}^{L}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right|^2\right] = \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}^2\left|\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right|^2 + \sum_{l=1}^{L}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right). \tag{A.45}$$

For the second summation in (A.31) (i.e., summation over users that are using a different pilot sequence), we obtain

$$\sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right] = \sum_{l=1}^{L}\sum_{k=1,k\neq i}^{K}\rho_{\mathrm{ul}}\mathbb{E}\left[\left(\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right)\left(\boldsymbol{g}_{jkl}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right)\right] \tag{A.46}$$

$$= \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \rho_{\text{ul}} \mathbb{E}\left[\text{tr}\left(\boldsymbol{g}_{jkl}\boldsymbol{g}_{jkl}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right)\right] \tag{A.47}$$

$$\overset{(e)}{=} \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \rho_{\text{ul}} \text{tr}\left(\mathbb{E}\left[\boldsymbol{g}_{jkl}\boldsymbol{g}_{jkl}^{\dagger}\right] \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\right) \tag{A.48}$$

$$= \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \rho_{\text{ul}} \text{tr}\left(\boldsymbol{R}_{jkl}\rho_{\text{p}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right) \tag{A.49}$$

$$= \sum_{l=1}^{L} \sum_{k=1,k\neq i}^{K} \rho_{\text{ul}}\rho_{\text{p}} \text{tr}\left(\boldsymbol{R}_{jkl}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right), \tag{A.50}$$

where (e) is due to the fact that $\boldsymbol{g}_{jkl}$ and $\hat{\boldsymbol{g}}_{jij}$ are uncorrelated for $k \neq i$, as these users are using different pilot sequences. Hence, (A.31) (the first term in the denominator of the lower bound in (3.13)) is simplified as below

$$\sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\text{ul}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right] = \sum_{l=1}^{L} \rho_{\text{ul}}\rho_{\text{p}}^2 \left|\text{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right|^2 + \sum_{l=1}^{L}\sum_{k=1}^{K} \rho_{\text{ul}}\rho_{\text{p}} \text{tr}\left(\boldsymbol{R}_{jkl}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right). \tag{A.51}$$

The computation of the second term in the denominator of the lower bound in (3.13) follows steps similar to those of the numerator, except that now the summation is over all $l = 1, ..., L$. We thus have

$$\sum_{l=1}^{L} \rho_{\text{ul}} \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2 = \sum_{l=1}^{L} \rho_{\text{ul}}\rho_{\text{p}}^2 \left|\text{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\right|^2. \tag{A.52}$$

Lastly, the third term in the denominator of the lower bound in (3.13) is calculated as follows

$$\mathbb{E}\left[\|\boldsymbol{v}_{jij}\|^2\right] = \mathbb{E}\left[\left\|\hat{\boldsymbol{g}}_{jij}\right\|^2\right] \tag{A.53}$$

$$= \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right] \tag{A.54}$$

$$= \mathrm{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\right) \tag{A.55}$$

$$= \rho_{\mathrm{p}}\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right). \tag{A.56}$$

Therefore, for the denominator of the lower bound in (3.13), we obtain

$$\sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\mathbb{E}\left[\left|\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jkl}\right|^2\right]-\sum_{l=1}^{L}\rho_{\mathrm{ul}}\left|\mathbb{E}\left[\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2+\mathbb{E}\left[\|\boldsymbol{v}_{jij}\|^2\right] \tag{A.57}$$

$$= \sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\mathrm{ul}}\rho_{\mathrm{p}}\,\mathrm{tr}\left(\boldsymbol{R}_{jkl}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)+\rho_{\mathrm{p}}\,\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right). \tag{A.58}$$

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound. This completes the proof.

A.5 Proof of Theorem 4

Following a similar approach to that of [64, Appendix C], for the numerator of the lower bound in (3.16), we compute $\sum_{j\in\Omega}\left(\rho_{\mathrm{dl}}/\lambda_j\right)\left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\boldsymbol{w}_{jij}\right]\right|^2$, with the precoding vector given by $\boldsymbol{w}_{jij}=\hat{\boldsymbol{g}}_{jij}$. Hence,

$$\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left|\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2=\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left|\mathbb{E}\left[\left(\hat{\boldsymbol{g}}_{jil}^{\dagger}+\boldsymbol{\epsilon}_{jil}^{\dagger}\right)\hat{\boldsymbol{g}}_{jij}\right]\right|^2 \tag{A.59}$$

$$\overset{(a)}{=}\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left|\mathbb{E}\left[\mathrm{tr}\left(\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jil}^{\dagger}\right)\right]\right|^2 \tag{A.60}$$

$$\overset{(b)}{=}\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left|\mathrm{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\boldsymbol{R}_{jij}^{-1}\boldsymbol{R}_{jil}\right)\right|^2 \tag{A.61}$$

$$=\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left|\mathrm{tr}\left(\rho_{\mathrm{p}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jij}^{-1}\boldsymbol{R}_{jil}\right)\right|^2 \tag{A.62}$$

$$=\sum_{j\in\Omega}\frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}}\left|\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\right|^2, \tag{A.63}$$

where (a) follows by the fact that the estimate and the estimation error are uncorrelated and have zero mean, and (b) follows from the identity $\hat{\boldsymbol{g}}_{jil} = \boldsymbol{R}_{jil}\boldsymbol{R}_{jij}^{-1}\hat{\boldsymbol{g}}_{jij}$.

Next, we need to compute the first and the second term in the denominator of the lower bound in (3.12). For the first term, similar to the proof of uplink MRC, we separately compute the summation over the users that are sharing the same pilot sequence as user i in cell l and the summation over those users that are using a different pilot sequence, i.e.,

$$\sum_{j=1}^{L}\sum_{k=1}^{K}\frac{\rho_{\mathrm{dl}}}{\lambda_j}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right] = \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right] + \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right]. \tag{A.64}$$

For the first term in the r.h.s of (A.64) (i.e., the summation over the users sharing the i^{th} pilot sequence), we obtain

$$\begin{aligned}
\sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right] &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right] && \text{(A.65)}\\
&= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left(\hat{\boldsymbol{g}}_{jil}^{\dagger}+\boldsymbol{\epsilon}_{jil}^{\dagger}\right)\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\left(\hat{\boldsymbol{g}}_{jil}+\boldsymbol{\epsilon}_{jil}\right)\right] && \text{(A.66)}\\
&\overset{(c)}{=} \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right]+\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\right]\right), && \text{(A.67)}
\end{aligned}$$

where (c) is obtained by noting that the estimate and the estimation error are uncorrelated and have zero mean. The first term in (A.67) is computed as follows

$$\begin{aligned}
\sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right] &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right] && \text{(A.68)}\\
&\overset{(d)}{=} \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\sqrt{\rho_{\mathrm{p}}}\boldsymbol{r}_{ji}^{\dagger}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}\right|^2\right] && \text{(A.69)}\\
&= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\boldsymbol{r}_{ji}^{\dagger}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}\right|^2\right] && \text{(A.70)}
\end{aligned}$$

$$\overset{(e)}{=} \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \left|\mathrm{tr}\left(\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\right)\right|^2$$
$$+ \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{\Lambda}_{ji}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\right) \quad \text{(A.71)}$$
$$= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \left|\mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\right)\right|^2$$
$$+ \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{R}_{jil}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\right) \quad \text{(A.72)}$$
$$= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \left|\mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\right|^2$$
$$+ \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right), \quad \text{(A.73)}$$

where (d) is obtained using the identities $\hat{\boldsymbol{g}}_{jij} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}$ and $\hat{\boldsymbol{g}}_{jil} = \sqrt{\rho_{\mathrm{p}}}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{r}_{ji}$, and (e) follows from applying Lemma 4 and noting that $\boldsymbol{r}_{ji} \sim \mathcal{CN}(\boldsymbol{0},\ \boldsymbol{\Lambda}_{ji})$.

The second term in (A.67) is obtained as follows

$$\sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathbb{E}\left[\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\right] = \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathbb{E}\left[\mathrm{tr}\left(\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right)\right] \quad \text{(A.74)}$$
$$= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right] \mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right]\right) \quad \text{(A.75)}$$
$$= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\left(\rho_{\mathrm{p}}\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\right)\left(\boldsymbol{R}_{jil} - \rho_{\mathrm{p}}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\right) \quad \text{(A.76)}$$
$$= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\right)$$
$$- \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right). \quad \text{(A.77)}$$

Thus, adding up (A.73) and (A.77), for the first summation in the r.h.s of (A.64) (i.e., the summation over the users sharing the i^{th} pilot sequence) we obtain

$$\sum_{j=1}^{L} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right] = \sum_{j=1}^{L} \frac{\rho_{\text{dl}}\rho_{\text{p}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jij}\boldsymbol{R}_{jil}\right) + \sum_{j=1}^{L} \frac{\rho_{\text{dl}}\rho_{\text{p}}^2}{\lambda_j^{\text{mrt}}} \left|\text{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\right|^2. \tag{A.78}$$

For the second summation in the r.h.s of (A.64), we have

$$\sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right] = \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\hat{\boldsymbol{g}}_{jkj}^{\dagger}\boldsymbol{g}_{jil}\right] \tag{A.79}$$

$$= \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \mathbb{E}\left[\text{tr}\left(\hat{\boldsymbol{g}}_{jkj}\hat{\boldsymbol{g}}_{jkj}^{\dagger}\boldsymbol{g}_{jil}\boldsymbol{g}_{jil}^{\dagger}\right)\right] \tag{A.80}$$

$$\overset{(f)}{=} \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jkj}\hat{\boldsymbol{g}}_{jkj}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{g}_{jil}\boldsymbol{g}_{jil}^{\dagger}\right]\right) \tag{A.81}$$

$$= \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\rho_{\text{p}}\boldsymbol{R}_{jkj}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jkj}\boldsymbol{R}_{jil}\right) \tag{A.82}$$

$$= \sum_{j=1}^{L}\sum_{k=1,k\neq i}^{K} \frac{\rho_{\text{dl}}\rho_{\text{p}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\boldsymbol{R}_{jkj}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jkj}\boldsymbol{R}_{jil}\right), \tag{A.83}$$

where (f) follows by the fact that $\hat{\boldsymbol{g}}_{jkl}$ and $\boldsymbol{g}_{jil}$ are uncorrelated for $k \neq i$. Hence, by adding (A.83) and (A.78), we obtain

$$\sum_{j=1}^{L}\sum_{k=1}^{K} \frac{\rho_{\text{dl}}}{\lambda_j} \mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right] = \sum_{j=1}^{L}\sum_{k=1}^{K} \frac{\rho_{\text{dl}}\rho_{\text{p}}}{\lambda_j^{\text{mrt}}} \text{tr}\left(\boldsymbol{R}_{jkj}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jkj}\boldsymbol{R}_{jil}\right) + \sum_{j=1}^{L} \frac{\rho_{\text{dl}}\rho_{\text{p}}^2}{\lambda_j^{\text{mrt}}} \left|\text{tr}\left(\boldsymbol{R}_{jij}\boldsymbol{\Lambda}_{ji}^{-1}\boldsymbol{R}_{jil}\right)\right|^2. \tag{A.84}$$

Lastly, the second summation in the denominator of the lower bound in (3.12), with the precoding vector given by $\boldsymbol{w}_{jij} = \hat{\boldsymbol{g}}_{jij}$, can be computed following steps similar to those

of the numerator in (A.59), except that now the summation is over all $j = 1, ..., L$. Hence,

$$\sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j} \left| \mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right] \right|^2 = \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}} \rho_{\mathrm{p}}^2}{\lambda_j^{\mathrm{mrt}}} \left| \mathrm{tr}\left(\boldsymbol{R}_{jij} \boldsymbol{\Lambda}_{ji}^{-1} \boldsymbol{R}_{jil} \right) \right|^2. \tag{A.85}$$

Therefore, the denominator of the lower bound in (3.12) is simplified as follows

$$\begin{aligned} &\sum_{j=1}^{L} \sum_{k=1}^{K} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathbb{E}\left[\left| \boldsymbol{g}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jkj} \right|^2 \right] - \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \left| \mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right] \right|^2 + 1 \\ &\qquad = \sum_{j=1}^{L} \sum_{k=1}^{K} \frac{\rho_{\mathrm{dl}} \rho_{\mathrm{p}}}{\lambda_j^{\mathrm{mrt}}} \mathrm{tr}\left(\boldsymbol{R}_{jkj} \boldsymbol{\Lambda}_{jk}^{-1} \boldsymbol{R}_{jkj} \boldsymbol{R}_{jil} \right) + 1. \end{aligned} \tag{A.86}$$

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound. This completes the proof.

A.6 Proof of Corollary 1

We start by computing the power of the desired signals, P_1^{mrc}. Note that

$$P_1^{\mathrm{mrc}} = \sum_{l \in \Omega} \rho_{\mathrm{ul}} \left| \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jil} \right] \right|^2 = \sum_{l \in \Omega} \rho_{\mathrm{ul}} \left| \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \left(\hat{\boldsymbol{g}}_{jil} + \boldsymbol{\epsilon}_{jil} \right) \right] \right|^2 \tag{A.87}$$

$$\overset{(a)}{=} \sum_{l \in \Omega} \rho_{\mathrm{ul}} \left| \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger} \hat{\boldsymbol{g}}_{jil} \right] \right|^2 \tag{A.88}$$

$$\overset{(b)}{=} \sum_{l \in \Omega} M^2 \rho_{\mathrm{ul}} \left(\frac{\beta_{jil}}{\beta_{jij}} \right)^2 \frac{\rho_{\mathrm{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1} \right)^2} \tag{A.89}$$

$$= M^2 \sum_{l \in \Omega} \rho_{\mathrm{p}} \rho_{\mathrm{ul}} \beta_{jil}^2 \alpha_{jij}^2, \tag{A.90}$$

where (a) follows from the fact that $\hat{\boldsymbol{g}}_{jij}$ and $\boldsymbol{\epsilon}_{jij}$ are independent, and (b) is obtained by noting that $\hat{\boldsymbol{g}}_{jil} = (\beta_{jil}/\beta_{jij})\hat{\boldsymbol{g}}_{jij}$. Note that as explained earlier, all terms in the effective noise are uncorrelated; thus $\mathrm{var}[z'_{jij}] = P_2^{\mathrm{mrc}} + P_3^{\mathrm{mrc}} + P_4^{\mathrm{mrc}}$, where P_2^{mrc} is the variance of interference due to the beamforming gain uncertainty, P_3^{mrc} is the variance caused by other users and P_4^{mrc} is the variance of noise.

To compute P_2^{mrc}, we write

$$P_2^{\mathrm{mrc}} = \sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{g}_{jil}\right]\right|^2\right] \tag{A.91}$$

$$\overset{(c)}{=} \sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right]\right|^2\right] + \sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\right|^2\right], \tag{A.92}$$

where (c) is due to the fact that $\hat{\boldsymbol{g}}_{jil} = (\beta_{jil}/\beta_{jij})\hat{\boldsymbol{g}}_{jij}$. For the first term in (A.92) we obtain

$$\sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jil}\right]\right|^2\right]$$

$$= \sum_{l=1}^{L} \rho_{\mathrm{ul}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \mathrm{var}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right] \tag{A.93}$$

$$= \sum_{l=1}^{L} M\rho_{\mathrm{ul}} \left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2 \frac{\rho_{\mathrm{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}\right)^2}. \tag{A.94}$$

Similarly, for the second term in (A.92) we obtain

$$\sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\right|^2\right]$$

$$= \sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathbb{E}\left[\mathrm{tr}\left(\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right)\right] \tag{A.95}$$

$$= \sum_{l=1}^{L} \rho_{\mathrm{ul}} \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right] \mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\right) \tag{A.96}$$

$$= M\rho_{\mathrm{ul}} \sum_{l=1}^{L} \left(\beta_{jil} - \frac{\rho_{\mathrm{p}} \beta_{jil}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right) \left(\frac{\rho_{\mathrm{p}} \beta_{jij}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right). \tag{A.97}$$

Therefore, using (A.94) and (A.97), one can verify that

$$P_2^{\mathrm{mrc}} = \sum_{l=1}^{L} M\rho_{\mathrm{ul}} \left(\frac{\beta_{jil}}{\beta_{jij}} \right)^2 \left(\frac{\rho_{\mathrm{p}}^2 \beta_{jij}^4}{\left(1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1} \right)^2} \right) \tag{A.98}$$

$$+ M\rho_{\mathrm{ul}} \left(\frac{\rho_{\mathrm{p}} \beta_{jij}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right) \sum_{l=1}^{L} \left(\beta_{jil} - \frac{\rho_{\mathrm{p}} \beta_{jil}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right)$$

$$= M\sqrt{\rho_{\mathrm{p}}} \beta_{jij} \alpha_{jij} \sum_{l=1}^{L} \rho_{\mathrm{ul}} \beta_{jil}. \tag{A.99}$$

For the power of the interference of other users, P_3^{mrc}, we write

$$P_3^{\mathrm{mrc}} = \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \mathbb{E} \left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{g}_{jkl} \right|^2 \right] \tag{A.100}$$

$$= \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \mathbb{E} \left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \hat{\boldsymbol{g}}_{jkl} \right|^2 \right] + \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \mathbb{E} \left[\left| \hat{\boldsymbol{g}}_{jij}^{\dagger} \boldsymbol{\epsilon}_{jkl} \right|^2 \right] \tag{A.101}$$

$$= \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \mathrm{tr} \left(\mathbb{E} \left[\hat{\boldsymbol{g}}_{jij} \hat{\boldsymbol{g}}_{jij}^{\dagger} \right] \mathbb{E} \left[\hat{\boldsymbol{g}}_{jkl} \hat{\boldsymbol{g}}_{jkl}^{\dagger} \right] \right) + \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \mathrm{tr} \left(\mathbb{E} \left[\hat{\boldsymbol{g}}_{jij} \hat{\boldsymbol{g}}_{jij}^{\dagger} \right] \mathbb{E} \left[\boldsymbol{\epsilon}_{jkl} \boldsymbol{\epsilon}_{jkl}^{\dagger} \right] \right) \tag{A.102}$$

$$= \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} M\rho_{\mathrm{ul}} \left(\frac{\rho_{\mathrm{p}} \beta_{jij}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right) \left(\frac{\rho_{\mathrm{p}} \beta_{jkl}^2}{1 + \rho_{\mathrm{p}} \sum_{l_2=1}^{L} \beta_{jkl_2}} \right)$$

$$+ \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} M\rho_{\mathrm{ul}} \left(\frac{\rho_{\mathrm{p}} \beta_{jij}^2}{1 + \rho_{\mathrm{p}} \sum_{l_1=1}^{L} \beta_{jil_1}} \right) \left(\beta_{jkl} - \frac{\rho_{\mathrm{p}} \beta_{jkl}^2}{1 + \rho_{\mathrm{p}} \sum_{l_2=1}^{L} \beta_{jkl_2}} \right)$$

$$= M\sqrt{\rho_{\mathrm{p}}} \beta_{jij} \alpha_{jij} \sum_{l=1}^{L} \sum_{k=1, k\neq i}^{K} \rho_{\mathrm{ul}} \beta_{jkl}. \tag{A.103}$$

Finally, for the power of the noise, P_4^{mrc}, we obtain

$$P_4^{\text{mrc}} = \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{n}_j\right|^2\right] \tag{A.104}$$

$$= \text{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{n}_j\boldsymbol{n}_j^{\dagger}\right]\right) \tag{A.105}$$

$$= M\frac{\rho_{\text{p}}\beta_{jij}^2}{1+\rho_{\text{p}}\sum_{l_1=1}^{L}\beta_{jil_1}} \tag{A.106}$$

$$= M\sqrt{\rho_{\text{p}}}\beta_{jij}\alpha_{jij}. \tag{A.107}$$

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound.

A.7 Proof of Theorem 5

We start by computing the variance of the desired signals, I_1^{zf}, in (3.22). We can easily verify that

$$I_1^{\text{zf}} = \sum_{l\in\Omega}\rho_{\text{ul}}\left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2. \tag{A.108}$$

Note that as explained earlier, all terms in the effective noise z''_{jij} in (3.22) are uncorrelated; thus

$$\text{var}\left[z''_{jij}\right] = I_2^{\text{zf}} + I_3^{\text{zf}}, \tag{A.109}$$

where I_2^{zf} is the variance of interference due to channel estimation error and I_3^{zf} is the variance of noise in (3.22). For I_2^{zf}, we obtain

$$I_2^{\text{zf}} = \sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\text{ul}}\mathbb{E}\left[\left|\boldsymbol{v}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jkl}\right|^2\right] \tag{A.110}$$

$$= \sum_{l=1}^{L}\sum_{k=1}^{K}\rho_{\text{ul}}\text{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jkl}\boldsymbol{\epsilon}_{jkl}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{v}_{jij}\boldsymbol{v}_{jij}^{\dagger}\right]\right) \tag{A.111}$$

$$= \sum_{l=1}^{L} \sum_{k=1}^{K} \rho_{\mathrm{ul}} \left(\beta_{jkl} - \sqrt{\rho_{\mathrm{p}}} \beta_{jkl} \alpha_{jkl} \right) \mathbb{E} \left[\boldsymbol{v}_{jij}^{\dagger} \boldsymbol{v}_{jij} \right] \tag{A.112}$$

$$= \sum_{l=1}^{L} \sum_{k=1}^{K} \rho_{\mathrm{ul}} \frac{\beta_{jkl} - \sqrt{\rho_{\mathrm{p}}} \beta_{jkl} \alpha_{jkl}}{(M-K) \sqrt{\rho_{\mathrm{p}}} \beta_{jij} \alpha_{jij}}, \tag{A.113}$$

where in the last step we have used the following standard result in random matrix theory [114]

$$\mathbb{E} \left[\boldsymbol{v}_{jij}^{\dagger} \boldsymbol{v}_{jij} \right] = \frac{1}{(M-K) \sqrt{\rho_{\mathrm{p}}} \beta_{jij} \alpha_{jij}}. \tag{A.114}$$

Finally, for the variance of noise, I_3^{zf}, we can write

$$I_3^{\mathrm{zf}} = \mathbb{E} \left[\left| \boldsymbol{v}_{jij}^{\dagger} \boldsymbol{n}_j \right|^2 \right] \tag{A.115}$$

$$= \mathrm{tr} \left(\mathbb{E} \left[\boldsymbol{v}_{jij} \boldsymbol{v}_{jij}^{\dagger} \right] \mathbb{E} \left[\boldsymbol{n}_j \boldsymbol{n}_j^{\dagger} \right] \right) \tag{A.116}$$

$$= \mathbb{E} \left[\boldsymbol{v}_{jij}^{\dagger} \boldsymbol{v}_{jij} \right] \tag{A.117}$$

$$= \frac{1}{(M-K) \sqrt{\rho_{\mathrm{p}}} \beta_{jij} \alpha_{jij}}. \tag{A.118}$$

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound. The proof is thus complete.

A.8 Proof of Corollary 2

Let P_1^{mrt} be the variance of the desired signals in (3.9). We start by computing P_1^{mrt} as follows

$$P_1^{\mathrm{mrt}} = \sum_{j \in \Omega}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \right) \left| \mathbb{E} \left[\boldsymbol{g}_{jil}^{\dagger} \hat{\boldsymbol{g}}_{jij} \right] \right|^2$$

$$
\begin{aligned}
&= \sum_{j\in\Omega}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right) \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right] + \mathbb{E}\left[\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2 \\
&\overset{\text{(a)}}{=} \sum_{j\in\Omega}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right) \left|\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2 \\
&\overset{\text{(b)}}{=} \sum_{j\in\Omega}^{L} \frac{\left(M\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}\right)^2 \rho_{\mathrm{dl}}\beta_{jil}^2}{\lambda_j^{\mathrm{mrt}}\beta_{jij}^2} \\
&= M^2 \sum_{j\in\Omega} \frac{\rho_{\mathrm{dl}}\rho_{\mathrm{p}}\beta_{jil}^2\alpha_{jij}^2}{\lambda_j^{\mathrm{mrt}}},
\end{aligned} \tag{A.119}
$$

where (a) is due to the fact that $\boldsymbol{\epsilon}_{jil}$ and $\hat{\boldsymbol{g}}_{jij}$ are independent and (b) follows from $\hat{\boldsymbol{g}}_{jil} = (\beta_{jil}/\beta_{jij})\hat{\boldsymbol{g}}_{jij}$ and the distribution of $\hat{\boldsymbol{g}}_{jkj}$ explained below (2.11). Since all three terms in the effective noise are uncorrelated, we have $\mathrm{var}\left[w'_{il}\right] = P_2^{\mathrm{mrt}} + P_3^{\mathrm{mrt}} + P_4^{\mathrm{mrt}}$, where P_2^{mrt} is the variance of interference due to beamforming gain uncertainty, P_3^{mrt} is the variance of interference caused by other users and P_4^{mrt} is the variance of noise in (3.9).

To compute P_2^{mrt}, we write

$$
\begin{aligned}
P_2^{\mathrm{mrt}} &= \sum_{j=1}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right) \mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij} - \mathbb{E}\left[\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2\right] \\
&= \sum_{j=1}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right) \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2\right] + \sum_{j=1}^{L} \left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right) \mathbb{E}\left[\left|\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right].
\end{aligned} \tag{A.120}
$$

For the first term in (A.120), it is obtained

$$
\begin{aligned}
\sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij} - \mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right]\right|^2\right] &= \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \mathrm{var}\left[\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right] \\
&= M \sum_{j=1}^{L} \frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}} \rho_{\mathrm{p}}\beta_{jil}^2\alpha_{jij}^2.
\end{aligned} \tag{A.121}
$$

Similarly, for the second term in (A.120), we obtain

$$\begin{aligned}\sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\left|\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right|^2\right] &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}^{\dagger}\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jij}\right] \\ &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\,\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right]\mathbb{E}\left[\hat{\boldsymbol{g}}_{jij}\hat{\boldsymbol{g}}_{jij}^{\dagger}\right]\right) \\ &= M\sum_{j=1}^{L}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right)\beta_{jil}\left(1-\sqrt{\rho_{\mathrm{p}}}\alpha_{jil}\right)\left(\sqrt{\rho_{\mathrm{p}}}\beta_{jij}\alpha_{jij}\right). \qquad \text{(A.122)}\end{aligned}$$

Adding up (A.121) and (A.122), P_2^{mrt} is thus obtained. Next, for P_3^{mrt} we write

$$\begin{aligned}P_3^{\mathrm{mrt}} &= \sum_{j=1}^{L}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right)\sum_{k=1,k\neq i}^{K}\mathbb{E}\left[\left|\boldsymbol{g}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right] \\ &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\sum_{k=1,k\neq i}^{K}\left(\mathbb{E}\left[\left|\hat{\boldsymbol{g}}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right]+\mathbb{E}\left[\left|\boldsymbol{\epsilon}_{jil}^{\dagger}\hat{\boldsymbol{g}}_{jkj}\right|^2\right]\right) \\ &= \sum_{j=1}^{L}\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\sum_{k=1,k\neq i}^{K}\left(\mathrm{tr}\left(\mathbb{E}\left[\hat{\boldsymbol{g}}_{jil}\hat{\boldsymbol{g}}_{jil}^{\dagger}\right]\mathbb{E}\left[\hat{\boldsymbol{g}}_{jkj}\hat{\boldsymbol{g}}_{jkj}^{\dagger}\right]\right)+\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right]\mathbb{E}\left[\hat{\boldsymbol{g}}_{jkj}\hat{\boldsymbol{g}}_{jkj}^{\dagger}\right]\right)\right) \\ &= \sum_{j=1}^{L}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right)\sum_{k=1,k\neq i}^{K}\left(M\left(\sqrt{\rho_{\mathrm{p}}}\beta_{jil}\alpha_{jil}\right)\left(\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\right)\right. \\ &\quad \left.+M\left(\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\right)\left(\beta_{jil}-\sqrt{\rho_{\mathrm{p}}}\beta_{jil}\alpha_{jil}\right)\right) \\ &= M\sum_{j=1}^{L}\left(\frac{\rho_{\mathrm{dl}}}{\lambda_j^{\mathrm{mrt}}}\right)\sum_{k=1,k\neq i}^{K}\sqrt{\rho_{\mathrm{p}}}\beta_{jkj}\alpha_{jkj}\beta_{jil}.\end{aligned}$$

Finally, the power of the noise w_{il} is $P_4^{\mathrm{mrt}}=1$.

Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound.

A.9 Proof of Theorem 6

We start by computing the variance of the desired signal, i.e., $\sum_{j\in\Omega}\sqrt{\rho_{\rm dl}/\lambda_j^{\rm zf}}\,(\beta_{jil}/\beta_{jij})\,s_j[i]$. It is readily verified that

$$\text{var}\,[\,\text{Desired signal}\,]=\sum_{j\in\Omega}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\left(\frac{\beta_{jil}}{\beta_{jij}}\right)^2. \tag{A.123}$$

For the variance of the effective noise, w_{il}'', we have

$$\text{var}\,[w_{il}''] \overset{(a)}{=} \text{var}\,[\,\text{Interference due to estimation error}\,]+\text{var}\,[w_{il}]\,, \tag{A.124}$$

where (a) is due to the fact that w_{il} is uncorrelated from the interference caused by the estimation error. To compute the first variance we can write

$$\begin{aligned}
&\text{var}\,[\,\text{Interference due to estimation error}\,]\\
&=\sum_{j=1}^{L}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\sum_{k=1}^{K}\mathbb{E}\left[\left|\boldsymbol{\epsilon}_{jil}^{\dagger}\boldsymbol{v}_{jkj}\right|^2\right] &\text{(A.125)}\\
&=\sum_{j=1}^{L}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\sum_{k=1}^{K}\mathbb{E}\left[\left(\boldsymbol{\epsilon}_{jil}^{\dagger}\boldsymbol{v}_{jkj}\right)\left(\boldsymbol{v}_{jkj}^{\dagger}\boldsymbol{\epsilon}_{jil}\right)\right] &\text{(A.126)}\\
&=\sum_{j=1}^{L}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\sum_{k=1}^{K}\mathbb{E}\left[\text{tr}\left(\boldsymbol{\epsilon}_{jil}^{\dagger}\boldsymbol{v}_{jkj}\boldsymbol{v}_{jkj}^{\dagger}\boldsymbol{\epsilon}_{jil}\right)\right] &\text{(A.127)}\\
&=\sum_{j=1}^{L}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\sum_{k=1}^{K}\text{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jil}\boldsymbol{\epsilon}_{jil}^{\dagger}\right]\mathbb{E}\left[\boldsymbol{v}_{jkj}\boldsymbol{v}_{jkj}^{\dagger}\right]\right) &\text{(A.128)}\\
&=\sum_{j=1}^{L}\left(\frac{\rho_{\rm dl}}{\lambda_j^{\rm zf}}\right)\sum_{k=1}^{K}\left(\beta_{jil}-\sqrt{\rho_{\rm p}}\beta_{jil}\alpha_{jil}\right)\frac{1}{(M-K)\sqrt{\rho_{\rm p}}\beta_{jkj}\alpha_{jkj}}, &\text{(A.129)}
\end{aligned}$$

where in the last step we have used the following standard result in random matrix theory [114]

$$\mathbb{E}\left[\boldsymbol{v}_{jkj}^{\dagger}\boldsymbol{v}_{jkj}\right]=\frac{1}{(M-K)\sqrt{\rho_{\rm p}}\beta_{jkj}\alpha_{jkj}}. \tag{A.130}$$

Lastly, for the variance of the zero-mean noise, we have $\mathbb{E}\left[|w_{il}|^2\right] = 1$. Therefore, since the effective noise is zero-mean and uncorrelated from the desired signals, one can directly apply Lemma 2 in conjunction with the computed variances above to obtain the required lower bound. This completes the proof of the lower bound in (3.32).

A.10 Proof of Corollary 3

Using (3.52)-(3.58), it can be verified that in case (i) we have $R_{\text{Sym}}^{\text{SD},1} = I(\hat{y}_{1i}; x_2[i]|x_1[i])$, $R_{\text{Sym}}^{\text{S-SND},1} = \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$, and $R_{\text{Sym}}^{\text{SND},1} = R_{\text{Sym}}^{\text{TIN},1} = I(\hat{y}_{1i}; x_1[i])$.

Furthermore, using (3.62), one can rewrite (3.59) as $I(\hat{y}_{1i}; x_2[i]|x_1[i]) < \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$, and conclude that $R_{\text{Sym}}^{\text{SD},1} < R_{\text{Sym}}^{\text{S-SND},1}$. Moreover, using (3.62), one can rewrite (3.59) as $\frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i]) < I(\hat{y}_{1i}; x_1[i])$, and conclude that $R_{\text{Sym}}^{\text{S-SND},1} < R_{\text{Sym}}^{\text{SND},1} = R_{\text{Sym}}^{\text{TIN},1}$. This is illustrated in Fig. 3.3b. Similarly, using (3.52)-(3.58), it can be verified that in case (ii) we have $R_{\text{Sym}}^{\text{TIN},1} = I(\hat{y}_{1i}; x_1[i])$, and $R_{\text{Sym}}^{\text{SD},1} = R_{\text{Sym}}^{\text{SND},1} = R_{\text{Sym}}^{\text{S-SND},1} = \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$. Also, using (3.62)-(3.63), one can rewrite (3.60) as $\max\{I(\hat{y}_{1i}; x_1[i]), I(\hat{y}_{1i}; x_2[i])\} \leq \frac{1}{2}I(\hat{y}_{1i}; x_1[i], x_2[i])$. Hence, when this condition holds it yields (3.67), which is also illustrated in Fig. 3.3c.

A.11 An achievable region for the generalized RS scheme

Here, to establish an achievable region, following a technique used in [72] we provide analysis of the probability of error for the proposed generalization of the RS scheme when applied to the case of $L = 2$. First, note that after dropping the index i, the cloud center and the satellite codeword generated at BS l, $l = 1, 2$, are given by $\boldsymbol{s}_l^{(b)}(m_l^{(b)})$, and $\boldsymbol{s}_l(m_l^{(a)}, m_l^{(b)})$, respectively. We only show the achievability proof at receiver 1, i.e., user of cell 1, as a similar analysis can be applied at receiver 2, i.e., user of cell 2.

Receiver 1 tries to *uniquely* recover both parts of its intended signal's message,

$(m_1^{(a)}, m_1^{(b)})$ and to *non-uniquely* recover messages from each layer of the interfering signal, $(m_2^{(a)}, m_2^{(b)})$. Therefore, receiver 1 finds the unique pair $(\hat{m}_1^{(a)}, \hat{m}_1^{(b)})$ such that

$$\left(\boldsymbol{s}_1^{(b)}(\hat{m}_1^{(b)}), \boldsymbol{s}_1(\hat{m}_1^{(a)}, \hat{m}_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } (m_2^{(a)}, m_2^{(b)}), \tag{A.131}$$

where $\mathcal{T}_\epsilon^n$ is the set of ϵ-typical n-sequences (see [115, Section 2.4] for formal description of typical sets).

Assume without loss of generality that the message pairs $(m_1^{(a)}, m_1^{(b)}) = (1,1)$ and $(m_2^{(a)}, m_2^{(b)}) = (1,1)$ are sent. Receiver 1 declares an error if one or both of the following error events happen:

$$E_1 = \left\{\left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(1,1), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(1,1), \boldsymbol{y}_1^{\text{dl}}\right) \notin \mathcal{T}_\epsilon^n\right\}, \tag{A.132}$$

$$E_2 = \left\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \right. \tag{A.133}$$
$$\left. \text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } (m_2^{(a)}, m_2^{(b)})\right\}.$$

By the law of large numbers, $P(E_1) \to 0$, as $n \to \infty$. We bound $P(E_2)$ in three different ways. As in [72], note that the joint typicality of the tuple $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right)$ implies that $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n$, i.e., the triple $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right)$ is jointly typical. Hence,

$$E_2 \subseteq \left\{\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1)\right\} = E_{21}. \tag{A.134}$$

The event E_{21} can be partitioned into the following 3 events:

$$E_{21}^{(1)} = \left\{\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(b)} \neq 1\right\}, \tag{A.135}$$

$$E_{21}^{(2)} = \left\{\left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(a)} \neq 1\right\}, \tag{A.136}$$

$$E_{21}^{(3)} = \left\{\left(\boldsymbol{s}_1^{(b)}(m_1(b)), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{y}_1^{\text{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1)\right\}, \tag{A.137}$$

leading to the following

$$P(E_{21}) \leq P(E_{21}^{(1)}) + P(E_{21}^{(2)}) + P(E_{21}^{(3)}). \tag{A.138}$$

By the packing lemma [115, Section 3.2], $P(E_{21}^{(1)})$, $P(E_{21}^{(2)})$ and $P(E_{21}^{(3)})$ tend to zero, as $n \to \infty$, if the following constraints are satisfied

$$R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}}\right) \tag{A.139}$$

$$R_1^{\mathrm{dl,(a)}} \leq I\left(s_1; y_1^{\mathrm{dl}} | s_1^{(b)}\right) \tag{A.140}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}}\right) \tag{A.141}$$

Notice that due to the codewords construction, the r.h.s in (A.139) and (A.141) are identical, however the former is not necessary since the latter is the tighter condition. Therefore, we are left only with two rate constraints, (A.140) and (A.141). Further note that the special structure of the codewords yields

$$I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}}\right) = I\left(s_1; y_1^{\mathrm{dl}}\right). \tag{A.142}$$

In addition, note that the joint typicality of the tuple $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right)$ implies that $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n$, i.e., the quadruple $\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right)$ is jointly typical. Consequently,

$$\begin{aligned} E_2 \subseteq \Big\{ &\left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ &\text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } m_2^{(b)} \Big\} = E_{22} \end{aligned} \tag{A.143}$$

The event E_{22} can be partitioned into the following 6 events:

$$E_{22}^{(1)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(b)} \neq 1 \right\}, \tag{A.144}$$

$$E_{22}^{(2)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(a)} \neq 1 \right\} \tag{A.145}$$

$$E_{22}^{(3)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{y}_1^{\mathrm{dl}} \right) \in \mathcal{T}_\epsilon^n, \text{ for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1) \right\} \tag{A.146}$$

$$E_{22}^{(4)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}} \right) \in \mathcal{T}_\epsilon^n, \text{for some } m_1^{(b)} \neq 1, \text{and some } m_2^{(b)} \neq 1 \right\} \tag{A.147}$$

$$E_{22}^{(5)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}} \right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(a)} \neq 1, \text{ and some } m_2^{(b)} \neq 1 \right\} \tag{A.148}$$

$$\begin{aligned} E_{22}^{(6)} = \Big\{ & \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}} \right) \in \mathcal{T}_\epsilon^n, \\ & \text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } m_2^{(b)} \neq 1 \Big\}, \end{aligned} \tag{A.149}$$

leading to the following

$$P(E_{22}) \leq P(E_{22}^{(1)}) + P(E_{22}^{(2)}) + P(E_{22}^{(3)}) + P(E_{22}^{(4)}) + P(E_{22}^{(5)}) + P(E_{22}^{(6)}). \tag{A.150}$$

It can be verified that by the packing lemma, the probabilities $P(E_{22}^{(1)})$, $P(E_{22}^{(2)})$, $P(E_{22}^{(3)})$, $P(E_{22}^{(4)})$, $P(E_{22}^{(5)})$, and $P(E_{22}^{(6)})$ all tend to zero, as $n \to \infty$, if the following constraints are satisfied

$$R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2^{(b)} \right) \tag{A.151}$$

$$R_1^{\mathrm{dl,(a)}} \leq I\left(s_1; y_1^{\mathrm{dl}} | s_1^{(b)}, s_2^{(b)} \right) \tag{A.152}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2^{(b)} \right) \tag{A.153}$$

$$R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2^{(b)}; y_1^{\mathrm{dl}} \right) \tag{A.154}$$

$$R_1^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_2^{(b)}; y_1^{\mathrm{dl}} | s_1^{(b)} \right) \tag{A.155}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2^{(b)}; y_1^{\mathrm{dl}} \right). \tag{A.156}$$

Notice that due to the codewords construction, the r.h.s of (A.151) and (A.153) are identical, however the latter is the tighter condition and thus the former can be omitted. Similarly, it is verified that (A.154) is not necessary, since the constraint of (A.156) is

the tighter condition. As such, by removing (A.151) and (A.154), we are left with only four necessary constraints, i.e., (A.152), (A.153), (A.155) and (A.156). Also, due to the structure of the codewords, we have

$$I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2^{(b)}\right) = I\left(s_1; y_1^{\mathrm{dl}} | s_2^{(b)}\right) \tag{A.157}$$

$$I\left(s_1, s_1^{(b)}, s_2^{(b)}; y_1^{\mathrm{dl}}\right) = I\left(s_1, s_2^{(b)}; y_1^{\mathrm{dl}}\right). \tag{A.158}$$

Lastly, the third way to bound $P(E_2)$ is to partition E_2 into the following 12 events:

$$E_2^{(1)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(1,1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(b)} \neq 1 \right\}, \tag{A.159}$$

$$E_2^{(2)} = \left\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(1,1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \text{ for some } m_1^{(a)} \neq 1 \right\}, \tag{A.160}$$

$$E_2^{(3)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(1,1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1) \Big\}, \tag{A.161}$$

$$E_2^{(4)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(1, m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } m_1^{(b)} \neq 1, \text{ and some } m_2^{(b)} \neq 1 \Big\}, \tag{A.162}$$

$$E_2^{(5)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(1, m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } m_1^{(a)} \neq 1, \text{ and some } m_2^{(b)} \neq 1 \Big\}, \tag{A.163}$$

$$E_2^{(6)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(1, m_2^{(b)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } m_2^{(b)} \neq 1 \Big\}, \tag{A.164}$$

$$E_2^{(7)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(m_2^{(a)}, 1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } m_1^{(b)} \neq 1, \text{ and some } m_2^{(a)} \neq 1 \Big\}, \tag{A.165}$$

$$E_2^{(8)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(m_2^{(a)}, 1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \\ \text{for some } m_1^{(a)} \neq 1, \text{ and some } m_2^{(a)} \neq 1 \Big\}, \tag{A.166}$$

$$E_2^{(9)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(1), \boldsymbol{s}_2(m_2^{(a)}, 1), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \tag{A.167}$$

$$\text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } m_2^{(a)} \neq 1\Big\},$$

$$E_2^{(10)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(1, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(a)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \tag{A.168}$$
$$\text{for some } m_1^{(b)} \neq 1, \text{ and some } (m_2^{(a)}, m_2^{(b)}) \neq (1,1)\Big\},$$

$$E_2^{(11)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(1), \boldsymbol{s}_1(m_1^{(a)}, 1), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(a)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \tag{A.169}$$
$$\text{for some } m_1^{(a)} \neq 1, \text{ and some } (m_2^{(a)}, m_2^{(b)}) \neq (1,1)\Big\},$$

$$E_2^{(12)} = \Big\{ \left(\boldsymbol{s}_1^{(b)}(m_1^{(b)}), \boldsymbol{s}_1(m_1^{(a)}, m_1^{(b)}), \boldsymbol{s}_2^{(b)}(m_2^{(b)}), \boldsymbol{s}_2(m_2^{(a)}, m_2^{(a)}), \boldsymbol{y}_1^{\mathrm{dl}}\right) \in \mathcal{T}_\epsilon^n, \tag{A.170}$$
$$\text{for some } (m_1^{(a)}, m_1^{(b)}) \neq (1,1), \text{ and some } (m_2^{(a)}, m_2^{(b)}) \neq (1,1)\Big\},$$

leading to the following

$$\begin{aligned} P(E_2) \leq & P(E_2^{(1)}) + P(E_2^{(2)}) + P(E_2^{(3)}) + P(E_2^{(4)}) + P(E_2^{(5)}) + P(E_2^{(6)}) \\ & + P(E_2^{(7)}) + P(E_2^{(8)}) + P(E_2^{(9)}) + P(E_2^{(10)}) + P(E_2^{(11)}) + P(E_2^{(12)}). \end{aligned} \tag{A.171}$$

Using the packing lemma, it can be shown that the probabilities $P(E_2^{(1)})$ through $P(E_2^{(12)})$ above tend to zero, as $n \to \infty$, if the following conditions are satisfied

$$R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2, s_2^{(b)}\right) \tag{A.172}$$

$$R_1^{\mathrm{dl,(a)}} \leq I\left(s_1; y_1^{\mathrm{dl}} | s_1^{(b)}, s_2, s_2^{(b)}\right) \tag{A.173}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2, s_2^{(b)}\right) \tag{A.174}$$

$$R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2, s_2^{(b)}; y_1^{\mathrm{dl}}\right) \tag{A.175}$$

$$R_1^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_2, s_2^{(b)}; y_1^{\mathrm{dl}} | s_1^{(b)}\right) \tag{A.176}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2, s_2^{(b)}; y_1^{\mathrm{dl}}\right) \tag{A.177}$$

$$R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(a)}} \leq I\left(s_1, s_1^{(b)}, s_2; y_1^{\mathrm{dl}} | s_2^{(b)}\right) \tag{A.178}$$

$$R_1^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(a)}} \leq I\left(s_1, s_2; y_1^{\mathrm{dl}} | s_1^{(b)}, s_2^{(b)}\right) \tag{A.179}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(a)}} \leq I\left(s_1, s_1^{(b)}, s_2; y_1^{\mathrm{dl}} | s_2^{(b)}\right) \tag{A.180}$$

$$R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2, s_2^{(b)}; y_1^{\mathrm{dl}}\right) \tag{A.181}$$

$$R_1^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_2, s_2^{(b)}; y_1^{\mathrm{dl}} | s_1^{(b)}\right) \tag{A.182}$$

$$R_1^{\mathrm{dl,(a)}} + R_1^{\mathrm{dl,(b)}} + R_2^{\mathrm{dl,(a)}} + R_2^{\mathrm{dl,(b)}} \leq I\left(s_1, s_1^{(b)}, s_2, s_2^{(b)}; y_1^{\mathrm{dl}}\right). \tag{A.183}$$

It can be readily seen that the constraints of (A.172), (A.175), (A.176), (A.177), (A.178) and (A.181) are redundant and can be removed. We are thus left with only 6 necessary constraints, i.e., (A.173), (A.174), (A.179), (A.180), (A.182) and (A.183). Also, the code construction yields the following

$$I\left(s_1, s_1^{(b)}; y_1^{\mathrm{dl}} | s_2, s_2^{(b)}\right) = I\left(s_1; y_1^{\mathrm{dl}} | s_2\right) \tag{A.184}$$

$$I\left(s_1; y_1^{\mathrm{dl}} | s_1^{(b)}, s_2, s_2^{(b)}\right) = I\left(s_1; y_1^{\mathrm{dl}} | s_1^{(b)}, s_2\right) \tag{A.185}$$

$$I\left(s_1, s_1^{(b)}, s_2, s_2^{(b)}; y_1^{\mathrm{dl}}\right) = I\left(s_1, s_2; y_1^{\mathrm{dl}}\right) \tag{A.186}$$

$$I\left(s_1, s_2, s_2^{(b)}; y_1^{\mathrm{dl}} | s_1^{(b)}\right) = I\left(s_1, s_2; y_1^{\mathrm{dl}} | s_1^{(b)}\right) \tag{A.187}$$

$$I\left(s_1, s_1^{(b)}, s_2; y_1^{\mathrm{dl}} | s_2^{(b)}\right) = I\left(s_1, s_2; y_1^{\mathrm{dl}} | s_2^{(b)}\right). \tag{A.188}$$

Hence, by bounding $P(E_2)$ using these three different approaches, an achievable region is established at receiver 1 (denoted by $\mathcal{R}_1^{\mathrm{RS}}$), which is the union of the three regions described above. One can similarly obtain the achievable region at receiver 2 (denoted by $\mathcal{R}_2^{\mathrm{RS}}$) by replacing y_1^{dl} with y_2^{dl} and swapping appropriate indices. The network-wide achievable region obtained by the generalized RS scheme in conjunction with *non-unique* decoding for a two-cell system can then be written in the following form

$$\mathscr{R}^{\mathrm{RS}} = \bigcap_{l} \mathcal{R}_l^{\mathrm{RS}}, \quad l = 1, 2, \tag{A.189}$$

where

$$\mathcal{R}_l^{\mathrm{RS}} = \bigcup_{\Omega_l \in \mathcal{S}_l} \mathcal{R}_{\mathrm{MAC}(\Omega_l, l)}^{\mathrm{RS}}, \quad l = 1, 2, \tag{A.190}$$

where $\mathcal{R}_{\mathrm{MAC}(\Omega_l, l)}^{\mathrm{RS}}$ is a modified MAC region, which has less than $2^{|\Omega_l|} - 1$ rate constraints,

as some of the constraints are removed from the regular MAC region (due to the codewords construction as explained above), and $\mathcal{S}_l, l = 1, 2$, are given by

$$\mathcal{S}_1 = \left\{ \left\{ m_1^{(a)}, m_1^{(b)} \right\}, \left\{ m_1^{(a)}, m_1^{(b)}, m_2^{(b)} \right\}, \left\{ m_1^{(a)}, m_1^{(b)}, m_2^{(a)}, m_2^{(b)} \right\} \right\} \tag{A.191}$$

$$\mathcal{S}_2 = \left\{ \left\{ m_2^{(a)}, m_2^{(b)} \right\}, \left\{ m_2^{(a)}, m_2^{(b)}, m_1^{(b)} \right\}, \left\{ m_2^{(a)}, m_2^{(b)}, m_1^{(a)}, m_1^{(b)} \right\} \right\}. \tag{A.192}$$

Specifically, if Ω_l contains messages of both layers $(m_j^{(a)}, m_j^{(b)})$, for some j, then those constraints that involve $R_j^{\mathrm{dl,(b)}}$ but not $R_j^{\mathrm{dl,(a)}}$ are not needed and will thus be removed from the rate region. In particular, the constraints that are removed from each of the three regions described above are as follows: (A.139) from the first region, (A.151) and (A.154) from the second region, (A.172), (A.175), (A.176), (A.177), (A.178) and (A.181) from the third region.

One can also re-write $\mathcal{S}_l$ as follows

$$\mathcal{S}_l = \left\{ \left\{ m_l^{(a)}, m_l^{(b)} \right\} \right\} \times \left\{ \emptyset, \left\{ m_j^{(b)} \right\}, \left\{ m_j^{(a)}, m_j^{(b)} \right\} \right\}, \quad l = 1, 2, \quad j \neq l. \tag{A.193}$$