MAXIMIZATION OF SPECTRAL EFFICIENCY AND ENERGY EFFICIENCY IN MASSIVE MIMO

A Project report submitted in partial fulfilment of the requirements for

the award of the degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

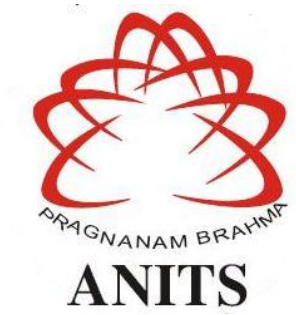
D.V.S. Kalyan (318126512016) P. Mohan Manoj (318126512038)

P. Madhusudhan (318126512041)

Under the guidance of

Dr. B. Somasekhar

Associate Professor



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

ANIL NEERUKONDA INSTITUTE OF TECHNOLOGY AND SCIENCES

(UGC AUTONOMOUS)

(*Permanently Affiliated to AU, Approved by AICTE and Accredited by NBA & NAAC with 'A' Grade*)

Sangivalasa, Bheemili Mandal, Visakhapatnam Dist. (A.P)

2021-2022

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

ANIL NEERUKONDA INSTITUTE OF TECHNOLOGY AND SCIENCES
(UGC AUTONOMOUS)
(Permanently Affiliated to AU, Approved by AICTE and Accredited by NBA & NAAC with 'A' Grade)
Sangivalasa, Bheemili Mandal, Visakhapatnam Dist. (A.P)

ANITS

CERTIFICATE

This is to certify that the project report entitled **"MAXIMIZATION OF SPECTRAL EFFICIENCY AND ENERGY EFFICIENCY IN MASSIVE MIMO"** submitted by **D.V.S. Kalyan (318126512016), P. Mohan Manoj (318126512038), P. Madhusudhan (318126512041)** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Electronics & Communication Engineering** of Andhra University, Visakhapatnam is a record of bonafide work carried out under my guidance and supervision.

Project Guide
Dr. B. Somasekhar

Associate Professor
Department of E.C.E
ANITS

Associate Professor
Department of E.C.E.
Anil Neerukonda
Institute of Technology & Sciences
Sangivalasa, Visakhapatnam-531 162

Head of the Department

Dr. V.Rajyalakshmi
Professor&HOD
Department of E.C.E
ANITS

Head of the Department
Department of E C E
Anil Neerukonda Institute of Technology & Sciences
Sangivalasa-531 162

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our project guide **Dr. B. Somasekhar** Associate professor, Department of Electronics and Communication Engineering, ANITS, for his guidance with unsurpassed knowledge and immense encouragement. We are grateful to **Dr. V. Rajyalakshmi**, Head of the Department, Electronics and Communication Engineering, for providing us with the required facilities for the completion of the project work.

We are very much thankful to the **Principal and Management, ANITS, Sangivalasa,** for their encouragement and cooperation to carry out this work.

We express our thanks to all **teaching faculty** of Department of ECE, whose suggestions during reviews helped us in accomplishment of our project. We would like to thank **all non-teaching staff** of the Department of ECE, ANITS for providing great assistance in accomplishment of our project.

We would like to thank our parents, friends, and classmates for their encouragement throughout our project period. At last, but not the least, we thank everyone for supporting us directly or indirectly in completing this project successfully.

PROJECT STUDENTS

D.V.S. Kalyan (318126512016)

P. Mohan Manoj (318126512038)

P. Madhusudhan (318126512041)

ABSTRACT

Massive MIMO is a promising technique to increase the spectral efficiency (SE) of cellular networks, by deploying antenna arrays with hundreds or thousands of active elements at the base stations and performing coherent transceiver processing. A common rule-of-thumb is that these systems should have an order of magnitude more antennas, M, than scheduled users, K, because the user's channels are likely to be near-orthogonal when M/K > 10. Here we analyse how the optimal number of scheduled users, K*, depends on M and other system parameters. To this end, new SE expressions are derived to enable efficient system-level analysis with power control, arbitrary pilot reuse, and random user locations. The value of K* in the large-M regime is derived in closed form, while simulations are used to show what happens at finite M, in different interference scenarios, with different pilot reuse factors, and for different processing schemes.

To improve the cellular energy efficiency, without sacrificing quality-of-service (QoS) at the users, the network topology must be densified to enable higher spatial reuse. We analyse a combination of two densification approaches, namely "massive" multiple-input multiple-output (MIMO) base stations and small-cell access points. If the latter are operator-deployed, a spatial soft-cell approach can be taken where the multiple transmitters serve the users by joint non-coherent multiflow beamforming. We minimize the total power consumption (both dynamic emitted power and static hardware power) while satisfying QoS constraints. Furthermore, we provide promising simulation results showing how the total power consumption can be greatly improved by combining massive MIMO and small cells; this is possible with both optimal and low-complexity beamforming.

CONTENTS

LIST OF SYMBOLS

B	Bandwidth
D	Average cell density
f	Frequency
Bc	Coherence bandwidth
Tc	Coherence time
τc	Complex-valued samples
M	Antennas at base station
K	User equipments in a cell
S	Transmission Symbols per Frame
w	Precoding vector
I	Interference term
f	Pilot reuse factor
μ	Propagation parameter

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

MIMO	Multiple Input Multiple Output
BS	Base Station
UE	User Equipment
DL	Downlink
UL	Uplink
SE	Spectral Efficiency
EE	Energy Efficiency
OFDM	orthogonal frequency-division multiplexing
FDD	Frequency Division Duplexing
TDD	Time Division Duplexing
SNR	Signal to Noise Ratio
ICI	Inter Carrier Interference
LS	Least-Square
MMSE	Minimum-Mean-Square-Error
MRT	Maximum Ratio Transmission
ZF	Zero Forcing
RZF	Regularized Zero Forcing
PC	Power Consumption
CP	Circuit Power
ETP	Effective Transmit Power

CHAPTER 1
INTRODUCTION

Wireless communication technology has fundamentally changed the way we communicate. The time when telephones, computers, and Internet connections were bound to be wired, and only used at predefined locations, has passed. These communications services are nowadays wirelessly accessible almost everywhere on Earth, thanks to the deployment of cellular wide area networks (e.g., based on the GSM1, UMTS2, and LTE3 standards), local area networks (based on different versions of the Wi-Fi standard IEEE 802.11), and satellite services. Wireless connectivity has become an essential part of the society—as vital as electricity—and as such the technology itself spurs new applications and services. We have already witnessed the streaming media revolution, where music and video are delivered on demand over the Internet. The first steps towards a fully networked society with augmented reality applications, connected homes and cars, and machine-to-machine communications have also been taken. Looking 15 years into the future, we will find new innovative wireless services that we cannot predict today.

The amount of wireless voice and data communications has grown at an exponential pace for many decades. This trend is referred to as Cooper's law because the wireless researcher Martin Cooper [1] noticed in the 1990s that the number of voice and data connections has doubled every two-and-a-half year, since Guglielmo Marconi's first wireless transmissions in 1895. This corresponds to a 32% annual growth rate. Looking ahead, the Ericsson Mobility Report forecasts a compound annual growth rate of 42% in mobile data traffic from 2016 to 2022 [2], which is even faster than Cooper's law. The demand for wireless data connectivity will definitely continue to increase in the foreseeable future; for example, since the video fidelity is constantly growing, since new must-have services are likely to arise, and because we are moving into a networked society, where all electronic devices connect to the Internet. An important question is how to evolve the current wireless communications technologies to meet the continuously increasing demand, and thereby avoid an imminent data traffic

crunch. An equally important question is how to satisfy the rising expectations of service quality. Customers will expect the wireless services to work equally well anywhere and at any time, just as they expect the electricity grid to be robust and constantly available. To keep up with an exponential traffic growth rate and simultaneously provide ubiquitous connectivity, industrial and academic researchers need to turn every stone to design new revolutionary wireless network technologies. This monograph explains what the Massive multiple-input multiple-output (MIMO) technology is and why it is a promising solution to handle several orders-of-magnitude4 more wireless data traffic than today's technologies.

1.1 Cellular Networks

Wireless communication is based on radio, meaning that electromagnetic (EM) waves are designed to carry information from a transmitter to one or multiple receivers. Since the EM waves propagate in all possible directions from the transmitter, the signal energy spreads out and less energy reaches a desired receiver as the distance increases. To deliver wireless services with sufficiently high received signal energy over wide coverage areas, researchers at Bell Labs postulated in 1947 that a cellular network topology is needed [3]. According to this idea, the coverage area is divided into cells that operate individually using a fixed-location base station; that is, a piece of network equipment that facilitates wireless communication between a device and the network. The cellular concept was further developed and analysed over the subsequent decades and later deployed in practice. Without any doubt, the cellular concept was a major breakthrough and has been the main driver to deliver wireless services in the last forty years (since the "first generation" of mobile phone systems emerged in the 1980s).

Definition 1.1: (Cellular networks). A cellular network consists of a set of base stations (BSs) and a set of user equipments (UEs). Each UE is connected to one of the BSs, which provides service to it. The downlink (DL) refers to signals sent from the BSs to their respective UEs, while the uplink (UL) refers to transmissions from the UEs to their respective BSs.

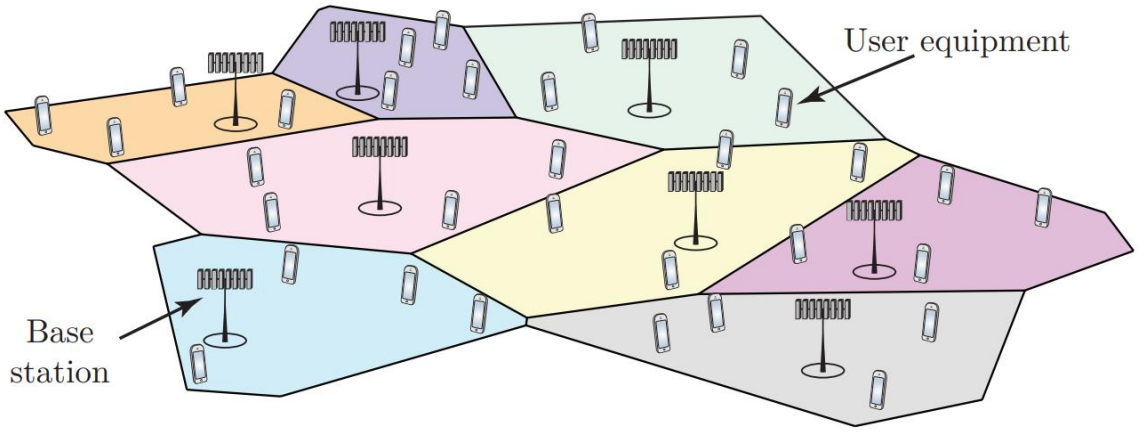


Figure 1.1: A basic cellular network, where each BS covers a distinct geographical area and provides service to all UEs in it. The area is called a "cell" and is illustrated with a distinct colour.

Cellular networks were originally designed for wireless voice communications, but it is wireless data transmissions that dominate nowadays [4]. Video on-demand accounts for the majority of traffic in wireless networks and is also the main driver of the predicted increase in traffic demand [5]. The area throughput is thus a highly relevant performance metric of contemporary and future cellular networks. It is measured in bit/s/km^2 and can be modelled using the following high-level formula:

Area throughput [bit/s/km^2] = B [Hz] · D [cells/km^2] · SE [bit/s/Hz/cell] (1.1)

where B is the bandwidth, D is the average cell density, and Spectral Efficiency (SE) is the SE per cell. The SE is the amount of information that can be transferred per second over one Hz of bandwidth.

These are the three main components that determine the area throughput, and that need to be increased in order to achieve higher area throughput in future cellular networks. Based on (1.1), one can think of the area throughput as being the volume of a rectangular box with sides B, D, and SE; There is an inherent dependence between these three components in the sense that the choice of frequency band and cell density affects the propagation conditions; for example, the probability of having a line-of-sight (LoS) channel between the transmitter and receiver (and between out-of-cell interferers and the receiver), the average propagation losses, etc. However, one can treat these three

components as independent as a first-order approximation to gain basic insights. Consequently, there are three main ways to improve the area throughput of cellular networks:

1. Allocate more bandwidth;
2. Densify the network by deploying more BSs;
3. Improve the SE per cell.

1.2 Evolving Cellular Networks for Higher Area Throughput

Suppose, for the matter of argument, that we want to design a new cellular network that improves the area throughput by a factor of 1000 over existing networks; that is, to solve "the 1000× data challenge" posed by Qualcomm [6]. Note that such a network can handle the three orders-of-magnitude increase in wireless data traffic that will occur over the next 15–20 years, if the annual traffic growth rate continues to be in the range of 41%–59%. How can we handle such a huge traffic growth according to the formula in (1.1)?

One potential solution would be to increase the bandwidth B by 1000×. Current cellular networks utilize collectively more than 1 GHz of bandwidth in the frequency range below 6 GHz. This is physically impractical since the frequency spectrum is a global resource that is shared among many different services, and also because it entails using much higher frequency bands than in the past, which physically limits the range and service reliability. There are, however, substantial bandwidths in the millimetre wavelength (mm Wave) bands (e.g., in the range 30–300 GHz) that can be used for short-range applications.

Higher cell density and larger bandwidth have historically dominated the evolution of the coverage tier, which explains why we are approaching a saturation point where further improvements are increasingly complicated and expensive. However, it might be possible to dramatically improve the SE of future cellular networks.

Mobile wireless communication system has gone through several evolution stages in the past few decades after the introduction of the first-generation mobile network in

early 1980s. Due to huge demand for more connections worldwide, mobile communication standards advanced rapidly to support more users. Let's take a look on the evolution stages of wireless technologies for mobile communication.

1.2.1 1G – First generation mobile communication system

The first generation of mobile network was deployed in Japan by Nippon Telephone and Telegraph company (NTT) in Tokyo during 1979. In the beginning of 1980s, it gained popularity in the US, Finland, UK and Europe. This system used analogue signals and it had many disadvantages due to technology limitations.

Key features (technology) of 1G system

- Frequency 800 MHz and 900 MHz
- Bandwidth: 10 MHz (666 duplex channels with bandwidth of 30 KHz)
- Technology: Analogue switching
- Modulation: Frequency Modulation (FM)
- Mode of service: voice only
- Access technique: Frequency Division Multiple Access (FDMA)

Disadvantages of 1G system

- Poor voice quality due to interference
- Poor battery life
- Large sized mobile phones (not convenient to carry)
- Less security (calls could be decoded using an FM demodulator)
- Limited number of users and cell coverage
- Roaming was not possible between similar systems

1.2.2 2G – Second generation communication system GSM

Second generation of mobile communication system introduced a new digital technology for wireless transmission also known as Global System for Mobile communication (GSM). GSM technology became the base standard for further development in wireless standards later. This standard was capable of supporting up to 14.4 to 64kbps (maximum) data rate which is sufficient for SMS and email services.

Code Division Multiple Access (CDMA) system developed by Qualcomm also introduced and implemented in the mid-1990s. CDMA has more features than GSM in terms of spectral efficiency, number of users and data rate.

Key features of 2G system

- Digital system (switching)
- SMS services is possible
- Roaming is possible
- Enhanced security
- Encrypted voice transmission
- First internet at lower data rate

Disadvantages of 2G system

- Low data rate
- Limited mobility
- Less features on mobile devices
- Limited number of users and hardware capability

2.5G and 2.75G system

In order to support higher data rate, General Packet Radio Service (GPRS) was introduced and successfully deployed. GPRS was capable of data rate up to 171kbps (maximum).

EDGE – Enhanced Data GSM Evolution also developed to improve data rate for GSM networks. EDGE was capable to support up to 473.6kbps (maximum). Another popular technology CDMA2000 was also introduced to support higher data rate for CDMA networks. This technology has the ability to provide up to 384 kbps data rate (maximum).

1.2.3 3G – Third generation communication system

Third generation mobile communication started with the introduction of UMTS – Universal Mobile Terrestrial / Telecommunication Systems. UMTS has the data rate of 384kbps and it support video calling for the first time on mobile devices.

After the introduction of 3G mobile communication system, smart phones became popular across the globe. Specific applications were developed for smartphones which handles multimedia chat, email, video calling, games, social media and healthcare.

Key features of 3G system

- Higher data rate
- Video calling
- Enhanced security, a greater number of users and coverage
- Mobile app support
- Multimedia message support
- Location tracking and maps
- Better web browsing
- TV streaming
- High quality 3D games

3.5G to 3.75 Systems

In order to enhance data rate in existing 3G networks, another two technology improvements are introduced to network. HSDPA – High Speed Downlink Packet access and HSUPA – High Speed Uplink Packet Access, developed and deployed to the 3G networks. 3.5G network can support up to 2mbps data rate. 3.75 system is an improved version of 3G network with HSPA+ High Speed Packet Access plus. Later this system will evolve into more powerful 3.9G system known as LTE (Long Term Evolution).

Disadvantages of 3G systems

- Expensive spectrum licenses
- Costly infrastructure, equipment and implementation
- Higher bandwidth requirements to support higher data rate
- Compatibility with older generation 2G system and frequency bands

1.2.4 4G – Fourth generation communication system

4G systems are enhanced version of 3G networks developed by IEEE, offers higher data rate and capable to handle more advanced multimedia services. LTE and LTE advanced wireless technology used in 4th generation systems. Furthermore, it has compatibility with previous version thus easier deployment and upgrade of LTE and LTE advanced networks are possible.

Simultaneous transmission of voice and data is possible with LTE system which significantly improve data rate. All services including voice services can be transmitted over IP packets. Complex modulation schemes and carrier aggregation is used to multiply uplink / downlink capacity. Wireless transmission technologies like WiMax are introduced in 4G system to enhance data rate and network performance.

Key features of 4G system

- Much higher data rates up to 1Gbps
- Enhanced security and mobility
- Reduced latency for mission critical applications
- High-definition video streaming and gaming
- Voice over LTE network VoLTE (use IP packets for voice)

Disadvantages of 4G system

- Expensive hardware and infrastructure
- Costly spectrum (most countries, frequency bands are too expensive)
- High end mobile devices compatible with 4G technology required, which is costly
- Wide deployment and upgrade is time consuming

5G – Fifth generation communication system

5G network is using advanced technologies to deliver ultra-fast internet and multimedia experience for customers. Existing LTE advanced networks will transform into supercharged 5G networks in future. In earlier deployments, 5G network will function in non-standalone mode and standalone mode. In non-standalone mode both LTE spectrum and 5G-NR spectrum will be used together. Control signaling will be connected to LTE core network in non-standalone mode.

There will be a dedicated 5G core network higher bandwidth 5G – NR spectrum for standalone mode. Sub 6-GHz spectrum of FR1 ranges are used in the initial deployments of 5G networks.

In order to achieve higher data rate, 5G technology will use millimeter waves and unlicensed spectrum for data transmission.

Key features of 5G technology

- Ultra-fast mobile internet up to 10Gbps
- Low latency in milliseconds (significant for mission critical applications)

Total cost deduction for data

- Higher security and reliable network
- Uses technologies like small cells, beam forming to improve efficiency
- Forward compatibility network offers further enhancements in future
- Cloud based infrastructure offers power efficiency, easy maintenance and upgrade of hardware.

Comparison of 1G to 5G technology

Table 1.1 Comparison between different communication technologies.

Generation	Speed	Technology	Key Features
1G (1970 –1980s)	14.4 Kbps	AMPS,NMT, TACS	Voice only services
2G (1990 to 2000)	9.6/ 14.4 Kbps	TDMA,CDMA	Voice and Data services
2.5G to 2.75G (2001-2004)	171.2 Kbps 20-40 Kbps	GPRS	Voice, Data and web mobile internet, low speed streaming services and email services.
3G (2004-2005)	3.1 Mbps 500- 700 Kbps	CDMA2000 (1xRTT, EVDO) UMTS and EDGE	Voice, Data, Multimedia, support for smart phone applications, faster web browsing, video calling and TV streaming.
3.5G (2006-2010)	14.4 Mbps 1-3 Mbps	HSPA	All the services from 3G network with enhanced speed and more mobility.
4G (2010 onwards)	100-300 Mbps. 3-5 Mbps 100 Mbps (Wi-Fi)	WiMax, LTE and Wi-Fi	High speed, high quality voice over IP, HD multimedia streaming, 3D gamming, HD video conferencing and worldwide roaming.
5G (Expecting at the end of 2019)	1 to 10 Gbps	LTE advanced schemes, OMA and NOMA	Super fast mobile internet, low latency network for mission critical applications, Internet of Things, security and surveillance, HD multimedia streaming, autonomous driving, smart healthcare applications.

Definition 1.2 (Spectral Efficiency). The SE of an encoding/decoding scheme is the average number of bits of information, per complex-valued sample, that it can reliably transmit over the channel under consideration.

From this definition, it is clear that the SE is a deterministic number that can be measured in bit per complex-valued sample. Since there are B samples per second, an equivalent unit of the SE is bit per second per Hertz, often written in short-form as bit/s/Hz. For fading channels, which change over time, the SE can be viewed as the average number of bit/s/Hz over the fading realizations, as will be defined below. In this monograph, we often consider the SE of a channel between a UE and a BS, which for simplicity we refer to as the "SE of the UE". A related metric is the information rate [bit/s], which is defined as the product of the SE and the bandwidth B. In addition, we commonly consider the sum SE of the channels from all UEs in a cell to the respective BS, which is measured in bit/s/Hz/cell.

SE of a cell can be increased by using more transmit power, deploying multiple BS antennas, or serving multiple UEs per cell. All these approaches inevitably increase the PC of the network, either directly (by increasing the transmit power) or indirectly (by using more hardware), and therefore may potentially reduce the Energy Efficiency (EE).

In a broad sense, EE refers to how much energy it takes to achieve a certain amount of work. This general definition applies to all fields of science, from physics to economics, and wireless communication is no exception [7]. Unlike many fields wherein the definition of "work" is straightforward, in a cellular network it is not easy to define what exactly one unit of "work" is. The network provides connectivity over a certain area and it transports bits to and from UEs. Users pay not only for the delivered number of bits but also for the possibility to use the network anywhere at any time. Moreover, grading the performance of a cellular network is more challenging than it first appears, because the performance can be measured in a variety of different ways and each such performance measure affects the EE metric differently Among the different ways to define the EE of a cellular network, one of the most popular definitions takes inspiration from the definition of SE, that is, "the SE of a wireless communication system is the number of bits that can be reliably transmitted per complex-valued

sample" (a formal definition of the SE was given in Definition 1.2 on pg. 4). By replacing "SE" with "EE" and "complex-valued sample" with "unit of energy", the following definition is obtained

Definition 1.3: (Energy Efficiency). The EE of a cellular network is the number of bits that can be reliably transmitted per unit of energy. According to the definition above, we define the EE as

$$\text{EE} = \frac{\text{Throughput [bit/s/cell]}}{\text{Power consumption [W/cell]}} \tag{1.2}$$

which is measured in bit/Joule and can be seen as a benefit-cost ratio, where the service quality (throughput) is compared with the associated costs (power consumption). Hence, it is an indicator of the network's bit-delivery efficiency.

The main goal of this project is to demonstrate how we can achieve major improvements in SE and EE and how Massive MIMO is considered the most promising technology for improving the SE and EE in future cellular networks.

CHAPTER 2

MASSIVE MIMO SYSTEM

Introduction

MIMO stands for ‘Multiple Input, Multiple Output’, and is a form of radio antenna that increases efficiency by increasing the number of transmitters and receivers. Perhaps more importantly, MIMO antennas can send and receive signals over the same channel, without the need to take turns, which increases capacity without sacrificing spectrum.

In today’s 4G and 5G networks, base station antennas are typically fitted with around 12 antenna ports that broadcast information in every direction at once. Which means that current transceivers have to take turns if they want to transmit and receive data on the same frequency, or the data has to be moved to another frequency to avoid hold-ups, and this causes congestion.

In radio, multiple-input and multiple-output, or MIMO, is a method for multiplying the capacity of a radio link using multiple transmission and receiving antennas to exploit multipath propagation. MIMO has become an essential element of wireless communication standards including IEEE 802.11n (Wi-Fi 4), IEEE 802.11ac (Wi-Fi 5), HSPA+ (3G), WiMAX, and Long-Term Evolution (LTE). More recently, MIMO has been applied to power-line communication for three-wire installations as part of the ITU G.hn standard and of the Home Plug AV2 specification.

At one time, in wireless the term "MIMO" referred to the use of multiple antennas at the transmitter and the receiver. In modern usage, "MIMO" specifically refers to a practical technique for sending and receiving more than one data signal simultaneously over the same radio channel by exploiting multipath propagation. Although the "multipath" phenomenon may be interesting, it is the use of orthogonal frequency-division multiplexing (OFDM) to encode the channels that is responsible for the increase in data capacity. MIMO is fundamentally different from smart antenna techniques developed to enhance the performance of a single data signal, such as beamforming and diversity.

MIMO systems are an indispensable piece of current wireless systems, and lately they have been utilized broadly to accomplish high spectral effectiveness and energy productivity. Prior to the presentation of MIMO, single-input-single-output systems were for the most part utilized, which had exceptionally low throughput and couldn't support countless clients with high dependability. To oblige this massive client interest, different new MIMO innovation like single-client MIMO (SU-MIMO), multi-client MIMO (MU-MIMO) and organization MIMO were created. Be that as it may, these new innovations are additionally insufficient to oblige the always expanding requests. The wireless clients have expanded dramatically over the most recent couple of years, and these clients create trillions of information that should be taken care of productively with greater dependability.

2.1 Massive MIMO

Massive MIMO is the most dazzling innovation for 5G and past the wireless access period. Massive MIMO is the headway of contemporary MIMO systems utilized in current wireless organizations, which groups together hundreds and even large number of antennas at the base station and serves many clients at the same time. The additional antennas that massive MIMO uses will help center energy into a more modest area of room to give better spectral proficiency and throughput.

Definition 2.1: (Massive MIMO), Massive MIMO is a type of wireless communications technology in which base stations are equipped with a very large number of antenna elements to improve spectral and energy efficiency. Massive MIMO systems typically have tens, hundreds, or even thousands of antennas in a single antenna array.

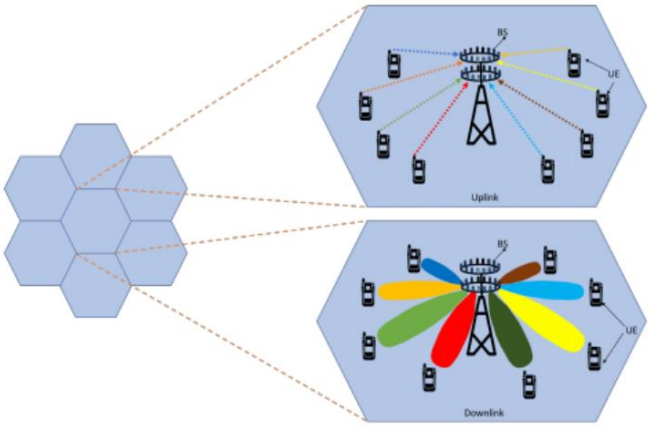


Figure 2.1 Massive MIMO Uplink and Downlink

- It uses SDMA to achieve a multiplexing gain by serving multiple UEs on the same time-frequency resources.
- It has more BS antennas than UEs per cell to achieve efficient interference suppression. If the anticipated number of UEs grows in a cell, the BS should be upgraded so that the number of antennas increases proportionally.
- It operates in TDD mode to limit the CSI acquisition overhead, due to the multiple antennas, and to not rely on parametrizable channel models.

2.2 Advantages of Massive MIMO

Massive MIMO fundamentally expands the capabilities of MU-MIMO through the inclusion of a higher number of antennas to bring drastic improvements in network performance. Hence, it has become one of the technological underpinnings of modern wireless cellular networks to include the 4G standard, LTE and LTE Advanced technologies, and 5G technologies. Placing a large number of antennas allow a particular access point to focus the transmission and reception of electromagnetic signals to specific regions or targeted areas, thus improving throughput, capacity, and efficiency. Note that a Massive MIMO system also coordinates the operation of these antennas through machine learning and algorithm.

2.2.1 Increases Network Capacity

Massive MIMO increases the capacity of a particular wireless communication network in two ways. First, it enables the deployment of higher frequencies, such as in the case of Sub-6 5G specification. Second, by employing multi-user MIMO, a cellular base station with Massive MIMO capability can send and receive multiple data streams simultaneously from different users using the same frequency resources.

Note that network capacity is determined by the number or amount of total data a particular network can serve to its end-users, as well as by the maximum number of end-users that can be served based on an expected service level.

2.2.2 Enhances Network Coverage

Another advantage of Massive MIMO is that it provides high spectral efficiency through the coordination of multiple antennas using simple processing and without intensive power consumption. When used in a 5G cellular network technology, it allows 10 times more spectral and network efficiency compared to fourth-generation networks. Furthermore, when applied in 4G technology, it improves the deep coverage of fourth-generation networks.

Because next-generation cellular network technologies use electromagnetic radiation with higher frequencies or more specifically, frequencies within the upper limits of radio waves and the range of microwaves, the signals they generate travel a short distance. Hence, enhancing network coverage is critical in modern and future cellular technologies.

2.2.3 Complements Beamforming

Beamforming technology works by focusing a signal toward a specific direction, rather than broadcasting in all directions, thus resulting in more direct communication between a transmitter and a receiver, more stable and reliable connectivity, and faster data transmission. As a signal processing technique and traffic-signaling system, this technology depends on advanced antenna technologies on both access points and end-user devices.

The large number of antennas in a Massive MIMO system enables three-dimensional beamforming in which a single beam of signal-bearing electromagnetic radiation travels through vertical and horizontal directions. The process increases data transmission rates further while reaching people in elevated areas such as buildings and those in moving vehicles

2.2.4 Enables Next-Gen Technologies

Massive MIMO is an essential component of 5G technology. For example, in Sub-6 5G specification, it allows the utilization of frequencies within the sub-6 GHz range. Moreover, in mmWave 5G specification, this technology increases frequency reach to expand network coverage, optimizes the propagation of signal-bearing electromagnetic radiation, and allows true multi-user wireless communication within a defined area.

2.3 Disadvantages of Massive MIMO

One of the biggest disadvantages of Massive MIMO is the cost associated with its implementation and deployment. The systems are several times more extensive than traditional base station units and antenna technologies. Furthermore, the design of multiple antenna systems for cellular networks is more complex and requires more effort and time during assembly and installation.

Furthermore, using frequency division duplex or FDD results in feedback overhead. This phenomenon transpires when a receiver sends out feedback signals to a transmitter. Increasing the antenna elements results in a further increase in the overhead. Hence, time-division duplex or TDD is more suitable for Massive MIMO implementation.

The placement of multiple antennas in a defined area within a base station means placing hardware components in a smaller space. An entire massive multiple-input and multiple-output system needs advanced components that are capable of delivering their intended level of performance despite their smaller size than their larger

counterparts. Remember that Massive MIMO is not simply about placing and using a large number of antennas. The entire technology also works using artificial intelligence and machine learning to complement frequency management, signal processing techniques, and data transmission. Doing so requires complex processing algorithms that further add to the cost and complexity of designing, implementing, and deploying an entire system.

2.4 Duplexing

A duplex communication system is a point-to-point system composed of two or more connected parties or devices that can communicate with one another in both directions. Duplex systems are employed in many communications networks, either to allow for simultaneous communication in both directions between two connected parties or to provide a reverse path for the monitoring and remote adjustment of equipment in the field.

It takes two forms:

1. Half duplex
2. Full duplex

In half duplex, the two communicating parties take turns transmitting over a shared channel. Two-way radios work this way. As one-party talks, the other listens. Speaking parties often say “Over” to indicate that they’re finished and it’s time for the other party to speak. In networking, a single cable is shared as the two computers communicating take turns sending and receiving data.

Full duplex refers to simultaneous two-way communications. The two communicating stations can send and receive at the same time. Landline telephones and cell phones work this way. Some forms of networking permit simultaneous transmit and receive operations to occur. This is the more desirable form of duplexing, but it is more complex and expensive than half duplexing.

There are two basic forms of full duplexing:

1. Frequency division duplex (FDD)
2. Time division duplex (TDD)

2.4.1 Frequency Division Duplexing

FDD requires two separate communications channels. In networking, there are two cables. Full-duplex Ethernet uses two twisted pairs inside the CAT5 cable for simultaneous send and receive operations. Wireless systems need two separate frequency bands or channels (Fig 2.2). A sufficient amount of guard band separates the two bands so the transmitter and receiver don't interfere with one another. Good filtering or duplexers and possibly shielding are a must to ensure the transmitter does not desensitize the adjacent receiver.

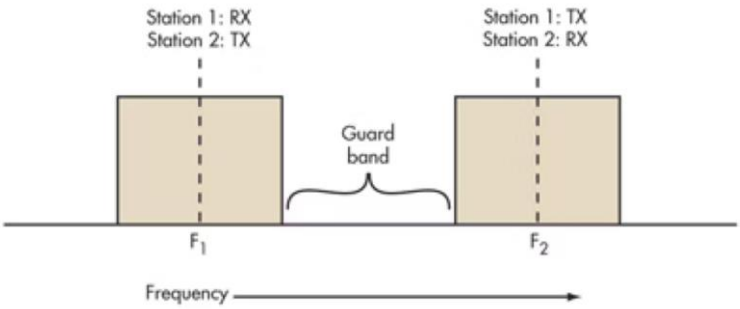


Figure 2.2 Two symmetrical segments of spectrum for the uplink and downlink channels

In a cell phone with a transmitter and receiver operating simultaneously within such close proximity, the receiver must filter out as much of the transmitter signal as possible. The greater the spectrum separation, the more effective the filters.

FDD uses lots of frequency spectrum, though, generally at least twice the spectrum needed by TDD. In addition, there must be adequate spectrum separation between the transmit and receive channels. These so-called guard bands aren't useable, so they're wasteful. Given the scarcity and expense of spectrum, these are real disadvantages.

However, FDD is very widely used in cellular telephone systems, such as the widely used GSM system. In some systems the 25-MHz band from 869 to 894 MHz is used as the downlink (DL) spectrum from the cell site tower to the handset, and the 25-MHz band from 824 to 849 MHz is used as the uplink (UL) spectrum from the handset to cell site.

Another disadvantage with FDD is the difficulty of using special antenna techniques like multiple-input multiple-output (MIMO) and beamforming. These

technologies are a core part of the new Long-Term Evolution (LTE) 4G cell phone strategies for increasing data rates. It is difficult to make antenna bandwidths broad enough to cover both sets of spectrum. More complex dynamic tuning circuitry is required.

2.4.2 Time Division Duplexing (TDD)

TDD uses a single frequency band for both transmit and receive. Then it shares that band by assigning alternating time slots to transmit and receive operations (Fig 2.3). The information to be transmitted—whether it's voice, video, or computer data—is in serial binary format. Each time slot may be 1 byte long or could be a frame of multiple bytes.

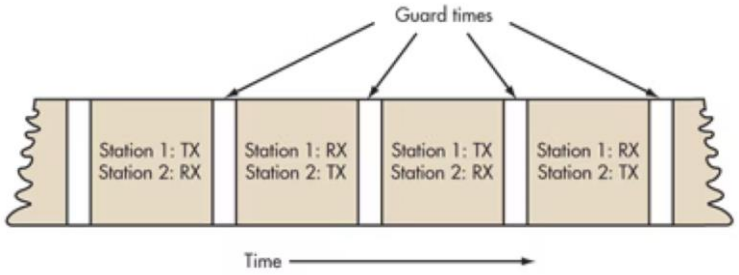


Figure 2.3 TDD alternates the transmission and reception of station data over time.

In some TDD systems, the alternating time slots are of the same duration or have equal DL and UL times. However, the system doesn't have to be 50/50 symmetrical. The system can be asymmetrical as required. For instance, in Internet access, download times are usually much longer than upload times so more or fewer frame time slots are assigned as needed. The real advantage of TDD is that it only needs a single channel of frequency spectrum. Furthermore, no spectrum-wasteful guard bands or channel separations are needed. The downside is that successful implementation of TDD needs a very precise timing and synchronization system at both the transmitter and receiver to make sure time slots don't overlap or otherwise interfere with one another.

2.5 Spatial Multiplexing and Spatial Diversity

Spatial multiplexing and spatial diversity are two radio communication techniques used in modern antenna systems in 4G LTE and 5G NR networks. Both these techniques play essential but separate roles in the MIMO (Multiple Input Multiple Output) antenna systems.

Definition 2.2: (Spatial Diversity). Spatial diversity is a technique in MIMO that reduces signal fading by sending multiple copies of the same radio signal through multiple antennas. Spatial diversity improves radio signal link quality by employing multiple antennas at the transmitter or receiver to communicate numerous copies of the same signal. That allows the antennas to overcome the negative impact of multipath fading by using the copies of the signal to reconstruct it.

Diversity is not a new concept in mobile communications and has been used for years to address the negative impact of signal fading. When a radio signal (e.g., a mobile signal) travels from the cellular base station to the receiver of a mobile phone, it can take many routes depending on the obstructions in its way. Obstructions can be things like buildings, trees, poles, mountains etc. When the signal encounters any obstructions, it can get scattered and become weak or "fade" by the time it reaches the receiver. Diversity in radio communications is the ability of an antenna system to create redundant network resources for the signal to minimise the overall impact of signal fading. In plain English, it means creating additional copies of the signal so that bits and pieces of the scattered signal can be picked up to reconstruct the signal. At a theoretical level, at least three types of diversity solutions are available, including frequency, time, and space diversity. Frequency diversity requires multiple frequency channels, each communicating a version or copy of the same signal. Time diversity does the same thing but uses different time-slots instead so that different copies of the signal are communicated at different time intervals. But the diversity type employed by MIMO antenna systems is space diversity, also known as spatial diversity. In MIMO systems, spatial diversity is achieved by multiple antennas at the transmitter and the receiver that communicate (transmit or receive) a different version of the same signal. These versions are essentially a replica of the original signal. If used at the receiver end,

the receiver can collect all the different versions of the signal to reconstruct it to overcome the negative impact of signal fading. In 4G LTE and 5G NR networks, spatial diversity is a critical part of MIMO systems that use multiple antennas at the transmitter and the receiver.

Definition 2.3 (Spatial Multiplexing). spatial multiplexing is a technique in MIMO that boosts data rates by sending the data payload in separate streams through spatially separated antennas. Spatial multiplexing improves data rates by allowing the overall data payload to be communicated to a user device in the form of multiple data streams that carry small portions of the overall information. The data streams can be targeted at a single user device or multiple user devices.

Spatial multiplexing or Space Division Multiplexing (SDM) is a multiplexing technique employed by MIMO antenna systems. It is an essential feature of MIMO and is the primary reason for introducing MIMO in 4G LTE and 5G NR networks. In spatial multiplexing, a transmitter or receiver can use several antennas separated in space by their angular direction. These antennas can send and receive multiple data streams using the same frequency and time resources and act as individual channels to communicate the information (e.g. a WhatsApp message) between the transmitter and receiver. The multiple data streams within a MIMO system can target a single user device or multiple user devices for simultaneous communication. When the data payload is sent towards a user device in the form of multiple concurrent streams, the data rate for the user device goes up. 5G NR networks use an enhanced version of MIMO, called Massive MIMO which consists of tens or even hundreds of antenna elements within a single antenna panel. Due to the sheer volume of antenna elements and the multi-user support capability, Massive MIMO can simultaneously offer higher data rates to several user devices.

MIMO systems in 4G LTE and 5G NR networks use both spatial multiplexing and spatial diversity to improve data rates whilst improving signal quality. In MIMO, spatial diversity is a technique that provides the ability to overcome the negative impact of multipath signal fading by communicating separate versions or copies of the same signal through multiple antennas. On the other hand, spatial multiplexing is a technique

that improves the achievable data rates for end-users by transmitting and receiving multiple streams of data through various spatially-separated antennas.

2.6 Importance of channel in wireless communication

Communicating data or an information signal from transmitter or receiver to receiver or transmitter requires some form of path way or medium called CHANNEL. Channel plays and important role in wireless communication since it can degrade the information signal by adding multipath fading and Doppler effects (if channel is mobile). Correct knowledge of channel is a fundamental perquisite for the design of a wireless communication system. A communication channel either to a physical transmission medium such as wire, or through a local connection over a multiplexed medium such as a radio channel.

A channel is used to convey an information signal from one of several senders to one of several receivers. A channel has a certain capacity E, for transmitting information often measure by its bandwidth in Hz or its data rate in bits per second.

Bandwidth is a limited resource used by different organization due to which widespread use of wireless networks are limited. The wireless channel is susceptible to a variety of transmission impediments, path loss, interference and blockage These factors restrict the range and he reliability of the wireless transmission the extent to which these factors affect the transmission depends upon the environmental conditions and the mobility of the transmitter and the receiver.

2.6.1 Additive white Gaussian Noise

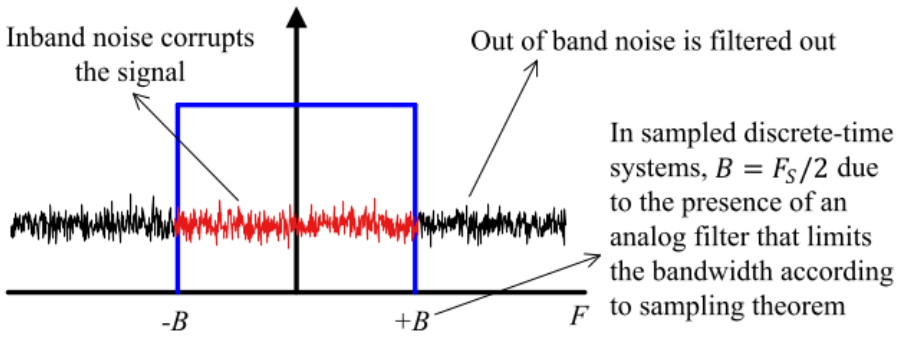


Figure 2.4 Additive white gaussian noise

The performance of a digital communication system is quantified by the probability of bit detection errors in the presence of thermal noise. In the context of wireless communications. the main source of thermal noise is addition of random signals arising from the vibration of atoms in the receiver electronics.

The term additive white Gaussian noise (AWGN) originates due to the following reasons Additive: The noise is additive, Le., the received signal is equal to the transmitted signal plus noise. This gives the most widely used equality in communication systems.

$$r(t)=s(t)+w(t)(1)(1)\ r(t)=s(t)+w(t) \qquad (2.1)$$

which is shown in Figure below. Moreover, this noise is statistically independent of the signal Remember that the above equation is highly simplified due to neglecting every single imperfection a Tx signal encounters, except the noise itself.

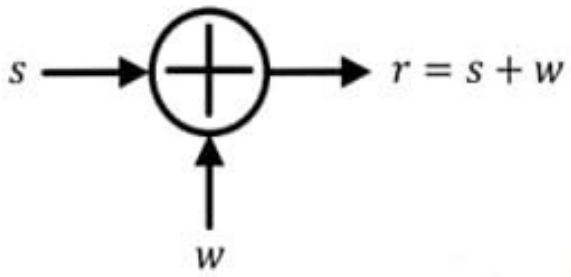


Figure 2.5 Additive Noise

White: Just like the white colour which is composed of all frequencies in the visible spectrum, white noise refers to the idea that it has uniform power across the whole frequency band. As a consequence, the Power Spectral Density (PSD) of white noise is constant for all frequencies ranging from to -∞ to +∞.

Gaussian: The probability distribution of the noise samples is Gaussian with a zero mean, e, in time domain, the samples can acquire both positive and negative values and in addition, the values close to zero have a higher chance of occurrence while the values

far away from zero are less likely to appear. This is shown in Figure below. As a result, the time domain average of large number of noise samples is equal to zero.

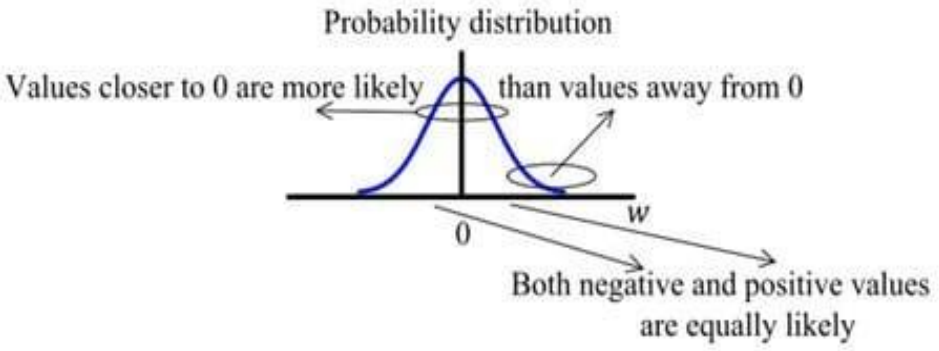


Figure 2.6 Gaussian Noise

In reality, the ideal flat spectrum from $-\infty$ to $+\infty$ is true for frequencies of interest in wireless communications (a few kHz to hundreds of GHz) but not for higher frequencies. Nevertheless, every wireless communication system involves filtering that removes most of the noise energy outside the spectral band occupied by our desired signal. Consequently, after filtering, it is not possible to distinguish whether the spectrum was ideally flat or partially flat outside the band of interest. To help in mathematical analysis of the underlying waveforms resulting in closed-form expressions — a holy grail of communication theory — it can be assumed to be flat before filtering. For a discrete signal with sampling rate FS, the sampling theorem dictates that the bandwidth of a signal is constrained by a lowpass filter within the range ±FS/2 to avoid aliasing. For the purpose of calculations, this filter is an ideal lowpass filter with

$$H(F) = \begin{cases} 1, & -\frac{\mathrm{FS}}{2} < \mathrm{F} < +\frac{\mathrm{FS}}{2} \\ 0, & elsewhere \end{cases} \quad (2.2)$$

The resulting in-band power is shown in red in the figure below, while the rest is filtered out.

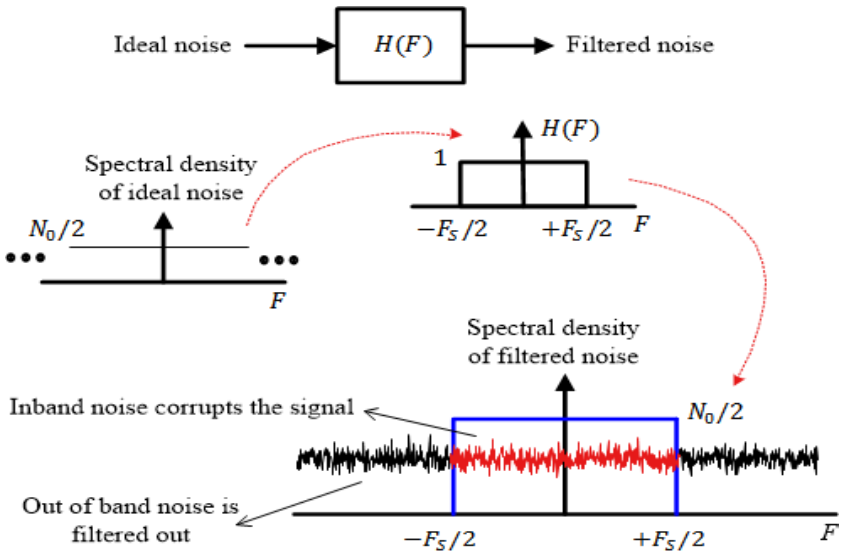


Figure 2.7 resulting in-band power

2.6.2 Rayleigh Channel

Rayleigh fading is a statistical model for the effect of a propagation environment on a radio signal, such as that used by wireless devices. Rayleigh fading models assume that the magnitude of a signal that has passed through such a transmission medium (also called a communication channel) will vary randomly, or fade, according to a Rayleigh distribution, the radial component of the sum of two uncorrelated Gaussian random variables.

Rayleigh fading is viewed as a reasonable model for tropospheric and ionospheric signal propagation as well as the effect of heavily built-up urban environments on radio signals. Rayleigh fading is most applicable when there is no dominant propagation along a line of sight between the transmitter and receiver. If there is a dominant line of sight, Rician fading may be more applicable Rayleigh fading is a special case of two-wave with diffuse power (TWDP) fading

Rayleigh fading is a reasonable model when there are many objects in the environment that scatter the radio signal before it arrives at the receiver. The central limit theorem holds that, if there is sufficiently much scatter, the channel impulse response will be

well-modelled as a Gaussian process irrespective of the distribution of the individual components. If there is no dominant component to the scatter, then such a process will have zero mean and phase evenly distributed between 0 and 2n radians. The envelope of the channel response will therefore be Rayleigh distributed Calling this random variable, it will have a probability density function:

$$p_R(r) = \frac{2r}{\Omega} e^{-r^2/\Omega},\ r \geq 0 \tag{2.3}$$

Where $\Omega = E(R^2)$.

Often, the gain and phase elements of a channel's distortion are conveniently represented as a complex number. In this case, Rayleigh fading is exhibited by the assumption that the real and imaginary parts of the response are modelled by independent and identically distributed zero-mean Gaussian processes so that the amplitude of the response is the sum of two such processes.

Rayleigh channel model: The Rayleigh fading environment is described by the many multipath components, each having relatively similar signal magnitude, and uniformly distributed phase, that means there is no line of sight (LOS) path between transmitter and receiver. The channel in which the signal takes various path to reach the receiver after getting reflect from various objects in the environment. The signal receiving at receiver is sum of the reflected signal and the main signal. The signal in the environment gets diffracted or reflected from the objects like tree, building, moving vehicle etc and imposes problem when the envelope of the individual signal is added up.

2.7 System Model

A Massive MIMO network is a multicarrier cellular network with L cells that operate according to a synchronous TDD protocol. BS j is equipped with $M_j >> 1$ antennas, to achieve channel hardening. BS j communicates with K_j single-antenna UEs simultaneously on each time/frequency sample, with antenna-UE ratio $M_j/K_j > 1$. Each

BS operates individually and processes its signals using linear receive combining and linear transmit precoding.

Definition 2.4 (Coherence block). A coherence block consists of a number of subcarriers and time samples over which the channel response can be approximated as constant and flat-fading. If the coherence bandwidth is Bc and the coherence time is Tc, then each coherence block contains τc = BcTc complex-valued samples.

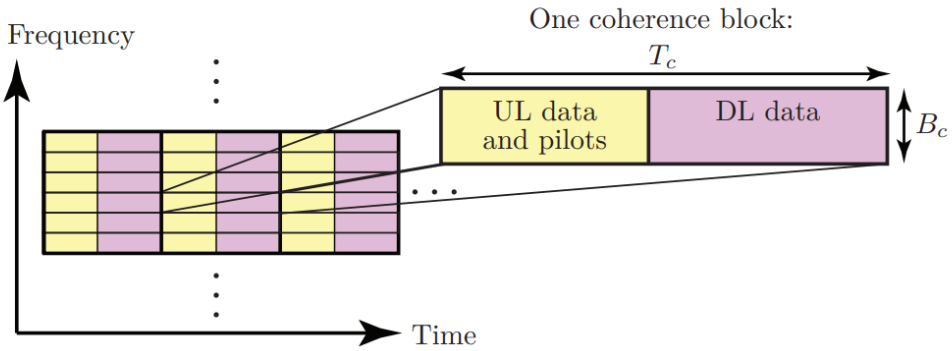
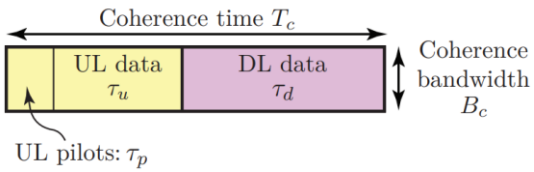
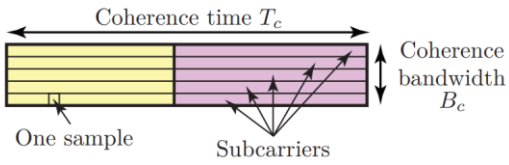


Fig 2.8 The TDD multicarrier modulation scheme of a Massive MIMO Network. The time frequency plane is divided into coherence blocks.



(a) The samples are used for UL pilots, UL data, and DL data.



(b) The samples can belong to different subcarriers.
Fig 2.9 Each coherence block contains τc = BcTc complex-valued samples.

Each coherence block is operated in TDD mode and Figure 2.9 illustrates how the τ_c samples are located in the time and frequency plane. The samples are used for three different things:

• τ_p UL pilot signals;

• τ_u UL data signals;

• τ_d DL data signals.

Clearly, we need $\tau p + \tau u + \tau d = \tau c$. The fraction of UL and DL data can be selected based on the network traffic characteristics, while the number of pilots per coherence block is a design parameter. Many user applications (e.g., video streaming and web browsing) mainly generate DL traffic, which can be dealt with by selecting $\tau d > \tau u$.

Each cell is assigned an index in the set L, where the cardinality |L| is the number of cells. The BS in each cell is equipped with an array of M antennas and communicates with K single-antenna UEs at the time, out of a set of Kmax UEs. We are interested in massive MIMO topologies where M and Kmax are large and fixed, while K is a design parameter and all UEs have unlimited demand for data. The subset of active UEs changes over time, thus the name UE $k \in \{1, \ldots, K\}$ in cell $l \in L$ is given to different UEs at different times. The geographical position $z_{lk} \in R_2$ of UE k in cell l is therefore an ergodic random variable with a cell-specific distribution. This model is used to study the average performance for a random rather than fixed set of interfering UEs. The time-frequency resources are divided into frames consisting of Tc seconds and Wc Hz, as illustrated in Fig. 2.8 This leaves room for S = TcWc transmission symbols per frame. We assume that the frame dimensions are such that Tc is smaller or equal to the coherence time of all UEs, while Wc is smaller or equal to the coherence bandwidth of all UEs. Hence, all the channels are static within the frame; $h_{jlk} \in C\ N$ denotes the channel response between BS j and UE k in cell l in a given frame. These channel responses are drawn as realizations from zero mean circularly symmetric complex Gaussian distributions:

$$\mathbf{h}_{jlk} \sim \mathcal{CN}\left(\mathbf{0}, d_j(\mathbf{z}_{lk})\mathbf{I}_M\right), \tag{2.4}$$

where I_M is the M × M identity matrix. This is a theoretical model for non-line-of-sight propagation that is known to give representative results with both few and many BS antennas. The deterministic function dj (z) gives the variance of the channel attenuation from BS j to any UE position z. The value of dj (z_{lk}) varies slowly over time

and frequency, thus we assume that the value is known at BS j for all l and k and that each UE knows its value to its serving BS. The exact UE positions z_{lk} are unknown.

We consider the time-division duplex (TDD) protocol shown in Fig 2.8, where $B \geq 1$ out of the S symbols in each frame are reserved for UL pilot signaling. There is no DL pilot signaling and no feedback of CSI, because the BSs can process both UL and DL signals using the UL channel measurements due to the channel reciprocity in TDD systems. The remaining $S - B$ symbols are allocated for payload data and are split between UL and DL transmission. We let $\zeta^{(ul)}$ and $\zeta^{(dl)}$ denote the fixed fractions allocated for UL and DL, respectively. These fractions can be selected arbitrarily, subject to the constraint $\zeta^{(ul)} + \zeta^{(dl)} = 1$ and that $\zeta^{(ul)}(S - B)$ and $\zeta^{(dl)}(S - B)$ are positive integers. Below, we define the system models for the UL and DL. The BSs are not exchanging any short-term information in this work, but we will see how the pilot allocation and transmission processing can be coordinated in a distributed fashion.

2.7.1 System Model for Uplink

The **uplink channel** is used to transmit data and the pilot signal from the user terminal to the base station. The received UL signal $y_j \in C^M$ at BS j in a frame is modeled,

$$\mathbf{y}_j = \sum_{l=1}^{L} \sum_{k=1}^{K_l} \mathbf{h}_{lk}^{j} s_{lk} + \mathbf{n}_j = \underbrace{\sum_{k=1}^{K_j} \mathbf{h}_{jk}^{j} s_{jk}}_{\text{Desired signals}} + \underbrace{\sum_{\substack{l=1 \\ l \neq j}}^{L} \sum_{i=1}^{K_l} \mathbf{h}_{li}^{j} s_{li}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{n}_j}_{\text{Noise}} \tag{2.5}$$

$s_{lk} \in C$ is the symbol transmitted by UE k in cell l.

The additive noise $nj \in C^M$ is modeled as $nj \sim C^N(0, \sigma^2 I_M)$, where σ^2 is the noise variance. $P_{lk} > 0$ is the uplink transmit power.

Contrary to most previous works on massive MIMO, which assume fixed UL power, we consider statistics-aware power control the symbols from UE k in cell l have the transmit power $p_{lk} = \rho / dl(z_{lk})$, where $\rho > 0$ is a design parameter. This power-control policy inverts the average channel attenuation $dl(z_{lk})$ and has the merit of making the average effective channel gain the same for all UEs: $E\{p_{lk}||h_{lk}||^2\} = M\rho$. Hence, this policy guarantees a uniform user experience, saves valuable energy at UEs, and avoids near-far blockage where weak signals drown in stronger signals due to the finite dynamic range of analog-to-digital converters (ADCs).

2.7.2 System Model for Downlink

The uplink channel is used to transmit data and the pilot signal from the user terminal to the base station. Based on channel reciprocity, The received **Downlink signal** $y_{jk} \in C$ at UE k in cell j in a frame is modeled as

$$y_{jk} = \sum_{l=1}^{L}\sum_{i=1}^{K_l}(\mathbf{h}_{jk}^{l})^{\mathrm{H}}\mathbf{w}_{li}\varsigma_{li} + n_{jk}$$

$$= \underbrace{(\mathbf{h}_{jk}^{j})^{\mathrm{H}}\mathbf{w}_{jk}\varsigma_{jk}}_{\text{Desired signal}} + \underbrace{\sum_{\substack{i=1 \\ i\neq k}}^{K_j}(\mathbf{h}_{jk}^{j})^{\mathrm{H}}\mathbf{w}_{ji}\varsigma_{ji}}_{\text{Intra-cell interference}} + \underbrace{\sum_{\substack{l=1 \\ l\neq j}}^{L}\sum_{i=1}^{K_l}(\mathbf{h}_{jk}^{l})^{\mathrm{H}}\mathbf{w}_{li}\varsigma_{li}}_{\text{Inter-cell interference}} + \underbrace{n_{jk}}_{\text{Noise}} \quad (2.6)$$

ς_{lm} is the symbol intended for UE m in cell l, The additive noise $n_{jk} \in CM$ is modeled as $n_{jk} \sim CN(0, \sigma^2 I_M)$, where σ^2 is the noise variance. $w_{li} \in C^M$ is the corresponding precoding vector.

2.8 Interference

2.8.1 Inter symbol Interference

Inter symbol interference is a form of distortion of a signal in which one symbol interferes with subsequent symbols. This is an unwanted phenomenon as the previous symbols have similar effect as noise, thus making the communication less reliable. ISI is usually caused by multipath propagation or the inherent non-linear frequency response of a channel causing successive symbols to blur together. The presence of ISI in the system introduces error in the decision device at the receiver output. Therefore, in the design of the transmitting and receiving filters, the objective is to minimize the effects of ISI and thereby deliver the digital data to its destination with the smallest error rate possible.

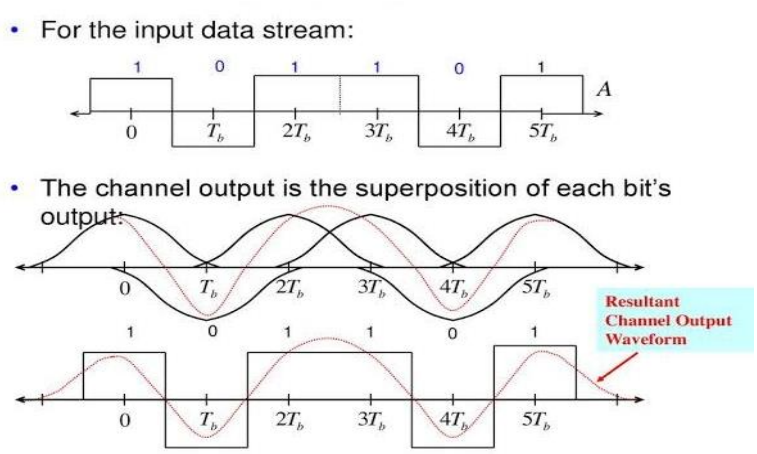


Figure 2.10 Inter symbol interference

2.8.2 Inter Carrier Interference

Presence of Doppler shifts and frequency and phase offsets in an OFDM system causes loss in orthogonality of the sub carriers. As a result, interference is observed between subcarriers This phenomenon is known as Inter Carrier Interference (ICI).

In OFDM, the spectra of sub carriers overlap but remain orthogonal to each other. This means that at the maximum of each subcarrier spectrum, all the spectra of other sub carriers are zero. The receiver samples data symbols on individual subcarriers at the maximum points and demodulates them free from any interference from the other subcarriers. Interference caused by data symbols on adjacent subcarriers is referred to inter carrier interference.

2.9 Summary

This chapter mainly focused on Massive MIMO, advantages and disadvantages of massive MIMO, frequency division duplexing, time division duplexing, importance of channel and types of channels, system model for both uplink and downlink, interference and types of interference. In the upcoming chapter we are going to discuss about spectral efficiency in massive MIMO and we will discuss various transmit or receive combining techniques in order to increase the spectral efficiency.

CHAPTER 3
MAXIMIZATION OF SPECTRAL EFFICIENCY IN MASSIVE MIMO

Introduction:

Cellular communication networks are continuously evolving to keep up with the rapidly increasing demand for wireless data services. Higher area throughput (in bit/s per km^2.) has traditionally been achieved by a combination of three multiplicative factors: more frequency spectrum (Hz), higher cell density (more cells per km^2), and higher spectral efficiency (bit/s/Hz/cell). This paper considers the latter and especially the massive multiple-input multiple-output (MIMO) concept, which has been identified as the key to increase the spectral efficiency (SE) by orders of magnitude over contemporary systems. The massive MIMO concept is based on equipping base stations (BSs) with hundreds or thousands of antenna elements which, unlike conventional cellular technology, are operated in a coherent fashion. This can provide unprecedented array gains and a spatial resolution that allows for multi-user MIMO communication to tens or hundreds of user equipment (UEs) per cell, while maintaining robustness to inter-user interference. The research on massive MIMO has so far focused on establishing the fundamental physical (PHY) layer properties; in particular, that the acquisition of channel state information (CSI) is limited by the channel coherence block (i.e., the fact that channel responses are only static in limited time/frequency blocks) and how this impacts the SEs and the ability to mitigate inter-cell interference. In addition, the aggressive multiplexing in massive MIMO has been shown to provide major improvements in the overall energy efficiency.

In this chapter, we consider a related resource allocation question: how many UEs should be scheduled per cell to maximize the spectral efficiency? This question has, to the best of our knowledge, not been answered for multi-cell systems. We show how the coherence block length, number of antennas, pilot allocation, hardware impairments, and other system parameters determine the answer. To this end, we derive new SE expressions which are valid for both uplink (UL) and downlink (DL)

transmission, with random user locations and power control that yields uniform UE performance. We consider both conventional linear processing schemes such as maximum ratio (MR) combining/transmission and zero-forcing (ZF), and a new full-pilot zero-forcing (P-ZF) scheme that actively suppresses inter-cell interference in a fully distributed coordinated beamforming fashion. In this chapter we refer to the system model discussed in chapter 2 (pg. 38).

3.1 Channel Estimation

In an OFDM system, the transmitter modulates the message bit sequence into PSK/QAM symbols, performs IFFT on the symbols to convert them into time-domain signals, and sends them out through a (wireless) channel. The received signal is usually distorted by the channel characteristics. In order to recover the transmitted bits, the channel effect must be estimated and compensated in the receiver. Each subcarrier can be regarded as an independent channel, as long as no ICI (Inter-Carrier Interference) occurs, and thus preserving the orthogonality among subcarriers. The orthogonality allows each subcarrier component of the received signal to be expressed as the product of the transmitted signal and channel frequency response at the subcarrier. Thus, the transmitted signal can be recovered by estimating the channel response just at each subcarrier. In general, the channel can be estimated by using a **preamble or pilot symbols** known to both transmitter and receiver.

3.1.1 Training based channel estimation

Training symbols are those symbols which are known to both transmitter and receiver. Training symbols can be used for channel estimation, usually providing a good performance. The least-square (LS) and minimum-mean-square-error (MMSE) techniques are widely used for channel estimation when training symbols are available. We assume that all subcarriers are orthogonal (i.e., ICI-free). Then, the training symbols for N subcarriers can be represented by the following diagonal matrix:

$$\mathbf{X} = \begin{bmatrix} X[0] & 0 & \cdots & 0 \\ 0 & X[1] & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & X[N-1] \end{bmatrix}$$

where X[k] denotes a pilot tone at the kth subcarrier, with E{X[k]} = 0 and Var{X[k]} = σ^2_x, k = 0,1,2,…, N-1.

The received training signal Y[k] can be represented as

$$\mathbf{Y} = \mathbf{XH} + \mathbf{Z} \tag{3.1}$$

Where H is a channel vector and Z is noise vector.

3.1.2 Least Square (LS) channel estimation

The least-square (LS) channel estimation method finds the channel estimate in such a way that the following cost function is minimized:

$$J(\hat{\mathbf{H}}) = \mathbf{Y}^H\mathbf{Y} - \mathbf{Y}^H\mathbf{X}\hat{\mathbf{H}} - \hat{\mathbf{H}}^H\mathbf{X}^H\mathbf{Y} + \hat{\mathbf{H}}^H\mathbf{X}^H\mathbf{X}\hat{\mathbf{H}} \tag{3.2}$$

By equating the derivative of this function with respect to $\widehat{H}$ to 0

We obtain

$$\begin{aligned} &\mathbf{X}^H\mathbf{X}\hat{\mathbf{H}} = \mathbf{X}^H\mathbf{Y} \\ &\hat{\mathbf{H}}_{LS} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{Y} = \mathbf{X}^{-1}\mathbf{Y} \end{aligned} \tag{3.3}$$

3.1.3 Minimum Mean Square Error (MMSE) Estimation

Consider the LS solution in equation (3.3). Using the weight matrix W, define, which corresponds to the MMSE estimate. MSE of the channel estimate H is given as

$$J(\hat{\mathbf{H}}) = E\{||\mathbf{e}||^2\} = E\{||\mathbf{H} - \hat{\mathbf{H}}||^2\} \tag{3.4}$$

$\widehat{\mathrm{H}} \longrightarrow$ MMSE channel estimate

$\widetilde{\mathrm{H}} \longrightarrow$ LS channel estimate

the MMSE channel estimation method finds a better (linear) estimate in terms of W in such a way that the MSE in Equation (3.4) is minimized.

The MMSE channel estimate follows as

$$\hat{\mathbf{H}} = \mathbf{W}\tilde{\mathbf{H}} = \mathbf{R}_{\mathbf{H}\tilde{\mathbf{H}}}\mathbf{R}_{\tilde{\mathbf{H}}\tilde{\mathbf{H}}}^{-1}\tilde{\mathbf{H}}$$
$$= \mathbf{R}_{\mathbf{H}\tilde{\mathbf{H}}}\left(\mathbf{R}_{\mathbf{H}\mathbf{H}} + \frac{\sigma_z^2}{\sigma_x^2}\mathbf{I}\right)^{-1}\tilde{\mathbf{H}} \tag{3.5}$$

Lemma 3.1 The MMSE estimate at BS j of the effective power controlled UL channel h^{eff}_{jlk}, for any UE k ∈ {1, . . . , K} in any cell l ∈ L, is

$$\hat{\mathbf{h}}_{jlk}^{\text{eff}} = \frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\mathbf{Y}_j(\mathbf{\Psi}_j^{\mathrm{T}})^{-1}\mathbf{v}_{i_{lk}}^* \tag{3.6}$$

where $(\cdot)*$ denotes the complex conjugate and the normalized covariance matrix $\Psi_j \in C^{B\times B}$ of the received signal is

$$\mathbf{\Psi}_j = \sum_{\ell\in\mathcal{L}}\sum_{m=1}^{K}\frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{\ell m}}\mathbf{v}_{i_{\ell m}}^{\mathrm{H}} + \frac{\sigma^2}{\rho}\mathbf{I}_B. \tag{3.7}$$

The estimation error covariance matrix $C_{jlk} \in C^{M\times M}$ is given by

$$\mathbf{C}_{jlk} = \mathbb{E}\left\{(\mathbf{h}_{jlk}^{\text{eff}} - \hat{\mathbf{h}}_{jlk}^{\text{eff}})(\mathbf{h}_{jlk}^{\text{eff}} - \hat{\mathbf{h}}_{jlk}^{\text{eff}})^{\mathrm{H}}\right\}$$
$$= \rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\left(1 - \frac{\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}B}{\sum_{\ell\in\mathcal{L}}\sum_{m=1}^{K}\frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{lk}}^{\mathrm{H}}\mathbf{v}_{i_{\ell m}} + \frac{\sigma^2}{\rho}}\right)\mathbf{I}_M \tag{3.8}$$

and the mean-squared error (MSE) is $MSE_{jlk} = tr(C_{jlk})$.

$$\text{SINR}_{jk}^{(\text{ul})} = \frac{p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}{\sum\limits_{l\in\mathcal{L}}\sum\limits_{m=1}^{K}p_{lm}\mathbb{E}_{\{\mathbf{h}\}}\{|\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jlm}|^2\} - p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2 + \sigma^2\mathbb{E}_{\{\mathbf{h}\}}\{\|\mathbf{g}_{jk}\|^2\}}. \tag{3.9}$$

3.2 Spectral Efficiency

The SE of an encoding/decoding scheme is the average number of bits of information, per complex-valued sample, that it can reliably transmit over the channel under consideration.

From this definition, it is clear that the SE is a deterministic number that can be measured in bit per complex-valued sample. Since there are B samples per second, an

equivalent unit of the SE is bit per second per Hertz, often written in short-form as bit/s/Hz. For fading channels, which change over time, the SE can be viewed as the average number of bit/s/Hz over the fading realizations, as will be defined below. In this monograph, we often consider the SE of a channel between a UE and a BS, which for simplicity we refer to as the "SE of the UE". A related metric is the information rate [bit/s], which is defined as the product of the SE and the bandwidth B. In addition, we commonly consider the sum SE of the channels from all UEs in a cell to the respective BS, which is measured in bit/s/Hz/cell.

3.3 Achievable UL Spectral Efficiencies

The channel estimates in Lemma 1 enable each BS to (semi)coherently detect the data signals from its UEs. In particular, we assume that BS j applies a linear receive combining vector $g_{jk} \in C^M$ to the received signal, as $g^H_{jk}y_j$, to amplify the signal from its kth UE and reject interference from other UEs in the spatial domain. We want to derive the ergodic achievable SE for any UE, where codewords span over both the Rayleigh fading and random locations of the interfering UEs—specific UE distributions are considered in Section IV. For notational convenience, we assume that $\beta = \frac{B}{K}$ is an integer that we refer to as the pilot reuse factor. The cells in L are divided into $\beta \geq 1$ disjoint subsets such that the same K pilot sequences are used within a set, while different pilots are used in different sets. We refer to this as fractional pilot reuse. The following lemma shows how the SE depends on the receive combining, for Gaussian codebooks where $x_{jk} \sim CN(0, 1)$.

Lemma 3.2 In the UL, an ergodic achievable SE of an arbitrary UE k in cell j is

$$\zeta^{(\mathrm{ul})} \left(1 - \frac{B}{S}\right) \mathbb{E}_{\{\mathbf{z}\}} \left\{ \log_2(1 + \mathrm{SINR}_{jk}^{(\mathrm{ul})}) \right\} \quad \textit{[bit/s/Hz]} \tag{3.10}$$

where the effective signal-to-interference-and-noise ratio (SINR), $SINR^{(ul)}_{jk}$, is given in (3.9) at the top of the page. The expectations $E_{\{z\}}\{\cdot\}$ and $E_{\{h\}}\{\cdot\}$ are with respect to UE positions and channel realizations, respectively.

3.4 Precoding and Beamforming

Precoding involves preprocessing of the transmit signal in an RF system. Precoding uses channel state information at the transmitter to improve performance and increase spectral efficiency. It is used to implement the superposition of multiple beams, including several different data streams of information for spatial multiplexing. Precoding is the transmitter signal processing needed to affect the received signal's maximization to specific receivers and antennas while reducing the interference to all other receivers and receiving antennas.

Precoding and beamforming are used together in Wi-Fi, 4G, and 5G systems, and the words are sometimes used interchangeably, but they are not identical. The word precoding refers more to a software implementation of communication theory, and beamforming refers more to the hardware implementation and the antennas in the system. And precoding generally refers to the transmitter side, while beamforming can be applied to both transmitters and receivers.

Precoding is a technique that exploits transmit diversity by weighting the information stream, i.e. the transmitter sends the coded information to the receiver to achieve pre-knowledge of the channel. The receiver does not have to know the channel state information. For example you are sending the information **x** and it will pass through the channel **h** and add Gaussian noise **n**. The received signal at the receiver front-end will be **y = hx + n.**

The receiver will have to know the information about **h** and **n**. It will suppress the effect of **n** by increasing SNR, but what about **h**?

Let us call the predicted channel $\mathbf{h_{est}}$ and for a system with precoder the information will be coded: $\mathbf{x/h_{est}}$. The received signal will be $\mathbf{y = (h/h_{est})x + n}$. If your prediction is perfect, $\mathbf{h=h_{est}}$ and **y = x + n** and it turns out to be the detection problem in Gaussian channels which is simple.

There are different types of precoding techniques

1. Linear precoding
2. Non-Linear precoding

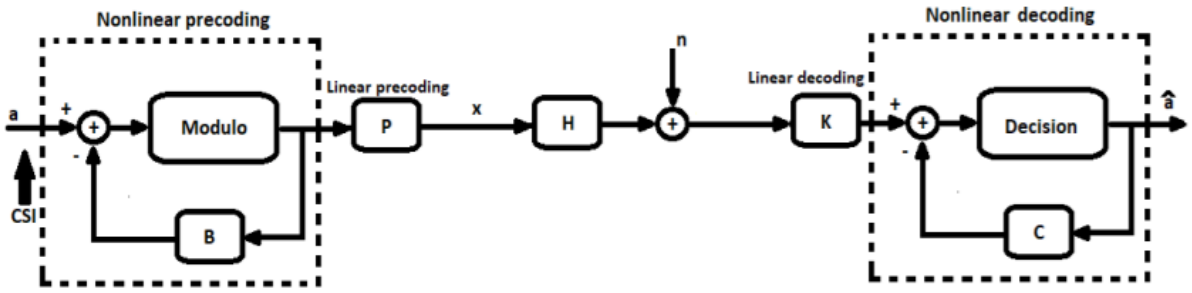


Figure 3.1 Generalized block diagram of communication systems with precoding and decoding techniques.

P is a feedforward matrix of linear precoding

B is a feedback matrix of linear precoding

K is a feedforward matrix of linear decoding

C is a feedback matrix of non-linear precoding.

3.4.1 Linear Precoding

From the block diagram, when B = 0 it acts as a linear precoding system.

Some of the linear precoding techniques:

1. Multicell – Minimum Mean Square Error (M-MMSE)
2. Singlecell – Minimum Mean Square Error (S-MMSE)
3. Maximum Ratio Transmission
4. Zero Forcing
5. Regularized Zero Forcing

3.4.2 M-MMSE Receive Combining

M-MMSE combining not only maximizes the instantaneous SINR but also minimizes the MSE in the data detection that is the average squared distance between the desired signal and the processed received signal.

By defining the diagonal matrix $P_l = \text{diag}(p_{l1}, \ldots, p_{lKl}) \in R^{Kl \times Kl}$ with the transmit powers of all UEs in cell l, we can collect the M-MMSE combining vectors for all UEs in cell j in a compact matrix form:

$$\mathbf{V}_j^{\text{M-MMSE}} = \left[\mathbf{v}_{j1} \ \cdots \ \mathbf{v}_{jK_j}\right] = \left(\sum_{l=1}^{L} \hat{\mathbf{H}}_l^j \mathbf{P}_l (\hat{\mathbf{H}}_l^j)^{\text{H}} + \sum_{l=1}^{L}\sum_{i=1}^{K_l} p_{li}\mathbf{C}_{li}^j + \sigma_{\text{UL}}^2 \mathbf{I}_{M_j}\right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j \quad (3.11)$$

where $\hat{H}^l_j$ is a matrix containing the estimates of all channels from UEs in cell l to BSj.

3.4.3 Alternative Receive Combining Schemes

Although M-MMSE combining is optimal, it is not so frequently used in the research literature. There are several reasons for this. One is the high computational complexity of computing the Mj × Mj matrix inverse in (3.13) when Mj is large. The complexity is also affected by the need to estimate the channels and acquiring the channel statistics of all UEs. Another reason is that the performance of M-MMSE is hard to analyze mathematically, while there are alternative schemes that can give more insightful closed-form SE expressions. A third reason is that receive combining schemes often are developed for single-cell scenarios and then applied heuristically in multicell scenarios.

We will now present the alternative receive combining schemes that are most common in the literature and explain how these are obtained as simplifications of M-MMSE combining. The alternative schemes are generally suboptimal and the conditions under which they are nearly optimal are generally not satisfied in practice. Hence, the alternative schemes provide lower SEs but are practically useful to reduce the computational complexity and/or the amount of channel estimates and channel statistics that are needed to compute the combining matrix Vj. If BS j only estimates the channels from its own UEs [8, 9, 10], we obtain the single-cell minimum mean-squared error (S-MMSE) combining scheme with

$$\mathbf{V}^{\text{S-MMSE}}_{\text{j}} = \left(\hat{\mathbf{H}}_j^j \mathbf{P}_j (\hat{\mathbf{H}}_j^j)^{\text{H}} + \sum_{i=1}^{K_j} p_{ji}\mathbf{C}_{ji}^j + \sum_{\substack{l=1 \\ l\neq j}}^{L}\sum_{i=1}^{K_l} p_{li}\mathbf{R}_{li}^j + \sigma_{\text{UL}}^2 \mathbf{I}_{M_j}\right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j. \quad (3.12)$$

If the channel conditions are good and the interfering signals from other cells are weak, we can neglect all the correlation matrices in (3.14) and obtain

$$\begin{aligned}\mathbf{V}_j^{\mathrm{RZF}} &= \left(\hat{\mathbf{H}}_j^j \mathbf{P}_j (\hat{\mathbf{H}}_j^j)^{\mathrm{H}} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j}\right)^{-1} \hat{\mathbf{H}}_j^j \mathbf{P}_j \\ &= \hat{\mathbf{H}}_j^j \mathbf{P}_j^{\frac{1}{2}} \left(\mathbf{P}_j^{\frac{1}{2}} (\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j \mathbf{P}_j^{\frac{1}{2}} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{K_j}\right)^{-1} \mathbf{P}_j^{\frac{1}{2}} \\ &= \hat{\mathbf{H}}_j^j \left((\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j + \sigma_{\mathrm{UL}}^2 \mathbf{P}_j^{-1}\right)^{-1}\end{aligned} \tag{3.13}$$

where the second equality follows from the first matrix identity in Lemma B.5[11]. We call this regularized zero-forcing (RZF) combining. The main benefit over S-MMSE is that a Kj × Kj matrix is inverted in (3.13) instead of an Mj × Mj matrix, which can substantially reduce the complexity since Mj >> Kj is typical in Massive MIMO. This benefit comes with a SE loss since, in general, the channel conditions will not be good to every UE and the interfering signals from other cells are non-negligible. The regularization terminology refers to the fact that (3.13) is a pseudo-inverse of the estimated channel matrix $\widehat{H}^j{}_j$ where the matrix that is inverted has been regularized by the diagonal matrix $\sigma^2{}_{UL}P^{-1}{}_j$. Regularization, a classic signal processing technique, improves the numerical stability of an inverse. In our case, it provides weighting between interference suppression (for small regularization terms) and maximizing the desired signals (for large regularization terms). The combining expression in (3.13) can be further approximated when the SNR is high. we can neglect the regularization term and obtain the zero-forcing (ZF) combining matrix

$$\mathbf{V}_j^{\mathrm{ZF}} = \hat{\mathbf{H}}_j^j \left((\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j\right)^{-1} \tag{3.14}$$

In low SNR conditions, we instead have $(\hat{\mathbf{H}}_j^j)^{\mathrm{H}} \hat{\mathbf{H}}_j^j + \sigma_{\mathrm{UL}}^2 \mathbf{P}_j^{-1} \approx \sigma_{\mathrm{UL}}^2 \mathbf{P}_j^{-1}$ and RZF in (3.15) is approximately equal to $\frac{1}{\sigma_{\mathrm{UL}}^2} \hat{\mathbf{H}}_j^j \mathbf{P}_j$. If we further remove the diagonal matrix 1 $\frac{1}{\sigma_{\mathrm{UL}}^2} \mathbf{P}_j$ (recall that the normalization of a combining vector does not affect the instantaneous UL SINR), we obtain

$$\mathbf{V}_j^{\mathrm{MR}} = \hat{\mathbf{H}}_j^j \tag{3.15}$$

which is known as MR combining. This scheme was considered already earlier, but the main difference is that we now use estimated channels instead of the exact ones (which are unknown in practice). Note that MR does not require any matrix inversion, in

contrast to the previously mentioned schemes. Since not every UE exhibits a low SNR in practice, it is expected that MR will provide lower SEs than RZF.

Theorem 3.1 Achievable SE, Uplink Let Lj (β) ⊂ L be the subset of cells that uses the same pilots as cell j. In the UL, an achievable SE in cell j is

$$\mathrm{SE}_j^{(\mathrm{ul})} = K\zeta^{(\mathrm{ul})}\left(1-\frac{B}{S}\right)\log_2\left(1+\frac{1}{I_j^{\mathrm{scheme}}}\right) \ \textit{[bit/s/Hz/cell]} \tag{3.16}$$

where the interference term is

$$I_j^{\mathrm{scheme}} = \sum_{l\in\mathcal{L}_j(\beta)\setminus\{j\}}\left(\mu_{jl}^{(2)}+\frac{\mu_{jl}^{(2)}-\left(\mu_{jl}^{(1)}\right)^2}{G^{\mathrm{scheme}}}\right) + \frac{\left(\sum_{l\in\mathcal{L}}\mu_{jl}^{(1)}Z_{jl}^{\mathrm{scheme}}+\frac{\sigma^2}{\rho}\right)\left(\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)}+\frac{\sigma^2}{B\rho}\right)}{G^{\mathrm{scheme}}} \tag{3.17}$$

depends on the receive combining scheme through G^{scheme} and $Z^{\mathrm{scheme}}{}_{jl}$. MR combining is obtained by $G^{\mathrm{MR}} = M$ and $Z^{\mathrm{MR}}{}_{jl} = K$, while ZF combining is obtained by $G^{\mathrm{ZF}} = M - K$

$$Z_{jl}^{\mathrm{ZF}} = \begin{cases} K\left(1-\frac{\mu_{jl}^{(1)}}{\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)}+\frac{\sigma^2}{B\rho}}\right) & \textit{if } l\in\mathcal{L}_j(\beta), \\ K, & \textit{if } l\notin\mathcal{L}_j(\beta). \end{cases} \tag{3.18}$$

Propagation Parameters

$$\mu_{jl}^{(\omega)} = \mathbb{E}_{\mathbf{z}_{lm}}\left\{\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^{\omega}\right\} \quad \textit{for } \omega = 1, 2. \tag{3.19}$$

The closed-form SE expressions in Theorem 1 are lower bounds on the ergodic capacity and slightly more conservative than the non-closed-form bound in Lemma 2 from [12]; We stress that the closedform SEs are only functions of the pilot allocation and the propagation parameters $\mu^{(1)}{}_{jl}$ and $\mu^{(2)}{}_{jl}$ defined in (3.19). The latter two are the average ratio between the channel variance to BS j and the channel variance to BS l, for an

arbitrary UE in cell l, and the second-order moment of this ratio, respectively. These parameters are equal to 1 for j = l and otherwise go to zero as the distance between BS j and cell l increases. The SE expression manifests the importance of pilot allocation, since the interference term in (3.17) contains summations that only considers the cells that use the same pilots as cell j. The first term in (3.17) describes the pilot contamination, while the second term is the inter-user interference. The difference between MR and ZF is that the latter scheme cancels some interference through Z scheme jl, at the price of reducing the array gain G^{scheme} from M to M − K. ZF combining only actively suppresses intra-cell interference, while the inter-cell interference is passively suppressed just as in MR combining. Further interference rejection can be achieved by coordinating the combining across cells, such that both intra-cell and inter-cell interference are actively suppressed by the receive combining. We propose a new fullpilot zero-forcing (P-ZF) combining.

Theorem 3.2 Let $L_l(\beta) \subset L$ be the subset of cells that uses the same pilots as cell l. In the UL, an achievable SE in cell j with P-ZF combining is given by for $G^{\text{P-ZF}}$ =M–B

$$Z_{jl}^{\text{P-ZF}} = K \left(1 - \frac{\mu_{jl}^{(1)}}{\sum_{\ell \in \mathcal{L}_l(\beta)} \mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}} \right). \tag{3.20}$$

The SE expressions were derived assuming that M and K are the same in all cells, for notational brevity. However, the results in this section are straightforward to extend to cell specific M and K values. The deterministic function $d_j(z)$ gives the variance of channel attenuation from BS j to any UE position z. More clearly $d_j(z_{lk})$ denotes channel variance between BS j and UE k in the cell j.

Corollary 3.1 Looking jointly at the UL and DL, an achievable SE in cell j is

$$\begin{aligned} \mathrm{SE}_j &= \mathrm{SE}_j^{(\mathrm{ul})} + \mathrm{SE}_j^{(\mathrm{dl})} \\ &= K\left(1 - \frac{B}{S}\right)\log_2\left(1 + \frac{1}{I_j^{\mathrm{scheme}}}\right) \textit{ [bit/s/Hz/cell]} \end{aligned} \tag{3.21}$$

where the interference term I scheme j for UE k is given Theorem 3.1 for MR and ZF and in Theorem 3.2 for P-ZF. This SE can be divided between the UL and DL arbitrarily using any positive fractions $\zeta^{(ul)}$ and $\zeta^{(dl)}$, with $\zeta^{(ul)} + \zeta^{(dl)} = 1$.

3.5 SE computation

1. Compute receive combining vectors v_{jk} and resulting $SINR^{UL}_{jk}$.
2. Compute "instantaneous" SE:

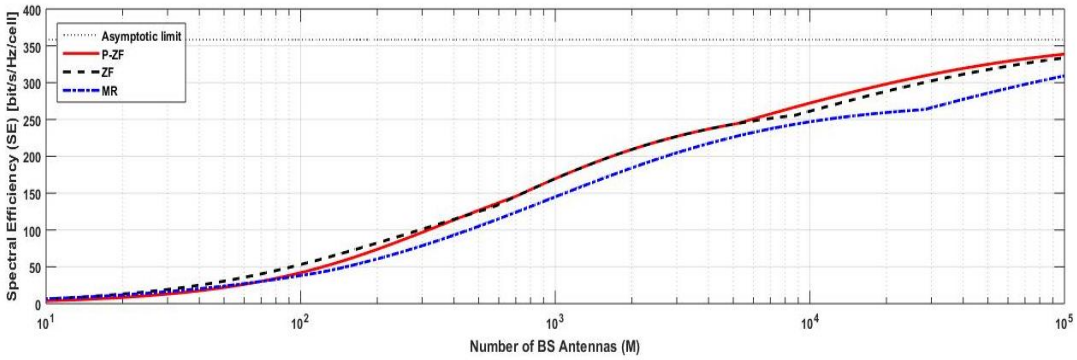
$$\mathsf{SE}_{jk}^{\mathrm{UL,inst.}} = \frac{\tau_u}{\tau_c}\log_2\left(1 + \mathsf{SINR}_{jk}^{\mathrm{UL}}\right) \tag{3.22}$$

3. Average $SE^{UL,inst.}_{jk}$ over estimated channels to obtain $SE^{UL}_{jk.}$
4. Obtain simulation results by considering the SEs of all UEs for different shadow fading realizations and UE locations.

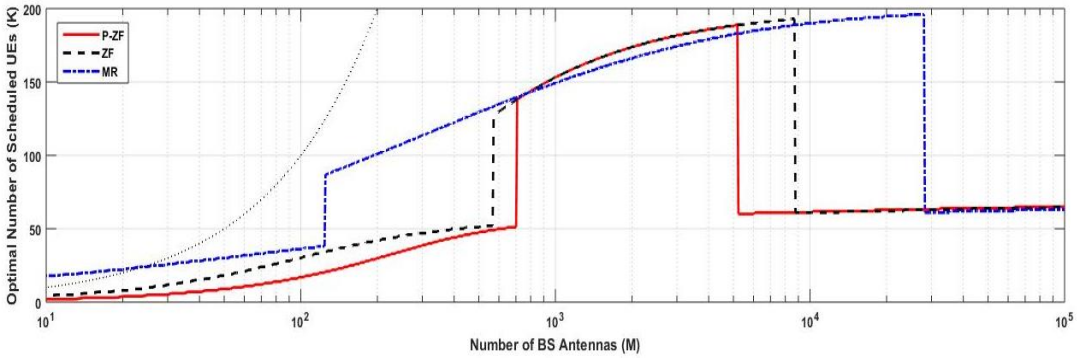
3.6 Optimizing SE for Different Interference Levels

We simulate the SE in an arbitrary cell on the hexagonal grid and take all non-negligible interference into account. The UEs can be anywhere in the cells, but at least 0.14r from the serving BS (this makes the analysis independent of r). Since the SE expressions from earlier are the same for the UL and DL, except for the fractions $\zeta^{(ul)}$ and $\zeta^{(dl)}$, we simulate the sum of these SEs and note that it can be divided arbitrarily between the UL and DL. The same linear processing schemes are used in both directions. The simulations consider MR, ZF, and P-ZF precoding/combining, and all results are obtained by computing the closed-form expressions from earlier in this chapter for different parameter combinations. The simulations were performed using Matlab.

For each number of antennas, M, we optimize the SE with respect to the number of UEs K and the pilot reuse factor β (which determine B = βK) by searching the range of all reasonable integer values. We set the coherence block length to S = 400 (e.g., 2 ms coherence time and 200 kHz coherence bandwidth), set the SNR to ρ/σ2 = 5 dB, and pick κ = 3.7 as pathloss exponent. The impact of changing the different system parameters.



(a) Optimized SE per cell.



(b) Corresponding optimal number of UEs: K*

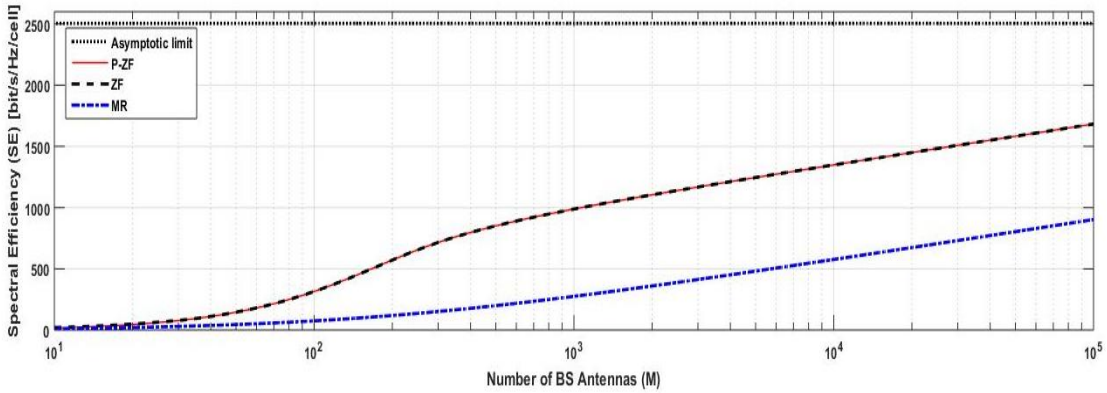
Figure 3.2 Simulation of optimized SE, as a function of M, with average inter-cell interference.

We consider three propagation environments with different severity of inter-cell interference:

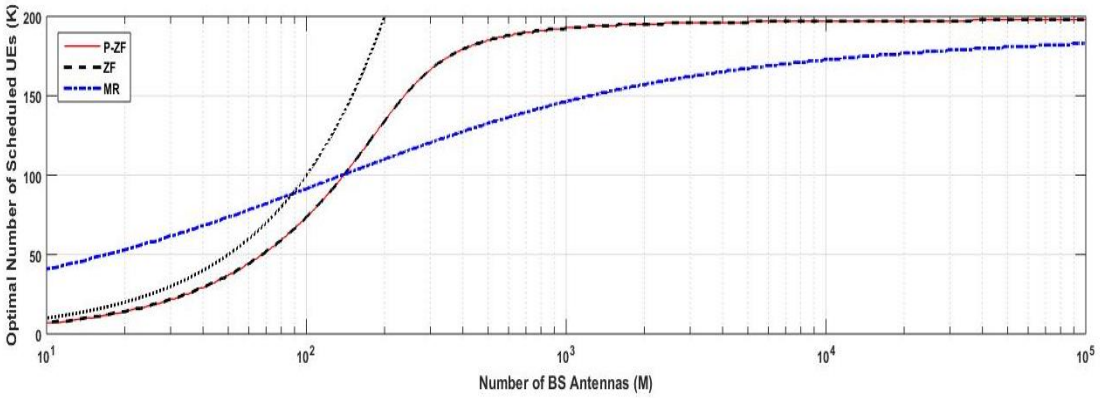
1. Average case: Averaging over uniform UE locations in all cells.
2. Best case: All UEs in other cells are at the cell edge furthest from BS j (for each j).
3. Worst case: All UEs in other cells are at the cell edge closest to BS j (for each j).

The corresponding values on the parameters $\mu^{(1)}_{jl}$ and $\mu^{(2)}_{jl}$ were computed by Monte-Carlo simulations with 106 UE locations in each cell.

The best case is overly optimistic since the desirable UE positions in the interfering cells are different with respect to different cells. However, it gives an upper bound on what is achievable by coordinated scheduling across cells. The worst case is overly pessimistic since the UEs cannot all be at the worst locations, with respect to all other cells, at the same time. The average case is probably the most applicable in practice, where the averaging comes from UE mobility, scheduling, and random switching of pilot sequences between the UEs in each cell. Results for the average case are shown in Fig. 3.2, the best case in Fig. 3.3, and the worst case in Fig. 3.4. The optimized SE and the corresponding K* are shown in (a) and (b), respectively.

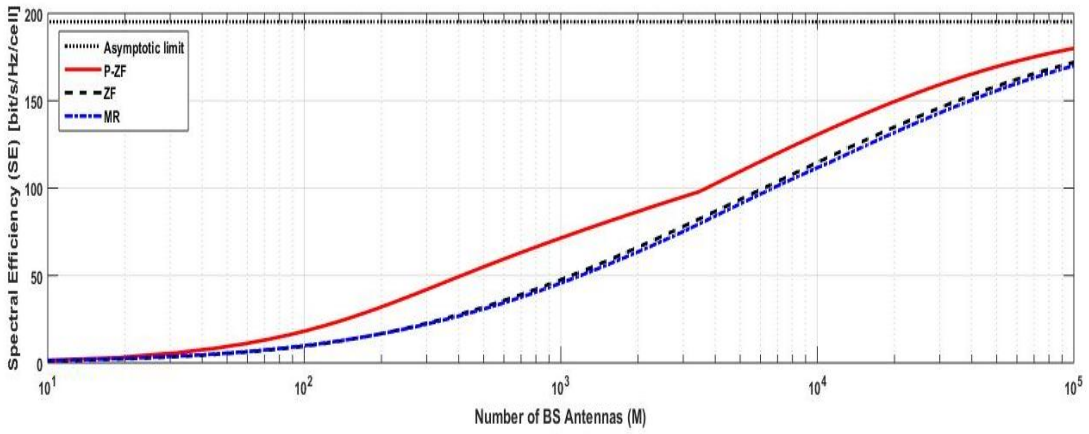


(a) Optimized SE per cell.

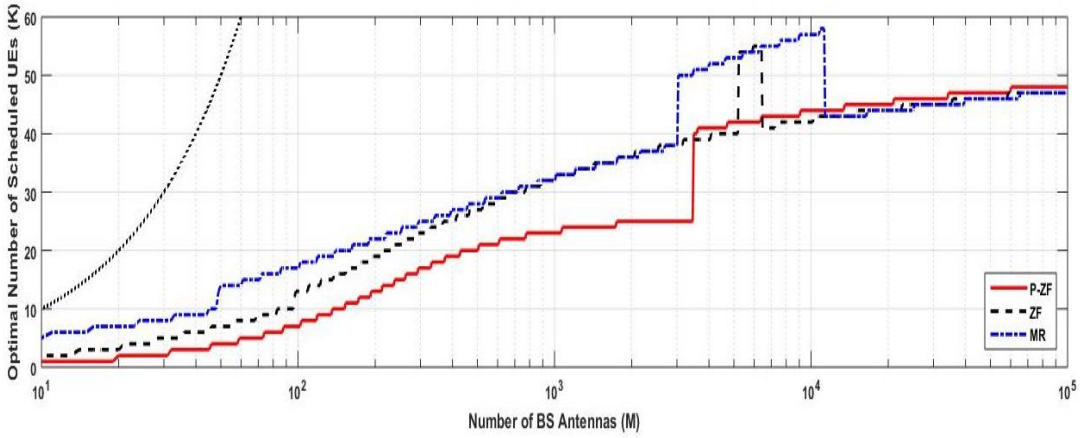


(b) Corresponding optimal number of UEs: K*

Figure 3.3 Simulation of optimized SE, as a function of M, with best-case inter-cell interference.



(a) Optimized SE per cell.



(b) Corresponding optimal number of UEs: K*

Figure 3.4 Simulation of optimized SE, as a function of M, with worst-case inter-cell interference.

The achievable SEs (per cell) are very different between the best-case interference and the two other cases—this confirms the fact that results from single-cell analysis of massive MIMO is often not applicable to multi-cell cases (and vice versa). ZF brings much higher SEs than MR under the best case intercell interference, since then the potential gain from mitigating intra-cell interference is very high. P-ZF is equivalent to ZF in the best case, but excels under worst case inter-cell interference since it can actively suppress also inter-cell interference. In the realistic average case, the optimized SEs are rather similar for MR, ZF, and P-ZF; particularly in the practical range of $10 \leq M \leq 200$ antennas. In all cases, the largest differences appear when the number of antennas is very large (notice the logarithmic M-scales). At least M = 105 is needed to

come close to the asymptotic limit, and many more antennas are required under best case interference. Clearly, the asymptotic limits should not be used as performance indicators since unrealistically many antennas are needed for convergence.

As seen from Figs. 3.2 - 3.4, the main difference between MR, ZF, and P-ZF is not the values of the optimized SE but how they are achieved; that is, which number of UEs K* and which pilot reuse factor β that are used. The general behaviour is that larger M implies a higher K* and a smaller β, because the channels become more orthogonal with M. Since the reuse factor is an integer, K* changes non-continuously when β is changed; smaller β allows for larger K*, and vice versa. MR schedules the largest number of UEs and switches to a smaller reuse factor at fewer antennas than the other schemes. In contrast, P-ZF schedules the smallest number of UEs and has the highest preference of large reuse factors, since this it can suppress more inter-cell interference in these cases. Simply speaking, MR gives low per-user SEs to many UEs (sometimes more than M), while ZF and P-ZF give higher per-user SEs to fewer UEs.

3.7 SE Comparison of Different Combining Schemes

We will now compare the different receive combining schemes. In this simulation, we consider K = 10 UEs per cell and a varying number of BS antennas. There are f_K pilots in each coherence block and the remaining $\tau_c - f_K$ samples are used for UL data transmission. We use Gaussian local scattering with ASD $\sigma_\phi = 10^\circ$ as channel model.

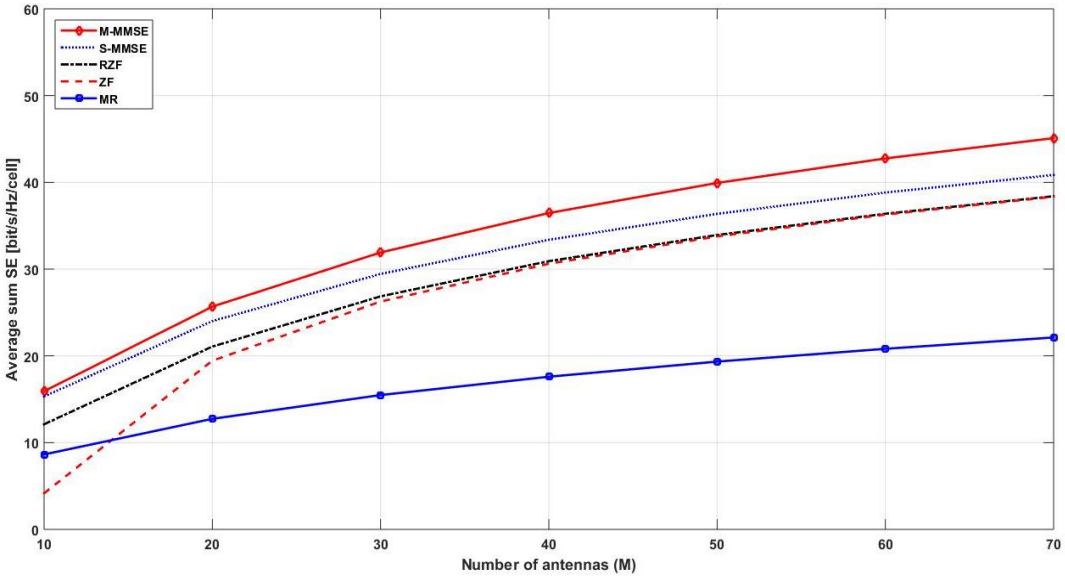
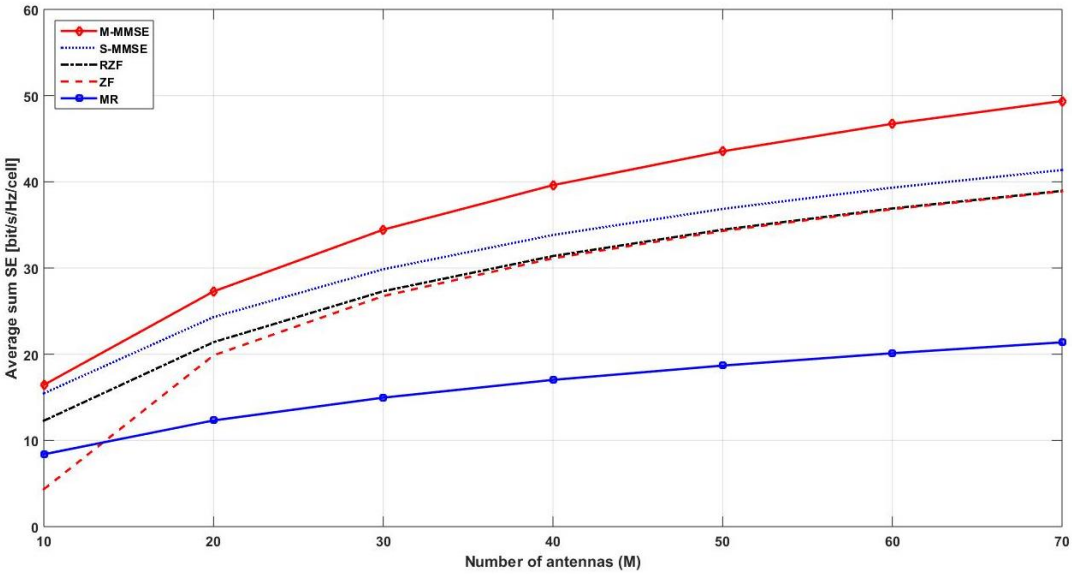
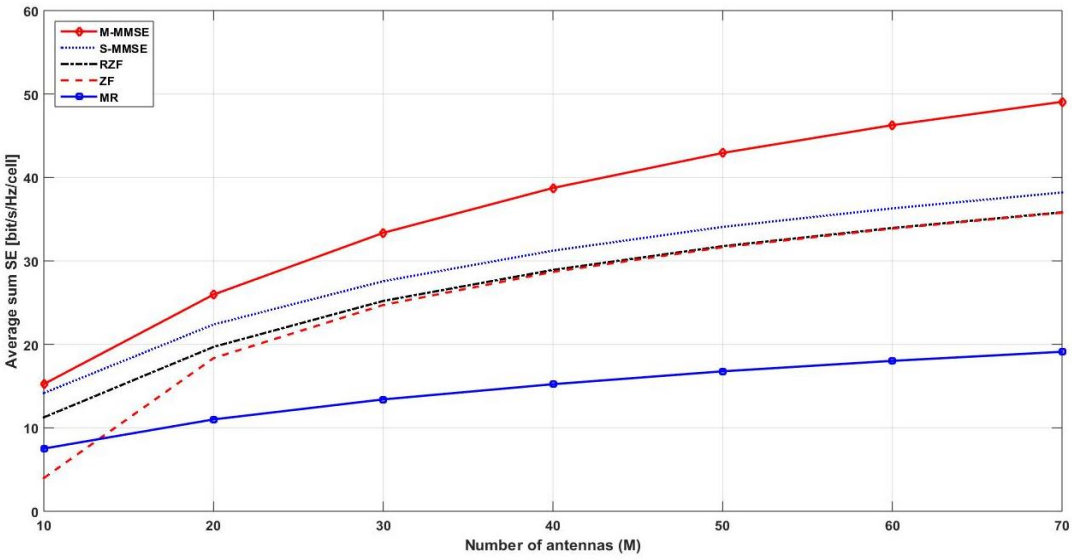


Figure 3.5 Average UL sum SE as a function of the number of BS antennas for different combining schemes. There are K = 10 UEs per cell and the same K pilots are reused in every cell.

Figure 3.5 shows the average UL sum SE as a function of the number of BS antennas for universal pilot reuse with f = 1. M-MMSE gives the largest SE in Figure 3.5. The SE reduces a little with every approximation that is made to obtain a scheme with lower complexity than M-MMSE. The S-MMSE scheme provides lower SE than M-MMSE, but 5%–10% higher SE than RZF and ZF. Note that RZF and ZF give essentially the same SE in the range $M \geq 20$ that is of main interest in Massive MIMO, but the SE with ZF deteriorates quickly for $M < 20$ since the BS does not have enough degrees of freedom to cancel the interference without also cancelling a large part of the desired signal. Hence, ZF should be avoided to achieve a robust implementation. Interestingly, MR provides only half the SE of the other schemes, it also reduces complexity by 10% as compared to RZF and requires no matrix inversions.



(a) Pilot reuse factor f = 2.



(b) Pilot reuse factor f = 4

Figure 3.6 Average UL sum SE as a function of the number of BS antennas for different combining schemes. There are K = 10 UEs per cell and either 2K or 4K pilots that are reused across cells.

Figure 3.6 shows the average sum SE with a non-universal pilot reuse f. In particular, we consider cases where each pilot is reused in every second or fourth cells. This is referred to as having a pilot reuse factor of f = 2 and f = 4, respectively.

Table 3.1 Average UL sum SE [bit/s/Hz/cell] for M = 100 and K = 10 for different pilot reuse factors f.

Scheme	f = 1	f = 2	f = 4
M-MMSE	50.32	55.10	55.41
S-MMSE	45.39	45.83	42.41
RZF	42.83	43.37	39.99
ZF	42.80	43.34	39.97
MR	25.25	24.41	21.95

The increased number of pilots reduces the pre-log factor, since $\tau_u = \tau_c - f$, but it also increases the instantaneous SINR since better channel estimates with less pilot contamination are obtained. M-MMSE benefits particularly much from having $f > 1$, because it can better suppress the interference from UEs in the surrounding cells when these UEs use other pilots. A reuse factor of 4 gives the highest SE with M-MMSE. S-MMSE, RZF, and ZF give comparable SE to each other for all f, and achieve the highest SE with f = 2. The SE of MR reduces when f is increased since the improved estimation quality does not outweigh the reduced pre-log factor when the estimate is only used to coherently combine the desired signal and not to cancel interference. These properties are quantified in Table 4.3, which summarizes the sum SE of all schemes with M = 100 and different f. The numbers can be compared with the SE 2.8 bit/s/Hz/cell achieved by a contemporary LTE system (see Remark 4.1 from [11]). With all pilot reuse factors, M-MMSE and RZF provide more than an order-of-magnitude higher SE per cell. With MR, the gain is a factor 7–9.

In summary, there are basically three combining schemes to choose from, if the running example is implemented in practice. M-MMSE provides the highest SE using the highest complexity, and should be implemented using non-universal pilot reuse. MR has the lowest complexity, but also delivers the lowest SE. Finally, RZF strikes a good balance between SE and complexity; it can double the SE as compared to MR while the computational complexity is only some tens of percentages higher. In practice, RZF is always a better choice than ZF, since it achieves similar or better SE and lacks ZF's robustness issues when $M \approx K$. However, ZF is a fairly common scheme

in the literature since it allows to compute closed-form SE expressions in the special case of spatially uncorrelated channels [13, 14, 15]. Approximate closed-form expressions can be computed with M-MMSE, S-MMSE, and RZF [8, 16]. The closed-form expressions predict the SE that is practically achievable with different schemes and are particularly useful for resource allocation and optimization.

3.8 Summary

This chapter mainly focussed on spectral efficiency in Massive MIMO and different precoding schemes to increase spectral efficiency. Out of all different schemes discussed, the computational complexity order and spectral efficiency order is given as

M-MMSE >S-MMSE >R-ZF> ZF> MR

There is always a trade-off between the computational complexity and achievable spectral efficiency. So out of all the schemes discussed above, it is better to go for R-ZF scheme which has good balance between complexity and spectral efficiency.

CHAPTER 4
MAXIMIZATION OF ENERGY EFFICIENCY IN MASSIVE MIMO

Introduction

In this section, we analyse the energy efficiency (EE) of Massive MIMO based on a realistic circuit power (CP) consumption model. Before looking into this, we explain in Section 4.1 why power consumption (PC) is a major concern for future cellular networks. In Section 4.1, we show that Massive MIMO can potentially improve the area throughput while providing substantial power savings. The asymptotic behaviour of the transmit power when the number of BS antennas grows towards infinity is also studied, which proves how quickly the transmit power can be reduced with the number of antennas while achieving a non-zero asymptotic SE. Section 4.3 formally introduces the EE metric and provides basic insights into the EE-SE trade-off, as a function of the key system parameters, such as the number of BS antennas and UEs. A tractable and realistic CP model for Massive MIMO networks is developed in Section 4.6.

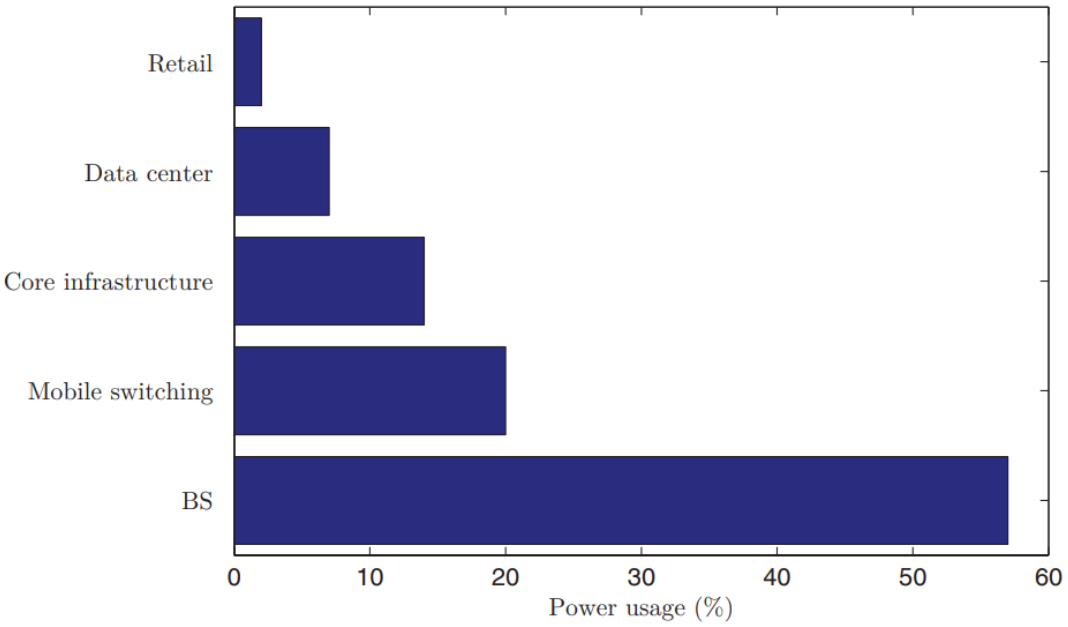


Figure 4.1 Breakdown of power consumed by cellular networks.

4.1 Motivation

If the annual traffic growth rate of cellular networks continues to be in the range of 41%–59%, the area throughput will have to increase by a factor of 1000 over the next 15–20 years [16]. If no active countermeasures are taken, the solution to the "1000× data challenge" will increase the PC prohibitively. This is because current networks are based on a rigid central infrastructure, that is powered by the electric grid and designed to maximize the throughput and the traffic load that each cell can handle. The PC is mainly determined by the peak throughput and varies very little with the actual throughput of the cell. This is problematic since the number of active UEs in a cell can change rapidly due to changes in user behaviours and the bursty nature of packet transmission. The measurements reported in [17] show that the daily maximum network load is 2–10 times higher than the daily minimum load. Hence, a lot of energy is wasted at the BSs in non-peak hours.

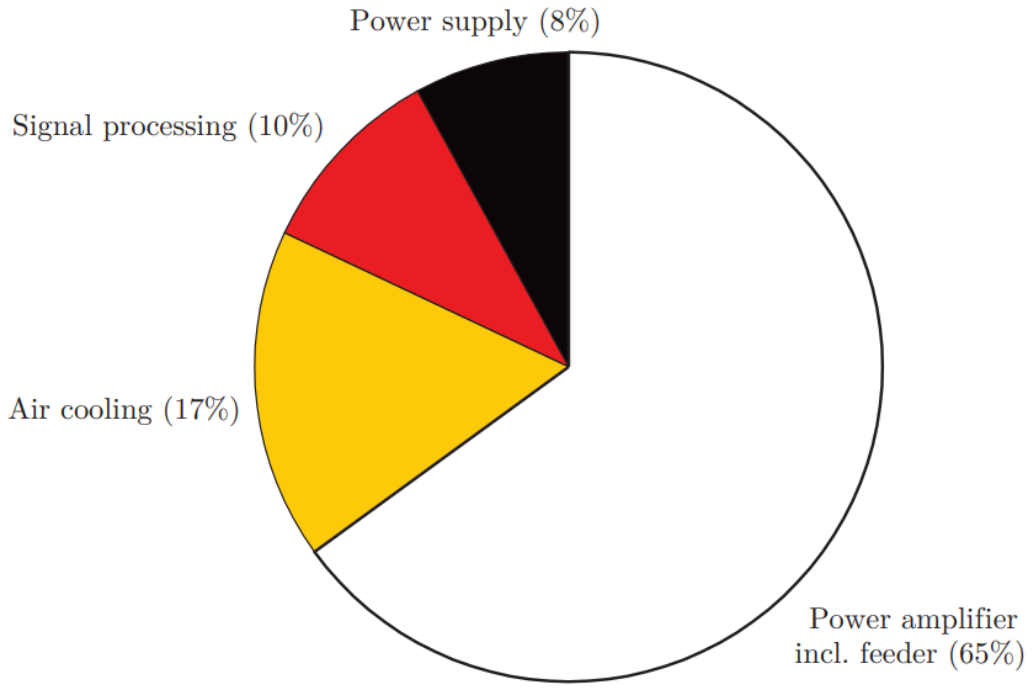


Figure 4.2 Percentage of power consumed by different components of a coverage tier BS

A quite remarkable effort has been devoted to reducing the PC of UEs, in order to enhance their battery lifetime. Academia and industry alike have recently shifted their attention towards the BSs. According to figures from Vodafone [18] shown in Figure 4.1, BSs account for almost 60% of the total power consumed by a cellular network, while 20% is consumed by mobile switching equipment, and around 15% by the core infrastructure. The rest is consumed by data centers and retail/offices. The total power consumed by a BS is composed of fixed (traffic-independent) and variable (traffic-dependent) parts. Figure 4.2 breaks down how different parts of a BS in the coverage tier contribute to the total PC. The fixed part, including control signaling and power supply, accounts for around one quarter of the total consumed power. This amount is not efficiently used during non-peak traffic hours or, even worse, it is completely wasted when no UE is active within the coverage area of a BS (as frequently happens in rural areas). The most significant portion of power is consumed in the power amplification process. Shockingly, 80%–95% of this power is dissipated as heat in the power amplifiers (PAs), since the total efficiency of currently deployed PAs is generally in the range of 5%–20% (depending on the communication standard and the equipment's condition). This is due to the fact that the modulation schemes used in contemporary communication standards, such as LTE, are characterized by strongly varying signal envelopes with peak-to-average power ratios that exceed 10dB. To avoid distortions of the transmitted signals, the PAs have to operate well below saturation.

Massive MIMO aims at evolving the coverage tier BSs by using arrays with a hundred or more antennas, each transmitting with a relatively low power. This allows for coherent multiuser MIMO transmission with tens of UEs being spatially multiplexed in both UL and DL of each cell. The area throughput is improved by the multiplexing gain. However, the throughput gains provided by Massive MIMO come from deploying more hardware (i.e., multiple RF chains per BS) and digital signal processing (i.e., SDMA combining/precoding) which, in turn, increase the CP per BS. Hence, the overall EE of the network, defined later as "how much energy it takes to achieve a certain amount of work", can be optimized only if these benefits and costs are properly balanced.

4.2 Transmit Power Consumption

A metric that is used to measure the transmit power consumed by a wireless network is the area transmit power (ATP), which is defined as the network-average power usage for data transmission per unit area. This metric is measured in W/km^2.

ATP = transmit power [W/cell] · D [cells/km2] (4.1)

where D is the average cell density. Consider the DL of a Massive MIMO network with L cells. BS j communicates with Kj UEs, BS j uses the precoding vector $w_{jk} \in C^{Mj}$ to transmit the data signal $ς_{jk} \sim NC(0, ρ_{jk})$ intended for UE k in cell j. Since the precoding vector is normalized as $E\{||w_{jk}||^2\} = 1$, the transmit power allocated to this UE is equal to the signal variance $ρ_{jk}$. The ATP of BS j is thus given by

$$\mathrm{ATP}_j^{\mathrm{DL}} = D \sum_{k=1}^{K_j} \rho_{jk}. \tag{4.2}$$

The corresponding UL expression is obtained if ρjk is replaced with pjk.

For sufficiently large number of BS antennas, Massive MIMO can achieve more than an order-of-magnitude higher area throughput than current networks, while also providing more than an order-of magnitude ATP savings. Notice that the division of the total transmit power among M antennas results into a low transmit power per antenna. With M = 100 and a total DL transmit power of 1W, we have only 10 mW per antenna in the considered scenario. This allows replacing the expensive high-power PAs used in current cellular networks (that consume most of the power in a BS) by hundreds of low-cost low power PAs with output power in the mW range. With a sufficiently low power per antenna, we might not even need to amplify the signal by a dedicated PA, but feed each antenna directly from a circuit. This can have very positive effects on the consumed power. It is important to note that these savings are obtained at the cost of deploying multiple RF chains per BS and using combining/precoding schemes, whose computational complexities depend on the number of BS antennas and UEs. This, in turn, increases the CP of the network, the ATP metric does not provide

the right insights into the net reduction in consumed power provided by Massive MIMO.

4.3 Definition of Energy Efficiency

In a broad sense, EE refers to how much energy it takes to achieve a certain amount of work. This general definition applies to all fields of science, from physics to economics, and wireless communication is no exception [19]. Unlike many fields wherein the definition of "work" is straightforward, in a cellular network it is not easy to define what exactly one unit of "work" is. The network provides connectivity over a certain area and it transports bits to and from UEs. Users pay not only for the delivered number of bits but also for the possibility to use the network anywhere at any time. Moreover, grading the performance of a cellular network is more challenging than it first appears, because the performance can be measured in a variety of different ways and each such performance measure affects the EE metric differently [19]. Among the different ways to define the EE of a cellular network, one of the most popular definitions takes inspiration from the definition of SE, that is, "the SE of a wireless communication system is the number of bits that can be reliably transmitted per complex-valued sample".

Definition 4.1 (Energy Efficiency). The EE of a cellular network is the number of bits that can be reliably transmitted per unit of energy. According to the definition above, we define the EE as

$$\text{EE} = \frac{\text{Throughput [bit/s/cell]}}{\text{Power consumption [W/cell]}} \tag{4.3}$$

which is measured in bit/Joule and can be seen as a benefit-cost ratio, where the service quality (throughput) is compared with the associated costs (power consumption). Hence, it is an indicator of the network's bit-delivery efficiency.

The transmit power only captures a part of the overall PC. Moreover, we notice that the transmit power does not represent the effective transmit power (ETP) needed for transmission since it does not account for the efficiency of the PA. The efficiency

of a PA is defined as the ratio of input power to output power. When the efficiency is low, a large portion of the power is dissipated as heat. To correctly evaluate the EE, the PC must be computed on the basis of the ETP (not of the radiated transmit power) and of the CP required for running the cellular networks.

$$\underbrace{\text{PC}}_{\text{Power consumption}} = \underbrace{\text{ETP}}_{\text{Effective transmit power}} + \underbrace{\text{CP}}_{\text{Circuit power}}. \tag{4.4}$$

A common model for CP is $CP = P_{FIX}$, where the term P_{FIX} is a constant quantity, which may account for the fixed power required for control signaling and load-independent power of baseband processors and backhaul infrastructure. However, this is not sufficiently accurate for comparing systems with different hardware setups (e.g., with a different number of antennas) and varying network loads2 because it does not account for the power dissipation in the analog hardware and in the digital signal processing. Therefore, there are many ways in which an overly simplistic CP model may lead to wrong conclusions. Detailed CP models are needed to evaluate the power consumed by a practical network and to identify the non-negligible components. Clearly, the complexity of this task makes a certain level of idealization unavoidable.

4.4 Wyner Model

We consider a two-cell network where the average channel gain between a BS and every UE in a cell is identical. This is a tractable model for studying the basic properties of cellular communications, due to the small number of system parameters. It is an instance of the Wyner model, initially proposed by Aaron Wyner. In the UL scenario shown in Figure 4.3, the UEs in cell 0 transmit to their serving BS, while the UL signals from the UEs in cell 1 leak into cell 0 as interference.

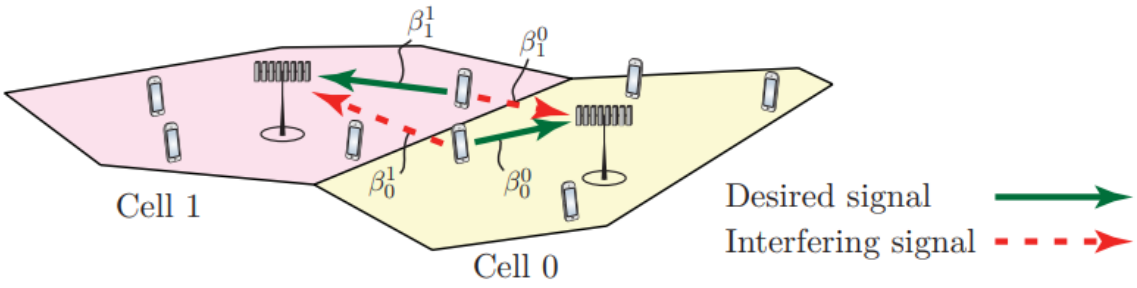


Figure 4.3 Illustration of the notion of desired and interfering UL signals in a two-cell network.

The average channel gain from a UE in cell 0 to its serving BS is denoted by $\beta^0{}_0$, while the interfering signals from UEs in cell 1 have an average channel gain of $\beta^0{}_1$. Similarly, the average channel gain from a UE in cell 1 to its serving BS is denoted by $\beta^1{}_1$, while the interfering signals from UEs in cell 0 have an average channel gain of $\beta^1{}_0$. Notice that the superscript indicates the cell of the receiving BS and the subscript indicates the cell that the transmitting UE resides in. The average channel gains are positive quantities that are often very small since the signal energy decays quickly with the propagation dimensionless distance; values in the range from −70 dB to −120 dB are common within the serving cell, while even smaller values appear for interfering signals. As shown later, it is not the absolute values that are of main importance when computing the SE, but the relative strength of the interference as compared to the desired signals. For simplicity, we assume that the intra-cell channel gains are equal (i.e., $\beta^0{}_0 = \beta^1{}_1$) and that the inter-cell channel gains are equal as well (i.e., $\beta^0{}_1 = \beta^1{}_0$); this is commonly assumed in the Wyner model. We can then define the ratio $\bar{\beta}$ between the inter-cell and intra-cell channel gains as

$$\bar{\beta} = \frac{\beta_1^0}{\beta_0^0} = \frac{\beta_0^1}{\beta_0^0} = \frac{\beta_1^0}{\beta_1^1} = \frac{\beta_0^1}{\beta_1^1}. \tag{4.5}$$

This ratio will be used in the analysis of both UL and DL. We typically have $0 \leq \bar{\beta} \leq 1$, where $\bar{\beta} \approx 0$ corresponds to a negligibly weak inter-cell interference and $\bar{\beta} \approx 1$ means that the inter-cell interference is as strong as the desired signals (which may happen for UEs at the cell edge).

4.5 Energy-Spectral Efficiency Trade-off

the SE of a cell can be increased by using more transmit power, deploying multiple BS antennas, or serving multiple UEs per cell. All these approaches inevitably increase the PC of the network, either directly (by increasing the transmit power) or indirectly (by using more hardware), and therefore may potentially reduce the EE. However, this is not necessarily the case. In fact, there exist operating conditions under which it is possible to use these techniques to jointly increase SE and EE. To explore this in more detail, the EE-SE trade-off is studied next and the impact of different network parameters and operating conditions are investigated. For simplicity, we focus on the UL of the two-cell Wyner model (i.e., L = 2) illustrated in Figure 4.3 (similar results can be obtained for the DL) and consider only uncorrelated Rayleigh fading channels over a bandwidth B, under the assumption that the BSs are equipped with M antennas, have perfect channel knowledge, and use MR combining.

4.5.1 Impact of Multiple BS antennas

Assume that there is only one active UE (i.e., K = 1) in cell 0 and that no interfering signals come from cell 1. an achievable SE of the UE in cell 0 is

$$\mathsf{SE}_0 = \log_2\left(1 + (M-1)\mathsf{SNR}_0\right) = \log_2\left(1 + (M-1)\frac{p}{\sigma^2}\beta_0^0\right) \tag{4.6}$$

where p is the transmit power, σ^2 is the noise power, and $\beta^0{}_0$ denotes the average channel gain of the active UE. We have omitted the superscript "NLoS", since we do not consider the LoS case here. To evaluate the impact of M on the EE, we distinguish between two different cases in the computation of the PC:

i) The CP increase due to multiple BS antennas is neglected;

ii) The CP increase is accounted for.

Assume, for the moment, that the CP of cell 0 consists only of the fixed power P_{FIX}; that is, $CP_0 = P_{FIX}$. Hence, the corresponding EE of cell 0 is

$$\mathsf{EE}_0 = \frac{B\log_2\left(1 + (M-1)\frac{p}{\sigma^2}\beta_0^0\right)}{\frac{1}{\mu}p + P_{\mathrm{FIX}}} \tag{4.7}$$

where B is the bandwidth and 1 μ p accounts for the ETP with $0 < \mu \leq 1$ being the PA efficiency. For a given SE, denoted as SE0, from (4.7) we obtain the required transmit power as

$$p = \frac{(2^{\mathrm{SE}_0} - 1)}{(M-1)} \frac{\sigma^2}{\beta_0^0} \tag{4.8}$$

$$\mathrm{EE}_0 = \frac{B\,\mathrm{SE}_0}{(2^{\mathrm{SE}_0} - 1)\frac{\nu_0}{M-1} + P_{\mathrm{FIX}}} \tag{4.9}$$

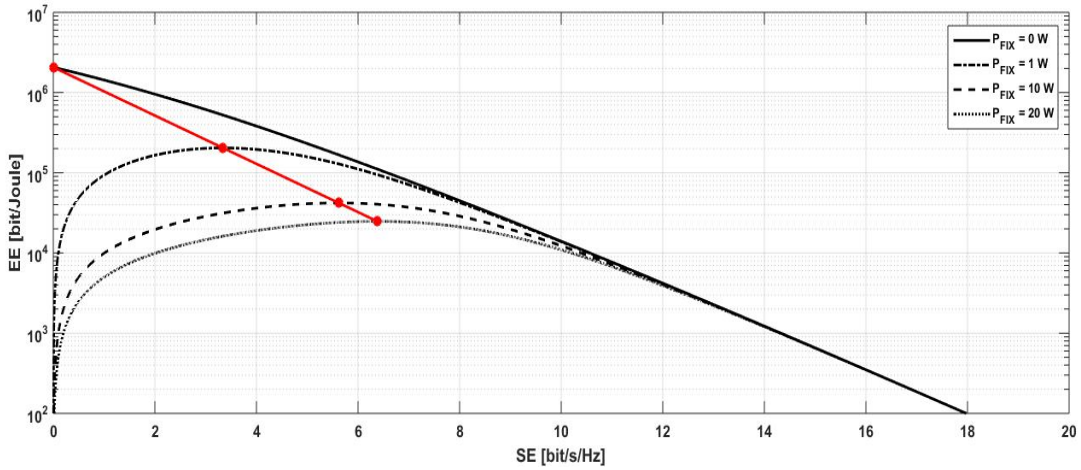


Figure 4.4 SE and EE relation in (4.12) for different values of CP = P_{FIX}

$$\nu_0 = \frac{\sigma^2}{\mu\beta_0^0}. \tag{4.10}$$

The above expression provides the relation between EE and SE for the UE in cell 0. Figure 4.4 illustrates the EE versus SE for M = 10, B = 100 kHz, $\sigma^2/\beta^0_0 = -6$ dBm, μ = 0.4, and $P_{FIX} \in \{0, 1, 10, 20\}$W. As we can see, if $P_{FIX} = 0$, there is a monotonic decreasing trade-off between EE and SE because (4.9) reduces to

$$\mathrm{EE}_0 = \frac{B\,\mathrm{SE}_0}{(2^{\mathrm{SE}_0} - 1)\frac{\nu_0}{M-1}}. \tag{4.11}$$

In other words, if the CP is not accounted for, an increased SE always comes at the price of a decreased EE. If, however, $P_{FIX} > 0$ (as it is in practice), then EE_0 is a unimodal function that increases for when SE_0 takes larger values. We can also see from Figure 4.4 that the EE-SE curve becomes flatter with increasing values of P_{FIX},

such that the range of SE values for which almost the same EE is achieved gets larger. values of SE_0 such that $(2^{SE0} - 1)\, \nu_0/(M-1) < P_{FIX}$ and decreases to zero as $SE_0/(2^{SE0}-1)$ when SE_0 takes larger values. We can also see from Figure 4.4 that the EE-SE curve becomes flatter with increasing values of P_{FIX}, such that the range of SE values for which almost the same EE is achieved gets larger.

To get some analytical insights into the EE-optimal point, we take the derivative of EE0 in (4.9) with respect to SE_0 and equate it to zero. We observe that the maximum EE (called EE*) and its corresponding SE (called SE*) satisfy the following identity:

$$\mathrm{SE}^{\star} = \frac{W\left((M-1)\frac{P_{\mathrm{FIX}}}{\nu_0 e} - \frac{1}{e}\right) + 1}{\log_e(2)} \tag{4.12}$$

$$\mathrm{EE}^{\star} = \frac{(M-1)Be^{-W\left((M-1)\frac{P_{\mathrm{FIX}}}{\nu_0 e} - \frac{1}{e}\right) - 1}}{\nu_0 \log_e(2)} \tag{4.13}$$

where W(·) is the Lambert function and e is Euler's number.

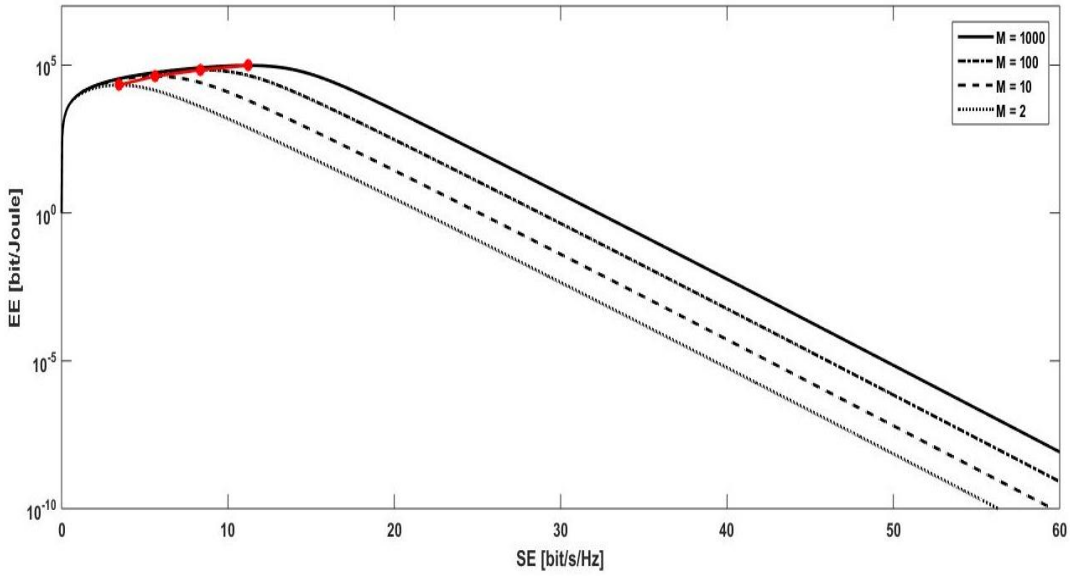


Figure 4.5 EE and SE relation in (4.12) for different values of M

The impact of M is illustrated in Figure 4.5 for P_{FIX} = 10W, B = 100 kHz, $\sigma^2/\beta^0_0 = -6$ dBm, and $\mu = 0.4$. As anticipated by the analysis, both EE and SE increase with M.

SE increases with M and P_{FIX}. On the other hand, EE grows with M and is an almost decreasing function of P_{FIX}. Therefore, it seems that an unbounded EE can be achieved by adding more and more antennas.

4.6 Circuit power consumption model

So far, we have used the simple two-cell Wyner network model to show that a PC model accounting for the transmit power as well as for the CP consumed by the transceiver hardware at the BS and UEs is necessary to avoid misleading conclusions about the EE. These are not the only contributions that must be taken into account to appropriately evaluate the CP of the UL and DL of Massive MIMO. We will show that we must also consider the power consumed by digital signal processing, backhaul signaling, encoding, and decoding [26]. a CP model for a generic BS j in a Massive MIMO network is:

$$\begin{aligned} \mathrm{CP}_j = & \underbrace{P_{\mathrm{FIX},j}}_{\text{Fixed power}} + \underbrace{P_{\mathrm{TC},j}}_{\text{Transceiver chains}} + \underbrace{P_{\mathrm{CE},j}}_{\text{Channel estimation}} + \underbrace{P_{\mathrm{C/D},j}}_{\text{Coding/Decoding}} \\ & + \underbrace{P_{\mathrm{BH},j}}_{\text{Load-dependent backhaul}} + \underbrace{P_{\mathrm{SP},j}}_{\text{Signal processing}} \end{aligned} \tag{4.14}$$

where $P_{FIX,j}$ was defined before as a constant quantity accounting for the fixed power required for control signaling and load-independent power of backhaul infrastructure and baseband processors. Furthermore, $P_{TC,j}$ accounts for the power consumed by the transceiver chains, $P_{CE,j}$ for the channel estimation process (performed once per coherence block), $P_{C/D,j}$ for the channel encoding and decoding units, $P_{BH,j}$ for the load dependent backhaul signaling, and $P_{SP,j}$ for the signal processing at the BS. Note that neglecting the power consumed by transceiver chains, channel estimation, precoding, and combining was previously the norm in multiuser MIMO. More precisely, the small numbers of antennas and UEs, before Massive MIMO was introduced, were such that the CP for all those operations was negligible compared to the fixed power. The CP associated with those operations was modelled for single-cell systems [21, 22, 23], while multicell systems were considered in [20]. Inspired by these works, we provide in what follows a tractable and realistic model for each term in (4.14), as a function of the main system parameters Mj and Kj. This is achieved by characterizing the hardware

setup using a variety of fixed hardware coefficients, which are kept generic in the analysis; typical values will be given later and strongly depend on the actual hardware equipment and the state-of-the-art in circuit implementation.

4.6.1 Comparison of CP with Different Processing Schemes

Table 4.1 Parameters in the CP model. Two different set of values are exemplified.

Parameter	Value set 1	Value set 2
Fixed power: P_{FIX}	10 W	5 W
Power for BS LO: P_{LO}	0.2 W	0.1 W
Power per BS antennas: P_{BS}	0.4 W	0.2 W
Power per UE: P_{UE}	0.2 W	0.1 W
Power for data encoding: P_{COD}	0.1 W/(Gbit/s)	0.01 W/(Gbit/s)
Power for data decoding: P_{DEC}	0.8 W/(Gbit/s)	0.08 W/(Gbit/s)
BS computational efficiency: L_{BS}	75 Gflops/W	750 Gflops/W
Power for backhaul traffic: P_{BT}	0.25 W/(Gbit/s)	0.025 W/(Gbit/s)

Two sets of CP parameters are given in Table 4.1. The first set is inspired by a variety of recent works: baseband power modeling from [24, 25], backhaul power according to [26], and the computational efficiencies from [21]. In the future, these parameters will take very different values that we cannot predict at the time of writing of this monograph. For the sake of the analysis, in what follows we also consider a setup in which the transceiver hardware's PC is reduced by a factor two whereas the computational efficiencies (which benefit from Moore's law) are increased by a factor ten. This leads to the second set of values in Table 4.1. We stress that these parameters tend to be extremely hardware-specific and thus may take substantially different values.

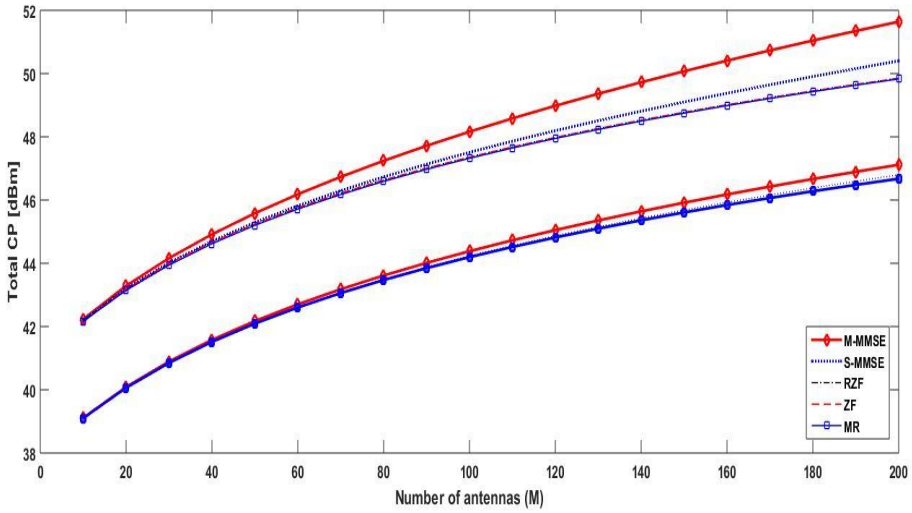


Figure 4.6(a) Total CP for K = 10 and varying M.

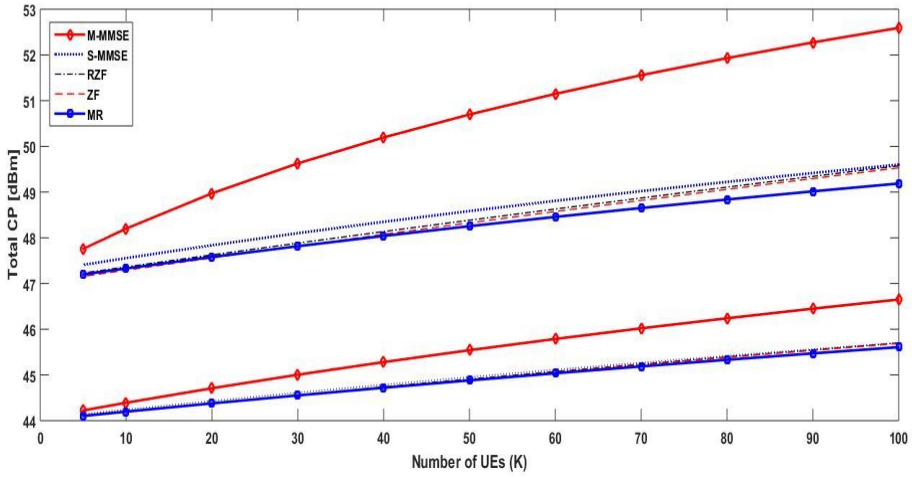


Figure 4.6(b) Total CP for M = 100 and varying K.

Figure 4.6 Total CP per cell of both UL and DL for the running example. The two sets of CP parameter values reported in Table 4.1 are considered.

Figure 4.6 illustrates the total CP per cell for the combined UL and DL scenario with different combining/precoding schemes. The MMSE estimator is used for channel estimation to fully exploit the spatial channel correlation. Note that the vertical axis is reported in dBm. In Figure 4.6(a), we consider K = 10 and let M vary from 10 to 200. The CP increases with M for all schemes and for both value sets. The highest CP is required by M-MMSE, followed by S-MMSE. For Value set 1, S-MMSE reduces the CP by 0.5%–25% since inter-cell channel estimates are not computed. Note, however, that M-MMSE provides higher SE than S-MMSE. Quantitatively speaking, the CP

required by M-MMSE for M = 100 and K = 10 is 48.16 dBm (65.48W) whereas S-MMSE needs 47.5 dBm (56.35W), which is roughly a 14% reduction. From Section 4, we know that this CP increase with M-MMSE is compensated by a 10% higher SE than with S-MMSE in both UL and DL. For Value set 2, the CP required by M-MMSE is only 0.1%–7% higher than with S-MMSE. This is mainly due to the increased computational efficiency. RZF and ZF consume less CP, since both invert matrices of dimensions K × K, rather than M × M. Compared to M-MMSE, when M = 100, this reduces the CP by 17% for Value set 1 and by 4% for Value set 2. MR is characterized by the lowest CP since no matrix inversions are required. However, the CP reduction compared to RZF and ZF is marginal for both value sets; MR only provides a substantial complexity reduction compared to RZF and ZF when the number of UEs is very large; we consider M = 100 and let K vary from 10 to 100. The CP increases with the number of UEs, but with a smaller slope than when M is changed (especially for Value set 2). Although the general trends for the two set of values are the same (e.g., M-MMSE requires the highest CP and MR the lowest), we see that for Value set 1 the CP required by M-MMSE is 8%–100% higher than with S-MMSE. This CP increase reduces to 2%–25% CP for Value set 2.

Table 4.2 CP per cell with M = 100 and K = 10 for different schemes and the two sets of values.

Scheme	Value Set 1	Value set 2
M-MMSE	65.48W	27.42W
S-MMSE	56.35W	26.51W
RZF	54.43W	26.32W
ZF	54.43W	26.32W
MR	53.96W	26.27W

The CP of all schemes, for the two different sets of parameter values, M = 100, and K = 10 are summarized in Table 4.2. As we can see, in this considered setup the CP required by the different schemes is marginally different. This happens because the CP of the transceiver hardware dominates over that of signal processing. Moreover,

comparisons in this section are made for given configurations of (M, K), which do not necessarily represent the optimal ones for maximizing the EE of the network.

4.7 Summary

This chapter mainly focused on energy efficiency, wyner model, transmit power that affects energy efficiency, power consumption model, trade-off between energy efficiency and spectral efficiency, energy efficiency for different precoding schemes.
Out of all different schemes discussed
Power consumption order:

M-MMSE >S-MMSE >R-ZF> ZF> MR

Energy efficiency order:

M-MMSE <S-MMSE <R-ZF< ZF< MR

CHAPTER 5
CONCLUSION

The main aim of the project is to maximize spectral efficiency (SE) and energy efficiency (EE) in massive MIMO networks. Massive MIMO is a type of wireless communications technology in which base stations are equipped with a very large number of antenna elements to improve spectral and energy efficiency. Massive MIMO systems typically have tens, hundreds, or even thousands of antennas in a single antenna array. We studied the massive MIMO system model for uplink and downlink in section 2.7.

We also studied that precoding can improve both spectral and energy efficiencies in section 3.4. Later we studied different precoding techniques like M-MMSE, S-MMSE, R-ZF, ZF, P-ZF, MR. Initially we studied the variation of spectral efficiency with an increasing number of BS antennas (M) for MR, ZF, P-ZF under best, average and worst cases of interference. Later we studied the SE variation for different combining schemes with different pilot reuse factors in fig 3.6. We need to choose a scheme that has less computational complexity and more spectral efficiency. So, it is best to choose R-ZF combining scheme which has medium computational complexity and medium spectral efficiency.

In Chapter 4, we focused only on energy efficiency (EE) in massive MIMO. We studied the wyner model in section 4.4 and circuit power consumption model in section 4.6. We understood the trade-off between energy and spectral efficiency. Later we studied EE vs SE plot for different Circuit Power (CP) = PFIX values in fig 4.4 and for different M values in fig 4.5. Based on CP model we considered two sets of power values in table 4.1. Based on the two set of values we have computed total CP and then plotted total CP vs number of antennas M (considering fixed value of K = 10) for different combining schemes in fig 4.6a. We also plotted total CP vs number of users K (considering fixed value of M = 100) for different combining schemes in fig 4.6b. In order to be more energy efficient, we need to reduce the power consumption (total CP).

The scheme having more computation complexity will have more computational power which in turn increases the total CP. Choosing a scheme with less computations may increase energy efficiency but it will reduce the spectral efficiency. So, it is best to choose the scheme with medium computational complexity like R-ZF which will offer acceptable energy and spectral efficiencies.

Paper Publications Details

1. Somasekhar Borugadda, Damarasingu Venkata Sai Kalyan, Mohan Manoj Pentakota, Madhusudhan Reddy Polamareddy "***Maximization of Spectral Efficiency and Energy Efficiency in Massive MIMO***" Springer International Publisher, [paper has been submitted to the editorial board]

REFERENCES

1. Cooper, M. 2010. “The Myth of Spectrum Scarcity”. Tech. rep. DYNA llc. url: https://ecfsapi.fcc.gov/file/7020396128.pdf.
2. Ericsson. 2017. “Ericsson mobility report”. Tech. rep. url: http://www.ericsson.com/mobility-report.
3. Ring, D. H. 1947. “Mobile Telephony - Wide Area Coverage”. Bell Laboratories Technical Memorandum.
4. Ericsson. 2017. “Ericsson mobility report”. Tech. rep. url: http://www.ericsson.com/mobility-report.
5. Cisco. 2016. “Visual networking index: Global mobile data traffic forecast update, 2015 – 2020”. Tech. rep.
6. Qualcomm. 2012. “Rising to meet the 1000x mobile data challenge”. Tech. rep. Qualcomm Incorporated.
7. Zappone, A. and E. Jorswieck. 2015. “Energy Efficiency in Wireless Networks via Fractional Programming Theory”. Foundations and Trends in Communications and Information Theory. 11(3-4):185–396.
8. Hoydis, J., S. ten Brink, and M. Debbah. 2013a. “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” IEEE J. Sel. Areas Commun. 31(2): 160–171.
9. Guo, K., Y. Guo, G. Fodor, and G. Ascheid. 2014. “Uplink power control with MMSE receiver in multi-cell MU-massive-MIMO systems”. In: Proc. IEEE ICC. 5184–5190.
10. Krishnan, N., R. D. Yates, and N. B. Mandayam. 2014. “Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: large system analysis”. IEEE Trans. Wireless Commun. 13(8): 4360–4373.
11. Emil Björnson, Jakob Hoydis and Luca Sanguinetti (2017), “Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency”, Foundations and Trends in Signal Processing: Vol. 11, No. 3-4, pp 154–655. DOI: 10.1561/2000000093.

12. Emil Bjornson, Erik G. Larsson and Merouane Debbah, "Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated?" DOI 10.1109/TWC.2015.2488634, IEEE Transactions on Wireless Communications.
13. Ngo, H. Q., E. G. Larsson, and T. L. Marzetta. 2013. "Energy and spectral efficiency of very large multiuser MIMO systems". IEEE Trans. Commun. 61(4): 1436–1449.
14. Yang, H. and T. L. Marzetta. 2013a. "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems". IEEE J. Sel. Areas Commun. 31(2): 172–179.
15. Marzetta, T. L., E. G. Larsson, H. Yang, and H. Q. Ngo. 2016. Fundamentals of Massive MIMO. Cambridge University Press.
16. Li, X., E. Björnson, E. G. Larsson, S. Zhou, and J. Wang. 2017. "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression". EURASIP J. Wirel. Commun. Netw. (117).
17. Qualcomm. 2012. "Rising to meet the 1000x mobile data challenge". Tech. rep. Qualcomm Incorporated.
18. Auer, G., V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske. 2011. "How much energy is needed to run a wireless network?" IEEE Wireless Commun. 18(5): 40–49.
19. Han, C., T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C. X. Wang, T. A. Le, M. R. Nakhai, J. Zhang, and L. Hanzo. 2011. "Green radio: Radio techniques to enable energy-efficient wireless networks". IEEE Commun. Mag. 49(6): 46–54.
20. Zappone, A. and E. Jorswieck. 2015. "Energy Efficiency in Wireless Networks via Fractional Programming Theory". Foundations and Trends in Communications and Information Theory. 11(3-4): 185–396.
21. Björnson, E., L. Sanguinetti, and M. Kountouris. 2016d. "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO". IEEE J. Sel. Areas Commun. 34(4): 832–847.

22. Yang, H. and T. L. Marzetta. 2013b. “Total energy efficiency of cellular large scale antenna system multiple access mobile networks”. In: Proc. IEEE Online GreenComm. 27–32.
23. Björnson, E., L. Sanguinetti, J. Hoydis, and M. Debbah. 2014d. “Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?” In: Proc. IEEE WCNC. 242–247.
24. Björnson, E., L. Sanguinetti, J. Hoydis, and M. Debbah. 2015c. “Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?” IEEE Trans. Wireless Commun. 14(6): 3059–3075.
25. Kang, D., D. Kim, Y. Cho, J. Kim, B. Park, C. Zhao, and B. Kim. 2011. “1.6 - 2.1 GHz broadband Doherty power amplifiers for LTE handset applications”. In: Proc. IEEE MTT-S. 1–4.
26. Kumar, R. and J. Gurugubelli. 2011. “How green the LTE technology can be?” In: Proc. Wireless VITAE
27. Tombaz, S., K. W. Sung, and J. Zander. 2012. “Impact of densification on energy efficiency in wireless access networks”. In: Proc. IEEE GLOBECOM Workshop. 57–62.