TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Nachrichtentechnik

Precoding Algorithms for Massive Multiple-Input, Multiple-Output Wireless Systems

Amir Mehdi Ahmadian Tehrani

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

Doktor–Ingenieurs

genehmigten Dissertation.

Vorsitzender:		Prof. Dr.-Ing. Eckehard Steinbach
Prüfer der Dissertation:	1.	Prof. Dr. sc. techn. Gerhard Kramer
	2.	Prof. Mari Kobayashi, Ph.D.

Die Dissertation wurde am 27.11.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 23.04.2020 angenommen.

Contents

List of Figures

List of Tables

Acknowledgements

This work was carried out at the Radio Research Group, Nokia Bell Labs, Munich in cooperation with the Chair for Communications Engineering (LNT), Technische Universität München between August 2015 and August 2018.

First of all, I am deeply grateful to Professor Gerhard Kramer, who gave me not only the opportunity pursuing my own research ideas, but also played a significant role between me and my supervisors at Nokia Bell Labs by contributing to the success of this work with the right comments. I do appreciate not only his patience on the progress of this work, but also the time he investigated correcting my thesis through multiple iterations. Honestly, this work would have not been successful without his support, guidance and encouragement.

I also would like to express my gratitude to my supervisors at Nokia Bell Labs include Wolfgang Zirwas, Dr. Rakash Sivasiva Ganesan providing me an opportunity to unleash my creativity through the industrial projects. I do appreciate your strong collaboration in developing me as a researcher. Special thanks go to Dr. Bernhard Wegmann , the Mobile Network Architecture Team Leader and Dr. Simone Redana, the Head of 5G Mobile Network Architecture at Nokia Bell Labs in Munich, who have been always there not only to hearing my requests, but also to provide an exciting working enviroment with all the internal colleagues. My gratitude to the doctoral students at the Radio Research Group for having been great friends and office partners sharing their experiences throughout last three years.

Moreover, I wish to thank all the colleagues at LNT for their valuable advices, ideas and collaborations during my weekly visits. I would like to acknowledge non-scientific colleagues at the Chair for their professional assistance throughout my work.

I also would like to acknowledge the examiner of this thesis, Professor Mari Kobayashi for sharing her feedback on my research. This thesis has been improved significantly based on her valuable suggestions. Additional gratitudes goes to Professor Eckehard Steinbach for organizing the committee chair via Zoom application considering the challenges concerned.

I would like to express my deepest sense of gratitude to my parents and sisters for their support being far away. I have been extremely lucky to have you as my family.

Above all, my warmest thanks go to my love, Atefeh for the ultimate love, support and constant encouragement I have gotten over the years working on this thesis. I am deeply indebted to you, who passionately encouraged me to finish my thesis. You taught me indeed to never give up. Thank you Schatz for staying late with me reviewing my thesis and giving your valuable feedbacks.

Munich, 22.10.2020 Amir Mehdi Ahmadian Tehrani

Abstract

Massive MIMO is a promising candidate technology to meet the ever-increasing wireless throughput demand in 5G mobile networks. To increase the spectral efficiency of orthogonal frequency division multiplexing (OFDM) massive MIMO, coordinated multipoint or cooperative MIMO has been identified as a potential solution. Joint transmission coordinated multipoint (JT CoMP) deals with constructive exploitation of interference through coherent eNodeB cooperation. To develop the frequency-division duplex (FDD) version of massive MIMO, the amount of channel state information (CSI) that needs to be fed back from the UEs to the eNodeB increases as compared to time-division duplex (TDD). To reduce the CSI reporting complexity in FDD-mMIMO, a predetermined grid of beams (GoB) is applied to make the channel matrix sparse when the UEs report only their preferred beams.

The computational complexity of zero-forcing precoding at the eNodeB benefits from the sparsity of the channel matrix when CoMP massive MIMO and GoB are combined. Each UE reports only the effective channel matrix relevant channel components (RCCs). Low-complexity Moore-Penrose inverse precoding gains a factor of 100 in terms of floating point operations (FLOPs) with respect to using the full pseudo inverse calculation when a MIMO-OFDM system with 100 physical resource blocks (PRBs) per time transmission interval (TTI) for single subcarrier and multiple subcarrier is considered. A trade-off between the computational complexity gain and SINR degradation as well as the spectral efficiency is analyzed.

Zusammenfassung

Massive MIMO ist eine vielversprechende Technologie, um die stets wachsende Nachfrage nach drahtlosem Datendurchsatz für 5G Mobilfunknetze zu leisten. Um die spektrale Effizienz von orthogonal frequency division multiplexing (OFDM) massive MIMO zu steigern, wurde coordinated multipoint oder cooperative MIMO als mögliche Lösung identifiziert. Joint transmission coordinated multipoint (JT CoMP) behandelt die konstruktive Nutzung von Interferenz durch kohärente eNodeB Kooperation. Um die frequency-division duplex (FDD) Version von massive MIMO zu entwickeln, lässt sich die Menge von channel state information (CSI), die Bedarf an einer Rückführung von den UEs zu dem eNodeB hat, im Vergleich zu time-division duplex (TDD) erhöhen. Ein prädeterminiertes Grid of Beam (GoB) wird angewendet, um eine dünnbesetzte Kanalmatrix zu erhalten, wenn jeder UE nur seine bevorzugten beams meldet. Dies reduziert die Komplexität vom CSI reporting.

Die Berechnungskomplexität von zero-forcing precoding am eNodeB profitiert von der dünnbesetzten Kanalmatrix, wenn CoMP massive MIMO und GoB kombiniert werden. Jeder UE meldet nur den betroffenen Teil seiner Kanalmatrix, der die jeweils relevant channel components (RCCs) enthält. Low-complexity Moore-Penrose inverse precoding verringert die Berechnungperformance in floating point operations (FLOPs) um Faktor 100 im Vergleich zu full pseudo inverse Berechnung bei 100 physical resource blocks (PRBs) pro time transmission interval (TTI) für single subcarrier und multiple subcarrier massive MIMO-OFDM Systeme. Es wird Analyse des Trade-offs zwischen der Reduzierung in FLOPs, der Verschlechterung von SINR und der spektralen Effizienz durchgeführt.

1

Introduction

1.1. Motivation

While the first Internet era (1985-2000) and the second Internet era (2000-2015) were defined to build the Internet and its new services, the third Internet wave is being defined by building the Internet into everything (2015+) which is distrupting many industry sectors [6]. Data traffic in the last five years of the second Internet era (2010-2015) witnessed growth dominated by sharing personal content delivered from and to mobile devices [7]. By introducing new mobile devices that support a wide range of applications, the gap of user experience between fixed and mobile environments becomes narrower. The new Internet era toward 2020 and beyond brings two significant trends in mobile services (see Figure 1.1):

1. Everything will be connected by wireless to enable information collection and control of devices. Remote monitoring and real-time services such as Vehicle-to-Vehicle (V2V) communications and smart metering which supports Machine-to-Machine (M2M) and Internet of Things (IoT) can be considered as the emerging service examples for this trend.

2. Wireless services, e.g., high resolution video streaming, tactile internet, low-latency safety systems and remote health consulting will be more extensive and enriched through content and are expected to be delivered in real-time.

Given these trends, the mobile data traffic has grown 24-fold in 2010 to 2015 [8]. According to the analysis of the growth in mobile network traffic rate published by Bell Labs Consulting, by 2020 the volume of mobile traffic will be 1000-fold larger compared to 2010 [2]. Furthermore, more diversified services will bring more variations in data rate, depending on the location, time, application and type of device [9].

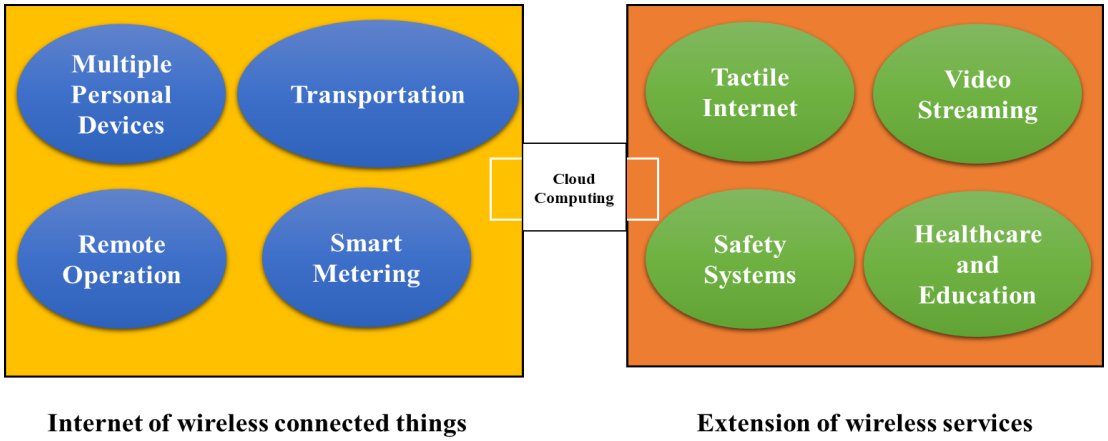


Figure 1.1.: New Internet era major trends toward 2020 and beyond [1]

The new Internet era imposes challenging requirements. For example, Figure 1.2 illustrates the latency and bandwidth requirements of the future Internet wave services [2]. Critical and immersive services such as Augmented Reality (AR) require ultra-low latency networks with single-digit milliseconds of delay. To carry high volumes of traffic from billions of mobile and IoT devices, the future Internet era requires ultra-high capacity. Services with ultra-high speed, e.g., 360-degree video, will demand peak data rates of up to 1 Gbps [2], [10].

Given these requirements, the fifth generation (5G) mobile networks will have to support more stringent latency and reliability requirements, a wide range of data rates and network scalability and flexibility [11]. The response to such requirements can be a combination of existing technologies and new radio concepts, as depicted in Figure 1.3. According to Bell Labs studies, in order to achieve 5G mobile network targets, six essential technologies should be considered [9]:

1. **Modular framing structure**: 5G transmission frame structure that enables very low latency and fast Hybrid Automatic Repeat Request (HARQ) coexistence with fourth generation (4G) and transmission of variable frame length to support ultra-broadband, ultra-narrowband and ultra-low latency applications.

2. **New air interface**: 5G New Radio (NR) that supports diverse use cases with extreme requirements, as well as a wide range of frequencies and deployment options using a flexible and scalable design.

3. **Massive MIMO**: to achieve a 1000-fold increase in mobile data traffic and enhanced Mobile Broadband (eMBB) targets, advances in Multiple-Input Multiple-Output (MIMO) technology is essential.

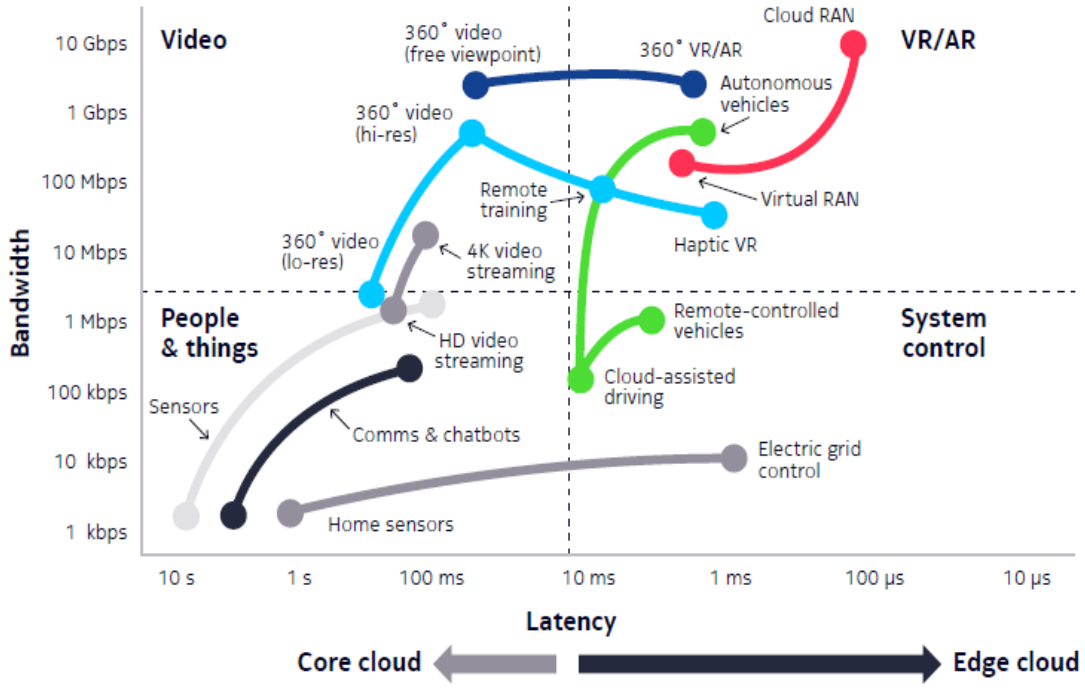


Figure 1.2.: New Internet era latency and bandwidth requirements [2]

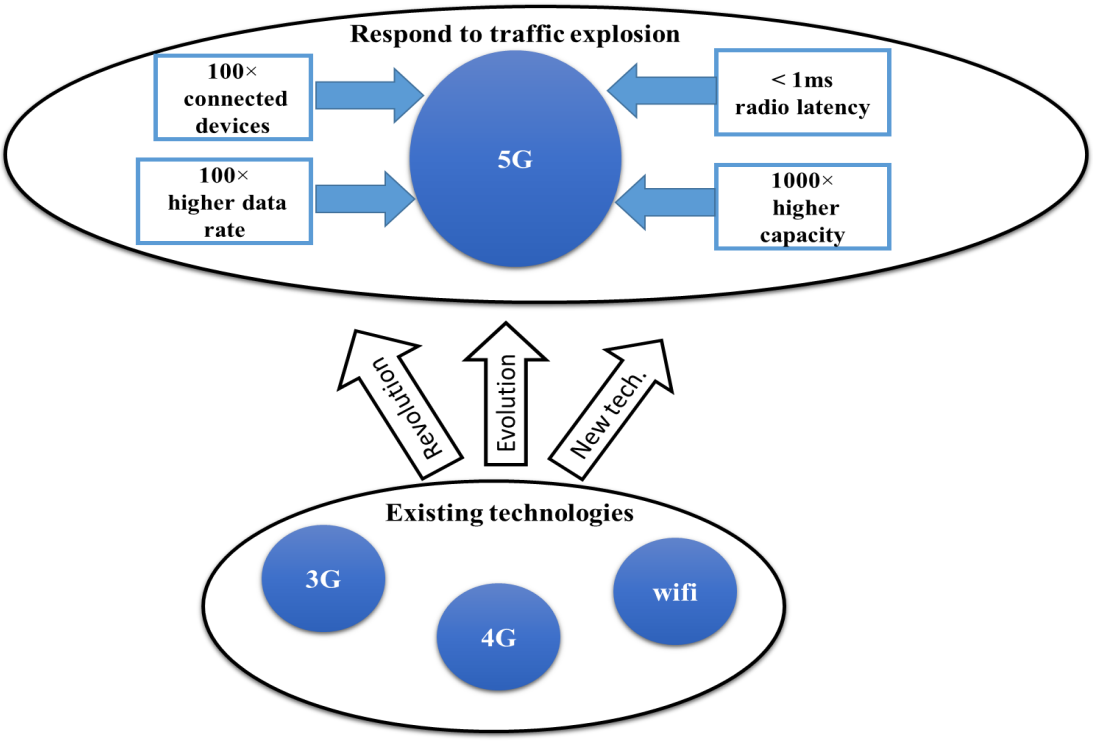


Figure 1.3.: 5G mobile network concept and requirements

4. **New spectrum**: to avoid the shortage of wireless carriers due to the rapid increase of mobile data growth, 5G is driving the exploration of millimeter wave frequency spectrum (>24 GHz).

5. **Multi-RAT**: to provide ultra-reliable broadband access to a wide range of applications, an integration of a variety of Radio Access Technologies (RATs) with network controlled traffic steering and cell-free architecture is required.

6. **New core**: to handle highly dynamic traffic profiles, cloudification of the core network is required. For any service mix, new vitualized and Software-Defined Networking (SDN) solutions are used to provide flexible routing.

1.2. Thesis Context

The benefits of using multiple antennas at both the Base Station (BS) and at the User Equipment (UE) was described by Foschini [12]. This multiple antenna approach is now used in every major wireless access standard in spite of the large gap between theoretical investigations and practical implementations. To achieve a 1000-fold increase in data rates for cell-edge UEs is a challenging task when the experienced rates are orders of magnitudes larger than the cell average. Coordinated Multi-Point (CoMP) can be used to reduce the interference from neighboring cells, but a lack of accurate Channel State Information (CSI) at the BS is still a challenge. Advanced cooperative precoder and beamforming design can improve interference mitigation on a multi-cell level [13], [14].

To reduce interference, increasing the number of antennas at the BS expands capacity from a single site through beamforming [15]. Considering a hundred antennas, simulation results [16], [17] have shown capacity increases of an order of magnitude. The deployment of large transmit antenna arrays serving a much smaller number of UEs is called Massive MIMO (mMIMO). Long Term Evolution (LTE) systems have solutions currently that use up to 64 atennas at the Evolved Node B (eNodeB) [18]. mMIMO can be used in heavy, scatter-propagation and Non-line-of-Sight (NLoS) conditions, to achieve much higher throughput at a given range. Antenna arrays in mMIMO systems scale with wavelength, which is inversely proportional to the frequency used.

Frequency Division Duplex (FDD) mMIMO is relevant compared to Time Division Duplex (TDD) as many of the below 6 GHz frequency bands are FDD bands and it provides better coverage than the high RF bands. Furthermore, FDD has coverage gains over TDD as FDD signals are transmitted constantly, while in TDD average transmit power is reduced due to switching between Uplink (UL) and Downlink (DL) signals. mMIMO with full cooperation (coherent joint transmission CoMP) has been a research topic in 3rd Generation Partnership Project (3GPP) due to some stringent requirements on time

and frequency synchronization, need for accurate CSI and precoding complexity. 3GPP Release 18 has recently identified CoMP mMIMO wireless systems as the main enabler for sixth generation (6G) mobile networks and it is getting more accepted by many companies.

To deploy below 6 GHz FDD mMIMO CoMP systems, one main aspect is to combat with the high overhead due to CSI feedback and the precoding complexity of large channel matrix at the eNodeB.

1.3. Overview of thesis and Contributions

In this thesis, we investigate the computational complexity of below 6 GHz FDD mMIMO systems. The deployment of a large number of antennas at the eNodeBs that are synchronized via a CoMP architecture greatly increases the downlink precoding complexity as compared to Multi-user MIMO (MU-MIMO) and Single-user MIMO (SU-MIMO) systems. To analyze the computational complexity of precoding at the eNodeB, Floating Point Operations (FLOPs) can be used as a performance metric to count the number of mathematical operations when different Zero-Forcing (ZF) precoding schemes are employed. One solution to reduce the complexity is to filter out significant beams received at the UE. To apply this idea, we use a joint analog beamformer and digital precoding. Analog beamformer or Grid of Beams (GoB) precoder converts physical antenna arrays at the eNodeB into a limited set of orthogonal narrow beams which are considered as signifant received beams at the UE. technique and a power threshold. We briefly summarize the contents and contributions of each chapter.

Chapter 2 introduces the principles of MIMO systems and briefly reviews mMIMO deployments in CoMP scenarios. By introducing the 3GPP 3D urban macro channel model, we motivate the use of an integrated mMIMO, GoB and CoMP system model. The chapter ends with an investigation of linear beamforming at the UE.

Chapter 3 reviews linear precoding at the eNodeB. First, we review the state-of-the-art methods and their complexity levels to calculate the pseudo inverse of the channel matrix for a single-subcarrier OFDM system. Second, we analyze Channel Component (CC) in the channel matrix which represents the combination of the beams formed at the eNodeB (Tx) and the UE Rx beamformers. As a result, channel matrix will be sparse due to the directivity of the generally narrow GoB beamformers. We will assume that the UEs receive those channel components with a power that is a predefined threshold above the Rx-power of the strongest beam. These limited number of CCs are called Relevent Channel Components (RCCs) and they will be used to adjust the mMIMO matrix precoder and a low-complexity pseudo inverse scheme. Utilizing a toolbox to count the FLOPs, we discuss the impact of a power threshold on the performance degradation and complexity. We also discuss the sensitivity of spectral efficiency and Signal to Interference plus Noise

Ratio (SINR) to a complexity gain for a single-subcarrier OFDM system.

Chapter 4 investigates linear precoding solutions for multiple-subcarriers FDD mMIMO systems. An approximate pseudo inverse scheme for the Strong Precoding Channel Components (SPCCs) is proposed. We define four major parameters and we discuss in detail the impact of each parameter on the performance metrics when a low-complexity pseudo inverse filter is applied to multiple-subcarriers.

Chapter 5 summarizes the main findings of this thesis and discusses open problems and possible future lines of work.

2

System Model

2.1. Introduction

This chapter lays the foundations for the thesis by giving an overview of MIMO communications. We first review the fundamental concepts of MIMO systems in section 2.2 Diversity and spatial multiplexing are considered as the main characteristics of MIMO to understand how link capacity and reliability can be increased. MIMO can be extended to scheduling a certain number of UEs at the same time and frequency. However, new challenges appear including accurate reporting of CSI and combating interference.

Beamforming and precoding at the eNodeB can reduce interference in Multi-user MIMO (MU-MIMO) systems, and such techniques also require CSI estimation. To serve many UEs, MIMO can be extended to massive MIMO (mMIMO) where the eNodeB is equipped with a large number of antennas. mMIMO's main features are discussed in section 2.3. Although mMIMO can provide higher spectral efficiency, the complexity of system design becomes challenging when handling high rate CSI for each antenna element and performing precoding at the eNodeB. A Grid of Beams (GoB) can reduce CSI reporting substantially, and we review this approach in section 2.3.1.

The complexity of system design is even more challenging when using Joint Transmission (JT) CoMP in combination with 3GPP urban macro channels. The computational complexity scales up to the entire Cooperation Area (CA) where each UE has a link not only to the dominant cell, but also to other eNodeBs. The main characteristics and challenges of CoMP are discussed in section 2.4.

The main features of the system model, i.e., the combination of mMIMO, GoB and JT CoMP, are provided in section 2.5. A 3GPP 3D urban macro channel model and the corresponding essentials are introduced in section 2.6.

2.2. MIMO Review

Wireless channels suffer from fading due to destructive addition of multipath components and interference from other UEs. Advanced antenna techniques can combat fading and interference. Single-Input Single-Output (SISO) is the basic antenna configuration that employs a single antenna at the transmitter and a single antenna at the receiver (see Figure 2.1). SISO can be expanded to using multiple antennas at the receiver which is

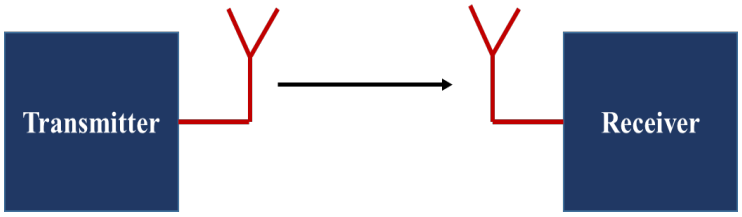


Figure 2.1.: SISO architecture

known as Single-Input Multiple-Output (SIMO) (see Figure 2.2). Multiple-Input Single-Output (MISO) occurs if multiple antennas are used at the transmitter and a single antenna at the receiver. Using multiple antennas at one side of the wireless link leads to the possibility of performing interference cancelation and to realize diversity gain.

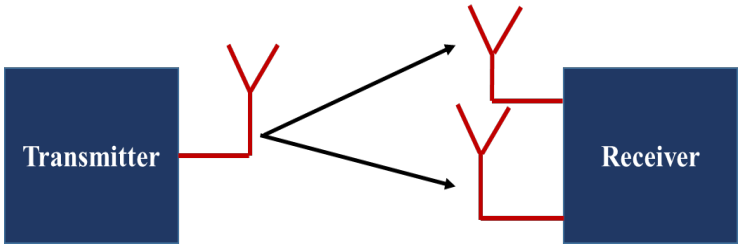


Figure 2.2.: SIMO architecture

Diversity equips the receiver with several (ideally independent) replicas of the transmitted signal leading to link reliability improvements. Diversity can be provided via time (due to Doppler spread) or frequency (due to delay spread). The use of *spatial* diversity is practically attractive as it does not incur an expenditure in transmission time or bandwidth.

Applying multiple antennas at both the transmitter and receiver is known as MIMO. The use of MIMO enables *spatial multiplexing* as an additional fundamental gain that increases spectral efficiency. If spatial multiplexing is used, the receiver exploits differences in the spatial signatures induced by the MIMO channel to separate data signals. This method can bring a linear capacity increase, as compared to a system with a single antenna at one or both sides of the communication link.

MIMO technology is logically divided into Single-user MIMO (SU-MIMO), MU-MIMO and mMIMO. SU-MIMO (see Figure 2.3) represents the simplest type of MIMO. Through

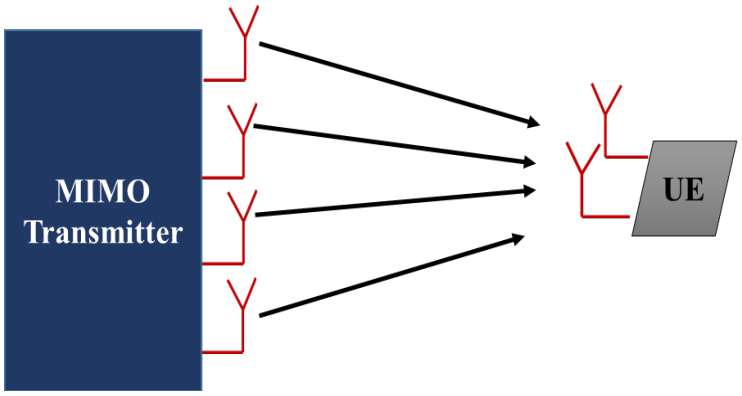


Figure 2.3.: Single-user MIMO

a combination of Time Division Multiplexing (TDM) and Frequency Division Multiplexing (FDM), different UEs are orthogonally multiplexed. In theory, the spectral efficiency for SU-MIMO can be increased by simultaneously applying large antenna arrays at the transmitter and the UE. However, in practice, considerations such as UE complexity in terms of the number of RF chains per antenna as well as separating data streams using digital signal processing seriously limit the implementation of SU-MIMO [19]. In contrast, MU-MIMO (see Figure 2.4) allows to serve more than one UE on the same time-frequency resource. One can view the gain of MU-MIMO as being achieved by breaking up the single-user UE antennas and treating them as multiple autonomous UEs. MU-MIMO technology generally has higher capacity than SU-MIMO under rich multipath conditions and low correlation between the antennas of different UEs. Further, a MU-MIMO eNodeB jointly processes the signals from each of the UEs which leads to higher spectral efficiency.

This thesis focuses on MU-MIMO and mMIMO. The main characteristics of MU-MIMO are explained briefly in section 2.2.2. mMIMO, which is the ultimate realization of MIMO technology, is introduced in section 2.3

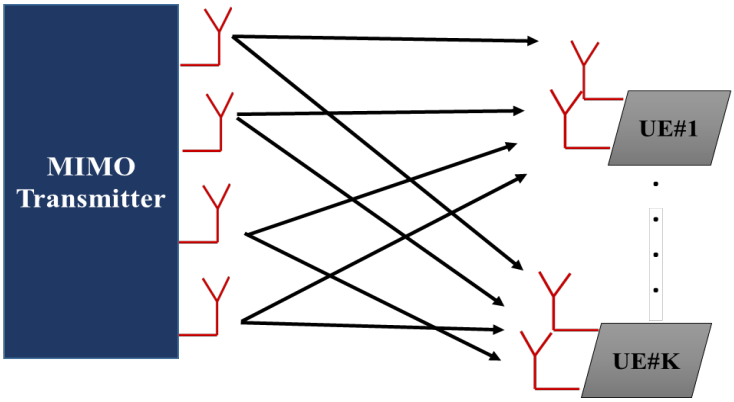


Figure 2.4.: Multi-user MIMO with K UEs

2.2.1. Time-Division Duplex versus Frequency-Division Duplex

CSI requirements at the UEs are different in SU-MIMO, MU-MIMO and mMIMO. CSI can be estimated from received pilot signals or it can be obtained through feedback from the UE to the eNodeB. If Time-Division Duplex (TDD) is applied, an eNodeB uses uplink pilots to estimate the uplink and downlink channels when the channel between the UE antennas and eNodeB antennas is reciprocal. In contrast, all the UEs in Frequency-Division Duplex (FDD) estimate the downlink channel from the pilots sent by the eNodeB. Estimated CSI known as *CSI feedback* is sent back to the eNodeB.

To ease channel estimation at the UE, each eNodeB antenna must be assigned a unique pilot waveform, and such pilots need to be mutually orthogonal. Thus, for M antennas that transmit orthogonal pilots in the downlink, at least M samples must be sent for pilots per coherence interval.

Depending on the power constraint, FDD can provide higher downlink rates than TDD [20]. For example, if B is the total bandwidth for both the uplink and downlink, P denotes the received power at the UE, and N_0 denotes the noise Power Spectral Density (PSD), then the TDD downlink rate under a noise-limited operation and a per-symbol power constraint is as follows:

$$TDD_{downlink} = \frac{\mathrm{B}}{2}\log_2(1 + \frac{\mathrm{P}}{\mathrm{BN}_0}) \quad (b/s). \tag{2.1}$$

The downlink rate in FDD operation can be expressed as follows:

$$FDD_{downlink} = \frac{\mathrm{B}}{2}\log_2(1 + \frac{\mathrm{P}}{(\mathrm{B}/2)\mathrm{N}_0}) \quad (b/s). \tag{2.2}$$

The expressions in (2.1) and (2.2) show the fact that transmission in TDD takes place

over the full bandwidth, but only half of the time. As such, with TDD the transmitter is silent half of the time and for a given P the received energy per unit time is half of FDD. Whereas, B in FDD is divided by 2 both inside and outside of the logarithm since transmission takes place continuously, but over half of the bandwidth. Under a peak power constraint and noise-limited operation, FDD provides 3 dB better SNR than TDD [20]. In this thesis, we focus on FDD MIMO technologies.

2.2.2. Multi-user MIMO

When MU-MIMO is employed, a certain number of UEs are scheduled on the same time and frequency resource. In the frequency domain, the transfer function of the channel between an eNodeB and a UE at a specified frequency is modeled by a complex matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$. The total number N of receive antennas is split between K UEs. For coherent detection in the downlink, CSI from all the UEs to the eNodeB antenna elements is needed. Utilizing estimated CSI, a joint precoding of data symbols can be applied prior to transmission.

We next introduce characteristics of MU-MIMO including channel estimation, precoding and beamforming, see Figure 2.5.

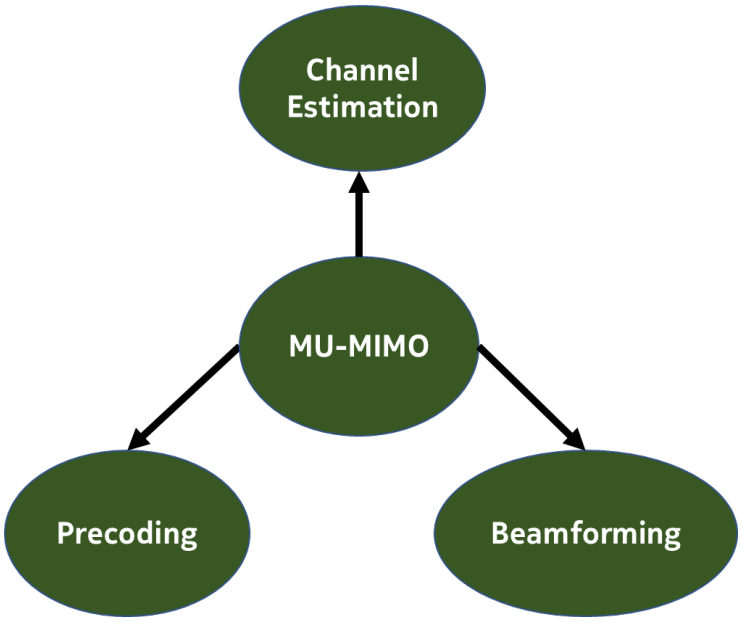


Figure 2.5.: MU-MIMO components

Channel Estimation

An eNodeB adapts the signal to the radio channel and each UE uses its CSI to decode. A major obstacle bringing MU-MIMO from theory into practice is how to obtain accurate CSI at the eNodeB. In general, accurate CSI is more readily available in TDD-based systems than in FDD-based systems due to the inherent channel reciprocity [21]. In TDD MU-MIMO, UEs send pilots on the uplink and all the channels between the UEs and the eNodeB can be estimated.

Although the number of required pilot sequences is independent of the number of antennas at the eNodeB, and feedback from the UEs is avoided, TDD restricts multi-cell MU-MIMO. Significantly, pilots in TDD in multi-cell MU-MIMO must be different from one cell to another cell. Thus, channel estimation in a "home" cell causes pilot contamination with channels from other cells which degrades channel estimation quality regardless of whether we add more antennas.

FDD MU-MIMO uses different uplink and downlink frequency bands which results in different CSI for both links. Downlink CSI is obtained by means of transmitting pilot symbols from an eNodeB to all UEs. The overhead due to reporting estimated CSI from the UEs to an eNodeB is one of the main challenges in FDD MU-MIMO.

Precoding

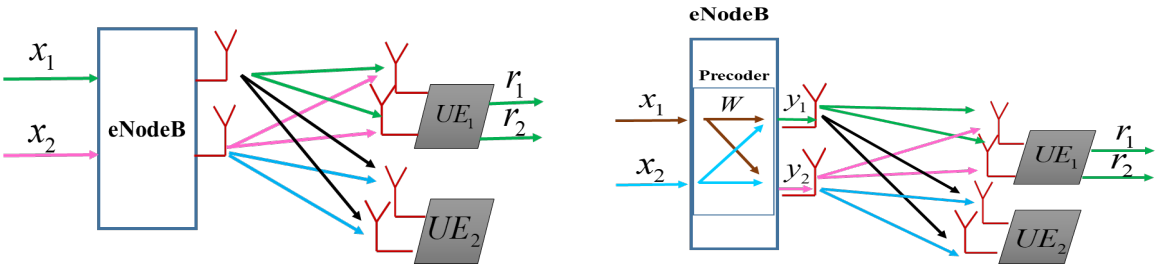
Precoding is a transmit processing operation in the spatial domain. Precoding is required to effectively utilize multiple channels between an eNodeB and the UEs. To illustrate the role of linear precoding in MU-MIMO, Figure 2.6(a) gives an overview of 2×2 MU-MIMO where both the UEs and the eNodeB are equipped with two antennas. x_1 and x_2 are the input signals at the eNodeB, and r_1 and r_2 are the received signals at the first UE. Figure 2.6(a) shows that the wireless channel provides four separate paths (shown as arrows) between the eNodeB and each UE. Transmission paths may comprise direct Line of Sight (LOS) and numerous paths created by reflection, scattering and diffraction from the channel environment.

If precoding is added as shown in Figure 2.6(b), the eNodeB must know the channel matrix $\mathbf{H}$ to a sufficient level of precision. The precoder can preprocess the streams before transmission. y_1 and y_2 in Figure 2.6(b) are the precoded signals applied to each transmit antenna. For a 2×2 MU-MIMO system with linear precoding, the precoding matrix $\mathbf{W}$ multiplies the input signals to generate:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{W} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{2.3}$$

The UE_1 output is then given by:

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \mathbf{H} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad = \mathbf{HW} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \tag{2.4}$$



(a) 2×2 MU-MIMO system (b) 2×2 MU-MIMO Precoder

Figure 2.6.: MU-MIMO precoding

MU-MIMO precoding can also be non-linear, e.g., precoding methods such as dirty-paper-coding (DPC) or Lattice-aided methods have better performance than linear precoding. However, linear precoding methods, e.g., Zero Forcing (ZF), Matched Filter (MF) or Minimum Mean Square Error (MMSE), generally have less complexity [22]. ZF precoding lets the eNodeB null out multiuser interference. In other words, ZF decouples the multiuser channel into independent single-user channels, which often achieves a large portion of the dirty paper coding capacity [23]. MF precoding is interference limited at high SINR, but it outperforms ZF at low SINR. In this thesis, we use ZF precoding to cancel out the interference.

Beamforming

Beamforming modifies the radiation pattern of an antenna array and can be applied in all antenna array systems [24]. The principle of beamforming is depicted in Figure 2.7. When the directions of the dominant propagation paths at the eNodeB(s) are identified, beamforming can be applied to adjust the eNodeB and UE beam patterns such that they have a high directivity towards the dominant angles of reception. Beamforming can lead to coverage gain, and it can reduce the delay spread caused by multipath signal propagation. Further, beamforming reduces CSI feedback since the number of beams received at the UEs (known as Rx beams) is less than the number of antenna elements located at the eNodeB.

A beamformer at the eNodeB is equipped with M antennas and a beamformer at the UE is equipped with N antennas. For the downlink, information bit sequences are

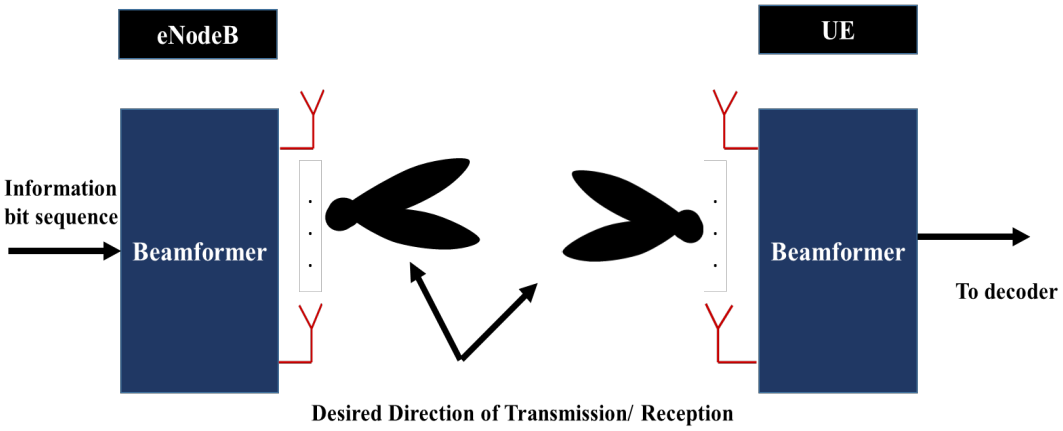


Figure 2.7.: Beamforming principle

first subjected to the eNodeB beamformer. For example, for M=2 the eNodeB can form two beams in the direction of the dominant multipath components. The UE beamformer adjusts its two beam patterns to provide high directivity towards the dominant angles of reception. Beamforming in MU-MIMO combines the signals by adjusting the phase and amplitude weights of the signals. We distinguish the following beamforming techniques:

1. **Fixed beamforming:** signals are combined applying fixed complex weights regardless of the channel conditions. These weights define the amplitude and phase shifts for each antenna element.

2. **Adaptive beamforming:** to give the desired peaks and nulls in the radiation pattern of the antenna array, the complex weights can be chosen adaptively. The weights are usually slowly changed to steer the beam until maximum signal strength is achieved and the "direction" to the signal source is found.

2.3. Massive MIMO Systems

mMIMO is an ultimate and scalable version of MU-MIMO [25]. There are other terms in literature used for mMIMO, such as Large Scale Antenna Systems, Very Large MIMO, Hyper MIMO and Full-Dimension MIMO, in which the number of antennas at the eNodeB is much larger than the number of UEs. A somewhat precise way to define mMIMO is to relate it to the ratio of active UEs to eNodeB antennas that serve those UEs. If this ratio is low, all UEs in the cell can be served simultaneously without any impacts in UE sepecific throughput. If number of UEs further increases the cell reaches its maximum capacity, the optimal point where the maximal cell resources are distributed among all Ues. Hence, mMIMO is understood as a vast over-provisioning of resources that presents

a fundamental paradigm shift to today's resource limited communication systems rather than a simple increase in number of eNodeB antennas.

Two main issues differentiate mMIMO from MU-MIMO. First, the number M of antennas at eNodeB in mMIMO is much larger than the number K of UEs and the total number N of receiver antennas. Second, mMIMO systems are more scalable than MU-MIMO systems. We consider the scenarios where the UEs are equipped either with a single antenna, i.e, $N = K$, or multiple antennas. A mMIMO system structure in the uplink and downlink with single UE antennas is shown in Figure 2.8 and Figure 2.9.

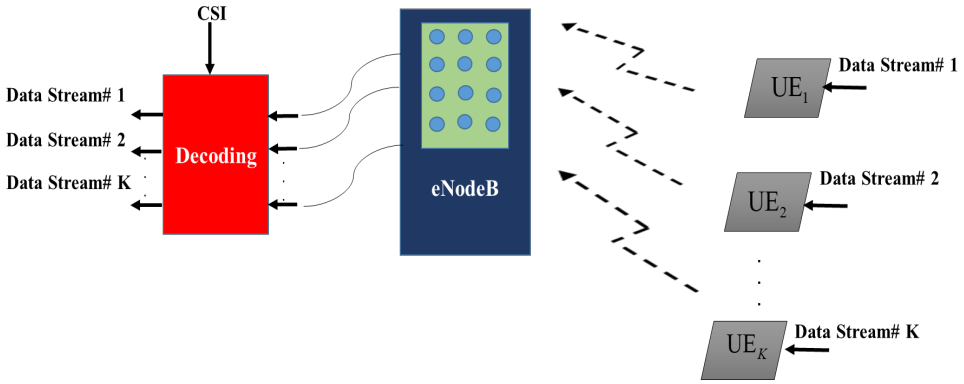


Figure 2.8.: mMIMO system uplink operation

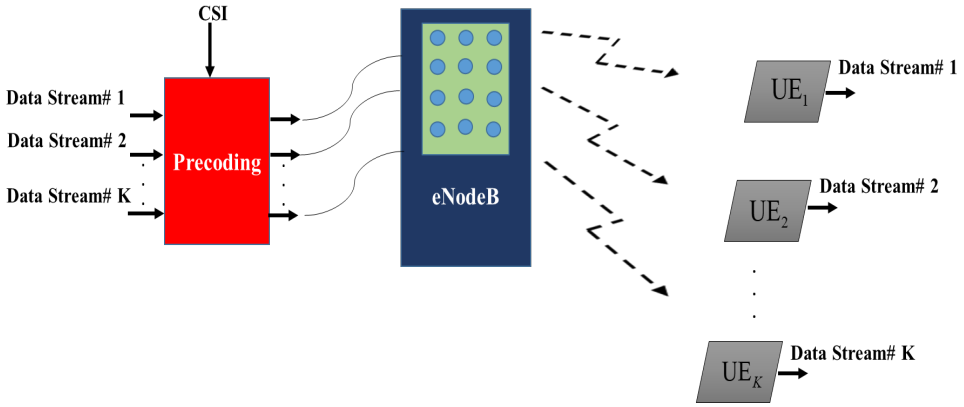


Figure 2.9.: mMIMO system downlink operation

When mMIMO is used, different data streams occupy the same frequencies due to spatial multiplexing. To perform multiplexing, the array needs to know the frequency response of the propagation channel between each of its elements and each of the UEs. If the number of antennas increases, the beams can be focused more specifically to the UEs.

In the uplink, the operation is substantially the reverse of the downlink. The UE data

streams are encoded and transmitted over the same frequencies. Prior to decoding, the eNodeB receives the sum of data streams which are affected by the propagation channels. Utilizing CSI, a demodulation scheme is applied at the eNodeB to produce individual data streams from the received signals.

The eNodeB ensures that each UE receives only its intended signal. A large number of eNodeB antennas ensures that the effects of small-scale fading disappear, as does intra-cell interference among UEs [26]. The only remaining impediment is pilot contamination from the transmissions that are associated with the same pilot sequence used in channel estimation.

To show how mMIMO can increase spectral efficiency, consider a Uniform Linear Antenna Array (ULA), see Figure 2.10.

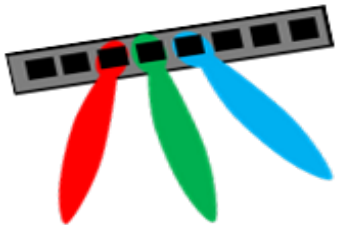


Figure 2.10.: Azimuth beamforming from a horizontal ULA

The antenna spacing is uniform at a distance of $\frac{\lambda}{2}$ where λ is the wavelength, and all the antennas have fixed radiation patterns and tilt all beams down with respect to the ground. Sending the same signal with different phase-shifts, the direction of the radiated signal will be different from the direction of individual antennas. Further, the elevation angle is the same for all the beams and the beamwidth in the azimuth domain shrinks with more antenna elements. In both MU-MIMO and mMIMO, multiple beams with different azimuth angles are created simultaneously.

To allow for steering beams in different elevation angles, the ULA array can be rotated so that the antennas are deployed at different heights above the ground, see Figure 2.11. To control both azimuth and elevation, multiple ULAs can be deployed on top of each other, see Figure 2.12.

The result is a Uniform Planar Array (UPA) with a massive number of antennas that give 2D beamforming gain wherever the UE is in the coverage area. As more antennas are added, the beams become narrower and easier to jointly steer in specific azimuth-elevation directions. As such, spectral efficiently in mMIMO can be increased by means of using narrow beams in specific azimuth-elevation directions.

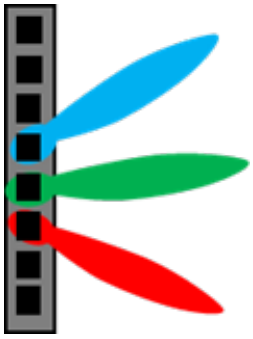


Figure 2.11.: Elevation beamforming from a vertical ULA

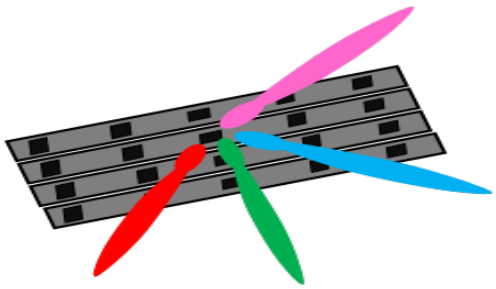


Figure 2.12.: Beamforming from a planar array

2.3.1. Grid of Beams and Relevant Channel Components

One major problem with FDD mMIMO is the need for more CSI when a large number of transmit antennas is used. Digital precoding, which requires CSI at the eNodeB, effectively compensates for the larger path loss in the higher frequency bands. Thus, accurate CSI for mMIMO must be obtained only near the eNodeB [27]. Digital precoding with fixed digital beamforming can be combined with a fixed analog beamformer to reduce the number of baseband signals that are passed to the digital precoder $\mathbf{W}$. Joint processing of the analog fixed beamforming and the digital precoding has been proposed to reduce the cost of simulteneous transmission of data streams in mMIMO systems [28].

A fixed analog beamformer is often called a Grid of Beams (GoB) precoder. This analog beamformer (GoB precoder) consists of phase shifters, power amplifiers and adders to reduce the number of baseband chains while keeping the throughput. GoB precoder limits UEs channel in the end to an effective channel via generating fixed wideband beams Figure 2.13 gives a structure of a mMIMO system at the eNodeB when a joint analog beamformer and digital precoding are employed. The GoB precoder $\mathbf{V}$ converts the M physical antenna arrays into a limited set of orthogonal narrow fixed GoB.

Considering an OFDM-mMIMO system with F baseband chains, K data streams at each subcarrier are precoded and converted into F precoded signals. The $M \times F$ GoB

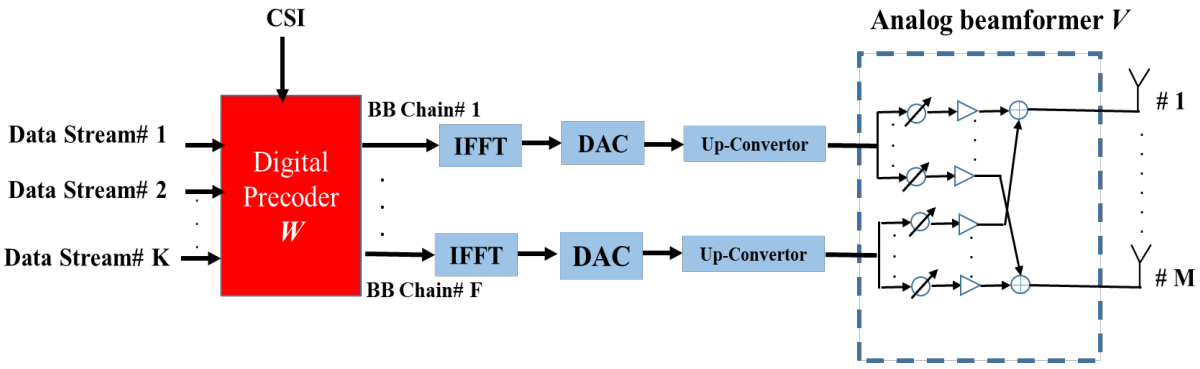


Figure 2.13.: Joint fixed GoB and digital precoding structure at the eNodeB

precoder $\mathbf{V}$ (comprised of $M \times F$ phase shifters and M power amplifiers) is applied after an Inverse Fast Fourier Transform (IFFT) and a Digital-to-Analog Converter (DAC).

The UEs thus receive only the effective channel matrices, whose number is relatively small compared to the number of entries of the mMIMO channel matrices [29]. In [30], GoB has been employed to subdivide a cell into radial subsectors in the azimuth direction. For the urban configuration considered in this thesis, an array of eight beams per azimuth and two beams per elevation directions generate 16 beams per cell (see Figure 2.14).

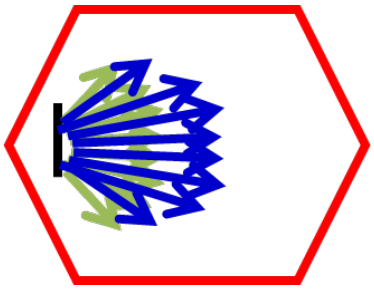


Figure 2.14.: GoB concept

Further, the number of fixed beams can be increased to 32 per cell if two polarizations per antenna element are applied. If $\mathbf{H}_i^b(t) = [\mathbf{H}_{i_1}^b(t), \mathbf{H}_{i_2}^b(t), ..., \mathbf{H}_{i_N}^b(t)]^T$ is the b_{th} matrix "tap" of the Channel Impulse Response (CIR) matrix $\mathbf{H}_i(t)$ of the link from eNodeB to the i_{th} UE at time instant t and N is the number of antennas at the UE, then (2.4) can be expressed as follows:

$$\begin{bmatrix} r_{i_1} \\ .. \\ .. \\ r_{i_N} \end{bmatrix}(t) = \begin{bmatrix} y_1 \\ .. \\ .. \\ y_M \end{bmatrix}(t) * \mathbf{H}_i(t) + \begin{bmatrix} Z_{1_b} \\ .. \\ .. \\ Z_{1_N} \end{bmatrix}(t) \tag{2.5}$$

or

$$\underline{r}_i(t) = \underline{y}(t) * \mathbf{H}_i(t) + \underline{Z}_i(t) = \sum_{b=0}^{N_{tap}-1} \overbrace{\mathbf{H}_i^b(t)}^{N\times M} \overbrace{\mathrm{y}(t-bT)}^{M\times 1}. \tag{2.6}$$

$\mathbf{H}_i(t)$ is then defined as:

$$\mathbf{H}_i(t) = \sum_{b=0}^{N_{tap}-1} \mathbf{H}_i^b(t)\delta(t-bT). \tag{2.7}$$

In (2.6) and (2.7), N_{tap} is total number of the taps and T is the sampling period. If we focus on a single link between the eNodeB and a specific UE, we drop the UE index i and (2.7) can be expressed as follows:

$$\mathbf{H}(t) = \sum_{b=0}^{N_{tap}-1} \mathbf{H}^b(t)\delta(t-bT). \tag{2.8}$$

In (2.8), $\mathbf{H}^b(t)$ is a matrix of size $N \times M$ where N is the number of antennas at the UE. The GoB azimuth and elevation characteristics influence the beamforming via a GoB precoder matrix $\mathbf{V}$ of size $M \times F$. The combination of GoB and mMIMO channel leads to [30]:

$$\mathbf{H}_{eff}(t) = \overbrace{\mathbf{H}(t)}^{N\times M} \overbrace{\mathbf{V}}^{M\times F} \tag{2.9}$$

where $\mathbf{H}_{eff}(t)$ represents the effective channel at time instant t. Considering (2.8) and (2.9), we have:

$$\mathbf{H}_{eff}(t) = \sum_{b=0}^{N_{tap}-1} \mathbf{H}^b(t)\mathbf{V}\delta(t-bT). \tag{2.10}$$

If we focus on a single matrix "tap" of $\mathbf{H}_{eff}(t)$ at a specific time instant, the effective channel with respect to the b_{th} "tap" of the channel impulse response matrix can be expressed as follows [31]:

$$\mathbf{H}_{eff}^b(t) = \mathbf{H}^b(t)\mathbf{V}. \tag{2.11}$$

A Channel Component (CC) in the matrix $\mathbf{H}_{eff}^b(t)$ represents the combination of the beams formed at the eNodeB (Tx) and the UE Rx beamformers at time t. As a result, $\mathbf{H}_{eff}^b(t)$ will be sparse due to the directivity of the generally narrow GoB beamformers. We will assume that the UEs receive those channel components with a power that is a predefined threshold above the Rx-power of the strongest beam [30]. These limited number of CCs are called Relevent Channel Components (RCCs) and they will be used to adjust the mMIMO matrix precoder, as discussed in Chapter 3. In the following, we will use $\mathbf{H}(t)$ to denote the effective matrix for simplicity.

2.3.2. OFDM Systems

Orthogonal Frequency-Division Multiplexing (OFDM) is a modulation scheme that uses a Discrete Fourier Transform (DFT) to decompose a frequency-selective channel into many parallel channels called *subcarriers*. The channel is converted to a cyclic convolution by prepending a cyclic prefix. This circular convolution is equivalent to multiplication in the frequency domain. The total bandwidth occupied by an OFDM symbol can be expressed as follows:

$$B_{symbol} = N_{PRBsub} B_{sep} = \frac{N_{PRBsub}}{T_{useful}} \tag{2.12}$$

where N_{PRBsub} denotes the number of subcarriers in one PRB, B_{sep} represents the frequency separation between neighboring subcarriers, and T_{useful} denotes the useful period of OFDM symbol transmission. B_{symbol} is greater than the channel coherence bandwidth B_{coh}, while the B_{sep} is smaller than B_{coh}. The total time period of an OFDM symbol is the sum of T_{useful} and the cyclic prefix time. Figure 2.15 demonstrates a Time-Frequency view of an OFDM symbol . Table 2.1 shows the sample OFDM parameters with a short cyclic prefix time that are employed in this thesis.

Parameter	Symbol	Value
OFDM symbol duration	T_{sym}	$\frac{1}{7}ms$
Useful OFDM symbol duration	T_{useful}	$\frac{1}{15}ms$
Cyclic prefix duration	T_{cp}	$\frac{1}{7\times15}ms$
Symbol bandwidth	B_{symbol}	180 kHz
Subcarrier spacing	B_{sep}	15 kHz
Coherence bandwidth	B_{coh}	105 kHz
Number of OFDM blocks per PRB	N_{OFDM}	7
Number of subcarriers in one PRB	N_{PRBsub}	12

Table 2.1.: OFDM parameters

To support MIMO systems with a large number of antennas and wide bandwidth, the combination of MIMO and OFDM was proposed [32]. MIMO-OFDM is based on transmitting OFDM signals through a number of antennas to achieve diversity. An example of a MIMO-OFDM system in the downlink is depicted in Figure 2.16. Each data signal is first passed through an OFDM modulator (OMOD) at the eNodeB. All data streams $x_1, x_2, ..., x_N$ are launched from their corresponding antennas simultaneously. At the receiver, each signal passes through an OFDM demodulator.

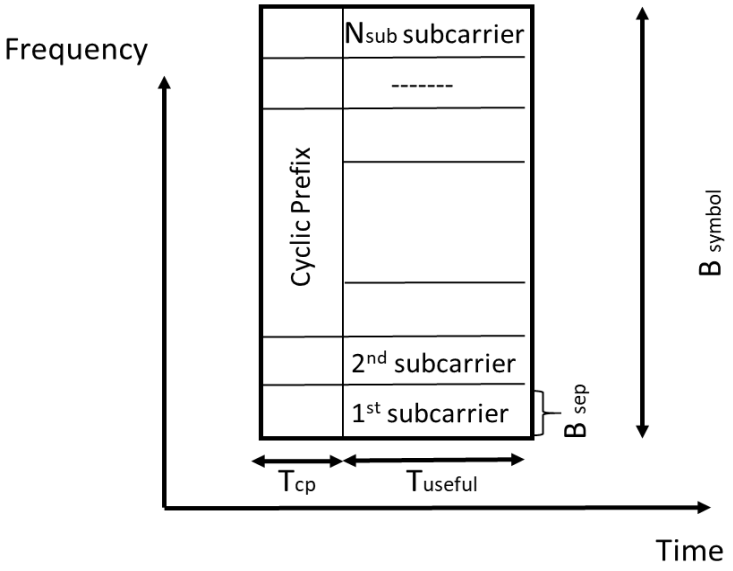


Figure 2.15.: OFDM structure

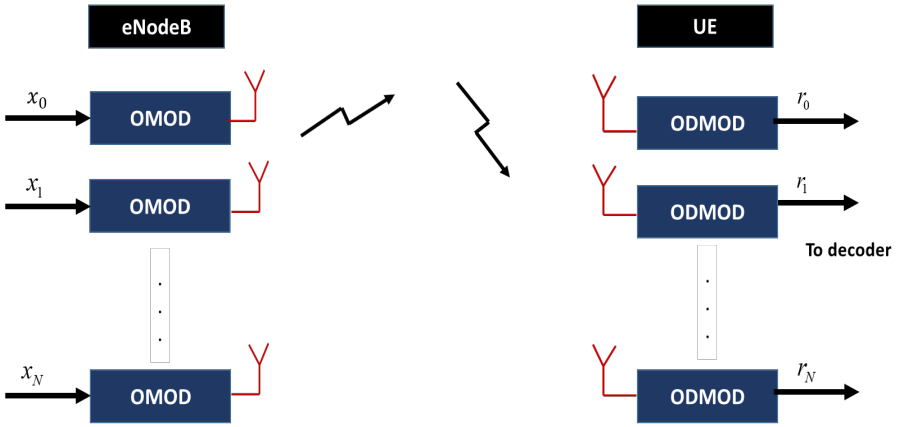


Figure 2.16.: MIMO-OFDM

2.4. Coordinated Multi Point

Coordinated multipoint (CoMP), or cooperative MIMO or Network MIMO, has been identified as a promising concept to improve spectral efficiency in MIMO-OFDM [32]. As presented in [32], [33], CoMP can help to mitigate inter-cell interference. 3GPP initiated a study item on CoMP for LTE-Advanced in March 2008 and this was followed by 3GPP Release 11 [34]. Since the same spectrum resources are used in CoMP multiple times due to spatial reuse, the interference signals at the edge between the cells are received with

similar power at the UEs. To exploit the interference, CoMP can be applied both in the uplink and downlink.

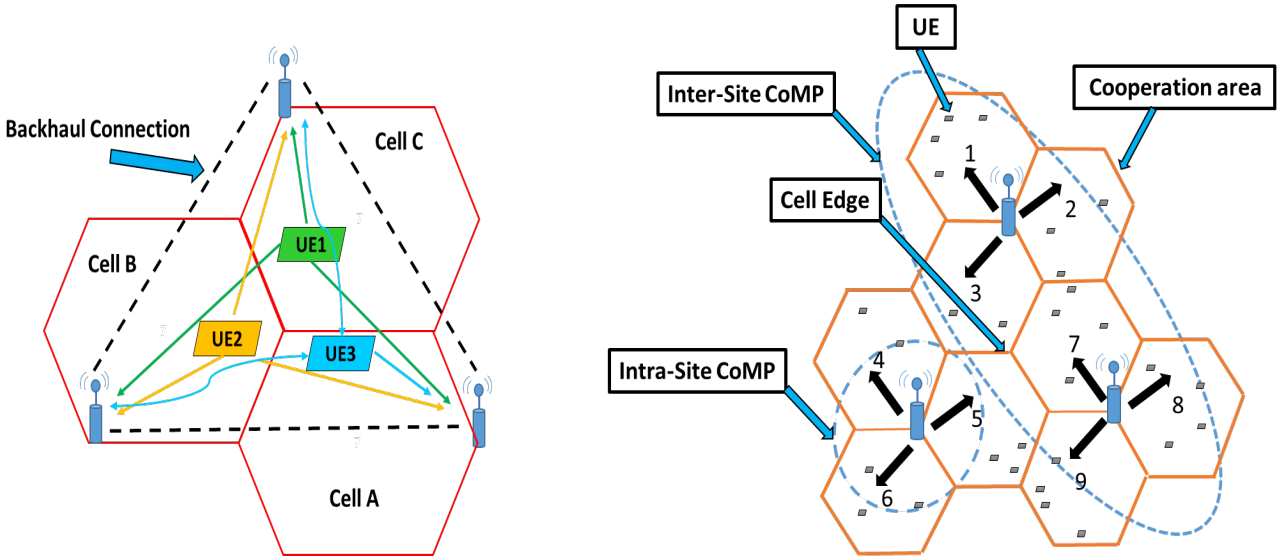
A typical setup for a CoMP area consisting of three cells is represented in Figure 2.17(a). Each of the UEs has a dominant link to its serving cell A, B or C. Further, there are strong links to the other two eNodeBs. EnodeBs are logically connected through a fast fiber link or a multi-hop connection involving different backhaul technologies. In this thesis, all the UEs use the same resources in frequency and time and the mutual interference is mitigated through multi-cell signal processing. Figure 2.17(b) depicts a cooperation architecture for CoMP in a Cooperation Area (CA) comprised of 9 cells or 3 sites. Multiple cells belonging to one eNodeB cooperate in *intra-site* CoMP. In contrast, *Inter-site* CoMP involves multiple eNodeBs. The cells at one site can have different links via fiber to a central unit.

Various levels of cooperation schemes in CoMP are studied in [35]. Two main downlink coordination categories are identified by 3GPP for LTE-Advanced [36], see Figure 2.18 and 2.19:

Joint Transmission (JT): CSI information and UE data are shared between the cooperation sites. This approach requires high backhaul bandwidth. Multiple sites can thus contribute to a single user by converting the interference signal to a useful signal. Transmission between the sites and UEs must be coherent for joint precoding and synchronized transmission.

Coordinated Scheduling/Beamforming (CS/CB): User data is not shared among all the cooperating sites. In other words, user data is available at one site, but scheduling and beamforming design is coordinated between sites. CSI is shared among all the cooperating sites. Moreover, beamforming vectors are selected such that the interfering joint transmissions are steered toward the null space of the interfered UE to minimize interference. In comparison with JT, CS/CB implementation requires less backhaul support.

This thesis concentrates on proposing solutions for JT CoMP. As expressed above, JT CoMP is a challenging technology where synchronization between cooperating sites and fast exchange of user data over the backhaul is essential. However, recent research validated that the concepts of JT CoMP including joint processing, CSI feedback and synchronous user data exchange can be feasible [35].



(a) CoMP setup (b) CoMP cooperation architecture

Figure 2.17.: CoMP

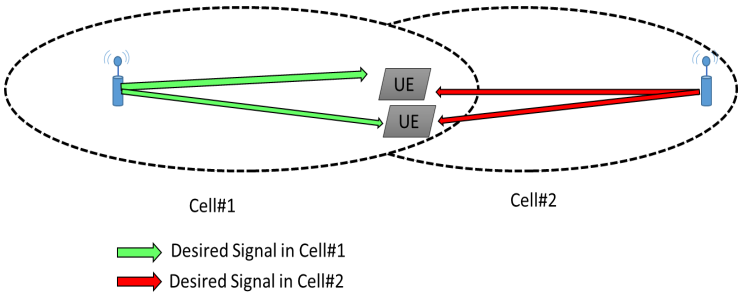


Figure 2.18.: Joint transmission

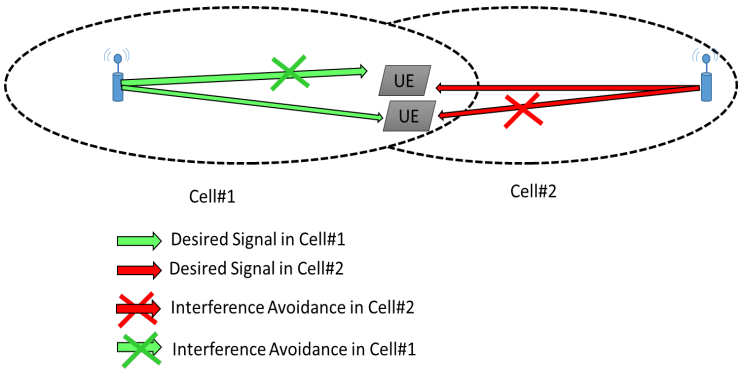


Figure 2.19.: Coordinated beamforming

2.5. Integrated mMIMO, GoBs, and JT CoMP

This section introduces the main concepts and components for integrating mMIMO, GoB and JT CoMP. An integrated mMIMO and JT CoMP provides a spatial multiplexing of a large number of UEs per cell combined with a centralized joint cooperation precoder to combat intra CA interference. Figure 2.20 shows the main characteristics of the integrated mMIMO, GoB and JT CoMP which are defined as follows [30], [37]:

1. Cooperation area :
   Forming cooperation areas varies from inter-site to intra-site JT CoMP. We concentrate on a JT CoMP scenario with a CA comprising of three sites (or nine cells).

2. Massive MU-MIMO:
   Due to the remarkable spatial multiplexing capability of mMIMO, many UEs can be served simultaneously in the CA.

3. mMIMO fixed GoB:
   As mentioned in section 2.3.1 the overhead due to the CSI reference signals can be limited by reducing the number of physical antenna elements to a small number of fixed GoBs. Thus, a limited number of channel components or beams will be received by the UEs. Due to the directivity of the analog GoB beamformer, UEs receive only a small number of beams and the channel matrix will become sparse. The number of beams and their direction depend on different scenarios and load conditions. The GoB precoding matrix $\mathbf{V}= [\underline{V}_1, \underline{V}_2, ..., \underline{V}_{32}]$ which was introduced in section 2.3.1, is used to generate the narrow beams, where the $\underline{V}_i$, i=1,2,...,32 are column vectors of length M.

4. CoMP Precoder:
   A flexible 5G solution adapts to different infrastructure availability and different UE capabilities. The precoding per UE for the MIMO precoding matrix $\mathbf{W}$ can be chosen by the eNodeB. Precoding can depend on the UE speed, UE reporting capabilities, reliability requirements, data packet size and load conditions.

5. Coded CSI:
   Availability of accurate CSI plays a vital role for JT CoMP MU-MIMO. As mentioned in section 2.2, high CSI overhead as well as the required uplink reporting makes FDD mMIMO design challenging. Employing GoB, hundreds of antenna elements can be reduced to more reasonable numbers. Since the channel matrix $\mathbf{H}$ is sparse and the number of RCCs is much smaller than the number of overall channel components, only a subset of RCCs will be received by each UE.

6. Multiple UE antennas:
   Employing beamforming at the UE can result in generating narrow Rx-beamformers

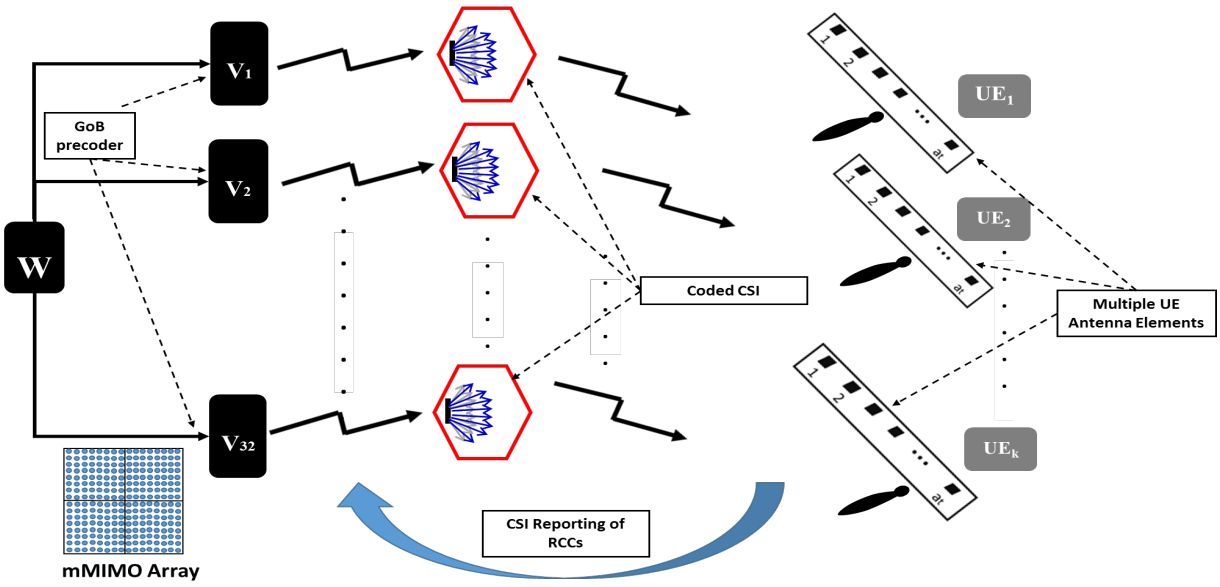


Figure 2.20.: Integrated GoB and mMIMO main components

and reducing RCCs or Relevant Multi Path Components (RMPCs). However, studies focusing on carrier frequencies below 6 GHz indicate that UE antenna elements are often limited to less than four or eight due to increased power consumption and other costs [3].

7. CSI Received Signals (RSs) estimation based on CIR:
   The CIR of one RCC in the time domain is denoted as CIR(t,γ) where γ is the delay at time t due to the MPCs. CSI RSs can be obtained in the frequency domain after an FFT operation. The MPCs define the frequency selectivity of the RCCs. We assume that the UE estimates the CIR based on the CSI RSs in the frequency domain Channel Transfer Function (CTF), i.e., CIR is obtained after an IFFT operation. The number of MPCs can be reduced to the number of relevant taps of the CIR. Each tap of the CIR represents the superposition of several MPCs in the time domain. If one Physical Resource Block (PRB) or multiple PRBs are selected for the frequency domain CTF, the number of relevant taps of the CIR is reduced [37]. As such, the CIR is reduced to a single tap corresponding to a single complex value per PRB. The required number of PRBs for PMI reporting in 3GPP LTE will be dramatically reduced if the wideband radio channel can be characterized by a reduced number of taps.

8. CSI reporting of RCCs:
   Accurate CSI at the eNodeB is required for advanced interference mitigation in integrated mMIMO and JT CoMP scenarios. One often restricts reporting to a relatively small number of RCCs above a certain power threshold.

2.6. 3GPP 3D Urban Macro Channel Model and Grid of Beams

To improve the spectral efficiency of a radio-link, one can exploit 3D spatial dimensions (azimuth and elevation) and use mMIMO configurations. An incorporation of a 3D channel model into the 3GPP evaluation methodology was started in January 2013 [3]. The 3D model is described as a Geometry-based Stochastic Channel Model (GSCM) which extends the 2D channel models from WINNERII and ITU [38], [39].

The 3GPP 3D Urban Macro Channel (3D-UMa) is considered a typical usage scenario for elevation beamforming in FDD-mMIMO systems [38]. The 3D-UMa applies to carrier frequencies between 2-6 GHz and bandwidths up to 100MHz. Higher carrier frequencies, e.g., up to 300 GHz, have a different set of challenges and are out of the scope of this thesis.

2.6.1. Antenna Modeling

We focus on a mMIMO 2D array of size 16×16 (16 elements per row and 16 elements per colunm). As discussed in section 2.3, to permit 3D beamforming in space, radio waves in both vertical (elevation) and horizontal (azimuth) directions in the structured 2D antenna array can be controlled. To increase the number of antenna elements up to 512, two cross polarized antennas are placed at each position (see Figure 2.21). Moreover, to split the power in the vertical and horizontal directions equally for all the UEs, a constant polarization model is used.

2.6.2. LOS Probability and Pathloss Modeling

LOS probability modeling in 3D-UMa is inspired by the stochastic modeling approach used in WINNER II [38]. To reuse the existing modeling parameters from 2D stochastic models and to reduce the processing time for system level simulations, building/street dimensions are not considered.

The 3D-UMa LOS probability is a function of height and distance, and is obtained by the sum of type-1 and type-2 LOS probabilities. As introduced in the ITU model [39], the type-1 LOS probability depends only on the horizontal distance between the eNodeB and the UEs. A UE on a high floor of a building is considered to be in type-2 LOS state if the UE on a first floor can never achieve a LOS state. In 3D-UMa, type-2 LOS states occur when the UEs are located on a floor higher than 12m.

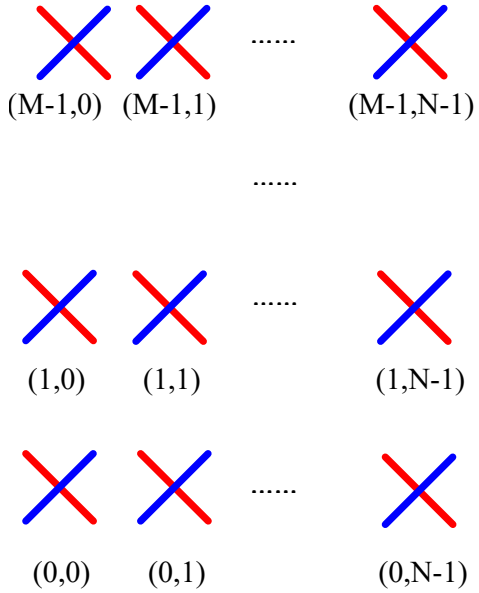


Figure 2.21.: Cross-polarized 2D antenna array structure at the eNodeB

LOS pathloss modeling in 3D-UMa assumes two-ray propagation resulting in a pathloss equation based on an environmental height, i.e., the height of a dominant reflection from the ground that can be added constructively or destructively to the direct ray received by the UE located at the street level. The environmental height is assumed to be 1m for a UE associated with a type-1 LOS condition. In case of type-2 condition, the environmental height is randomly determined from a discrete uniform distribution [40].

Non-Line of Sight (NLOS) pathloss modeling in 3D-UMa is dominated by the paths traveling via multiple diffraction over the rooftops followed by diffraction of edges of buildings. To model this phenomenon, a linear height gain term is given by $-\alpha(h-1-5)$ where α (in dB/m) represents the gain coefficient and h is the UE height.

2.6.3. Fast Fading Model

Each path between a transmitter and receiver is defined by a number of rays. The rays at the transmitter are characterized by their angles-of-departure (AoD), power and delay, and the rays at the receiver are characterized by their angles-of-arrival (AoA). The coordinate system for 3D-UMa is depicted in Figure 2.22 where θ denotes the zenith angle and φ is the azimuth angle. If $\theta = 0$ the direction points to the vertical, and if $\theta = 90$ the direction points to the horizon. The direction of propagation $\hat{n}$ defines the spherical basis vectors $\hat{\theta}$ and $\hat{\varphi}$. The procedure for generating 3D-UMa channel coefficients is shown in Figure 2.23 and has the following steps [41]:

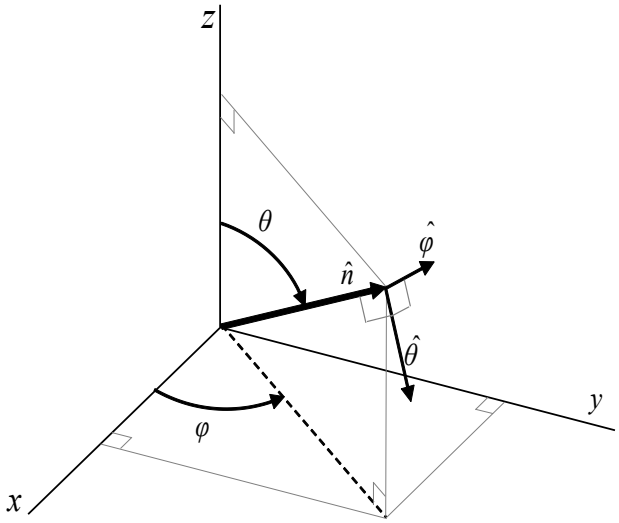


Figure 2.22.: 3D-UMa channel coordinate system

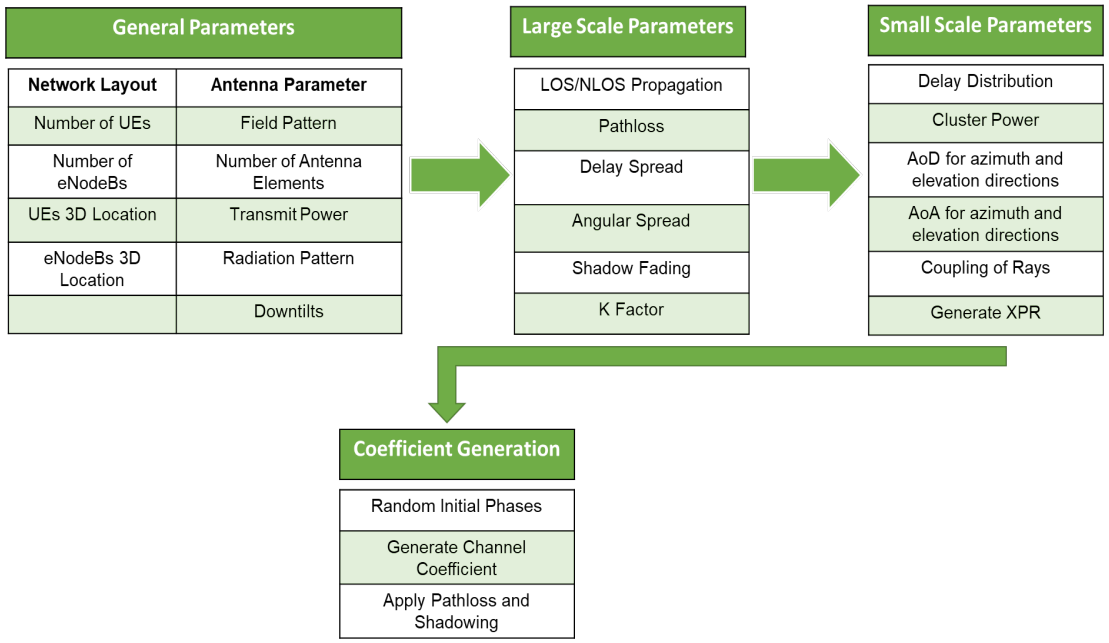


Figure 2.23.: Procedure to generate 3D-UMa channel coefficients [3]

1. General Parameters :
   Network layout is specified, including the number of UEs, eNodeBs and their 3D locations and antenna parameters like field patterns, number of antenna elements, transmit power, radiation pattern, downtilts.

2. Large Scale Parameters :
   Once the general parameters are defined, large scale parameters are specified between

the transmitter and receiver. Large scale parameters include LOS/NLOS propagation conditions, path loss, delay spread, angular spread, Ricean K factor and shadow fading. The maximum RMS azimuth arrival and azimuth departure spread values are set to 104 degrees, i.e., σ_{ASA}=min($\sigma_{ASA}, 104°$), σ_{ADA}=min($\sigma_{ADA}, 104°$). The maximum RMS zenith arrival and zenith departure spread values are set to to 52 degrees, i.e., σ_{ZSA}=min($\sigma_{ZSA}, 52°$), σ_{ZDA}=min($\sigma_{ZDA}, 52°$)

3. Small Scale Parameters :
   Small scale parameters are obtained randomly based on the 3D-UMa stochastic distribution functions defined in [3] and randomly drawn large scale parameters. The main small scale parameters are delay distribution and cluster powers. AoA and AoD for both azimuth and elevation directions of each ray are obtained randomly. Once the AoA and AoD are defined, the rays within a cluster are coupled randomly.

   Eventually, the Cross Polarization Power Ratios (XPR) for each ray of each cluster are calculated. Note that the difference in the distances travelled by the plane wave between different antennas leads to a difference in the phases of the corresponding fast fading channel components.

4. Coefficient Generation :
   As the next step, a third level of randomness is introduced by the random initial phases of the scatterers in each path. Channel coefficients are generated for each cluster and for each transmitter and receiver pair. As discussed, each ray with different AoA and AoD travels different distances to different antenna elements. As such, the sum of the channel coefficients corresponding to each ray leads to different channel coefficients corresponding to each path.

The stochastic parameters of the integrated 3D-UMa and GoB channel are independent of the antenna elements. Further, GoB is independent of the individual realizations of the stochastic parameters of the channel when it is applied to the antenna elements.

2.6.4. UE Cell Selection

For the simulation evaluations described in Chapter 3 and 4, a JT CoMP cooperation area of size 1320m × 1160m comprised of three adjacent sites or nine cells with three sector sites is considered. Each UE is allocated to a cell if the average received power of 32 beams from the corresponding cell is the highest received power among all the existing 9 eNodeBs. Figure 2.24 depicts the UEs' distribution in the JT CoMP cooperation area. UEs in each cell are specified with the same color. Inter-site distance is taken as 500m.

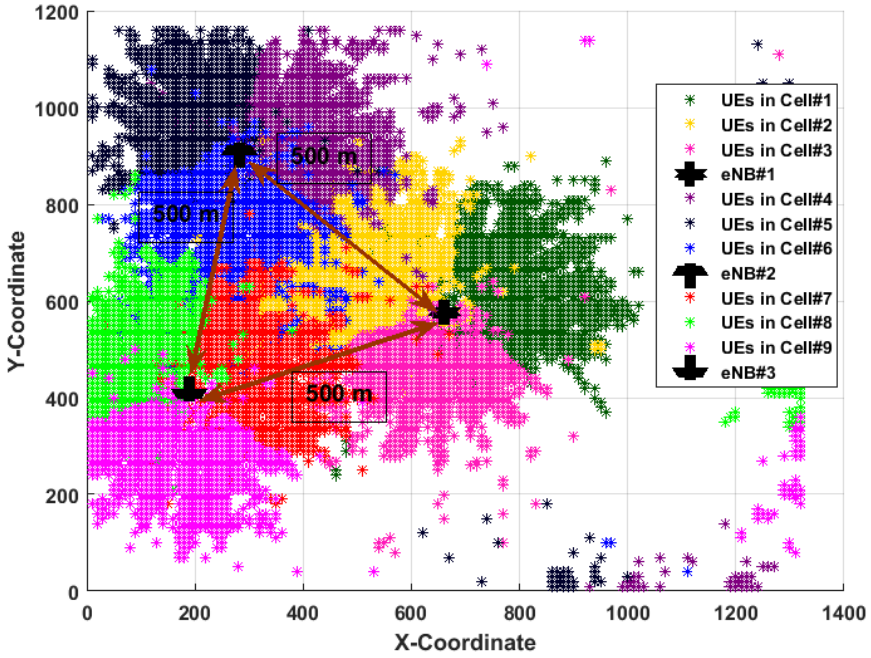


Figure 2.24.: UE distribution [4]

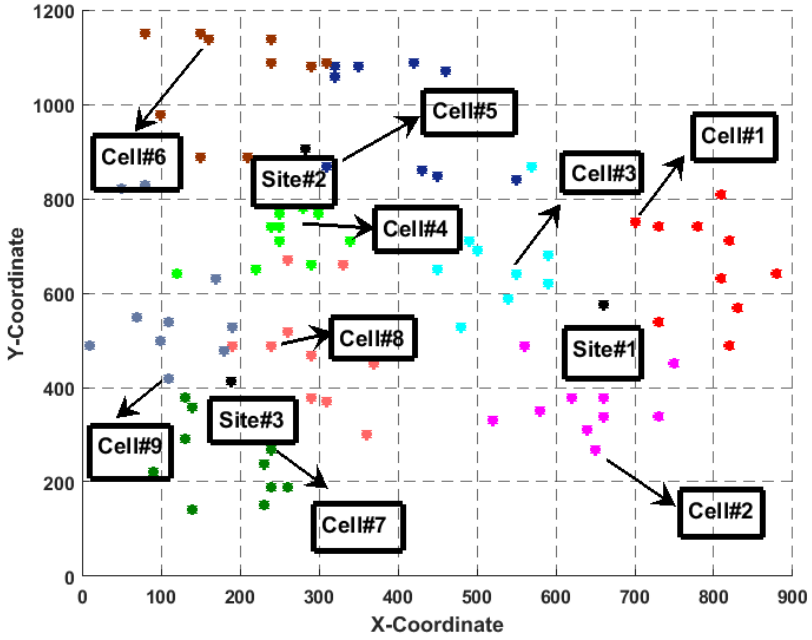


Figure 2.25.: Randomly selected UE distribution [4]

To specify the simulation results, 10 UEs per cell are selected though a uniform random distribution so that for a three-site JT CoMP cooperation there will be 90 UEs distributed in 9 cells. Figure 2.25 shows a sample of the randomly selected 90 UE positions in the cooperation area.

2.6.5. SINR and Spectral Efficiency Models

This section describes the SINR and spectral efficiency models that we use for our simulation results detailed in Chapter 3 and 4. The ZFBF precoder $\mathbf{W}$ is given by the $M \times N$ matrix

$$\mathbf{W} = \overbrace{\mathbf{H}^{\mathsf{H}}}^{\mathrm{M \times N}} \overbrace{(\mathbf{H}\mathbf{H}^{\mathsf{H}})^{-1}}^{\mathrm{N \times N}} \tag{2.13}$$

where $\mathbf{H}^{\mathsf{H}}$ is the complex-conjugate transpose of $\mathbf{H}$. The received signal is

$$\underline{y} = \mathbf{H}\mathbf{W}\underline{x} + \underline{n} \tag{2.14}$$

where $\underline{x}$ indicates the symbol vector. We define the noise-free signal

$$\underline{\widetilde{y}} = \mathbf{H}\mathbf{W}\underline{x}. \tag{2.15}$$

The essential simulation parameters are summarized in Table 2.2 The transmit power per cell is equally distributed over 1200 subcarriers and 32 beams with a bandwidth of 18 MHz around a carrier frequency of 2.1 GHz. The derived SINR per UE can be expressed as [5]:

$$\mathrm{SINR[dB]} = P_{tx}[\mathrm{dBm}] - \mathrm{P}_n[\mathrm{dBm}] - \mathrm{I}_{UE(i)} - \mathrm{PNL}_{UE(i)}[\mathrm{dB}] \tag{2.16}$$

where

$$P_{tx} = \mathrm{Tx\ Power}_{dBm}/N_{UE} - 10\log_{10} N_{Totalsub} \tag{2.17}$$

$$P_n = -173.83 + 10\log_{10} N_{SubSp} + \mathrm{Receiver\ NF\ [dB]}. \tag{2.18}$$

P_{tx} specifies the transmit power distributed equally among all the UEs. $\mathrm{I}_{UE(i)}$ denotes the received interference at UE i in dBm. N_{UE} is number of UEs in one cell, $N_{Totalsub}$ is the total number of subcarriers in 100 PRBs. In order to observe the effect of noise on the channel, the noise formulation referred to as P_n has been used. B_{sep} denotes the subcarrier spacing and Receiver NF refers to the receiver noise figure in dB. N is either 8 (for UE beamforming experiments) or 1.

The spectral efficiency per cell in bit/sec/Hz for each UE is [5]:

$$SE_{bit/sec/Hz/cell} = \frac{90 \times C_{LTE} \times 1200 \times 7}{1_{ms} \times 20_{MHz} \times 9}. \tag{2.19}$$

Symbols	PHY Layer Parameters Value	Value
f_c	Carrier Frequency (GHz)	2.1
TB	Total Bandwidth (MHz)	20
UB	Used Bandwidth (MHz)	18
B_{sep}	Subcarrier Spacing (KHz)	15
$N_{Totalsub}$	Number of Subcarriers	1200
N_{PRB}	Number of PRBs	100
Tx Power$_{dBm}$	Tx Power per UE per cell (dBm)	46
$ReceiverNF$	Receiver NF (dB)	7
N_{UE}	Number of UEs	90
N	Number of Antennas per UE	1 or 8
N_{cell}	Number of Cells	9
N_{Beams}	Number of Tx Beams	288
N_{OFDM}	Number of OFDM blocks per PRB	7
TTI	Duration of one subframe (ms)	1

Table 2.2.: Simulation Parameters

C_{LTE} in (2.19) is average number of bits per modulation symbol based on the 3GPP physical layer procedures described in [42], where the SINR is mapped to an LTE Modulation and Coding Scheme (MCS). N_{UE}=90 in (2.19) is the number of UEs, $N_{Totalsub}$=1200 is the number of subcarriers, N_{OFDM}=7 is the number of OFDM symbols per PRB, TTI= 1_{ms} is the duration of one subframe, TB =20_{MHz} is the used bandwidth and N_{cell}=9 is the number of cells.

2.6.6. Scenario and Network Development Setup

When the UE cell selection (explained in section 2.6.4) is completed, the following steps are taken (see Figure 2.26):

Figure 2.26.: Channel matrix generation

1. Path loss, e.g. Reference Signal Received Power (RSRP), measurements for all Tx-beams are collected per UE. With 32 beams per cell and nine cells per cooperation area there are up to $32 \times 9 = 288$ channel components in total per UE.

2. Once the channel components are collected at the UE, an FFT operation of size 2048 is applied to the Rx-signals of the UE antenna element.

3. The frequency domain signal which is created after the FFTs is limited to 18 MHz bandwidth with 1200 sub carriers, or 100 PRBs with 12 sub carriers, times 7 OFDM symbols.

2.7. UE Beamforming

2.7.1. Linear Receiver Beamformer for JT CoMP Scenarios

Low complexity linear precoding with perfect CSI for MIMO beamforming at the UE has received attention in recent studies [43], [44], [45]. The maximum SINR subject to a total power constraint per eNodeB was studied in [46], [47]. Decentralized beamforming algorithms, which achieve Pareto-boundary rates, were proposed for the MIMO interference channel [48].

Investigating the FDD mMIMO JT CoMP scenario, the feasibility of handling the overhead due to the CSI estimation is essential. GoB can reduce feedback by limiting the transmit signals to a set of narrow beams. When little CSI feedback is reported, the sensitivity of channel estimation errors to perform MU-MIMO precoding will be reduced [33], of course at the expense of accuracy. GoB provides efficient feedback by reporting only the RCCs within a threshold with respect to the strongest channel matrix component observed by the UE. Focusing on the JT CoMP urban macro scenarios detailed in Chapter 2, eight Tx-beams are generated through GoB in the azimuth direction and two in the elevation direction per polarization. Figure 2.27 depicts the approach of multiple receive antennas when using linear beamforming at the receiver.

2.7.2. Maximum Ratio Combining

A common method to combat multipath fading is antenna diversity. A classic combining technique is maximum-ratio-combining (MRC), where the signals from the receive antenna elements are weighted such that the SINR of their sum is maximized. When MRC is used, the receiver and transmitter can combat severe fading effects. There have been various transmit diversity techniques proposed in literature. For example, a delay transmit diversity scheme was proposed by [49] where the signal replicas are transmitted through multiple antennas at different times.

A simple but effective example of transmit diversity was proposed by [50], where first

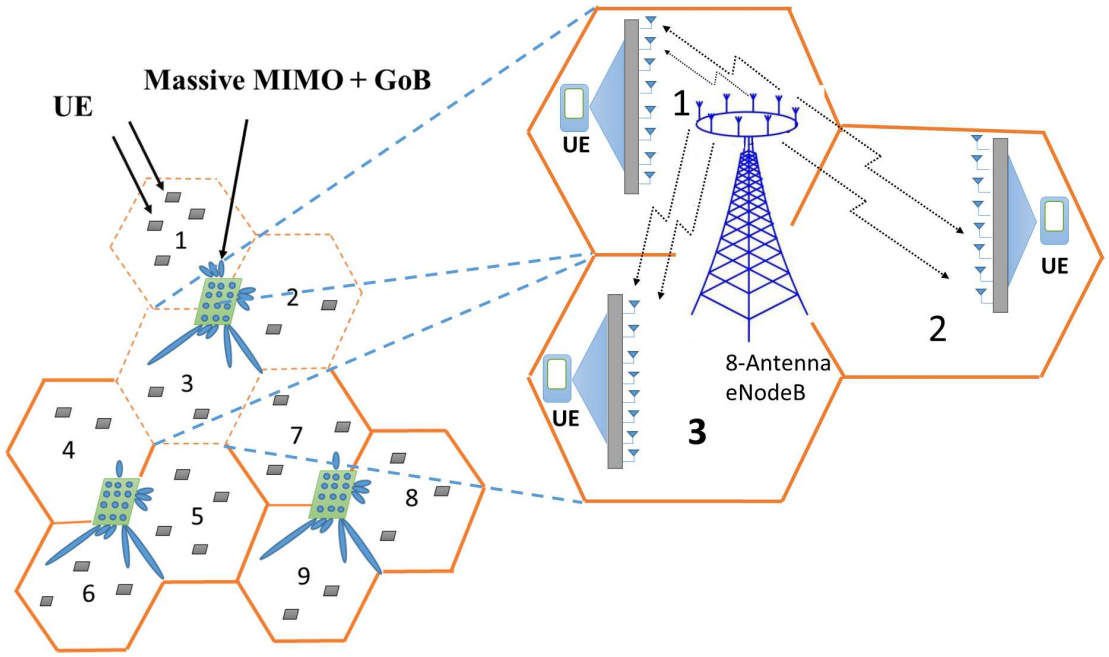


Figure 2.27.: UE MRC beamforming [4]

a pair of symbols is transmitted using two antennas, and then the transformed version of the pair is transmitted to obtain MRC-like diversity. These transmit diversity techniques maximize the SINR, but not the rate. To establish MRC aiming at maximizing SINR of the combiner output with multiple receive antennas, we assume the UEs are equipped with eight antenna elements and they employ an MRC Rx-beamformer with respect to the strongest Tx-beam. If $be_d(k)$ denotes the gain of the strongest Tx-beam of the k_{th} UE at the d_{th} antenna and $h_d(k)$ represents the k_{th} channel component of the d_{th} antenna, the MRC weight is given by a complex correlation between the two signals [4]:

$$\omega_d(k) = be_d(k)^{\mathsf{H}} h_d(k) \tag{2.20}$$

where $\omega_d(k)$ is the k_{th} MRC weight for the related Tx-beam $be_d(k)$. The MRC output for the d_{th} antenna at the k_{th} UE is then given by the sum of all the weighted signals:

$$h_{MRC}(k) = \sum_{d=1}^{8} \omega_d(k) be_d(k). \tag{2.21}$$

The same beamforming weights will affect the reception of all other beams from the same UE.

MRC UE beamforming can be applied to the system models in section 2.5 and section 2.6, where the channel precoding matrix $\mathbf{W}$ is calculated through the pseudo inverse

algorithm proposed in [4]. The sparse effective channel matrix **H** used is based on the RCCs above the power threshold relative to the strongest beam. Figure 2.28 illustrates the system architecture when linear beamforming is applied at the UE. The GoB is generated through a fixed wideband uniform beamforming vector **V** being combined into the GoB precoding matrix. If each UE is equipped with eight Rx antennas, the 256 channel components can be reduced when an MRC beamformer of size $\mathbb{C}^{8\times1}$ is employed.

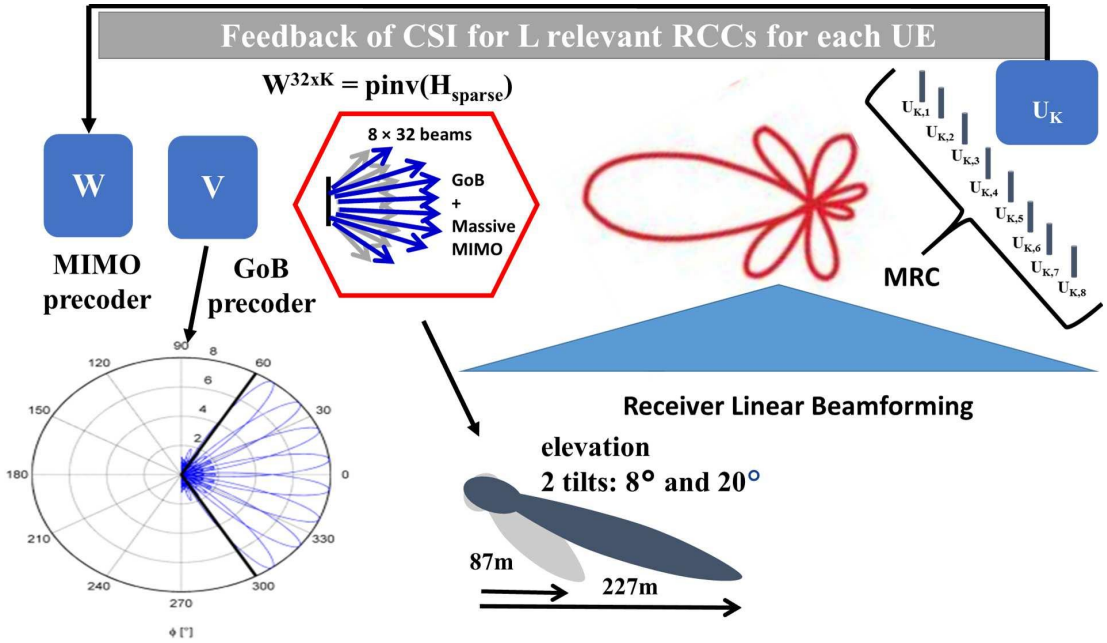


Figure 2.28.: UE beamforming per user per cell [4]

Using a geometry-based 3D channel model, the GoB is obtained by a first fixed precoder applying two different antenna tilt values (8° and 20°) to 8 azimuth beams with two polarizations each. In order to apply the proposed scheme, we take the following steps:

1. UEs are allocated to cells based on the strongest received beam. In other words, UEs located in the first cell have the beam with the highest received power in the first cell.

2. 10 UEs per cell are selected randomly so that for a three site cooperation area there will be 90 UEs distributed in the 9 cells.

3. RSRP measurements for all Tx-beams are collected by each UE. With 32 beams per cell and nine cells per cooperation area there are $32 \times 9 = 288$ channel components in total per UE.

4. Once the channel components are collected at the UE, an FFT of size 2048 is applied to each Rx-signal for each of the eight UE antenna elements.

5. The frequency domain signal is created by the FFT at the Transceiver (TRx) and is limited to a 18 MHz bandwidth with 1200 subcarrier or 100 PRBs of size 12 subcarriers with 7 OFDM symbols.

6. Applying the proposed MRC method at the UEs over 8 receiver antenna elements, the effective RCCs of the channel matrix are obtained.

2.7.3. Power Normalization Loss

As a result of applying GoB to the RCCs, some of the UEs receive the same beam or a subset of the same beams as their strongest beam(s). As such, a high correlation of these channels will often lead to a *high* Power Normalization Loss (PNL) [29]. PNL for ZF precoders degrades the overall performance. To combat this, the UEs that generate a very high PNL even with all the active beams should be removed from the served multi-user set. Apart from allocating a separate subband with different setups of active beams to such UEs, all the active beams for the remaining UEs must be considered.

However, the overhead for CSI reporting is still rather high. To find the best beam deactivation pattern for a set of UEs, the number of RCCs should not impose too high PNL to the network as compared with the case that all the beams are active, and the number of beams that contribute to the interference floors of different UEs must be reduced. Characterizing the performance of the overall solution, the PNL of UE i can be defined as

$$PNL_{dB} = 10\log_{10}\sum_{i=1}^{N_{UE}}\sum_{j=1}^{N_{Beams}}\|\varphi\|_{ij}^{2}; \quad \|\varphi\|_{ij} = W_{ij} \tag{2.22}$$

where

$$\mathbf{W} = \text{pinv}(\mathbf{H}_{beam}(\underline{B})). \tag{2.23}$$

W_{ij} in (2.22) is the i_{th} row and j_{th} column element of matrix $\mathbf{W}$. Vector $\underline{B}$ in (2.23) represents the beams $[be_1, be_2, ..., be_{N_{AC}beam}]$. If $P_{Rx}(H_{ij})$ denotes the Rx power of beam j at UE i, the total number of RCCs can be calculated by applying an RCC power threshold P_{TH} where $P_{Rx}(H_{ij}) > P_{TH}$. As such, those channel components with $P_{TH} > P_{Rx}(H_{ij})$ are considered as interference floor components. Reducing the number of interference floor channel components not only decreases the inter-stream interference, but also decreases additional PNL as compared to the full beam activation case.

3

mMIMO JT CoMP Precoding for a Single Subcarrier

3.1. Introduction

This chapter provides precoding analysis of JT CoMP focusing on a single subcarrier of an OFDM signal. As introduced in Chapter 2, the idea of MU-MIMO, where the base station is equipped with hundreds of antenna elements, is widely believed to be a key technology for future (below 6GHz) 5G mobile networks. When MU-MIMO is combined with GoB and JT CoMP, spectral efficiency increases even more compared to small-scale MIMO technology. In order to benefit from such integration, precoding algorithms are essential to avoid inter-cell interference.

Existing works show that linear precoding techniques such as Zero-Forcing (ZF) can approach capacity [51], [52], [53]. ZF precoding involves the inverse of the *Gram* matrix of the UEs' channel vectors. The computational complexity of inverting the Gram matrix is on the order of $O(K^3)$ with respect to the number K of UEs. To reduce complexity, there have been many efforts recently to use, e.g., SVD or QR decomposition. Furthermore, simple matrix multiplication via transforming the inverse of a Gram matrix into that of a simple precondition matrix using *Neumann* series can be applied [54], [55], [56]. Besides, to address the complexity challenges, an accurate analysis of complexity is essential.

In this chapter, the main focus is on ZF precoding methods that reduce the complexity of JT CoMP with sparse channel matrices. Moreover, we develop a flexible toolbox that provides routines to count the number of operations. This chapter is organized as

follows. In section 3.2, we introduce the main computational complexity characteristics of linear precoding for single subcarrier JT CoMP channel matrices. section 3.3 provides an overview of the conventional methods to compute the pseudo inverse of a channel matrix. section 3.4 proposes a new approach to exploit sparsity to reduce the computational complexity. The SINR performance of the proposed pseudo inverse algorithm is compared with a reference method [57], [58], [59]. In section 3.5, we propose linear beamforming at the UE to increase the spectral efficiency for single-subcarrier channel matrices. We show that linear beamforming at the UE can outperform the non-beamformed performance by reducing the number of RCCs. The SINR, spectral efficiency and the computational complexity of the proposed beamforming technique are compared with the reference model in [59].

3.2. Single-subcarrier Linear Precoding

OFDM is commonly used in multi user wireline and wireless communications. By dedicating to each UE a fraction of the available number of subcarriers, multiple-access interference is mitigated which leads to a larger system capacity. As discussed in Chapter 2, ZFBF enables decoupling into a system with a small number of parallel independent subchannels. Basically, each substream in a ZFBF system is independently coded and a linear processing is applied at the transmitter to eliminate interference among different substreams utilizing the channel spatial dimensions. Further, ZFBF can be carried out by means of beamforming weight vectors obtained from the columns of the Moore-Penrose pseudo inverse of the $m \times n$ channel matrix $\mathbf{H}$ as follows:

$$\begin{aligned} \mathbf{H}^{\dagger} &= \mathbf{H}^{\mathsf{H}}(\mathbf{H}\mathbf{H}^{\mathsf{H}})^{-1} \\ \mathbf{W} &= [\frac{\mathbf{H}_1^{\dagger}}{\|\mathbf{H}_1^{\dagger}\|}, \frac{\mathbf{H}_2^{\dagger}}{\|\mathbf{H}_2^{\dagger}\|}, ... \frac{\mathbf{H}_K^{\dagger}}{\|\mathbf{H}_K^{\dagger}\|}] \end{aligned} \tag{3.1}$$

where $\mathbf{H}^{\dagger}$ is the pseudo inverse, $\mathbf{H}^{\mathsf{H}}$ denotes complex conjugate transpose, and $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, ...\mathbf{H}_K]$. Note that we are now using $m \times n$ matrices $\mathbf{H}$, rather than $n \times M$ or $N \times M$ matrices $\mathbf{H}$ in Chapter 2. The inversion in (3.1) is possible if $\mathbf{H}$ has $m \leq n$ and has full rank.

In JT CoMP scenarios where both the UEs' CSI and data information are shared, precoding design faces numerous challenges. Most of these challenges are dominated by the computation of the inverse of the so-called Gram Matrix $\mathbf{G}$ which can be expressed as follows:

$$\begin{aligned} \text{case} \quad & m \leq n : \mathbf{G} = \mathbf{H}\mathbf{H}^{\mathsf{H}} \\ \text{case} \quad & m \geq n : \mathbf{G} = \mathbf{H}^{\mathsf{H}}\mathbf{H}. \end{aligned} \tag{3.2}$$

We will usually consider the former case $m \leq n$. In OFDM, we must apply a matrix

G for each subcarrier. Thus, developing efficient precoding algorithms is essential. The Hermitian structure of **H** can reduce the complexity due to the possibility of ignoring its lower triangular part. As such, the m main diagonal entries as well as the $\frac{m^2-m}{2}$ upper off-diagonal elements of **H** are taken into account. Computing each element of $\mathbf{G}$ requires n multiplications and $n-1$ additions, so there will be $\frac{1}{2}mn(m+1)$ multiplications and $\frac{1}{2}(n-1)m(m+1)$ additions in total. This gives $nm^2+mn-\frac{m^2}{2}-\frac{m}{2}$ FLOPs. Further, computing $\mathbf{G}^{-1}$ yields a total of m^3+m^2+m FLOPs using efficient methods like Cholesky factorization [60]. Since the Transmission Time Interval (TTI) is 1_{ms} for 7 OFDM symbols and we have 100 subdivided PRBs per TTI, the total number of FLOPs per symbol per second for a single-subcarrier will be on the order of 10^5 [5]. To investigate the significance of reducing the computational complexity in JT CoMP single subcarrier scenarios, we compare the complexity of existing methods to the proposed scheme.

3.3. Pseudo Inverse of Channel Matrices

We next give essential definitions and review the state-of-the-art methods to compute the Moore-Penrose pseudo inverse of a matrix. The matrix $\mathbf{H}^\dagger$ is the unique matrix satisfying the four Penrose equations [61]:

$$\begin{aligned}
&(i) \quad \mathbf{H}\mathbf{H}^\dagger\mathbf{H} &&= \mathbf{H}\\
&(ii) \quad \mathbf{H}^\dagger\mathbf{H}\mathbf{H}^\dagger &&= \mathbf{H}^\dagger\\
&(iii) \quad (\mathbf{H}\mathbf{H}^\dagger)^{\mathsf{H}} &&= \mathbf{H}\mathbf{H}^\dagger\\
&(iv) \quad (\mathbf{H}^\dagger\mathbf{H})^{\mathsf{H}} &&= \mathbf{H}^\dagger\mathbf{H}.
\end{aligned}$$

If **H** is square and non-singular, then its inverse satisfies all defining properties of a Moore-Penrose pseudo inverse. When the matrix **H** is full rank, then the Moore-Penrose pseudo inverse can be directly calculated as follows:

$$\begin{aligned}
&\text{case} \quad m \leq n : \mathbf{H}^\dagger = \mathbf{H}^{\mathsf{H}}\mathbf{G}^{-1}\\
&\text{case} \quad m \geq n : \mathbf{H}^\dagger = \mathbf{G}^{-1}\mathbf{H}^{\mathsf{H}}.
\end{aligned} \tag{3.3}$$

,

3.3.1. SVD Decomposition

One of the most commonly used methods to compute $\mathbf{H}^\dagger$ is Singular Value Decomposition (SVD). For any $\mathbf{H} \in \mathbb{C}^{m\times n}$, there exist unitary matrices $\mathbf{U} \in \mathbb{C}^{m\times m}$ and $\mathbf{V} \in \mathbb{C}^{n\times n}$ such that:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}} \tag{3.4}$$

where Σ is an $m \times n$ diagonal matrix having the form:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \dots 0 & 0 & 0 \\ 0 & \sigma_2 & 0 \dots 0 & 0 & 0 \\ 0 & 0 & \sigma_m & 0 \dots 0 & 0 \end{bmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0, \quad \text{for} \quad m \leq n. \tag{3.5}$$

The columns of $\mathbf{U}$ are the left singular vectors, while the columns of matrix $\mathbf{V}$ are the right singular vectors. As the first step to implement an SVD algorithm, the matrix $\mathbf{H}$ will be upper diagonalized. By applying unitary Householder reflections $\mathbf{U}_B$ and $\mathbf{V}_B$, a bidiagonal matrix $\mathbf{B}$ is derived as [62]:

$$\begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} = \mathbf{U}_B^{\mathsf{H}} \mathbf{H} \mathbf{V}_B. \tag{3.6}$$

Householder bidiagonalization of $\mathbf{H}$ requires $8(m-j)(n-j)$ FLOPs for iteration j computing the product of the transpose of $\mathbf{H}(j:m, j:n)$ and a Householder vector $\underline{v}$ of $\mathbf{H}$. The complexity of scaling this product by β to compute the Householder reflector $\underline{w}$ can be ignored. Computing $\underline{v}\ \underline{w}^{\mathsf{H}}$ needs $6(m-j)(n-j)$ FLOPs and subtracting $\underline{v}\ \underline{w}^{\mathsf{H}}$ from $\mathbf{H}$ requires $2(m-j)(n-j)$ FLOPs. Additional computations are taken into account if $j \leq n-2$, where computing the product of $\mathbf{H}(j:m, j+1:n)$ and $\underline{v}$ requires $8(m-j)(n-j)$ FLOPs. Considering $6(m-j)(n-j-1)$ FLOPs for $\underline{v}\ \underline{w}^{\mathsf{H}}$ and its subtraction from $\mathbf{H}$ requires $2(m-j)(n-j-1)$ FLOPs, so the total number of FLOPs for iteration j is approximately $32(m-j)(n-j)$. Thus, the total estimated FLOPs count for the complex Householder bidiagonalization can be expressed as follows:

$$\begin{aligned} FLOP_{SVD} &= \sum_{j=1}^{n} 32(m-j)(n-j) \\ &= 32 \sum_{j=1}^{n} [mn - (m+n)j + j^2] \\ &= 32 \sum_{j=1}^{n} mn - 32 \sum_{j=1}^{n} (m+n)j + 32 \sum_{j=1}^{n} j^2 \\ &= 32mn^2 - 32(m+n)\frac{n(n+1)}{2} + 32\frac{n(n+1)(2n+1)}{6} \\ &= 32mn^2 - 16(mn^2) - 16n^3 + \frac{16}{3}(2n^3 + 3n^2 + n) \\ &= 16mn^2 - 16n^3 + \frac{32}{3}n^3 + 16n^2 + \frac{16}{3}n. \end{aligned} \tag{3.7}$$

To achieve the canonical FLOPs count, lower order terms are neglected. We then have:

$$FLOP_{SVD} = 16mn^2 - \frac{16}{3}n^3. \tag{3.8}$$

Using the SVD, the pseudo inverse of a matrix can be easily computed as follows:

$$\mathbf{H}^{\dagger} = \mathbf{V}\mathbf{\Sigma}^{\dagger}\mathbf{U}^{\mathsf{H}}. \tag{3.9}$$

where for $m \leq n$, we have the $n \times m$ matrix

$$\mathbf{\Sigma}^{\dagger} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & 0 & \dots & 0 \\ 0 & 0 & 1/\sigma_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/\sigma_m \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0 \tag{3.10}$$

for all the non-zero singular values. If any of the σ_i are zero, then a zero is placed in the corresponding entry of $\mathbf{\Sigma}^{\dagger}$. If $\mathbf{H}$ is rank deficient, then one or more of its singular values will be zero.

The computation of the SVD is non-trivial. It suffices to know that all the respectable software packages for doing mathematics (such as MATLAB or Mathematica) contain functions for computing the SVD. A MATLAB pseudo-code to compute the SVD of a complex matrix $\mathbf{H}$ is given in section 3.5.5.

3.3.2. QR Decomposition

Another method to compute the Moore-Penrose pseudo inverse applies the so-called *QR factors* or *QR decomposition*. The QR decomposition breaks $\mathbf{H}$ into two sub matrices, $\mathbf{Q}$ and $\mathbf{R}$, where $\mathbf{Q}$ is unitary and $\mathbf{R}$ is upper triangular. This decomposition is practical for a square matrix of full rank, because it is automatically invertible. The QR decomposition is rather effective for computing the pseudo inverse. If $\mathbf{H} \in \mathbb{C}^{m \times n}$ where $m \leq n$ and $\mathbf{H}$ has rank equal to m, then the factorized matrix $\mathbf{H}$ with linearly independent columns can be written as:

$$\mathbf{H}_{m \times n} = \mathbf{Q}_{m \times n}\mathbf{R}_{n \times n} \tag{3.11}$$

where $\mathbf{Q}$ satisfies $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$ and we have

$$\begin{aligned}
\mathbf{R} &= \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ 0 & 0 & \dots & r_{nn} \end{bmatrix} \\
\mathbf{Q} &= \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1m} \\ \dots & \dots & \dots & \dots \\ q_{m1} & q_{m2} & \dots & q_{mn} \end{bmatrix} \\
\mathbf{H} &= \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ \dots & \dots & \dots & \dots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{bmatrix}.
\end{aligned} \tag{3.12}$$

Thus, the pseudo inverse of $\mathbf{H} \in \mathbb{C}^{m\times n}$ (where $m \leq n$) can be expressed as:

$$\begin{aligned}
\mathbf{H}^\dagger &= ((\mathbf{QR})^\mathsf{H}(\mathbf{QR}))^{-1}(\mathbf{QR})^\mathsf{H} \\
&= (\mathbf{R}^\mathsf{H}\mathbf{Q}^\mathsf{H}\mathbf{QR})^{-1}\mathbf{R}^\mathsf{H}\mathbf{Q}^\mathsf{H} \\
&= \mathbf{R}^{-1}\mathbf{Q}^\mathsf{H}
\end{aligned} \tag{3.13}$$

where the last step follows by $\mathbf{Q}^\mathsf{H}\mathbf{Q} = I$. Gram-Schmidt Orthogonalization (GSO) [63] is a direct method to compute the $\mathbf{Q}$ and $\mathbf{R}$ matrices. The Gram-Schmidt process for an $m \times n$ matrix $\mathbf{H}$ proceeds as:

$$\begin{aligned}
\underline{q}_j &= (\underline{h}_j - \sum_{i=1}^{j-1} \underline{r}_{ij}\underline{q}_i)/\underline{r}_{jj} \\
\underline{u}_j &= \underline{h}_j - \sum_{i=1}^{j-1} \underline{q}_i^H \underline{h}_j \underline{q}_i, \quad u_1 = h_1 \\
q_k &= \frac{u_k}{\|u_k\|} \\
r_j &= q_j^H \mathbf{H}
\end{aligned} \tag{3.14}$$

where $j = 1....n$. The $\underline{h}_j$ and $\underline{q}_j$ are the j_{th} column vectors of $\mathbf{H}$ and $\mathbf{Q}$, respectively. $\underline{r}_j$ denotes the j_{th} row vector of $\mathbf{R}$.

The Gram-Schmidt algorithm can be applied through either Classical Gram-Schmidt (CGS) or Modified Gram-Schmidt (MGS) [64]. The CGS algorithm allows a memory efficient implementation due to its inherent parallelism and because $\mathbf{Q}$ can build the original columns of $\mathbf{H}$.

The superior MGS algorithm overcomes inaccuracies of CGS by subtracting linear combinations of $\underline{q}_j$ not directly from $\mathbf{H}$ but from an intermediate column vector $\underline{t}_j$ of $\mathbf{Q}$ before constructing the orthogonal vectors (see Figure 3.1).

```
for j=1:n
    t(:,j)=h(:,j);
for i=1:j-1
r(i,j)=q(i,t(i,j));
t(:,j)=t(:,j)-(r(i,j)*t(:,j));
q(:,j)=t(:,j)./norm(t(:,j),2);
r(j,j)=norm(t(:,j),2);
end
end
```

Figure 3.1.: MGS algorithm

To compute $\mathbf{H}^{\dagger}$ using MGS, the QR decomposition involves the computation of norms and inner products of n-dimensional vectors, division of n-dimensional vectors by norms, and scaling the normalized columns of $\mathbf{Q}$ by the inner product. Further, the computation of $\mathbf{R}^{-1}$ requires multiplication and additions of scalars.

The multiplication of $\mathbf{R}^{-1}$ and $\mathbf{Q}^{\mathsf{H}}$ in (3.13) is done using an array to achieve maximum parallelism. The MGS QR decomposition FLOPs count for $\mathbf{H}^{\dagger}$ consists of counting additions, subtractions, multiplications and divisions as follows:

$$\begin{aligned}\mathrm{FLOP}^{\mathrm{QR}}_{\mathrm{MGS}}(\text{Addition}) &= \sum_{i=1}^{n}(m-1+\sum_{j=i+1}^{n} m-1) \\ &= \overbrace{n(m-1)}^{\text{real}} + \overbrace{\frac{n(n-1)(m-1)}{2}}^{complex} \\ &= \frac{1}{2}n(n+1)(m-1)\end{aligned} \tag{3.15}$$

$$\mathrm{FLOP}^{\mathrm{QR}}_{\mathrm{MGS}}(\text{Subtraction}) = \sum_{i=1}^{n}\sum_{j=i+1}^{n} m = \sum_{i=1}^{n} m(n-i) = \frac{1}{2}mn(n-1) \tag{3.16}$$

$$\mathrm{FLOP}^{\mathrm{QR}}_{\mathrm{MGS}}(\text{Division}) = \sum_{i=1}^{n} m = mn \tag{3.17}$$

$$\mathrm{FLOP}^{\mathrm{QR}}_{\mathrm{MGS}}(\text{Multiplication}) = \overbrace{mn^2 - mn}^{\text{real}} + \overbrace{nm^2}^{\text{complex}}\ . \tag{3.18}$$

3.3.3. Geninv Pseudo Inverse

We describe a low-complexity algorithm to compute the Moore-Penrose inverse. This algorithm is based on a reverse order applicable to a full rank singular symmetric positive matrix. By applying Cholesky factorization, the computational complexity can be reduced especially when we are dealing with large systems. According to [57], if we consider the matrix $\mathbf{H}$ of size $m \times n$ with $m < n$ and the positive matrix $(\mathbf{H}^{\mathsf{H}}\mathbf{H})$ of size $n \times n$, there is a unique upper triangular matrix $\mathbf{O}$ such that $\mathbf{O}^{\mathsf{H}}\mathbf{O} = \mathbf{H}^{\mathsf{H}}\mathbf{H}$. The computation of $\mathbf{O}$ is a simple Cholesky factorization of non-singular matrices. If we remove the zero rows from $\mathbf{O}$, a lower-diagonal matrix $\mathbf{L}$ of size $n \times r$ with rank of $r \leq n$ is obtained where we have

$$\mathbf{H}^{\mathsf{H}}\mathbf{H} = \mathbf{L}\mathbf{L}^{\mathsf{H}}. \tag{3.19}$$

To compute the Moore-Penrose inverse based on factorizing, we consider matrix $\mathbf{A}$ of size $n \times r$ and matrix $\mathbf{C}$ of size $r \times t$ where $r \leq n$ and $r \leq t$. The Moore-Penrose inverse of the matrix product $\mathbf{AC}$ can be expressed as [57]

$$(\mathbf{AC})^{\dagger} = (\mathbf{C})^{\dagger}(\mathbf{A})^{\dagger} = \mathbf{C}^{\mathsf{H}}(\mathbf{A}^{\mathsf{H}}\mathbf{A}\mathbf{C}\mathbf{C}^{\mathsf{H}})^{\dagger}\mathbf{A}^{\mathsf{H}}. \tag{3.20}$$

If $\mathbf{C}$= $\mathbf{I}$ in (3.20) we have

$$(\mathbf{A})^{\dagger} = (\mathbf{A}^{\mathsf{H}}\mathbf{A})^{\dagger}\mathbf{A}^{\mathsf{H}}. \tag{3.21}$$

Since $\mathbf{A}$ has size $n \times r$ and rank r, the following exists

$$(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{\dagger} = (\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}. \tag{3.22}$$

If we set $\mathbf{C}$ to $\mathbf{A}^{\mathsf{H}}$ and $\mathbf{C}^{\mathsf{H}} = \mathbf{A}$ then, (3.20) can be rewritten as

$$(\mathbf{AC})^{\dagger} = \mathbf{A}\Big(\mathbf{A}^{\mathsf{H}}\mathbf{A}\mathbf{A}^{\mathsf{H}}\mathbf{A}\Big)^{\dagger}\mathbf{A}^{\mathsf{H}} \tag{3.23}$$

$$= \mathbf{A}\Big(\big(\mathbf{A}^{\mathsf{H}}\mathbf{A}\big)^{\dagger}\big(\mathbf{A}^{\mathsf{H}}\mathbf{A}\big)^{\dagger}\Big)\mathbf{A}^{\mathsf{H}}. \tag{3.24}$$

According to (3.22), we have

$$(\mathbf{A}\mathbf{A}^{\mathsf{H}})^{\dagger} = \mathbf{A}(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{H}}. \tag{3.25}$$

Proposition 1. *The pseudo inverse of* $\boldsymbol{H}$ *can be obtained by:*

$$\boldsymbol{H}^{\dagger} = \Big(\boldsymbol{L}(\boldsymbol{L}^{\mathsf{H}}\boldsymbol{L})^{-1}(\boldsymbol{L}^{\mathsf{H}}\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathsf{H}}\Big)\boldsymbol{H}^{\mathsf{H}} \tag{3.26}$$

where $\boldsymbol{L}^{\mathsf{H}}$ *is conjugate transpose of* $\boldsymbol{L}$ *of size* $n \times r$.

Proof. Using (3.22) and (3.25) , we have

$$\mathbf{L}^{\dagger} = (\mathbf{L}^{\mathsf{H}}\mathbf{L})^{\dagger}\mathbf{L}^{\mathsf{H}}. \tag{3.27}$$

$$(\mathbf{L}\mathbf{L}^{\mathsf{H}})^{\dagger} = \mathbf{L}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}\mathbf{L}^{\mathsf{H}}. \tag{3.28}$$

From equation (3.29) and (3.28) we have:

$$(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{\dagger} = (\mathbf{L}\mathbf{L}^{\mathsf{H}})^{\dagger} = \mathbf{L}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}\mathbf{L}^{\mathsf{H}}. \tag{3.29}$$

by multiplying $\mathbf{H}^{\mathsf{H}}$ with (3.29), we have the Moore-penrose inverse of matrix $\mathbf{H}$ as follows

$$(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{\dagger}\mathbf{H}^{\mathsf{H}} = \mathbf{H}^{\dagger} = \Big(\mathbf{L}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}\mathbf{L}^{\mathsf{H}}\Big)\mathbf{H}^{\mathsf{H}}$$

□

This algorithm effectively relies on two principle operations, namely the full rank Cholesky factorization of $(\mathbf{H}^{\mathsf{H}}\mathbf{H})$ and the inverse of $(\mathbf{L}^{\mathsf{H}}\mathbf{L})$. In order to count the FLOPs for $(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}$, the first step is to analyse $\mathbf{L}^{H}$. Let $\mathbf{X} = [x_1, ..., x_n] = \mathbf{L}^{H}$ denote the inverse of the lower triangular matrix $\mathbf{L}$. For $b < s$ where $s = 2..., n$ we have $[X]_{b,s} = 0$. Thus, we have

$$\mathbf{L}x_s = e_s \tag{3.30}$$

which can be solved via *forward substitution*. For $t \leq b < s$, we can express (3.30) as:

$$\sum_{a=t}^{b} l_{b,a}x_{a,s} = \delta_{b,s} \tag{3.31}$$

where $l_{b,a} = [\mathbf{L}]_{b,a}$ for the a_{th} main diagonal element, $\delta_{b,s}$ denotes the *Kronecker* delta and $x_{a,s} = [X]_{a,s} = [x_s]_{a,1}$. If $s = b$, then $x_{s,s} = \frac{1}{l_{s,s}}$ and a single multiplication is required. For $b > s$, we need $(b - s + 1)$ multiplications and $(b - s - 1)$ additions. Thus, the number of FLOPs for the lower left off-diagonal entries consist of

$$\begin{aligned}
\sum_{t=1}^{n-1}\sum_{b=t+1}^{n}(b-s+1) &= \sum_{t=1}^{n-1}[(1-t)(n-t) + \sum_{b=t+1}^{n} b] \\
&= \sum_{t=1}^{n-1}[n + t^2 - t(n+1) + \frac{n^2+n-t^2-t}{2}] \\
&= \sum_{t=1}^{n-1}[\frac{n^2}{2} + \frac{3n}{2}\frac{t^2}{2} - t(n+\frac{3}{2})] \\
&= (n-1)\frac{n}{2}(n+3) + \frac{(n-1)n(2n-1)}{2.6} - (n+\frac{3}{2})\frac{(n-1)n}{2} \\
&= \frac{1}{6}n^3 - \frac{1}{2}n^2 - \frac{2}{3}n
\end{aligned} \tag{3.32}$$

multiplications and

$$\sum_{t=1}^{n-1}\sum_{b=t+1}^{n}(b-n+1)=\frac{1}{6}n^3-\frac{1}{2}n^2+\frac{1}{3}n \tag{3.33}$$

additions. Considering n multiplications for the main-diagonal entries, we have $\frac{1}{6}n^3-\frac{1}{2}n^2+\frac{1}{3}n$ multiplications and $\frac{1}{6}n^3-\frac{1}{2}n^2+\frac{1}{3}n$ additions. Thus, the total number of FLOPs is $\frac{1}{3}n^3+\frac{2}{3}n$.

To count the FLOPs to compute the Gram matrix $\mathbf{L}^H\mathbf{L}$, only the main diagonal elements and upper right off-diagonal elements have to be evaluated. Focusing on the main diagonal elements, we have:

$$[\mathbf{L}^H\mathbf{L}]_{a,a}=\sum_{t=a}^{N}|l_{n,a}|^2 \tag{3.34}$$

where $l_{n,a}=[\mathbf{L}]_{n,a}$ for a_{th} main diagonal element. As such, $n-a+1$ multiplications and $n-a$ additions are needed for the main diagonal elements. Hence, all main diagonal elements need $\sum_{t=1}^{n}(n-t+1)=\frac{1}{2}n^2+\frac{1}{2}n$ multiplications and $\sum_{t=1}^{n}(n-t)=\frac{1}{2}n^2-\frac{1}{2}n$ additions.
For the upper right off-diagonal element $[\mathbf{L}^H\mathbf{L}]_{a,b}$ in row a and column b, we have:

$$[\mathbf{L}^H\mathbf{L}]_{a,b}=\sum_{t=b}^{n}{l^*}_{t,a}l_{t,b} \tag{3.35}$$

where l^* denotes the conjugate. Summing expressions in (3.30), we have

$$\begin{aligned}\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}(n-b+1)&=\sum_{a=1}^{n-1}[(n-a)(n+1)-\sum_{b=a+1}^{n}b]\\&=\sum_{a=1}^{n-1}[n^2+n-a(n+1)-\frac{n(n+1)-a(a+1)}{2}]\\&=\sum_{a=1}^{n-1}\frac{n^2+n}{2}+\frac{a^2}{2}-a(n+\frac{1}{2})\\&=\frac{(n-1)(n+1)n}{2}+\frac{(n-1)n(2n-1)}{2.6}-(n+\frac{1}{2})\frac{n(n-1)}{2}\\&=\frac{1}{6}n^3-\frac{1}{6}n\end{aligned} \tag{3.36}$$

as the number of multiplications and

$$\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}(n-b)=\frac{1}{6}n^3-\frac{1}{2}n^2+\frac{1}{3}n \tag{3.37}$$

as the number of additions. According to (3.36) and (3.37), all the essential elements of the

Gram matrix $\mathbf{L}^H\mathbf{L}$ require $\frac{1}{6}n^3 + \frac{1}{2}n^2 + \frac{1}{3}n$ multiplications and $\frac{1}{6}n^3 - \frac{1}{6}n$ additions, which leads to $\frac{1}{3}n^3 + \frac{1}{6}n^2 + \frac{1}{6}n$ FLOPs. The total number of FLOPs accounting for $(\mathbf{L}^\mathsf{H}\mathbf{L})^{-1}$ is $n^3 + n^2 + n$.

3.3.4. Other Approaches to Compute Pseudo Inverse

Greville Pseudo Inverse

The Greville method is a common approach to obtain the pseudo inverse with low-complexity and it calculates the Moore-Penrose inverse $\mathbf{H}^\dagger$ in an iterative fashion. In the i^{th} iteration $(i = 1, 2, ..., n)$, we compute $\mathbf{H}_i^\dagger$, where $\mathbf{H}_i$ is the submatrix of $\mathbf{H}$ that includes its first i columns. Thus, the process begins with a column vector to calculate the Moore-Penrose inverse and then extends with one more vector in the following iteration. The Greville algorithm has the following steps [58].

1. Decompose $\mathbf{H}$ into a set of vectors $\underline{h}_i$ as follows:

$$\mathbf{H} = [\underline{h}_1, \underline{h}_2, ..., \underline{h}_n] \tag{3.38}$$

   where $\underline{h}_i$ is the i^{th} column of $\mathbf{H}$.

2. Form the $\mathbf{H}_i$ matrices of size $i \times n$ as follows:

$$\begin{aligned}\mathbf{H}_i &= [\mathbf{H}_{i-1} \quad \underline{h}_i] \\ \mathbf{H}_1 &= \underline{h}_1 \quad (i = 1) \\ \mathbf{H}_i &= \begin{bmatrix} H_{i-10} \\ \underline{h}_i \end{bmatrix} \quad (i = 2, ..., n).\end{aligned}$$

3. Set the vectors $\underline{d}_i$, $\underline{c}_i$ as:

$$\begin{aligned}\underline{d}_i &= \mathbf{H}_{i-1}^\dagger \underline{h}_i \\ \underline{c}_i &= \underline{h}_i - \mathbf{H}_{i-1}\underline{d}_i \\ &= \underline{h}_i - \mathbf{H}_{i-1}\mathbf{H}_{i-1}^\dagger \underline{h}_i\end{aligned} \tag{3.39}$$

4. To calculate the Moore-Penrose matrix $\mathbf{H}_{i-1}^\dagger$ of size $n \times i$ we have :

$$[\mathbf{H}_{i-1} \quad \underline{h}_i]^\dagger = \begin{bmatrix} H_{i-1}^\dagger - \underline{d}_i \underline{g}_i^\mathsf{T} \\ \underline{g}_i^\mathsf{T} \end{bmatrix} \quad (i = 2, ..., n) \tag{3.40}$$

where the vector $\underline{g}_i^{\mathsf{T}}$ can be obtained as :

$$\underline{g}_i^{\mathsf{T}} = \underline{c}_i^{\mathsf{T}} \quad (c_i \neq 0) \tag{3.41}$$

$$\underline{g}_i^{\mathsf{T}} = (1 + \underline{d}_i^{\mathsf{T}} \underline{d}_i)^{-1} \underline{d}_i^{\mathsf{T}} \mathbf{H}_{i-1}^{\mathsf{T}} \quad (c_i = 0) \tag{3.42}$$

where $\underline{g}_i^{\mathsf{T}}$ is the i^{th} row of $\mathbf{H}_i^{\dagger}$.

Neumann Series for Matrix Inversion

A non-conventional method to reduce the complexity of ZF precoding uses *Neumann* series to transform the inverse of the Gram matrix $\mathbf{G}$ into a simple precondition matrix [56]. The Neumann series expansion for $\mathbf{G}^{-1}$ is as follows [65]:

$$\mathbf{G}^{-1} \approx \sum_{r=1}^{E} (\mathbf{I}_K - \mathbf{DG})^r \mathbf{D} \tag{3.43}$$

and as we assume

$$\lim_{r \to \infty} (-\mathbf{D}^{-1}\mathbf{F})^r \mathbf{D}^{-1} = 0. \tag{3.44}$$

In (3.44), $\mathbf{F} = \mathbf{G} - \mathbf{D}$ and $\mathbf{I}_K$ is an identity matrix of size $n \times n$ in (3.43). Three different designs can be studied for the precondition matrix $\mathbf{D}$. The first design, denoted as $\mathbf{D}_d$, is a scaled identity matrix where only diagonal inputs are inverted. The second design, denoted as $\mathbf{D}_{fc}$, is a precondition matrix where the diagonals and the first column from matrix $\mathbf{G}$ are inverted. The complexity of the second approach is smaller than the first due to the diagonal dominance of $\mathbf{G}$. The third design, denoted as $\mathbf{D}_t$, adds secondary diagonal lines of the matrix $\mathbf{G}$ to reduce the degradation affected by the high channel correlation. To investigate the computational complexity, a three-term Neumann series expansion of $\mathbf{G}$ is given as follows:

$$\mathbf{G}_3^{-1} \approx (\mathbf{I} + (\mathbf{I} - \mathbf{DG}) + (\mathbf{I} - \mathbf{DG})^2)\mathbf{D}. \tag{3.45}$$

From (3.45), there are four steps that consist of computing the precondition matrix $\mathbf{D}$, the term $(\mathbf{I} - \mathbf{DG})$ and the square of the previous term. Table 3.1 lists the computational complexity of each step, if any of the three above preconditioning matrices is applied. As shown in Table 3.1, the complexity of the third step grows cubically with n. The exact matrix inverse can be replaced with a Neumann series. Further, the middle step (third step in the given example) can be computed partially.

	$\mathbf{D}$	$(\mathbf{I}-\mathbf{DG})$	$(\mathbf{I}-\mathbf{DG})^2$	$(\mathbf{I}-\mathbf{DG})\mathbf{D}$
$\mathbf{D}_d$	$3n$	n^2-n	$\frac{1}{2}n^3-\frac{1}{2}n^2$	$\frac{1}{2}n^2+\frac{1}{2}n$
$\mathbf{D}_{fc}$	$5n-2$	$2n^2-2$	$\frac{1}{2}n^3-\frac{3}{2}n+1$	$\frac{1}{2}n^2+\frac{3}{2}n-\frac{1}{2}$
$\mathbf{D}_t$	$8n-5$	$3n^2-2n$	$\frac{1}{2}n^3+\frac{1}{2}n^2$	$\frac{3}{2}n^2+\frac{1}{2}-1$

Table 3.1.: FLOPs count for each step corresponding to three precondition matrix designs

3.4. Proposed Scheme

In this section, we propose a Sparse Geninv (SGeninv) algorithm to calculate a Moore-Penrose pseudo inverse based on the state-of-the-art Geninv method described in section 3.3.4. SGeninv utilizes the sparsity of the combined mMIMO and JT CoMP channel matrices at multiple coordination sites and increases the feasibility of FDD-MIMO. In order to take advantage of sparsity, in section 3.4.1 we introduce the main properties of Relevant Channel Components (RCCs). Further, a reordering scheme to reduce the computational complexity is detailed in section 3.4.2.

3.4.1. Relevant Channel Components

FDD requires an explicit estimation of downlink channel components, as well as CSI reporting using a feedback channel. The limited GoB concept discussed in Chapter 2 can be used to subdivide the cells into subsectors. This reduces the overhead of CSI reporting.

As a simple example, the set of channel components within a certain power of the strongest channel component is often a small subset of all channel components. Thus, the channel matrix which contains CSI between all the eNodeB beams and the served UE will be sparse. Each UE estimates and reports the CSI for its sparse set of RCCs to the eNodeB. To enforce sparsity, a power threshold window P_{TH} with respect to the strongest channel component per UE is applied. An appropriate value for P_{TH} depends on the performance metrics and the desired impact of sparsity on the proposed pseudo inverse scheme.

3.4.2. Reverse Cuthill-Mckee Reordering

A supportive technique to compute the pseudo inverse of a sparse matrix is the Reverse Cuthill-Mckee (RCM) method. This technique reorders a matrix $\mathbf{A}$ to reduce the bandwidth defined as:

$$B_d = \max(Lab_i - Lab_j), \quad i,j = 1...k \tag{3.46}$$

where Lab_i is the label of node i and Lab_j is the label of node j when nodes i and j are neighbours in $\mathbf{A}$ among k nodes. The bandwidth reduction problem can be considered as a graph labling problem. We wish to find the node labling that minimizes B_d of the adjacency matrix $\mathbf{A}$ of the graph $G(k)$.

To reduce the bandwidth, one permutes the columns and rows so that all the non-zero elements are moved to a band near the diagonal. Thus, we will have a sparse matrix where all the non-zero elements are re-located.

The impact of RCM matrix reordering on a sparse matrix $\mathbf{A}$ is illustrated in Figure 3.2, where $\mathbf{A}^*$ denotes the reordered matrix. As can be seen, $\mathbf{A}^*$ is obtained through row and column permutations. For a graph $G(k)$, the RCM algorithm uses the following steps [66]:

1. Prepare an empty queue Qu and an empty result array Re.
2. Select the Parent node $G(k)_P$ with the lowest degree that has not been inserted in the result array Re.
3. Add $G(k)_P$ in the first position of Re.
4. Add to the queue all the nodes adjacent to $G(k)_P$ with respect to an increasing order of their degree.
5. Extract the first node from the queue and examine it as the child node $G(k)_C$.
6. If $G(k)_C$ has not been previously inserted in Re, add it in the first free position and add to Qu all the neighbours of $G(k)_C$ that are not in Re in increasing order of their degrees.
7. If Qu is not empty, repeat from (1).
8. If there are unexplored nodes repeat from (1).
9. Reverse the order of the elements in Re. Element $Re[i]$ is swapped with element $Re[n+1-i]$.

The computational complexity related to the RCM algorithm is estimated as $O(\log(f)|E|)$ where f and E denote the maximum degree in the adjacency graph and the total number of edges in the considered sparse matrix, respectively [67].

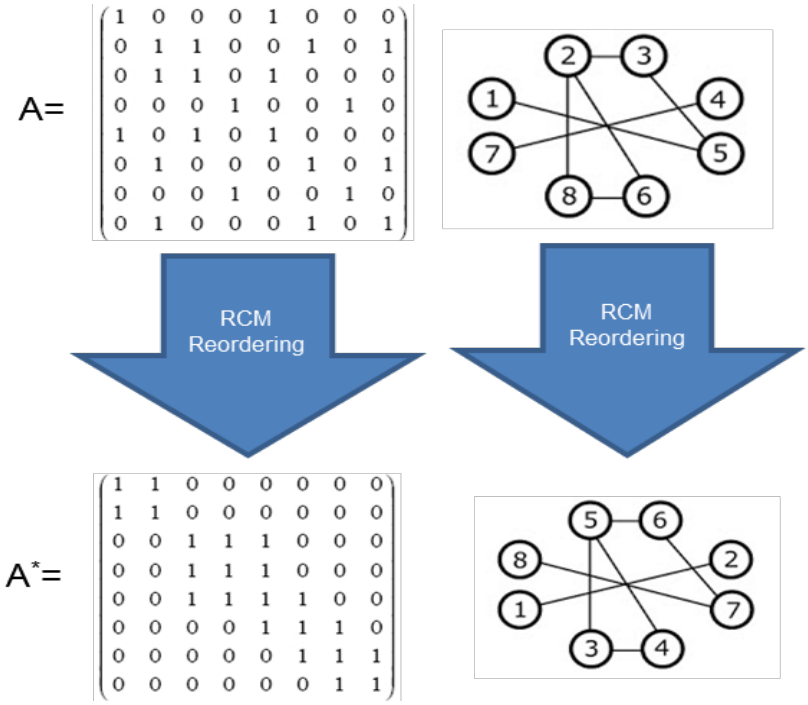


Figure 3.2.: RCM algorithm and adjacency graph [5]

3.4.3. Complexity Analysis of Sparse Channel Matrices

In order to design and benchmark a low-complexity pseudo inverse algorithm, a detailed analysis using FLOPs is required. As a baseline, the Lightspeed MATLAB toolbox which features a set of routines for accurate FLOPs counting was used [68]

The FLOPs count routines provided by Lightspeed are significantly more accurate than the FLOPs functions that were included in MATLAB up to version 5. However, the range of applicable functions in Lightspeed is limited and restricted to real-valued operations.

In order to analyze the number of FLOPs using a reliable approach, an evolved Lightspeed toolbox is proposed that considers complex-valued matrix operations. Moreover, the techniques for sparse pseudo inverse computation that were discussed in previous sections are used. Figure 3.3 gives an overview of how Lightspeed is used by the proposed toolbox. The toolbox basically scans each line and searches for the requested function in Lightspeed.

Figure 3.4 shows the proposed toolbox's main components. The mathematical functions and operations used in the toolbox along with their exact FLOPs count are detailed in the Appendix A. Our sparsity routines include the following:

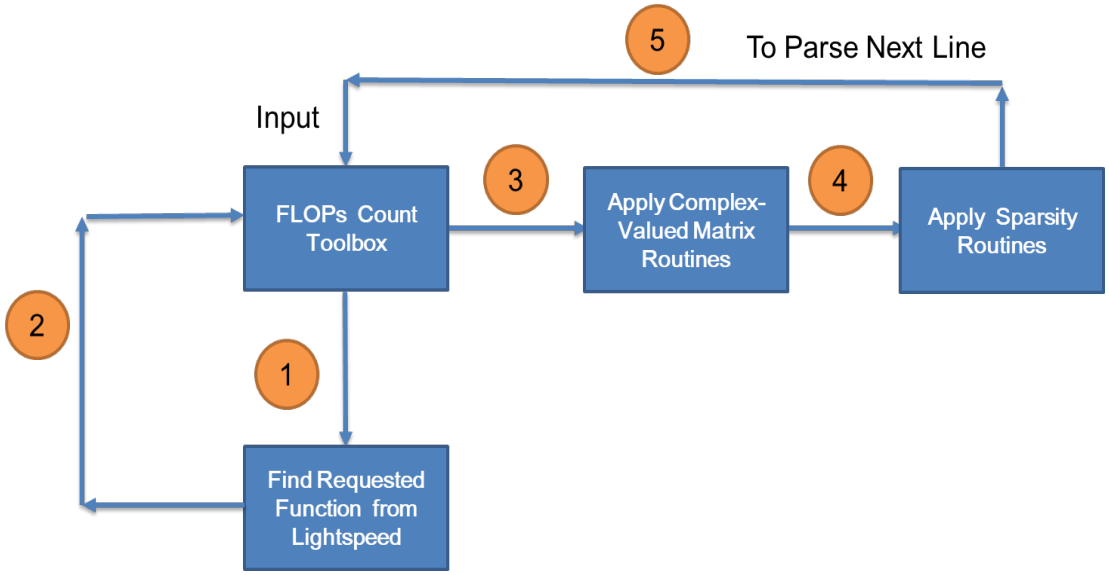


Figure 3.3.: Proposed scheme for FLOPs count

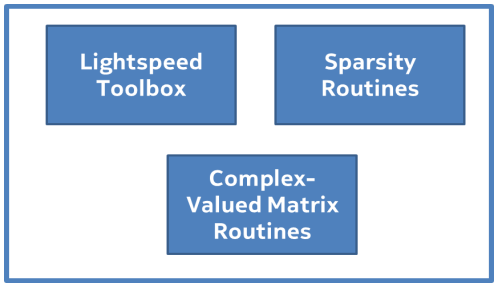


Figure 3.4.: Proposed FLOPs count scheme components

1. Store the locations of the non-zero elements in the matrix before reordering: this could be done by using the adjacency graph of the matrix.

2. Reorder the sparse matrix by applying the Reverse Cuthill-Mckee algorithm: to produce a matrix with much smaller bandwidth, the adjacency graph of the sparse matrix is used as input.

   We obtain the sparse matrix elements, where the non-zeros of the initial sparse matrix are relabeled. Figure 3.2 shows the effect of the RCM reordering process.

3. Multiply new reordered matrices while avoiding unnecessary multiplications and additions: Obtaining the reordered sparse matrix, where all the non-zero elements of each row are close to the diagonal, we can specify the position of the non-zero elements in the matrix. As such, unnecessary multiplications (zero multiplications) and non-essential additions can be avoided.

3.5. Evaluation Results

This section compares the deployments and linear precoding schemes introduced in this chapter. We first present the impact of P_{TH} on the channel matrix $\mathbf{H}$. Three pseudo inverse schemes including SVD, Greville and Sgeninv are employed to evaluate the average SINR, spectral efficiency and the computational complexity of the ZF precoding in FLOPs against the reference method. Evaluation results are generated based on the 3GPP 3D channel model which is explained in section 2.6. We presented preliminary results in [4], [5] and we add the following results:

- ▷ We analyze the impact of applying P_{TH} window both on the performance and complexity.
- ▷ We demontrate the effect of increasing the number of RCCs on the complexity.
- ▷ We analyze the impact of applying P_{TH} window on the RCCs.

3.5.1. Channel Matrix Power

Before evaluating the precoding methods, we analyze the effect of applying P_{TH} with respect to the strongest channel component on the channel coefficients. The power of the i_{th} UE and j_{th} Tx beam in the channel $\mathbf{H}$ with respect to the strongest Tx beam can be defined as follows:

$$\overline{P}_{REL} = P_{\mathbf{H}(i,j)} - P_{strbe_{(i)}} \tag{3.47}$$

where $P_{\mathbf{H}(i,j)}$ is the Tx power of the i_{th} UE and j_{th} Tx beam in the channel $\mathbf{H}$. $P_{strbe_{(i)}}$ denotes the strongest Tx beam power at the i_{th} UE:

$$P_{strbe_{(i)}} = \arg\max(be_{(i)}). \tag{3.48}$$

Figure 3.5 depicts the average P_{REL} of the $\mathbf{H}$ over 100 PRBs. As can be seen, the strongest Tx beams of the UEs in a same cell are the ones transmitted from the eNodeBs in the same cell. Figure 3.6 depicts the P_{REL} in one PRB.

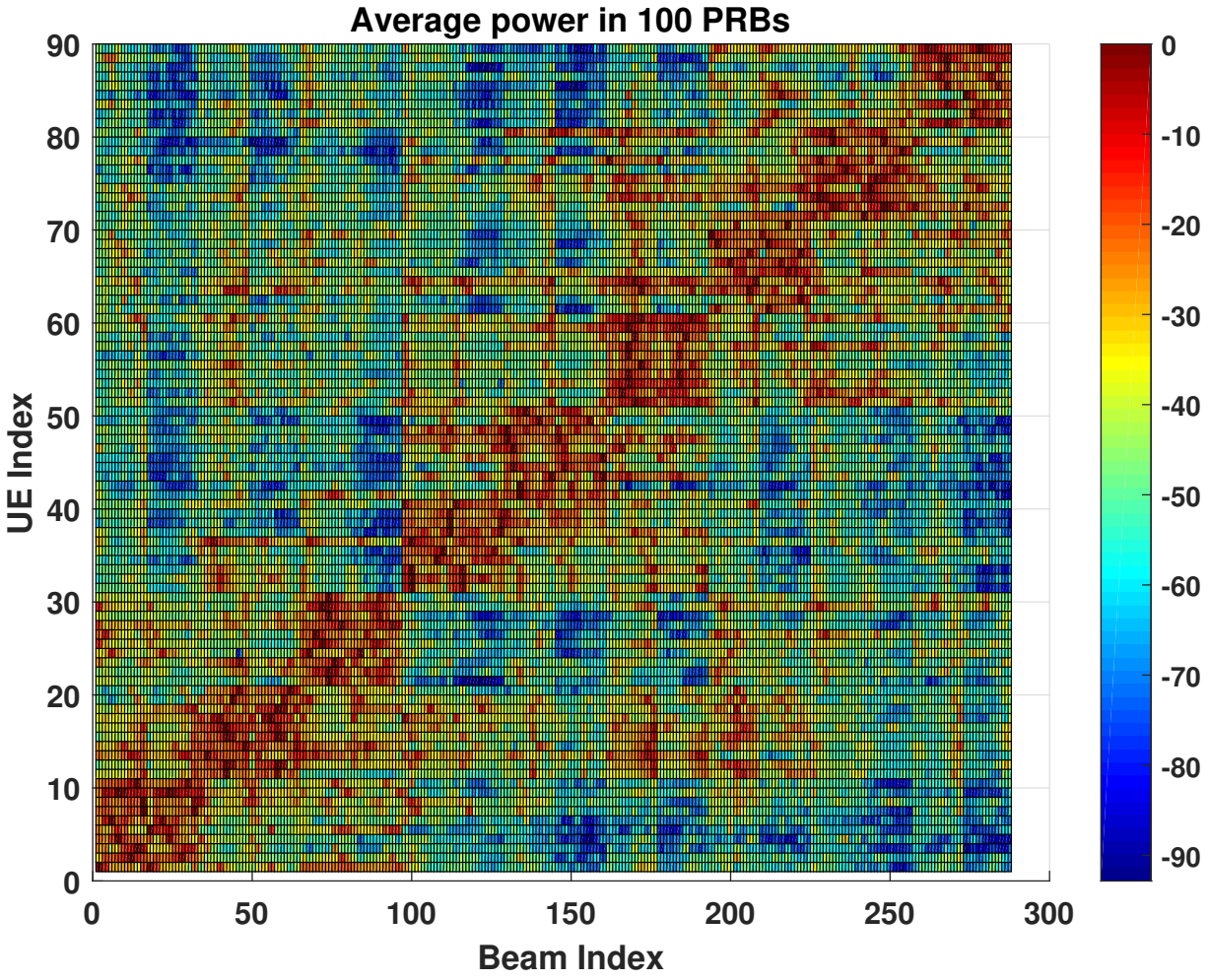


Figure 3.5.: Average P_{REL} of the $\mathbf{H}$ over 100 PRBs

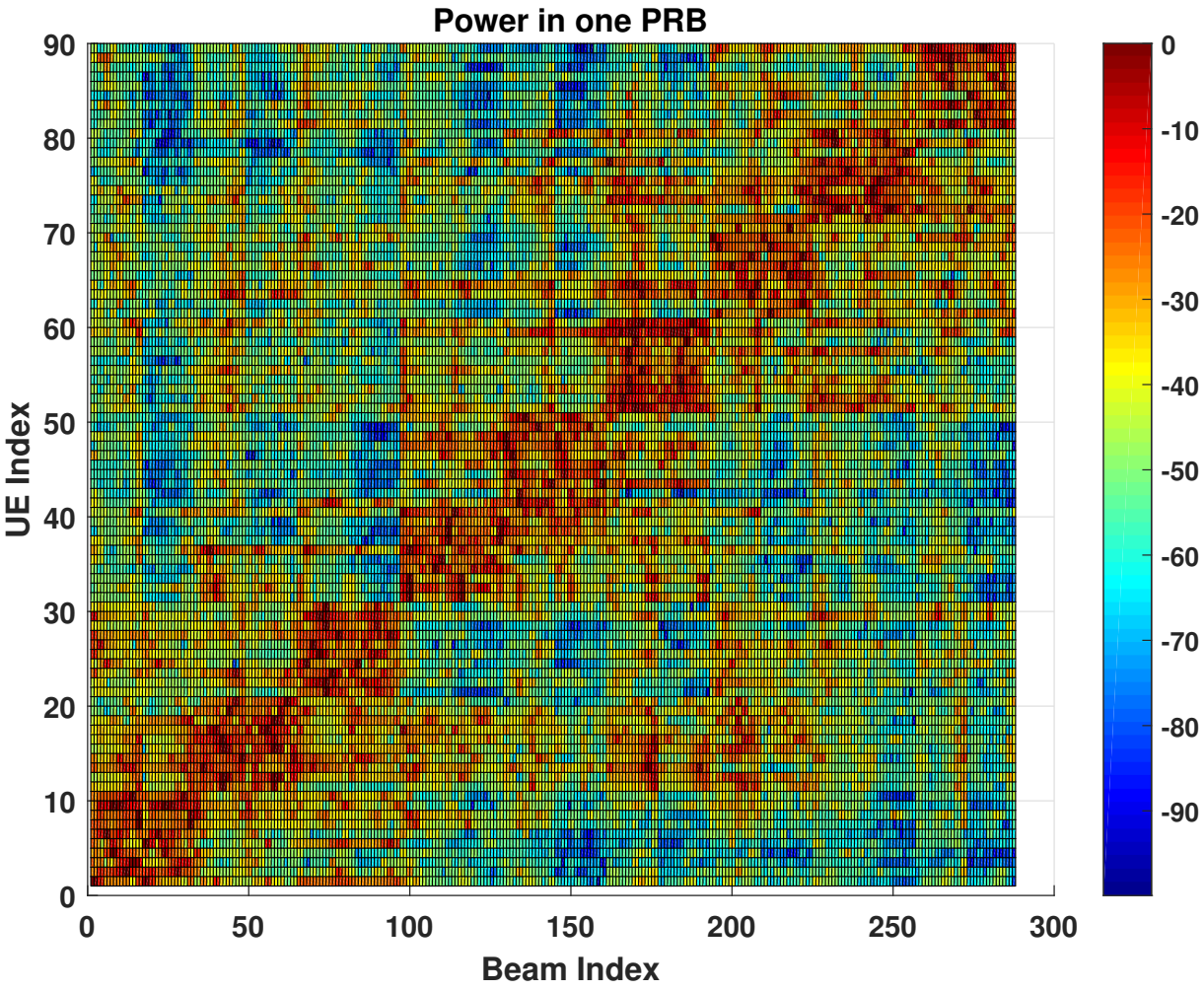


Figure 3.6.: P_{REL} of the $\mathbf{H}$ in one PRB

3.5.2. RCC Analysis

To distinguish the RCCs from the non-RCCs in **H**, Figure 3.7 shows the P_{REL} distribution when P_{TH}= -20 dB with respect to P_{strbe} is applied on the **H** channel components. RCCs that receive Tx beams greater than -20 dB with respect to P_{strbe} are depicted in red. The impact of P_{TH} window on the number of RCCs over 100 PRBs is given in Figure 3.8.

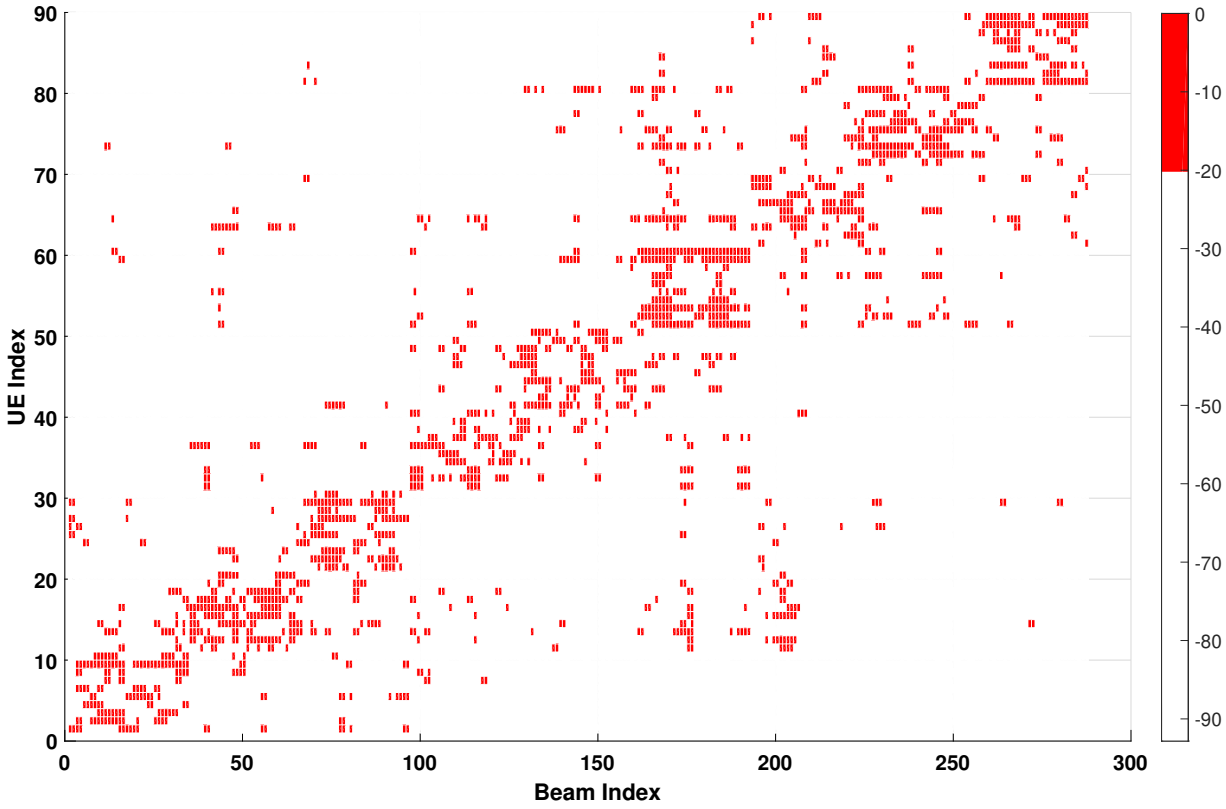


Figure 3.7.: RCC distribution in **H** when P_{TH}= -20 dB

For a small P_{TH} window, e.g., -15 dB, the RCCs percentage is less than 5% for all the PRBs. Considering 288 beams received at each UE (as discussed in Chapter 2), the number of RCCs when a larger P_{TH} window, e.g., -30 dB is applied, is not more than 70 or 25% for each UE over 100 PRBs. For performance evaluation, a P_{TH} window between -15 dB and -25 dB is applied.

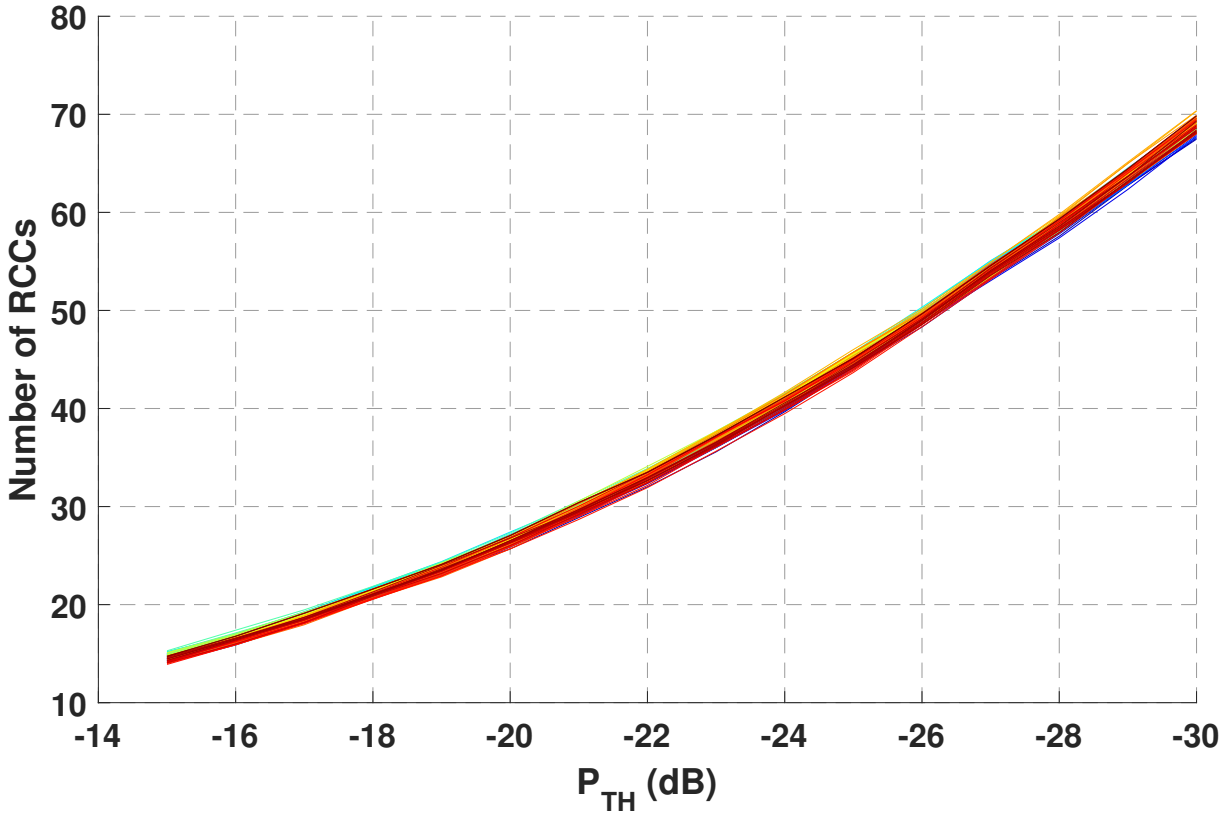


Figure 3.8.: P_{TH} impact on the RCC distribution per UE

3.5.3. Comparison of SINR

We first analayze the SINR under the 3D channel modeling and the system model explained in Chapter 2 (see section 2.6). Figure 3.9 shows the CDF of the SINR for Sgeninv when applying three P_{TH}s. The SINR degradation for P_{TH}= -20 dB reaches 2 dB on average. The term "on average" for the CDF plot means median.

In Figure 3.9 Sgeninv with P_{TH}= -20 dB has SINR loss of 2 dB on average since the point where 50 percent of the data that is at or below SINR= 13.5 dB (with P_{TH}= -20 dB) from Sgeninv and the point where 50 percent of the data that is at or below SINR= 15.5 dB from the MATLAB Pinv curve differs by 2 dB degradation. The SINR degradation can be reduced to 1 dB on average with P_{TH}= -25 dB. Figure 3.10 shows the CDF of the SINR for N_{UE}=90 (or 9 UEs per cell) and $N_{Antenna}$=1. For this deployment, the MATLAB Pinv function was used as a benchmark where the sparsity of the channel matrix is not taken into account.

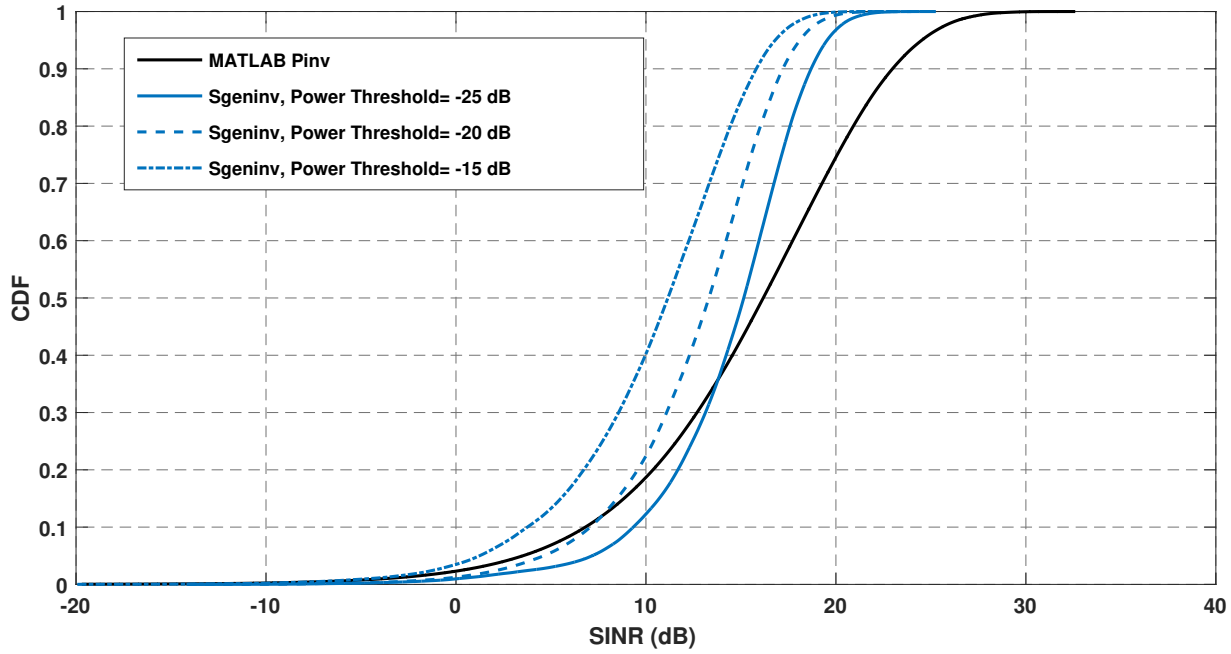


Figure 3.9.: Sgeninv SINR performance

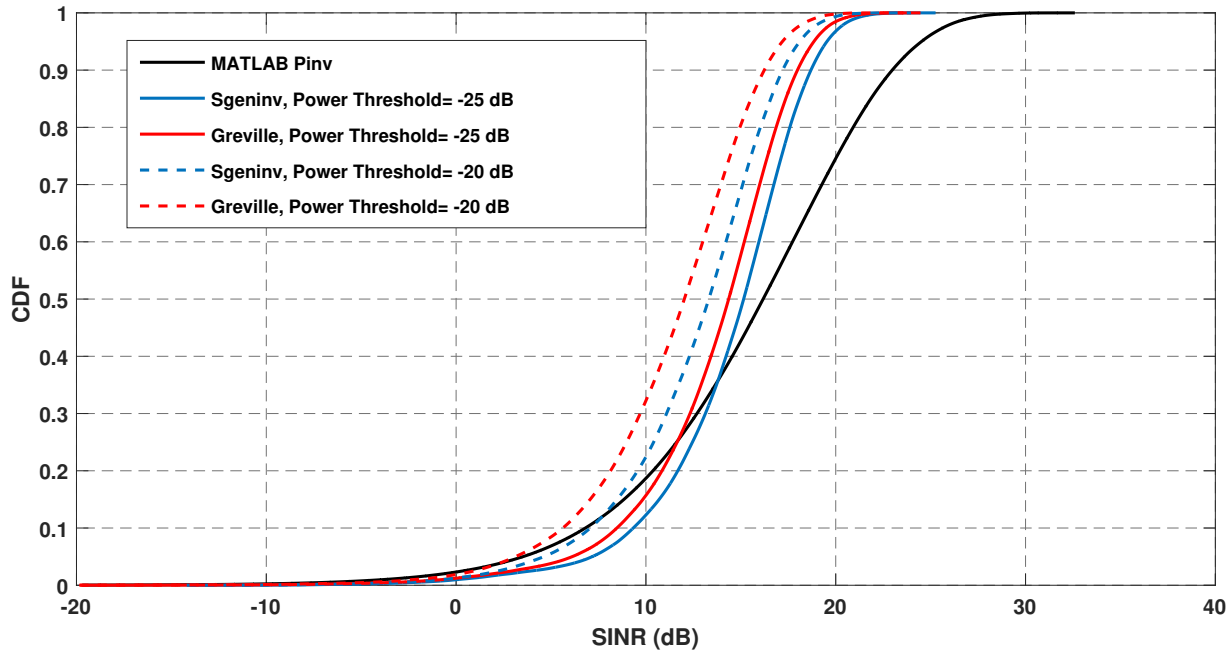


Figure 3.10.: Sgeninv and Greville SINR performance

The SINR loss caused by Sgeninv with P_{TH}= -25 dB is nearly 1 dB on average. The SINR loss for the state-of-the-art Greville with the same P_{TH} window is around 2 dB. The SINR performance of Sgeninv for a small window P_{TH} performs better as compared to

Greville. Figure 3.11 depicts the CDF of the SINR for Greville when applying P_{TH}= -15 dB, -20 dB and -25 dB. For a small P_{TH} window, when the RCCs changes between 5% and 10% per UE, the SINR loss changes between 6 dB and 4 dB. If the number of RCCs increases up to 20% per UE, the Greville SINR degradation reduces by 2 dB as compared to the benchmark. By increasing the number of RCCs to 20% per UE, Sgeninv SINR degradation drops by 1 dB. The Sgeninv algorithm performs better than Greville.

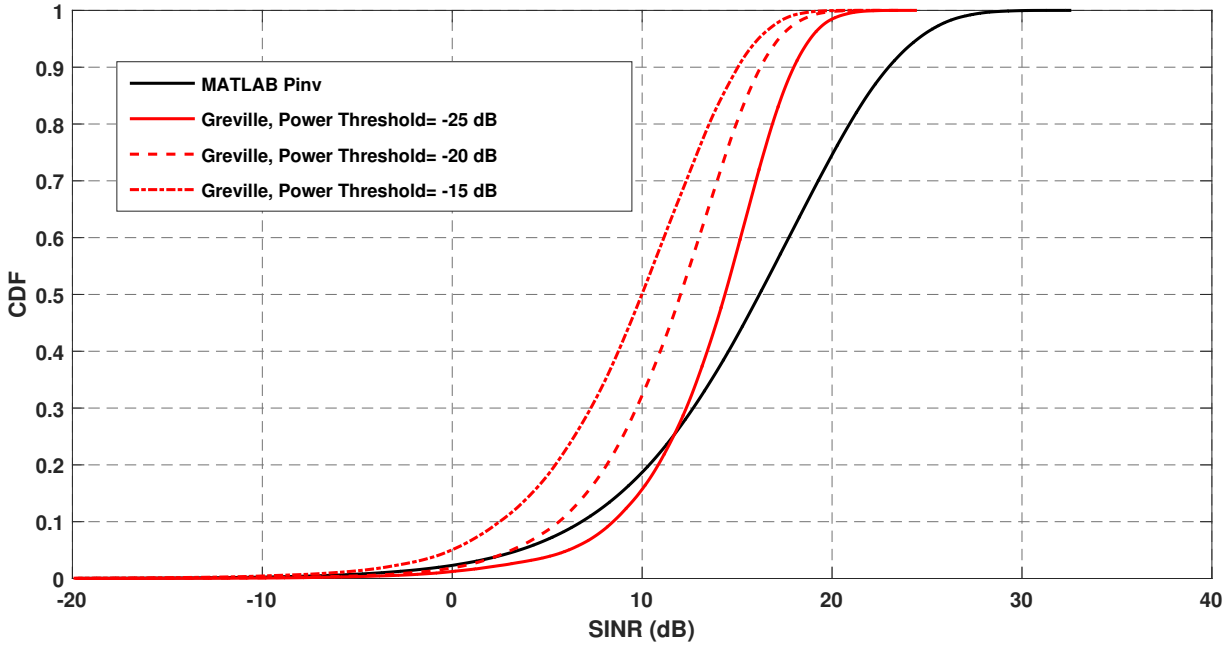


Figure 3.11.: Greville SINR performance

3.5.4. Comparison of Spectral Efficiency

Here we compare the spectral efficiency (as discussed in section 2.6.5) for both Sgeninv and Greville. Similar to the SINR, the MATLAB Pinv function was used as a benchmark to evaluate the spectral efficiency for both Sgeninv and Greville. The CDF of the spectral efficiency for Greville with P_{TH}=-15 dB, -20 dB and -25 dB is given in Figure 3.12. If a small P_{TH} window, e.g., -15 dB is applied, the degradation is greater than 10 bit/sec/Hz on average. For a larger P_{TH} window, e.g., -25 dB, the spectral efficiency loss drops to 6 bit/sec/Hz on average. Figure 3.13 shows the spectral efficiency of Sgeninv with P_{TH}= -15 dB, -20 dB and -25 dB.

Similar to Greville, the degradation is higher on average for a small P_{TH} window, e.g., -15 dB (9 bit/sec/Hz) when the Sgeninv algorithm is used. For a larger P_{TH} window, e.g., -25 dB, spectral efficiency loss can be reduced to 3 bit/sec/Hz on average. A spectral

efficiency comparison of Greville and Sgeninv is depicted in Figure 3.14. Similar to the SINR performance, the Sgeninv algorithm performs better than Greville. By increasing the number of RCCs to 20% (P_{TH}= -25 dB) per UE, the Sgeninv spectral efficiency performs well with a degradation of 3 bit/sec/Hz on average.

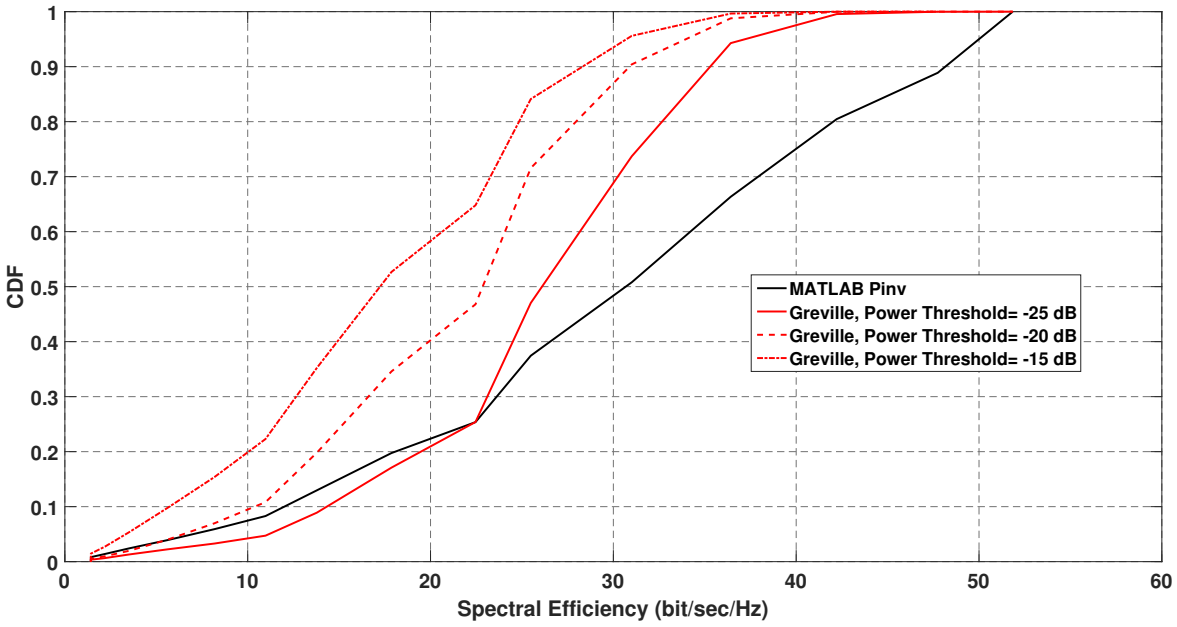


Figure 3.12.: Greville spectral efficiency

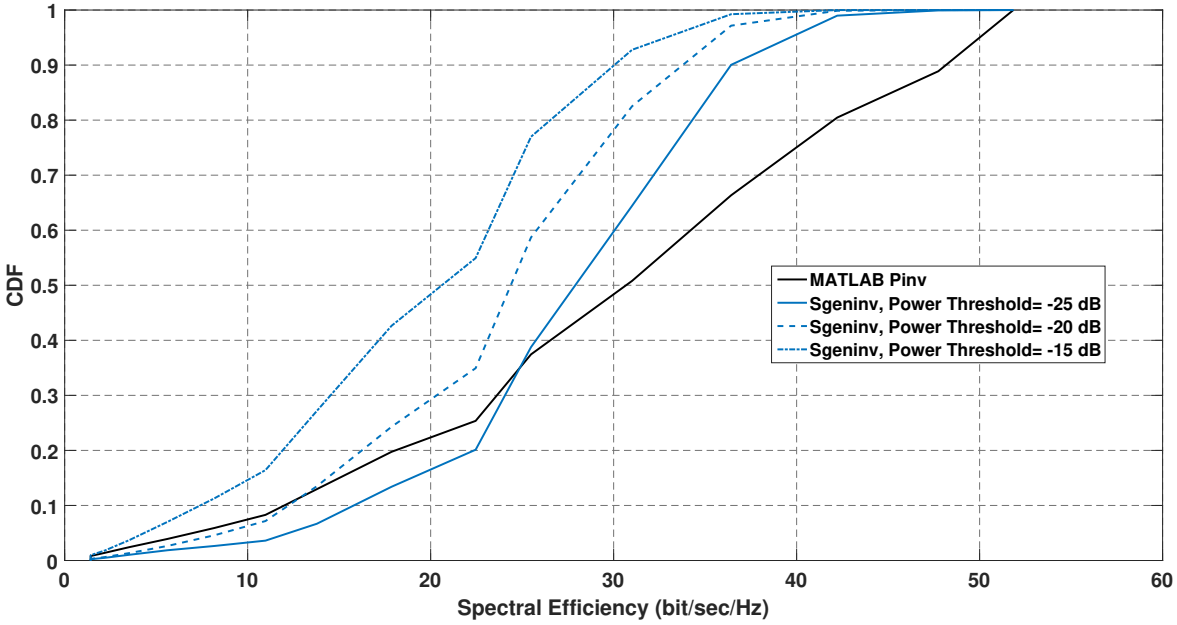


Figure 3.13.: Sgeninv spectral efficiency

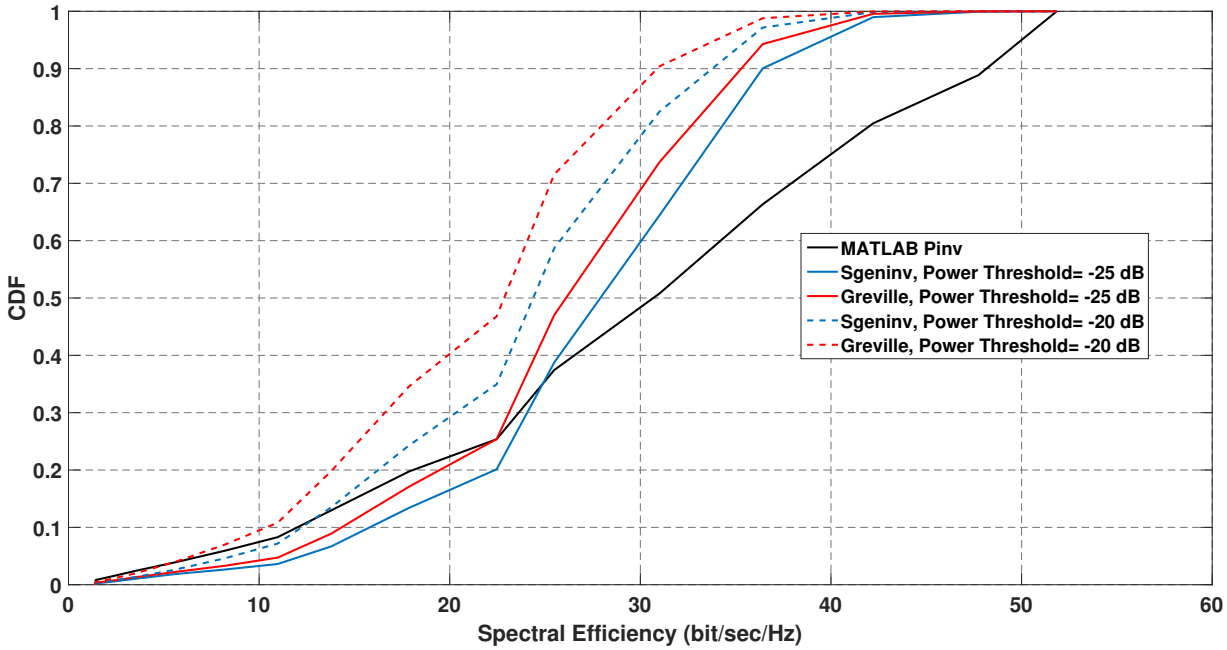


Figure 3.14.: Sgeninv and Greville spectral efficiency

3.5.5. Comparison of FLOPs

We next analyze the FLOPs count using the complexity toolbox introduced in section 3.4. We first compare the Greville, geninv and MATLAB Pinv performance as shown in Table 3.2 when no complexity reduction technique is used. The FLOPs count of each approach is calculated for a channel matrix $\mathbf{H}$ of size 90×288 based on the complexity analysis discussed in section 3.3.

The SVD based pseudo inverse function used in MATLAB is given in Figure 3.15. The FLOPs count for the MATLAB Pinv function includes counting FLOPs for the SVD decomposition function (known as *svdecon*) and the element wise binary operation handle function (indicated by *bsxfun* in Table 3.2).

Pseudo Inverse Method	**FLOPs**
geninv	17,617,371
MATLAB Pinv	$\overbrace{13,936,860}^{\text{svdecon}} + \overbrace{4,717,623}^{\text{bsxfun}} = 18,654,483$
Greville	27,248,466

Table 3.2.: FLOPs count for $\mathbf{H}_{90\times288}$

```
function [complexity_Total,W] = MATLABpinv(H,tol)


   [complexity_part1,U,S,V] = svdecon(H);
s = diag(S);
if nargin < 2
    tol = max(size(H)) * eps(norm(s,inf));
end
r1 = sum(s > tol)+1;
V(:,r1:end) = [];
U(:,r1:end) = [];
s(r1:end) = [];
s = 1./s(:);
W = bsxfun(@times,V,s.')*U';

complexity_part2=FLOPS(bsxfun(@times,V,s.'),U);
complexity_Total=complexity_part1+complexity_part2;
end
```

Figure 3.15.: MATLAB Pinv function

Figure 3.16 and Figure 3.17 depict the impact of the proposed scheme (discussed in section 3.4) on the Sgeninv and Greville performance. To investigate the impact, the average SINR loss per UE over 1200 subcarriers and computational complexity in terms of FLOPs are evaluated. As shown in Figure 3.16, both the FLOPs count and average SINR loss per UE over 1200 subcarriers follow an exponential behavior when P_{TH} window changes between -15 dB and -39 dB.

The computational complexity of the state-of-the-art Greville can be reduced by nearly five times (5.5×10^6 vs. 27.2×10^6) and by three times (5.5×10^6 vs. 18.6×10^6) as compared to the SVD based MATLAB Pinv function if P_{TH}= -15 dB. However, the average SINR loss becomes large (around 8 dB). If a P_{TH} window of -27 dB is taken, then the SINR loss of Greville can be reduced to 1 dB.

The complexity gain in this case will be around four times (6.5×10^6 vs. 27.2×10^6) as compared to the number of Greville FLOPs shown in Table 3.2. Similarly, the Sgeninv performance (as depicted in Figure 3.17) follows an exponential behavior when P_{TH} window changes between -15 dB and -39 dB. For low $P_{TH}s$ window, e.g, -15dB, the complexity reduction is around five times (3.5×10^6 vs. 18.6×10^6) as compared to the MATLAB Pinv function. However, the impact on the SINR will be approximately 6.5 dB. To decrease the SINR loss to 1 dB, a P_{TH} window of -25 dB can be employed. The FLOPs in this case will increase to 4.2×10^6 which is four times (4.5×10^6 vs. 18.6×10^6) less than the reference MATLAB Pinv function complexity.

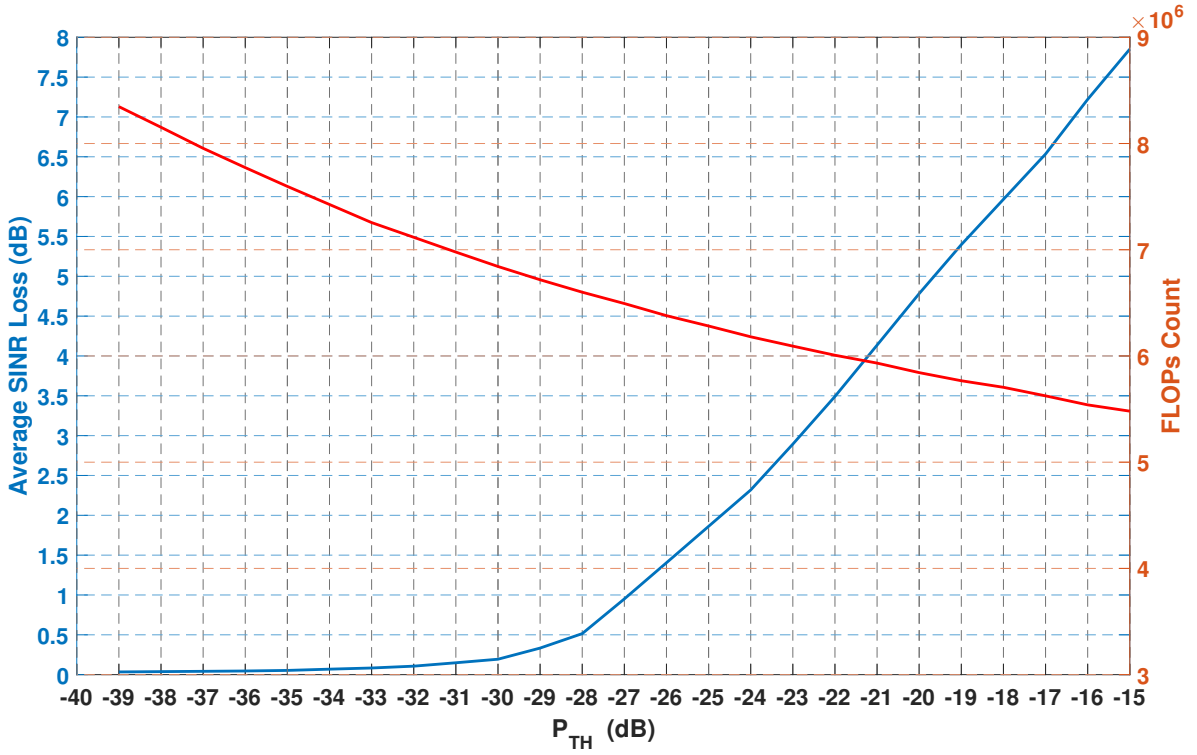


Figure 3.16.: Greville SINR Loss and FLOPs count a single subcarrier

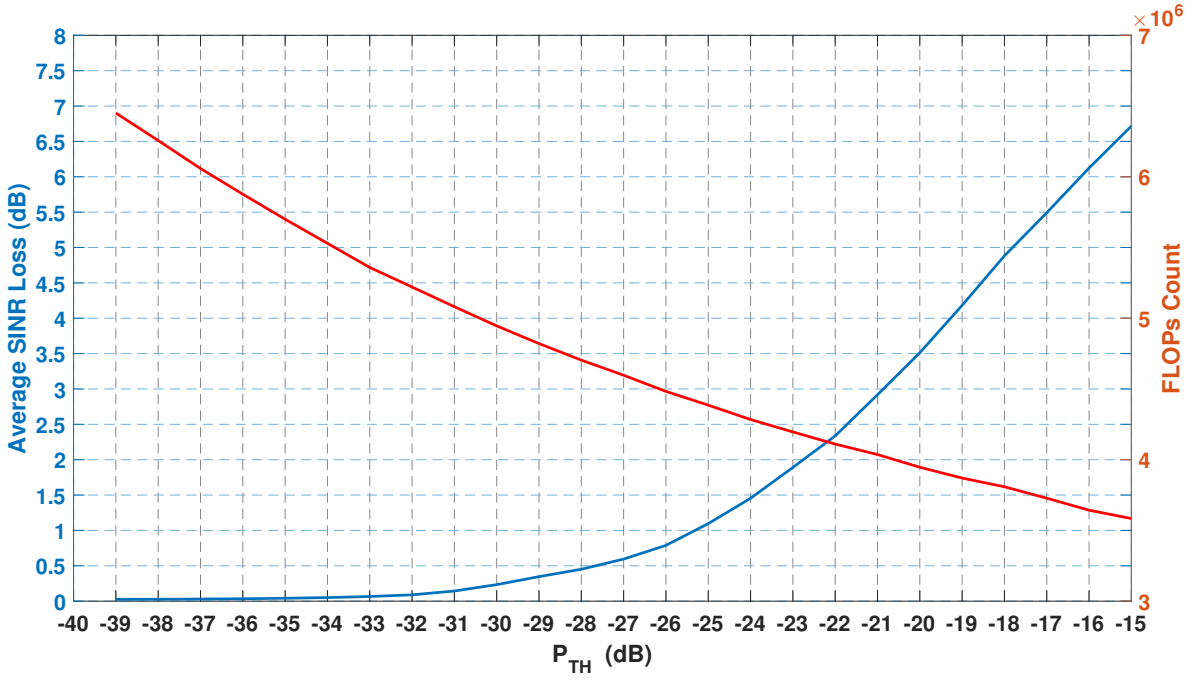


Figure 3.17.: Sgeninv SINR Loss and FLOPs count for a single subcarrier

Figure 3.18 compares the average SINR loss per UE over 1200 subcarriers with the percentage of RCCs for Sgeninv when P_{TH} window changes between -15 dB and -39 dB. If P_{TH}= -15 dB, the RCC percentage will be limited to only 15%. To achieve 1 dB loss on the average SINR, the number of RCCs will need to increase to 45%.

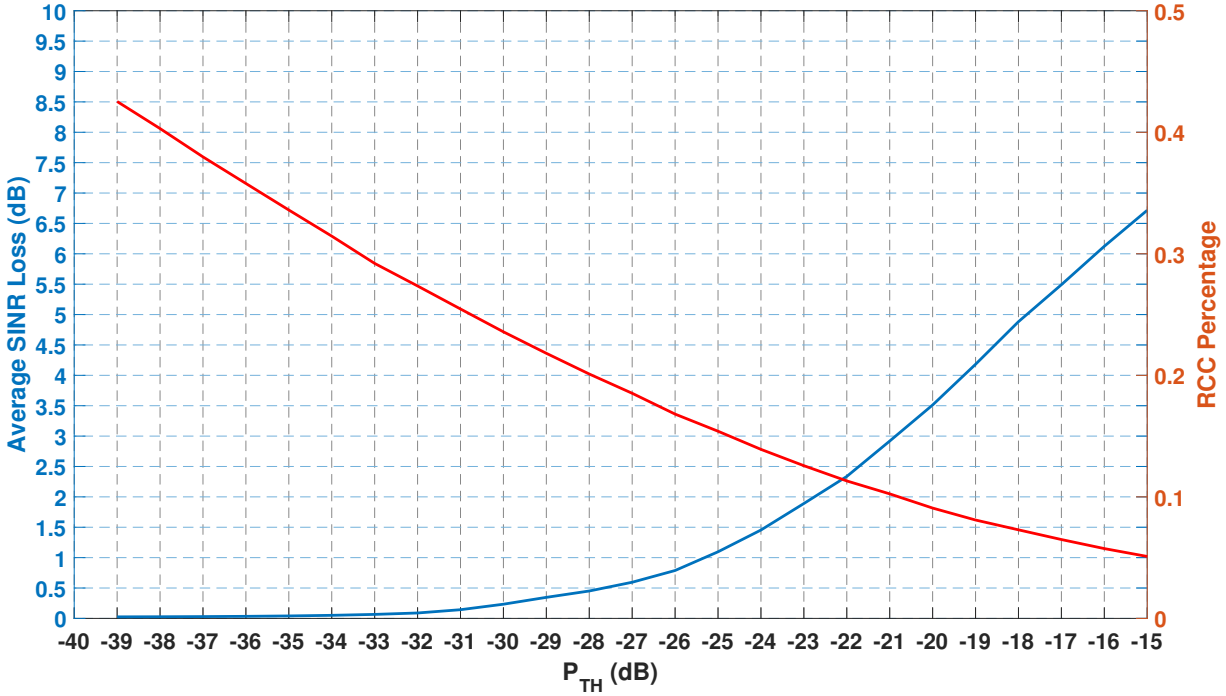


Figure 3.18.: Sgeninv SINR Loss and RCC percentage

3.5.6. UE Beamforming Performance Analysis

As discussed in section 2.7, beamforming at the UE can be considered. This section presents the evaluation results based on the linear receiver beamformer introduced in section 2.7.2 and section 2.7.3 and [4]. To exemplify the impact of changing the number of RCCs when each UE is equipped with linear beamforming using 8 received antennas in the cooperation area, two power level thresholds (-20 dB, -25 dB) with respect to the strongest channel component per UE have been employed. Figure 3.19 contrasts the number of reported RCCs to the eNodeB for a single antenna UE against the case when the UEs employ MRC beamformers with eight UE antennas. Utilizing UE beamformers leads to maximizing the received power on one hand and reducing the number of reported RCCs down to approximately 50 percent on average on the other hand. Further, this impact more pronounced for the larger threshold window of -25dB.

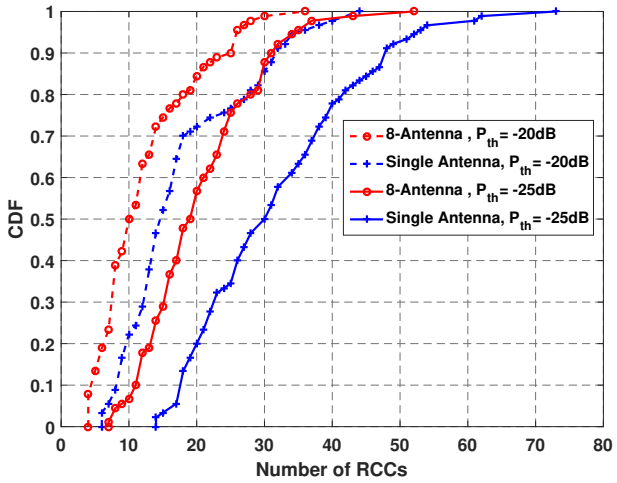


Figure 3.19.: CDF plot vs. RCCs per UE [4]

Figure 3.20 provides the CDF for the PNL over 100 PRBs. Taking the scheduling aspects of user grouping in MU-MIMO into account, the UEs are scheduled only if their PNL is below 5dB. The PNL for the case of an ideal reporting of all channel components is approximately 1 dB smaller compared to the case of limited reporting of RCCs. The MRC beamformer maximizes the received power and limits reporting to a smaller number of RCCs. Choosing an appropriate power threshold is important to balance both effects.

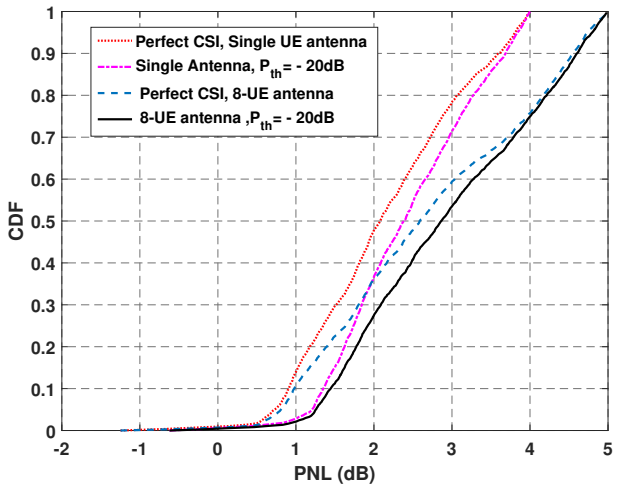


Figure 3.20.: CDF plot vs. PNL [4]

4

mMIMO JT CoMP Precoding for Multiple Subcarriers

4.1. Introduction

Precoding solutions for multiple-subcarrier systems can obtain the maximum possible gains of FDD mMIMO technology. Accurate CSI reporting to the eNodeB and the computational complexity of the ZF precoder are significant challenges in such systems. One approach to reduce the complexity for calculating the precoder matrices is to select a fraction of subcarriers and to perform interpolation across subcarriers.

MIMO-OFDM interpolated precoding based on QR decomposition was proposed in [69], [70] where the QR decomposition of Laurent Polynomials (LPs) approximates the precoding. Precoding based on unitary precoders was discussed in [71] where three interpolation approaches were applied. In the first approach, the interpolation is done by rotating a construction trajectory between two consecutive pilot subcarriers. In the second approach, a conditional interpolation is used that identifies an inherited precoder interpolation. In the third approach, an interpolation in the Grassmann manifold is discussed [71], [72].

In this chapter, the problem of computational complexity in multiple-subcarrier JT CoMP scenarios is formulated. The main focus is on reducing the number of FLOPs without degrading performance too much. A low-complexity pseudo inverse approximation scheme is proposed based on the *SGeninv* algorithm (see section 4.3). We demonstrate the benefits of precoding when multiple approaches are applied to the receiver's channel matrix beams with respect to the average received power.

This chapter is organised as follows: existing methods to approximate the matrix pseudo inverse are discussed in section 4.2. Details of a pseudo inverse approximation approach are provided in section 4.3. In section 4.4, a comprehensive implementation of the proposed multi-approach ZF precoding is presented. Further, the impact of using a multi-approach scheme applying both prior art and the *Sgeninv* pseudo inverse approximation to the RCCs is discussed. Simulation results are presented in section 4.5.

4.2. Interpolation-Based QR Decomposition in MIMO-OFDM Systems

A standard approach applies subcarrier-by-subcarrier QR decomposition, where one estimates the channel matrix $\mathbf{H}$ by interpolation, and performs the QR decomposition on each subcarrier [73]. The basic idea is to represent the inverse of a polynomial matrix by a matrix with rational entries. This approach provides complexity savings since the number of OFDM tones is much larger than the channel order.

Let $\mathbb{C}^{P\times D}$ denote the set of complex-valued $P \times D$ matrices and $T \triangleq \{s \in \mathbb{C} : |s| = 1\}$ indicate the unit circle. We define a matrix-valued function $\mathbf{E} : T \to \mathbb{C}^{P\times D}$ for integers j_1 and $j_2 \geq 0$ as the Laurent Polynomial (LP) matrix [73]:

$$\mathbf{E}(s) = \sum_{e=-j_1}^{j_2} E_e s^{-e}, s \in T \tag{4.1}$$

where $E_e \in \mathbb{C}^{P\times D}$. Considering QR decomposition of the LP matrices, the interpolation-based algorithms are formulated through LP matrices. Figure 4.1 and Figure 4.2 depict QR decomposition interpolation examples for both a general subcarrier-by-subcarrier approach, as well as the LP matrices interpolation-based approach. Using the brute-force scheme, $\mathbf{H}_1$, $\mathbf{H}_5$ and $\mathbf{H}_9$ are selected as the pilot subcarriers (see Figure 4.1). However, the computational complexity will be high if we apply QR decomposition on every subcarrier. To avoid that, an interpolation-based QR preprocessing can be used (see Figure 4.2).

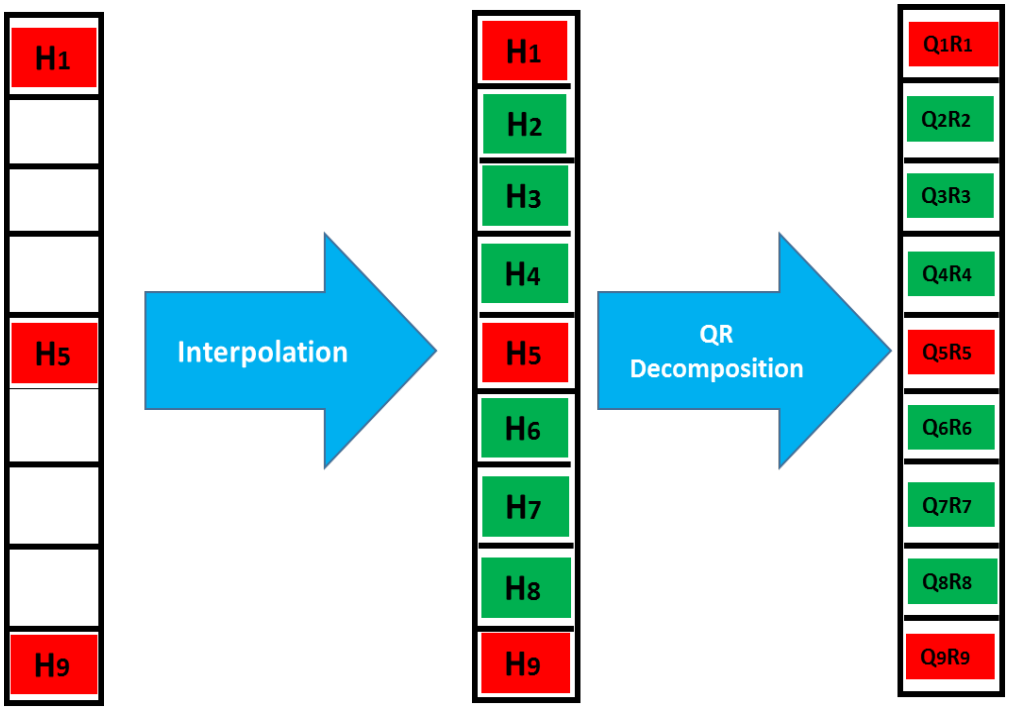


Figure 4.1.: QR decomposition general interpolation scheme

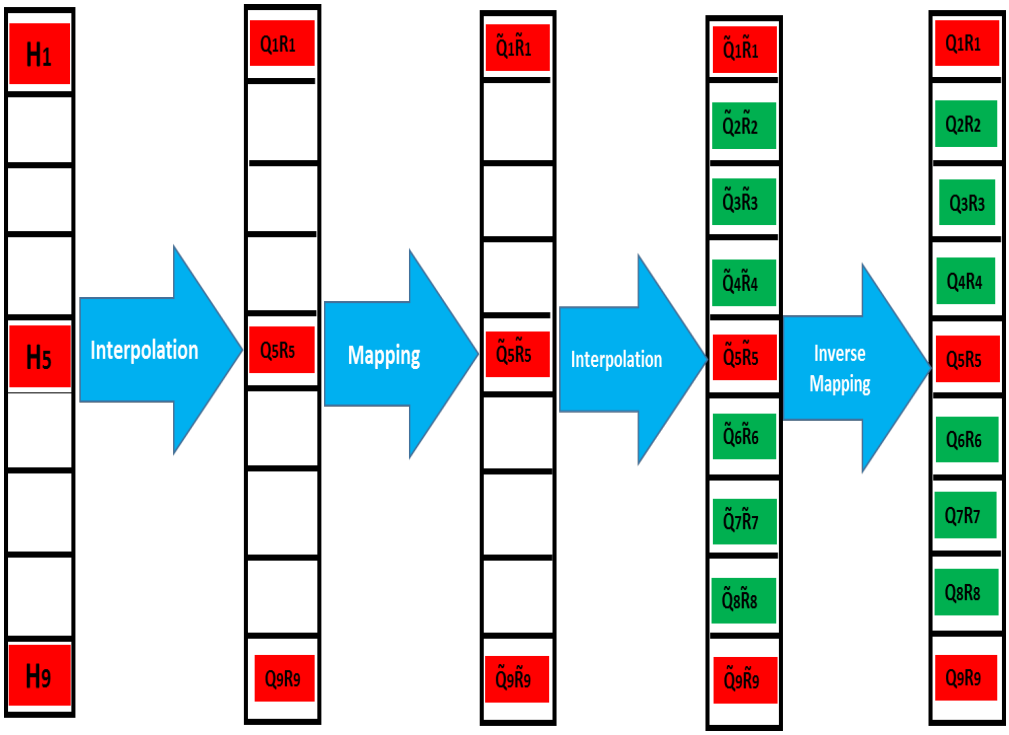


Figure 4.2.: QR decomposition interpolation with LP matrices

If $\underline{q}_k$ denotes k_{th} column of $\mathbf{Q}$ and $\underline{r}_k$ denotes k_{th} row of $\mathbf{R}$, the auxiliary variable Δ_k is defined as follows:

$$\Delta_k \triangleq \Delta_{k-1}[\mathbf{R}]_{k,k}^2 \quad k = 1, 2, 3, ..., M \tag{4.2}$$

with $\Delta_0 \triangleq 1$. $[\mathbf{R}]_{k,k}$ in (4.2) indicates the entry of k_{th} column and k_{th} row of $\mathbf{R}$, and M denotes the number of columns of $\mathbf{R}$. The mapped LP matrices $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$ are obtained as follows [73]:

$$\tilde{\mathbf{Q}} = \mathbf{Q}\Delta \qquad \tilde{\mathbf{R}} = \Delta\mathbf{R} \tag{4.3}$$

$$\mathbf{Q} = \tilde{\mathbf{Q}}\Delta^{-1} \qquad \mathbf{R} = \Delta^{-1}\tilde{\mathbf{R}} \tag{4.4}$$

where Δ is a diagonal matrix with (k, k) entry Δ_k, and

$$\tilde{\mathbf{Q}} \triangleq [\tilde{\underline{q}}_1 \tilde{\underline{q}}_2 ... \tilde{\underline{q}}_M] \tag{4.5}$$

$$\tilde{\mathbf{R}} \triangleq [\tilde{\underline{r}}_1^T \underline{r}_2^T ... \tilde{\underline{r}}_M^T]^T \tag{4.6}$$

where

$$\tilde{\underline{q}}_k \triangleq \Delta_{k-1}[\mathbf{R}]_{k,k}\underline{q}_k \tag{4.7}$$

$$\tilde{\underline{r}}_k \triangleq \Delta_{k-1}[\mathbf{R}]_{k,k}\underline{r}_k. \tag{4.8}$$

If we denote the ordered column rank of $\mathbf{E}$ by K, for $K > 0$ and $k = 1, 2, .., K$, then $\underline{r}_k$ and $\underline{q}_k$ can be computed using:

$$\underline{r}_k = (\Delta_{k-1}[\mathbf{R}]_{k,k})^{-1}\tilde{\underline{r}}_k \tag{4.9}$$

$$\underline{q}_k = (\Delta_{k-1}[\mathbf{R}]_{k,k})^{-1}\tilde{\underline{q}}_k. \tag{4.10}$$

$(\Delta_{k-1}[\mathbf{R}]_{k,k})$ is calculated from the elements of the main diagonal of $\tilde{\mathbf{R}}$ as follows:

$$(\Delta_{k-1}[\mathbf{R}]_{k,k}) = \sqrt{\tilde{\mathbf{R}}_{k,k}}, \qquad \text{if} \quad k = 1 \tag{4.11}$$

$$(\Delta_{k-1}[\mathbf{R}]_{k,k}) = \sqrt{\tilde{\mathbf{R}}_{k-1,k-1}\tilde{\mathbf{R}}_{k,k}}, \quad \text{if} \quad k = 2, 3, .., K. \tag{4.12}$$

The $\mathbf{Q}$ and $\mathbf{R}$ matrices of the data subcarriers are obtained by the inverse of their interpolated $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$ matrices. If $K > 0$, then for $k = 1, 2, ..., K$ the inverse of the mapped matrices $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}$ can be obtained through calculating the scaling factor $(\Delta_{k-1}[\mathbf{R}]_{k,k})^{-1}$ using (4.12) and scaling $\tilde{\underline{q}}_k$ and $\tilde{\underline{r}}_k^T$ according to (4.9) and (4.10). Employing LP matrices in QR decomposition for an $M \times M$ matrix can reduce the number of FLOPs to $O(M^2)$ as compared to the QR decomposition per subcarrier, for which the number of FLOPs is $O(M^3)$.

4.3. Proposed Scheme

4.3.1. Approximate Pseudo Inverse for Relevant Precoding Matrix Components

We next discuss the proposed approximate pseudo inverse scheme for multiple-subcarrier ZF precoding. According to (3.26), the Moore-Penrose pseudo inverse of $\mathbf{H}_{90\times288}$ can be expressed as:

$$\mathbf{H}^{\dagger} = \mathbf{H}^{\mathsf{H}}\mathbf{L}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}(\mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}\mathbf{L}^{\mathsf{H}} \tag{4.13}$$

where $\mathbf{H}^{\dagger}$ is a matrix size of 288×90 and $\mathbf{L}$ is a matrix size of 90×89. We subdivide (4.13) into three steps of multiplications as follows:

$$\begin{aligned} \mathbf{P_1} &= \mathbf{L} \times \left(\mathbf{L}^{\mathsf{H}}\mathbf{L}\right)^{-\mathbf{1}} \\ \mathbf{P_2} &= \mathbf{P_1} \times \ \mathbf{P_1^{\mathsf{H}}} \\ \mathbf{P_3} &= \mathbf{H}^{\dagger} = \mathbf{H}^{\mathsf{H}} \times \mathbf{P_2}. \end{aligned} \tag{4.14}$$

Let $\mathbf{W}^{(j,i)}_{Element-wise}$ be the element-wise precoding matrix element of the i_{th} row and j_{th} column of $\mathbf{H}^{\dagger}$. We write:

$$\mathbf{W}^{(j,i)}_{Element-wise} = \mathbf{P}_3^{(j,i)} \tag{4.15}$$

$$\mathbf{W}^{(j,i)}_{Element-wise} = \mathbf{H}^{\mathsf{H}}(j,:) \times \mathbf{P}_2^{(:,i)}. \tag{4.16}$$

We apply a power threshold with respect to the strongest element on the jth row of $\mathbf{H}^{\mathsf{H}}$; the elements below this threshold will be set to zero. Thus, in computing $\mathbf{P}_1$ the number of rows in the matrix $\mathbf{L}$ to be multiplied by $(\mathbf{L}^{\mathbf{H}}\mathbf{L})^{-\mathbf{1}}$ will be limited to the number of non zero elements in the jth row of $\mathbf{H}^{\mathsf{H}}$ while the power threshold window is applied.

4.4. Implementation of Multi-approach Precoding

This section presents a practical implementation of the proposed precoding with multiple subcarriers. The scheme is based on dividing the Precoding Channel Components (PCCs) into strong, weak and very weak classes based on the average power precoding matrix. To reduce the complexity, an element-wise pseudo inverse approximation is applied to the Strong Precoding Channel Components (SPCCs).

Linear interpolation is used to compute the pseudo inverse of the Weak Precoding

Channel Components (WPCCs) with low computational complexity. PCCs with the lowest average power, denoted as Very Weak Precoding Channel Components (VWPCCs), are set to zero.

4.4.1. Non-Interpolated ZF Precoding Matrices

Non-interpolated ZF precoding matrices are considered as precoding matrices which are obtained either through the SVD decomposition function *svdecon* (as introduced in section 3.5.5), state-of-the-art *geninv*, or the proposed *Sgeninv* algorithm, introduced in Chapter 3. W_{int} denotes the non-interpolated precoding matrix of size 288×90.

4.4.2. Average Power Precoding Matrix

In order to apply the proposed scheme, we must first distinguish the SPCCs from the WPCCs. If $N_{W_{int}}$ denotes the number of pilot precoding matrices W_{int}, the average power of the precoding matrix $P_{W_{int}}$ is calculated as follows:

$$P_{W_{int}(dB)} = 10\log_{10}\frac{1}{N_{W_{int}}}\sum_{i=1}^{288}\sum_{j=1}^{90}\sum_{\delta=1}^{N_{W_{int}}}\left|{W_{int}}^{(i,j,\delta)}\right|^2 \tag{4.17}$$

where $W_{int}^{(i,j,\delta)}$ denotes the i_{th} row and j_{th} column of the δ_{th} pilot precoding matrix W_{int}. The SPCCs and WPCCs can be distinguished using appropriate power thresholds.

4.4.3. Large Subcarrier Spacing

For all the W_{int} matrices, Large Subcarrier Spacing ($LargeSCS$) specifies the subcarriers for which the pilot ZF precoding is applied. Furthermore, the W_{int} matrix elements spaced by the $LargeSCS$ can be used as the WPCCs interpolation intervals.

4.4.4. Small Subcarrier Spacing

Similarly to the WPCCs, the SPCCs can be interpolated using a Small Subcarrier Spacing ($SmallSCS$). It is assumed that $SmallSCS \leq LargeSCS$ since the SPCCs have a stronger effect on the performance when the multi-approach scheme is applied.

4.4.5. Strong PCCs Power Threshold

To filter out the SPCCs from the WPCCs in the matrix W_{int}, a strong component power threshold window $Strong_{P_{TH}}$ with respect to the strongest power element $SPCC_{max}$ on W_{int} is employed. For each SPCC at the j_{th} column and i_{th} row of $P_{W_{int}}$, we have

$$P_{W_{int}(i,j)} \geq SPCC_{max} - Strong_{P_{TH}} \tag{4.18}$$

where

$$SPCC_{max} = \text{argmax}(P_{W_{int}}). \tag{4.19}$$

4.4.6. Weak PCCs and Very Weak PCCs

To identify Tx-beams that are sufficiently below the SPCCs, the WPCCs and VWPCCs are introduced. $P_{W_{int}(i,j)}$ is specified as follows:

$$P_{W_{int}(i,j)} \geq SPCC_{max} - Weak_{P_{TH}} \tag{4.20}$$

$$P_{W_{int}(i,j)} < SPCC_{max} - Strong_{P_{TH}}. \tag{4.21}$$

$Weak_{P_{TH}}$ in (4.20) denotes the weak PCC threshold window that identifies the WPCC matrix elements in $P_{W_{int}}$. $P_{W_{int}(i,j)}$ identifies as a VWPC at the j_{th} column and i_{th} row, if neither (4.18) nor (4.20), (4.21) is fulfilled. As such, VWPCCs are set to zero for all the $N_{W_{int}}$ number of subcarriers. Combining (4.18), (4.20) and (4.21), $P_{W_{int}(i,j)}$ can be defined as

$$P_{W_{int}(i,j)} = \begin{cases} SPCC, & \text{if } P_{W_{int}(i,j)} \geq SPCC_{max} - Strong_{P_{TH}} \\ WPCC, & \text{if } SPCC_{max} - Weak_{P_{TH}} \leq P_{W_{int}(i,j)} < SPCC_{max} - Strong_{P_{TH}} \\ VWPCC, & \text{otherwise.} \end{cases} \tag{4.22}$$

4.4.7. Interpolation-based Approximate Multiple-subcarrier ZF Precoding

Figure 4.3 illustrates the proposed scheme for interpolation. First, the *Sgeninv* algorithm (or other examined methods) are applied to obtain the pseudo inverse of channel matrices on pilot subcarriers spaced by the $LargeSCS$. In order to specify the SPCCs and WPCCs, the average power precoding matrix $P_{W_{int}}$ over all the pilot precoding matrices is calculated. By applying (4.22) on each element (i, j) of $P_{W_{int}(i,j)}$, all the SPCCs, WPCCs

and VWPCCs are specified.

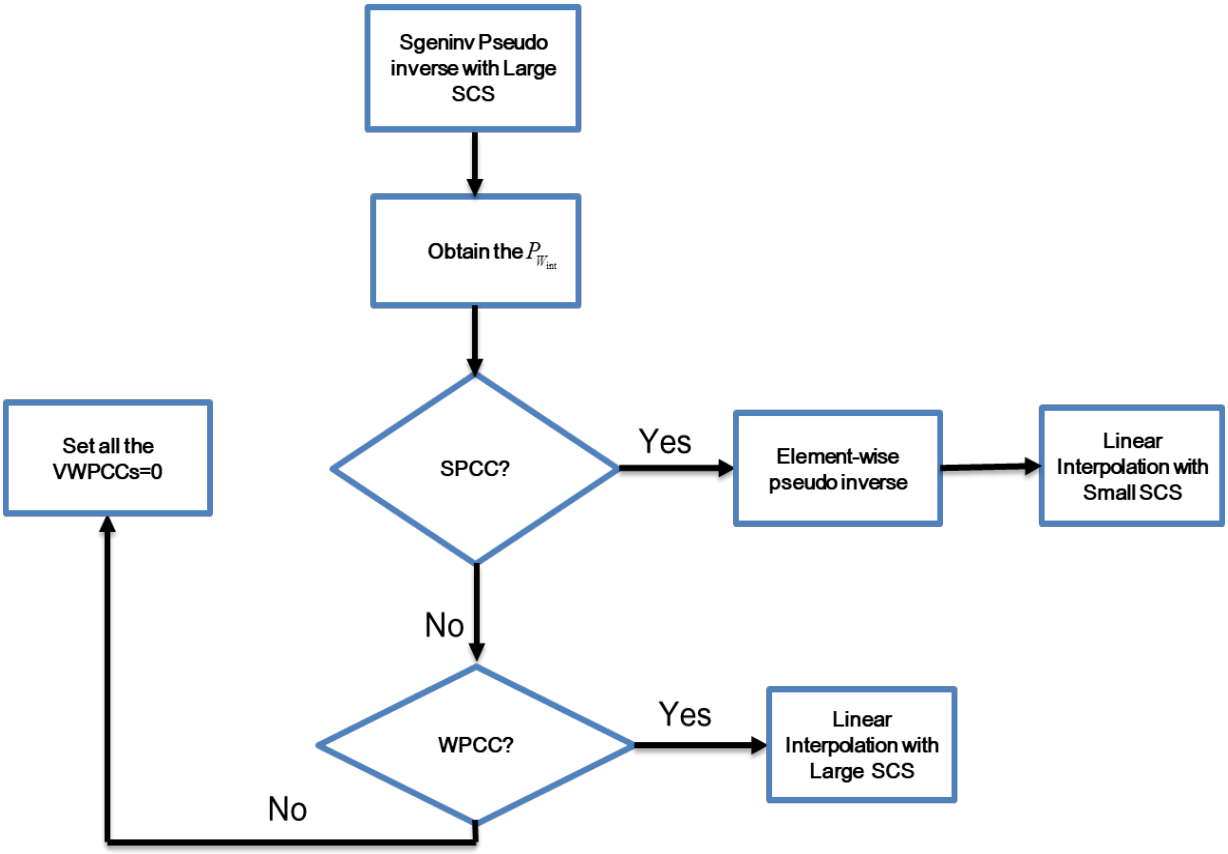


Figure 4.3.: Multi-approach interpolation-based scheme

An approximate pseudo inverse, as discussed insection 4.3, is applied to the SPCCs. Considering the approximated SPCCs as the main intervals, linear interpolation for all the SPCCs between the intervals spaced by $SmallSCS$ is applied. If $P_{W_{int}(i,j)}$ identifies a WPCC, linear interpolation with respect to $LargeSCS$ is employed. Eventually, all the specified VWPCCs are set to zero within all the subcarriers.

If $SmallSCS < LargeSCS$ then we define

$$N_{smallSCS} = \lfloor LargeSCS/SmallSCS \rfloor + (SmallSCS \bmod LargeSCS) \tag{4.23}$$

As depicted in Figure 4.4, α and β denote the subcarrier indices for the SPCCs and the WPCCs respectively. The number of SPCCs between $\alpha + N_{smallSCS} \times SmallSCS$ and $\alpha + \beta$ are approximated by calculating the remaining SPCCs as given in (4.23). If $SmallSCS = LargeSCS$, then α and β are equal. Figure 4.5 illustrates an example for the subcarrier spacing interpolation when $SmallSCS = LargeSCS$=72.

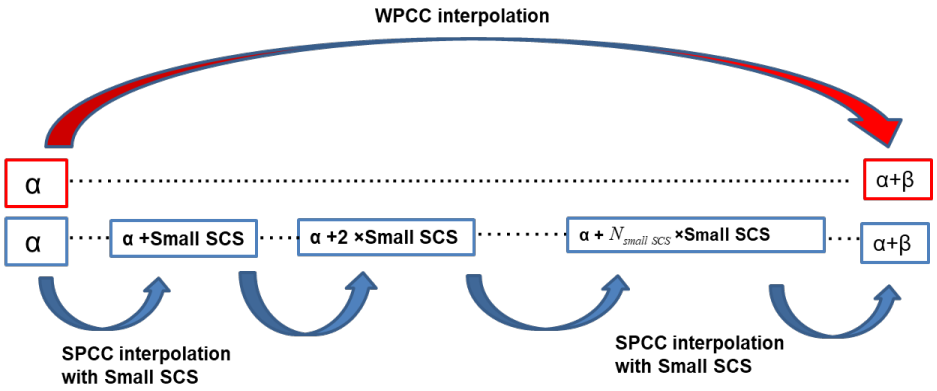


Figure 4.4.: Subcarrier interpolation when $SmallSCS < LargeSCS$

.

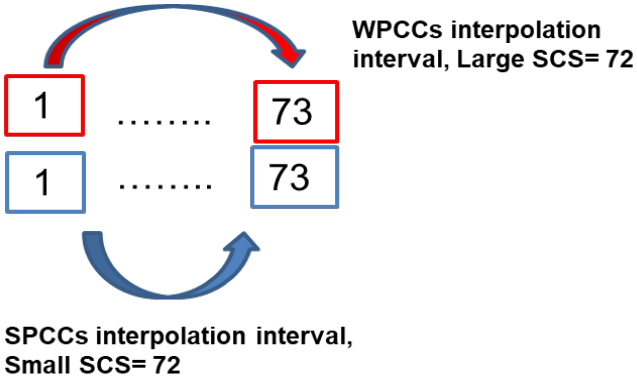


Figure 4.5.: Subcarrier interpolation when $SmallSCS = LargeSCS$

Interpolation-based Precoding Using State-of-the-Art Pseudo Inverse

This section details the proposed interpolation when the state-of-the-art pseudo inverse algorithm is applied to the pilot subcarriers. Figure 4.6 gives an overview of four pilot subcarriers among 1200 subcarriers when $LargeSCS = 72$.

If the state-of-the-art algorithm is applied to calculate the precoding matrices of the pilot subcarriers, then the PCCs, WPCCs and VWPCCs are found by first taking the average power precoding matrix $P_{W_{int}}$ of the pilot subcarriers. According to (4.19), both $Strong_{P_{TH}}$ and $Weak_{P_{TH}}$ are applied with respect to $SPCC_{max}$ which identifies the components in $P_{W_{int}}$.

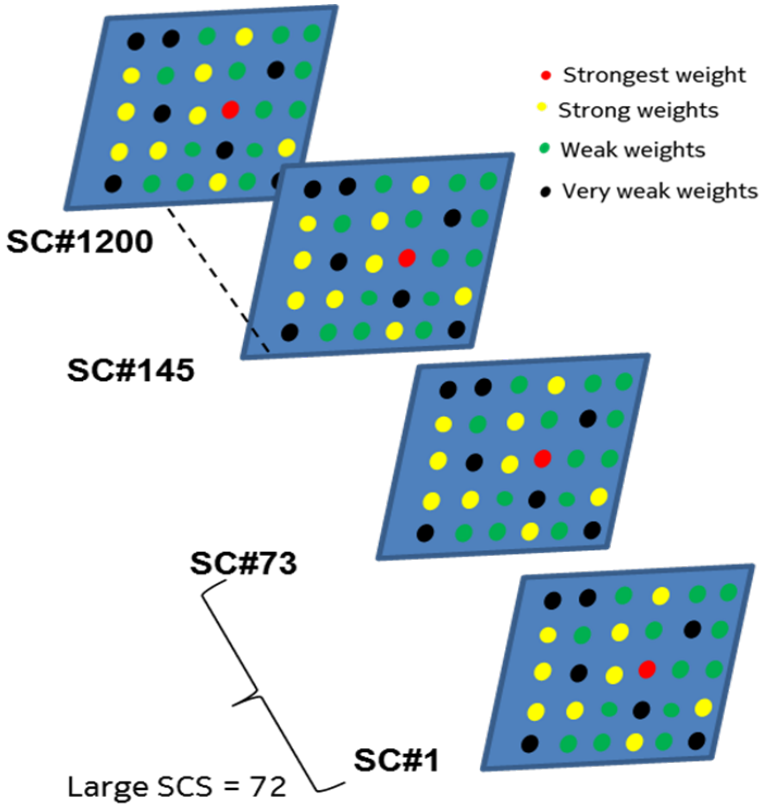


Figure 4.6.: Pilot subcarriers and precoding channel components

SPCCs in the $P_{W_{int}}$ matrix are employed as the intervals to approximate the SPCCs of the non-pilot subcarriers which are spaced by $SmallSCS$. Clearly, the number of SPCCs identified in $P_{W_{int}}$ depends on the selected $Strong_{P_{TH}}$ with respect to $SPCC_{max}$. The larger threshold value with respect to $SPCC_{max}$ results in applying the proposed approximate pseudo inverse method to a larger number of PCCs.

- Figure 4.7 depicts the SPCCs linear interpolation process when $SmallSCS = 4$. SPCCs on SC_5 are calculated using the proposed approximation method. However, all the SPCCs between SC_1 and SC_5 are obtained through linear interpolation between the approximated SPCCs on SC_1 and SC_5.

- Figure 4.8 represents the WPCCs linear interpolation process for a total number of 1200 subcarriers when $LargeSCS = 72$. All the WPCCs on SC_1 and SC_{73} are obtained by applying the state-of-the-art pseudo inverse. For WPCCs located between SC_1 and SC_{73} linear interpolation is applied.

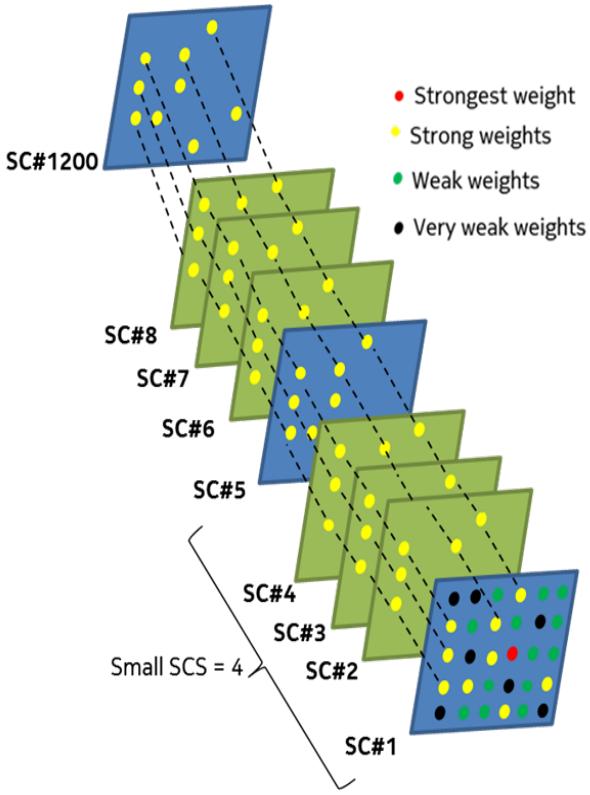


Figure 4.7.: Strong precoding channel components interpolation

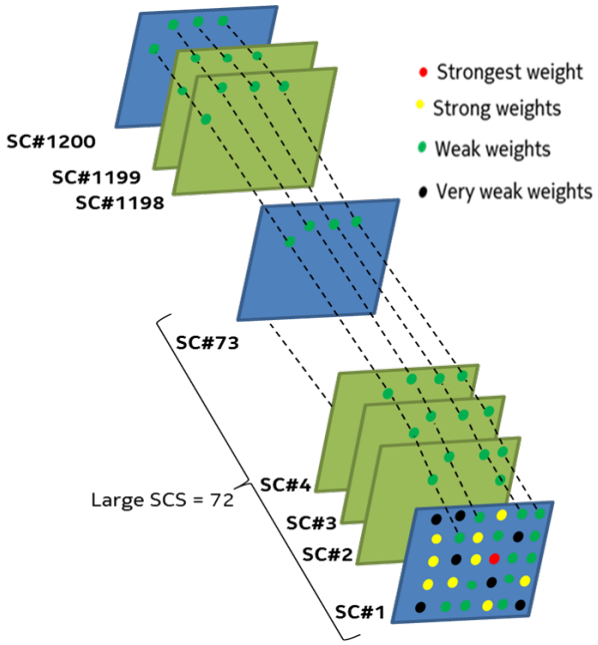


Figure 4.8.: Weak precoding channel components interpolation

▷ Figure 4.9 illustrates the process of zeroing out the VWPCCs as described in section 4.4.6. As discussed above, the number of VWPCCs which are set to zero for all the subcarriers depends on $Weak_{P_{TH}}$ and $Strong_{P_{TH}}$.

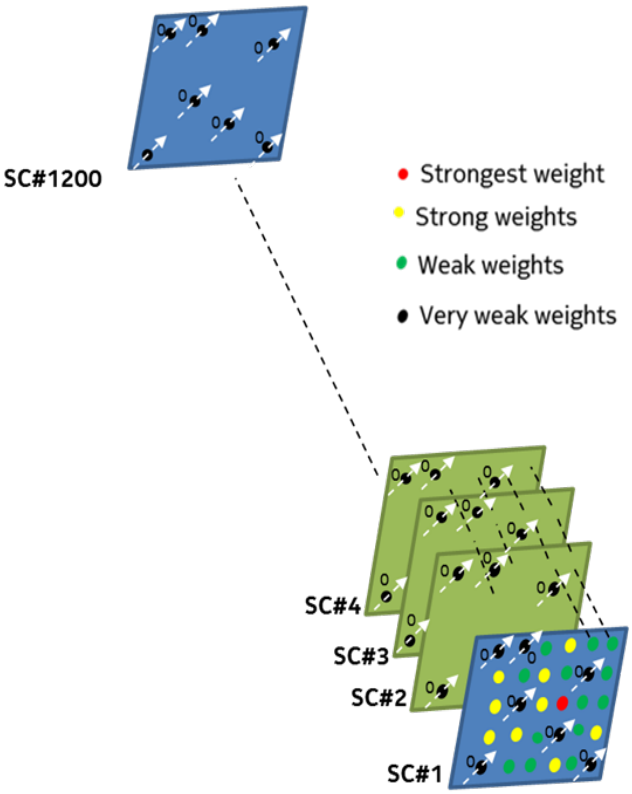


Figure 4.9.: Very weak precoding channel components interpolation

Interpolation-based Precoding using Sparse Geninv Pseudo Inverse

To illustrate the computational complexity reduction using the proposed pseudo inverse method, an average power matrix can be defined as follows:

$$P_{W_{int}(dB)} = 10\log_{10}\frac{1}{N_{W_{int}}}\sum_{i=1}^{288}\sum_{j=1}^{90}\sum_{\delta=1}^{N_{W_{int}}}\left|W_{SGeninv}{}^{(i,j,\delta)}\right|^2 \tag{4.24}$$

where

$$W_{Sgeninv}^{(,,\delta)} = SGeninv(H_\delta). \tag{4.25}$$

The complexity reduction is influenced not only by approximating the SPCCs, but an additional gain is obtained while the pseudo inverse of W_{int} is calculated. However, the performance degrades due to applying $Strong_{P_{TH}}$ window and $Weak_{P_{TH}}$ window to the

channel components of **H**. The main impact of employing $Strong_{P_{TH}}$ and $Weak_{P_{TH}}$ is to set some channel components to zero before calculating the $P_{W_{int}}$ and specifying the SPCCs, WPCCs and VWPCCs. An evaluation of the proposed approach is discussed next.

4.5. Evaluation Results

We analyze the performance of our multiple-subcarrier ZF precoding methods by examining the average SINR. In the following, first the performance degradation of the proposed approximate pseudo inverse with desired SPCCs is investigated. Following this, different $Weak_{P_{TH}}$ are considered to analyse the degradation of the average SINR when the reference ZF precoding is applied.

Finally, the performance loss as well as the computational complexity gain of the schemes introduced in section 4.3 and section 4.4.7 are compared with the state-of-the-art and reference algorithms.

4.5.1. Element-wise Approximate Pseudo Inverse

This section investigates the impact of the proposed approximate pseudo inverse on the average SINR for the multiple subcarrier case. In order to employ an appropriate range for the $SmallSCS$ with respect to $Strong_{P_{TH}}$, the interpolation based element-wise ZF precoding for SPCCs is evaluated.

Figure 4.10 shows the average SINR performance of 90 UEs over 1200 subcarriers as the $SmallSCS$ changes between 2 and 20. Moreover, the SPCCs are distinguished from the other coefficients in $P_{W_{int}}$. $Strong_{P_{TH}}$ changes from -3 dB to -21 dB. The average SINR for $Strong_{P_{TH}}$= -3 dB remains constant irrespective of $SmallSCS$. The reference average SINR= 15.5 dB is calculated over 1200 subcarriers using ZF precoding based on SVD decomposition.

The SINR loss will increase by up to 0.5 dB with respect to the reference when $Strong_{P_{TH}}$= -9 dB. However, $SmallSCS$ has no such significant impact. The SINR loss with respect to the reference point starts to increase slightly when $Strong_{P_{TH}}$= -12 dB. Performance degradation due to increasing $SmallSCS$ is moderate when -15 dB $\leq Strong_{P_{TH}} \leq$ -18 dB.

If $Strong_{P_{TH}}$= -18 dB, then the SINR loss increases for $SmallSCSs$ between 2 and 8. However, when the $SmallSCSs$ are between 8 and 17, the average SINR degradation reduces to 0.8 dB. If $Strong_{P_{TH}}$= -18 dB, then there is 1 dB loss in the average SINR when $SmallSCS$= 20. If $Strong_{P_{TH}}$ window increases to -21 dB and $SmallSCS$=14, then the

average SINR loss of 1 dB affects the performance moderately.

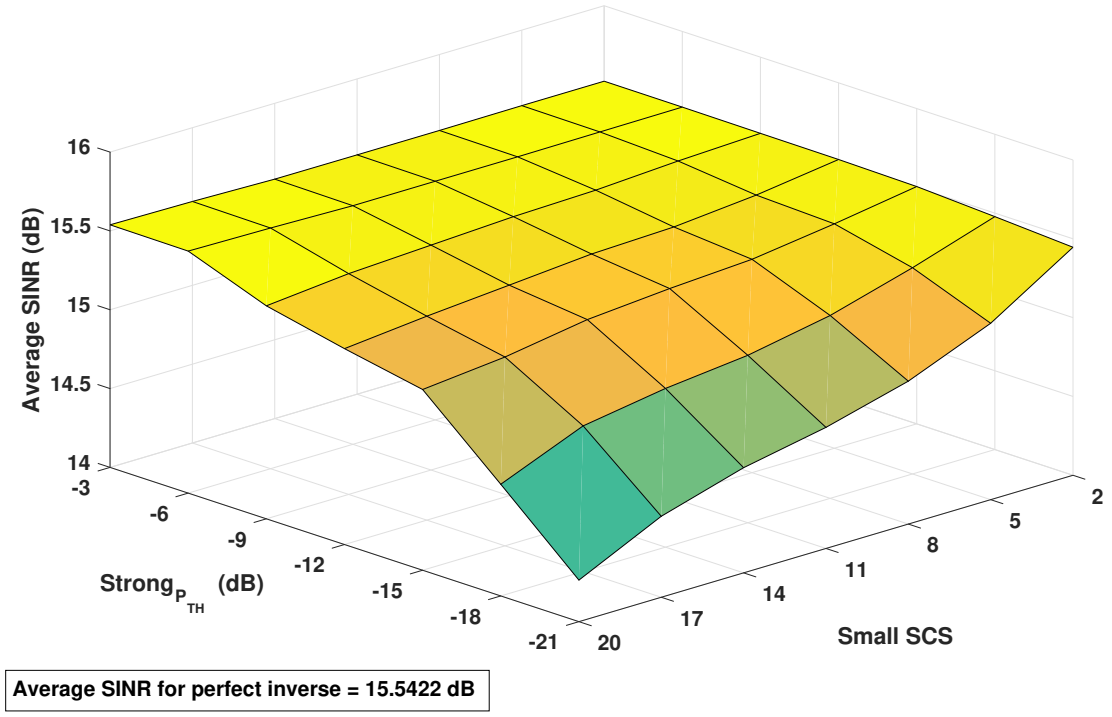


Figure 4.10.: Element-wise ZF precoding performance

4.5.2. VWPCC Impact on ZF Precoding with Multi-Subcarriers

Before investigating the complexity reduction, we first evaluate the performance degradation of the reference method of ZF precoding for multiple subcarriers when VWPCCs with respect to the SPCC are zeroed out.

Figure 4.11 shows the CDF for SINR when $VeryWeak_{P_{TH}}$ varies between -30 dB and -39 dB with respect to the SPCC of $\mathbf{W}_{sub}$ where sub ranges from 1 to 1200. For each experiment, we first obtain the precoding $\mathbf{W}$ of all the subcarriers using *svdecon* as a pseudo inverse method.

Consequently, the VWPCCs corresponding to each subcarrier are set to zero when $SPCC_{sub}$ is defined. $VeryWeak_{P_{TH}}$ defines the power threshold with respect to the VWPCC. The number of VWPCCs increases as a larger $VeryWeak_{P_{TH}}$ is employed. However, the SINR loss increases as a larger number of VWPCCs are set to zero. For $VeryWeak_{P_{TH}} \leq$ -35 dB, the SINR loss is less than 1 dB. For higher thresholds, the SINR loss increases up to 2 dB ($VeryWeak_{P_{TH}}$= -31 dB).

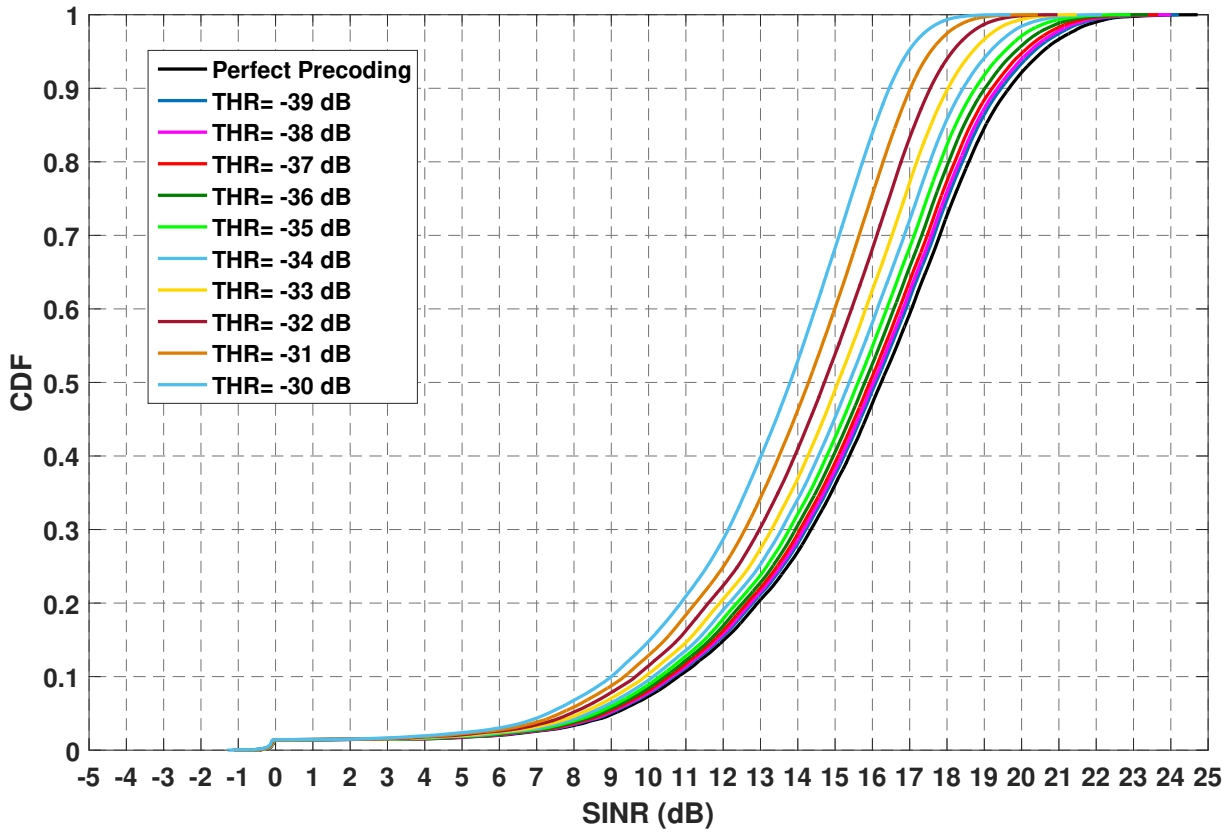


Figure 4.11.: VWPCCs cancellation effect

4.5.3. Comparison of SINR Loss and FLOPs Count

Simulation results for the proposed interpolation-based multiple-subcarrier ZF precoding is analyzed in this section. We choose $SmallSCS$=17 to keep the average SINR loss caused by the element-wise ZF precoding close to 1 dB. Furthermore, $VeryWeak_{P_{TH}}$= -35 dB is applied to minimize the computational complexity and to reduce the average SINR loss per UE over 1200 subcarriers to below 1 dB with respect to the reference example (as detailed in section 3.5.5). Simulation parameters that were used to calculate the number of FLOPs and the SINR loss for ZF precoding with multiple subcarriers are summarized in Table 4.1.

Each PRB in LTE consists of 12 consecutive subcarriers in the frequency domain, where each subcarrier comprises 6 resource elements if a long cyclic prefix is applied. Each PRB thus consists of $6 \times 12 = 72$ resource elements. To examine the approximate ZF precoding with multiple subcarriers with respect to the LTE physical transmission resources, we assume $LargeSCS$= 72. If the MATLAB Pinv function is used as the reference method to calculate ZF precoding of 1200 subcarriers without approximation, then the total com-

putational complexity approaches $1200 \times 18654483 = 2.2385 \times 10^{11}$ FLOPs. To exemplify

Symbols	PHY Layer Parameters Value	Value
f_c	Carrier Frequency (GHz)	2.1
TB	Total Bandwidth (MHz)	20
UB	Used Bandwidth (MHz)	18
B_{sep}	Subcarrier Spacing (KHz)	15
N_{sub}	Number of Subcarriers	1200
N_{PRB}	Number of PRBs	100
Tx Power$_{dBm}$	Tx Power per per cell (dBm)	46
$ReceiverNF$	Receiver NF (dB)	7
N_{UE}	Number of UEs	90
N	Number of Antennas per UE	1
N_{cell}	Number of Cells	9
N_{Beams}	Number of Tx Beams	288
N_{OFDM}	Number of OFDM blocks per PRB	6
TTI	Duration of one subframe (ms)	1
$LargeSCS$	Large subcarrier spacing	72
$SmallSCS$	Small subcarrier spacing	17
$VeryWeak_{P_{TH}}$	Very weak threshold	-35 dB
$SGeninvP_{TH}$	SGeninv power threshold	-20 dB
$GrevilleP_{TH}$	Greville power threshold	-20 dB

Table 4.1.: ZF Precoding with simulation parameters

the effect of the approximate ZF precoding on the SINR loss and complexity gain, Sgeninv, Greville and the MATLAB pinv function are employed to calculate the pilot subcarrier precoding.

Figure 4.12 depicts the average SINR loss per UE over 1200 subcarriers against the FLOPs count when the Greville algorithm considering $GrevilleP_{TH}$= -20 dB is applied. While $Strong_{P_{TH}}$ changes between -15 dB and -30 dB with respect to the SPCC, the SINR loss varies between 1.8 dB and 2.2 dB. The FLOPs count reduces linearly from 3.6×10^9 (with 1.8 dB average SINR loss) to 0.23×10^9 for 1200 subcarriers with 2.2 dB average SINR loss.

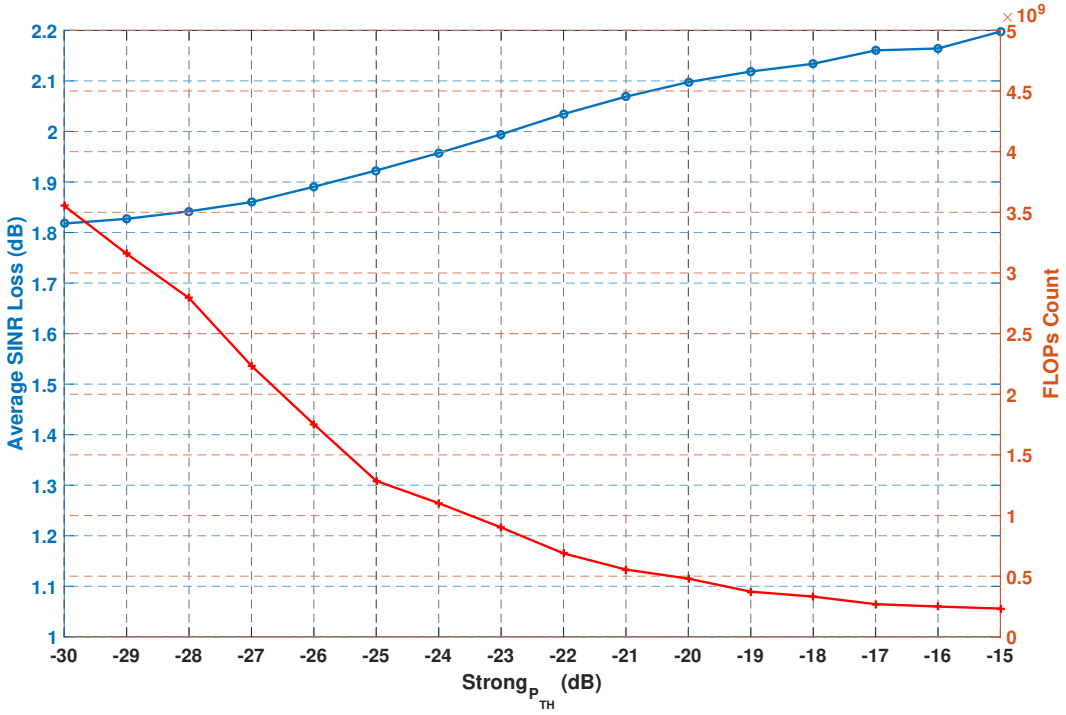


Figure 4.12.: Greville performance for multiple subcarriers

Figure 4.13 shows the approximate ZF precoding performance when the Sgeninv algorithm is used to calculate the pilot precoding matrices. The average SINR loss decreases by 0.4 dB as compared to Greville. Similar to the Greville approach, the computational complexity reduces by a factor of 100 as compared to the SVD based pseudo inverse when the approximate ZF precoding and an element-wise pseudo inverse for the SPCCs are applied.

Figure 4.14 demonstrates the approximate ZF precoding performance when the MATLAB optimized SVD based function is applied as the baseline to calculate pilot precoding matrices. In contrast with Sgeninv and Greville, no power threshold is employed to specify the RCCs before precoding. $Strong_{P_{TH}}$ is applied to the $P_{W_{int}}$ matrix with respect to the $SPCC_{max}$. The average SINR loss decreases by 0.4 dB as compared to Sgeninv and 0.8 dB as compared to Greville for $Strong_{P_{TH}}$= -15 dB, -16 dB. The approximate ZF precoding SINR loss difference between Pinv and Sgeninv is reduced when a lager $Strong_{P_{TH}}$ window of, e.g., -30 dB is applied as the performance loss drops by 0.26 dB.

The FLOPs count difference between Pinv and Sgeninv is 0.25×10^9 when $Strong_{P_{TH}}$= -15 dB. This difference increases to 1.14×10^9 when a larger $Strong_{P_{TH}}$, e.g., -30 dB is applied. Similar to the ZF precoding single subcarrier analysis, choosing an appropriate power threshold window, e.g., $Strong_{P_{TH}}$, for multiple subcarrier precoding improves the perfomance loss as well as the complexity gain.

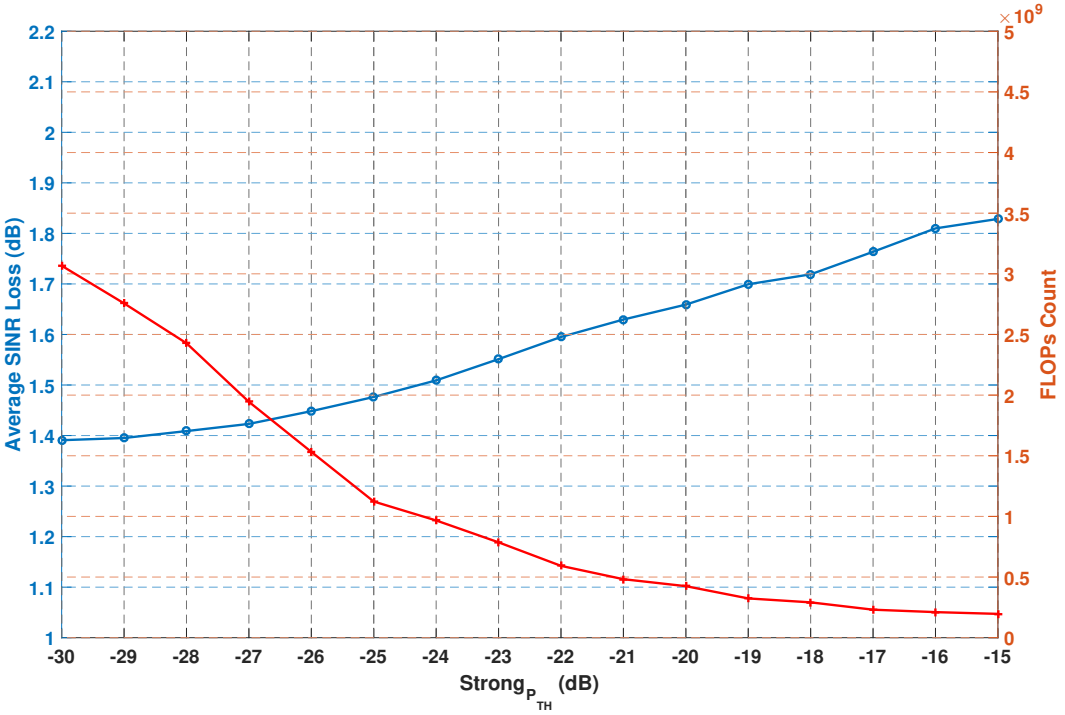


Figure 4.13.: Sgeninv performance for multiple subcarriers

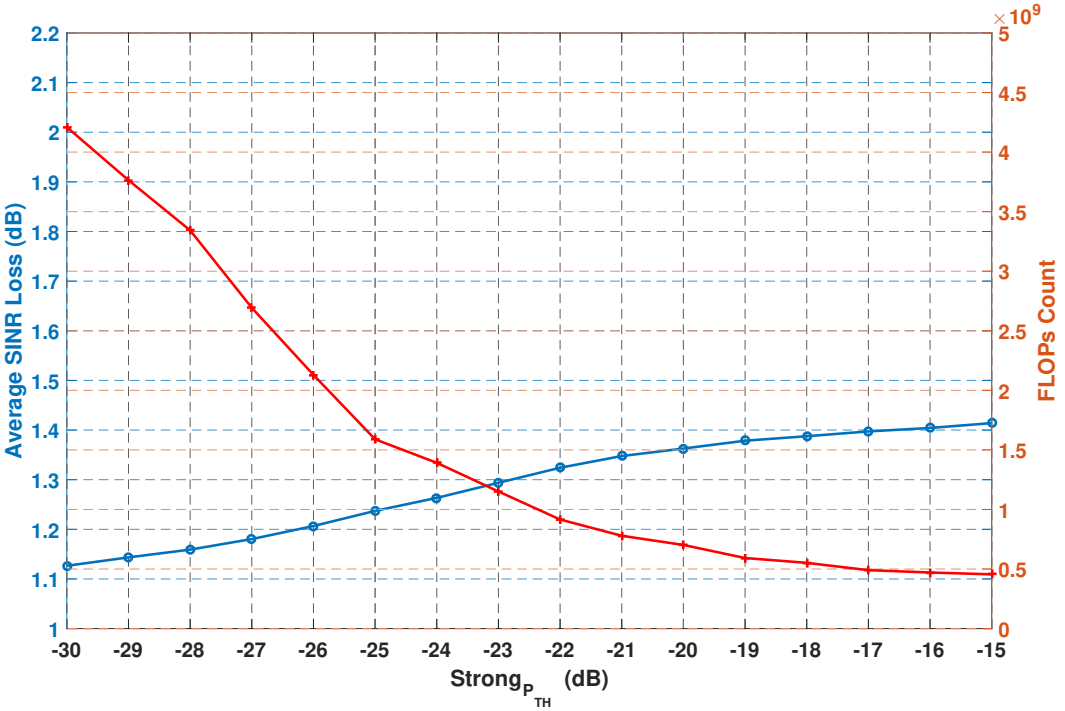


Figure 4.14.: MATLAB Pinv performance for multiple subcarriers

5

Conclusion and Outlook

This dissertation has presented precoding algorithms for mMIMO systems, including a computational complexity analysis. One of the main outcomes of this work is to present a toolbox to count the precoding computational complexity in floating point operations (FLOPs) for large channel matrices. This toolbox adds more mathematical features to calculate FLOPs as compared to the existing Lightspeed.

To present the low-complexity beamforming solutions, we focused on joint transmission coordinated multipoint (JT CoMP) or network MIMO scenarios where the interference is mitigated substantially but we must deal with a large channel matrix. Low-complexity beamforming methods were presented by employing a grid of beams (GoB) at the eNodeB to deactivate non-relevant beams towards the UEs.

Channel state information (CSI) feedback overhead is another challege in frequency division duplex (FDD) massive MIMO systems. A power threshold window P_{TH} was used at the UE to reduce the CSI reporting requirements to the eNodeB.

For MIMO-OFDM with multiple subcarriers, we extended the Sgeninv scheme by introducing two power thresholds that are employed after precoding. $Strong_{P_{TH}}$ and $Weak_{P_{TH}}$ were applied to distinguish the Strong Precoding Channel Components (SPCCs) and Weak Precoding Channel Components (WPCCs) of the average precoding matrix. Linear interpolation in combination with defining pilot subcarriers via Large Subcarrier Spacing (SCS) and Small SCS demonstrated a good performance. An element-wise pseudo inverse was used to approximate the SPCCs of the precoding matrices. Linear interpolation was employed for the WPCCs approximation over multiple subcarriers. To evaluate the performance, LTE physical transmission resources were considered. In contrast with single

subcarrier MIMO-OFDM, a complexity gain of a factor of 100 as compared to the reference method was observed (see section 4.5.3) when Sgeninv or Greville are used to generate the pilot subcarriers. The average SINR degradation increases by only 0.5 dB as compared to the performance for a single subcarrier. To reduce the SINR loss and achieve complexity gain for multiple subcarriers, the reference method was employed. The average SINR loss for a small $Strong_{P_{TH}}$ window, e.g, -15 dB, is about 0.4 dB as compared to Sgeninv if the reference MATLAB Pinv function is used to calculate the pilot precoding matrices.

Further research topics include the following:

- ▷ Extend the complexity analysis to other pseudo inverse methods for single and multiple subcarriers.
- ▷ Extend the analysis to larger cooperation sites with more UEs per cell.
- ▷ Evaluate the effect of expanding the size of the channel matrix on the complexity gain and performance degradation.
- ▷ Analyze the optimization problem where the number of FLOPs and SINR degradation are defined, for example, per $LargeSCS$, per $SmallSCS$, per $Strong_{P_{TH}}$ and per $Weak_{P_{TH}}$.
- ▷ Evaluate the precoding schemes using other interpolation methods to approximate the WPCCs.
- ▷ Include the CSI estimation techniques at the UE and consider the impact of the non-perfect CSI reporting on the precoding for single and multiple subcarriers.
- ▷ To enhance the use of low-complexity beamforming solutions, the analysis of the Relevant Multi Path Components (RMPCs) as well as the Relevant Channel Components (RCCs) can be considered. Reducing the number of MPCs affects the number of the relevant taps of the Channel Impulse Response (CIRs) per RCC. One could consider the use of RMPCs as well as the RCCs on the precoding complexity gain and performance degration.

A

FLOPs Count for Mathematical Operations

Basic Mathematical Operations

Basic complex-valued mathematical operations for the FLOPs count analysis discussed in Chapter 3 are given as follows:

FLOPs count for the real part:

$$\text{FLOP}_{\text{real}}(\text{Addition}) = 1$$

$$\text{FLOP}_{\text{real}}(\text{Subtraction}) = 1$$

$$\text{FLOP}_{\text{real}}(\text{Division}) = 1$$

$$\text{FLOP}_{\text{real}}(\text{Multiplication}) = 1$$

$$\text{FLOP}_{\text{real}}(\text{Comparison}) = 1$$

$$\text{FLOP}_{\text{real}}(\text{Fast Inverse Square Root}) = 8$$

FLOPs count for complex part:

$$\begin{aligned}
\text{FLOP}_{\text{complex}}(\text{Addition}) &= 2 \times \text{FLOP}_{\text{real}}(\text{Addition}) \\
\text{FLOP}_{\text{complex}}(\text{Subtraction}) &= 2 \times \text{FLOP}_{\text{real}}(\text{Subtraction}) \\
\text{FLOP}_{\text{complex}}(\text{Multiplication}) &= 4 \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \\
&\quad + 2 \times \text{FLOP}_{\text{real}}(\text{Addition}) \\
\text{FLOP}_{\text{complex}}(\text{Division}) &= \text{FLOP}_{\text{complex}}(\text{Multiplication}) \\
&\quad + 2 \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \\
&\quad + \text{FLOP}_{\text{real}}(\text{Multiplication}) \\
&\quad + 2 \times \text{FLOP}_{\text{real}}(\text{Division})
\end{aligned}$$

Matrix Operations

Basic complex-valued matrix operations used in the FLOPs count toolbox are given as follows:
If $\mathbf{A}$ is a matrix of size $m \times n$ and $\mathbf{B}$ is a matrix of size $n \times p$, then we have

$$\begin{aligned}
\text{FLOP}_{\mathbf{A}\times\mathbf{B}} = \Big[& m \times n \times p \times \text{FLOP}_{\text{complex}}(\text{Multiplication}) \\
& + m \times (n-1) \times p \times \text{FLOP}_{\text{complex}}(\text{Addition}) \Big]
\end{aligned}$$

If $\mathbf{A}$ is a matrix of size $m \times n$ and $\mathbf{B}$ is a matrix of size $p \times q$, then we have

$$\text{FLOP}_{\text{Kronecker Product}} = m \times n \times p \times q \times \text{FLOP}_{\text{complex}}(\text{Multiplication})$$

If $\mathbf{A}$ is a semi-definite matrix of size $m \times m$ and $\mathbf{B}$ is a semi definite matrix of size $n \times n$, then we have

$$\begin{aligned}
\text{FLOP}_{\text{Kronecker Product}} = \Bigg[& m \times \Big((n^2 - n) \times 0.5 \times 2 \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \\
& + n \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \Big) \\
& + (m^2 - m) \times 0.5 \times \Big((n^2 - n) \times 0.5 \times \text{FLOP}_{\text{complex}}(\text{Multiplication}) \\
& + 2 \times n \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \Big) \Bigg]
\end{aligned}$$

If $\mathbf{A}$ is a positive-definite matrix of size $m \times m$ then we have

$$\begin{aligned}\text{FLOP}_{(\mathbf{A}^{-1})} = \Bigg[&(\frac{1}{2}m^3 + \frac{3}{2}m^2) \times \text{FLOP}_{\text{complex}}(\text{Multiplication}) \\ &+ (\frac{1}{2}m^3 - \frac{1}{2}m^2) \times \text{FLOP}_{\text{complex}}(\text{Addition}) \\ &+ m \times \text{FLOP}_{\text{real}}(\text{Fast Inverse Square Root})\Bigg]\end{aligned}$$

If $\mathbf{A}$ is a matrix of size $m \times n$ then the Square Frobenius Norm can be calculated as follows:

$$\begin{aligned}\text{FLOP}_{SquareFrobNorm_{(\mathbf{A})}} = \Bigg[&2 \times m \times n \times \text{FLOP}_{\text{real}}(\text{Multiplication}) \\ &+ (n \times m - 1) \times \text{FLOP}_{\text{real}}(\text{Addition})\Bigg]\end{aligned}$$

B

Abbreviations

List of Abbreviations

$3GPP$	3rd Generation Partnership Project
$3D-UMa$	3D urban macro channel
$4G$	4th generation mobile networks
$5G$	5th generation mobile networks
AoA	angle of arrival
AoD	angle of departure
AR	augmented reality
BS	base station
CA	cooperation area
CB	coordinated beamforming
CC	channel component
CIR	channel impulse response
$CoMP$	coordinated multi point
CS	coordinated scheduling
CSI	channel state information
CTF	channel transmission function
DAC	digital analog convertor
DFT	discrete fourier transform
DPC	dirty-paper coding
$eMMB$	enhanced mobile broadband

$eNodeB$	evolved node B
FFT	fast fourier transform
FDD	frequency division duplex
FDM	frequency division multiplexing
$FLOPs$	floating point operations
$FLOP_{SVD}$	floating point operations for SVD
$FLOP^{QR}_{MGS}(Addition)$	floating point for number of additions in MGS QR
$FLOP^{QR}_{MGS}(Multiplication)$	floating point for number of subtractions in MGS QR
$FLOP^{QR}_{MGS}(Subtraction)$	floating point for number of subtractions in MGS QR
GoB	grid of beam
GSO	gram-schmidt orthogonalization
$GSCM$	geometric-based stochastistic channel model
$HARQ$	hybrid automatic repeat request
$IFFT$	hybrid automatic repeat request
IoT	internet of things
ITU	International Telecommunication Union
JT	joint transmission
LoS	line of sight
LP	laurent polynomials
LTE	long term evolution
$M2M$	machine-2-machine
MCS	modulation and coding scheme
MF	matched filter
MGS	modified gram-schmidt
$MIMO$	multiple-input multiple-output
$MISO$	multiple-input single-output
$mMIMO$	massive multiple-input multiple-output
$MMSE$	minimum mean square error
MPC	multi path component
MRC	maximal ratio combining
$NLoS$	non-line of sight
NR	new radio
$MU-MIMO$	multi-user multiple-input multiple-output
$OFDM$	orthogonal frequency division multi-point
$ODMOD$	OFDM demodulator
$OMOD$	OFDM modulator
PCC	precoding channel component
PNL	power normalization loss
PRB	physical resource block
PSD	power spectral density
RAT	radio access technology
RCC	relevant channel component

RCM	reverse cuthill-mckee
$RMPC$	relevant multi path component
RS	received signal
$RSRP$	reference signal received power
Rx	received beams at UE
SC	sub-carrier spacing
SDN	software defined network
$SINR$	signal-to-interference plus noise ratio
$SIMO$	single-input multiple-output
$SISO$	single-input single-output
$SPCC$	strong precosing channel component
$SU-MIMO$	single-user multiple-input multiple-output
SVD	singular value decomposition
TDD	time division duplex
TDM	time division multiplexing
TRx	Transceiver
Tx	received beams at the eNodeB
TTI	transmission time interval
UE	user equipment
ULA	uniform linear array
UPA	uniform planar array
$V2V$	vehicle-to-vehicle
$WPCC$	weak precoding channel component
$VWPCC$	very weak precoding channel component
XPR	cross polarization ratio
ZF	zero forcing
$ZFBF$	zero forcing beamforming

Bibliography

[1] NTT-DOCOMO, "5G radio access: requirements, concept and technologies," Tech. Rep., 2014.

[2] Bell-Labs-Consulting, "Going wireless: the next frontier in cable transformation," Tech. Rep., 2018.

[3] 3GPP, "Technical specification group radio access network; study on 3D channel model for LTE Release 12," TR 36.873, Tech. Rep., 2015.

[4] A. M. Ahmadian, R. SivasivaGanesan, and W. Zirwas, "Performance evaluation of linear beamforming receiver for large CoMP sparse massive MIMO channel matrices," in *Proc. of the 85th IEEE International Vehicular Technology Conference*, June 2017.

[5] A. M. Ahmadian, R. SivasivaGanesan, W. Zirwas, and B. Panzner, "Low complexity moore-penrose inverse for large CoMP areas with sparse massive MIMO channel matrices," in *Proc. of the 27th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2016.

[6] S. A. O. Brien. Steve case: We're at a pivotal point in the internet's history. [Online]. Available: https://money.cnn.com/2015/03/14/technology/steve-case-sxsw/index.html

[7] M. Meeker. Internet trends 2015. [Online]. Available: https://www.kleinerperkins.com/perspectives/2015-internet-trends/

[8] CiscoSystems, "Cisco visual networking index: global mobile data traffic forecast," Tech. Rep., 2016.

[9] M. K. Weldon, *The future X network a Bell Labs perspective*, 1st ed. CRC Press, 2016, ch. Enabling a new digital era, pp. 12–37.

[10] Ericsson, "5G systems: enabling industry and society transformation," Tech. Rep., 2015.

[11] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schnellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and

M. Fallgren, "Scenarios for 5G mobile and wireless communictions: The vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

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

[13] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," vol. 9, no. 5, 2005, pp. 1748–1759.

[14] X. Zhang, C. He, L. Jiang, and J. Xu, "Inter-cell interference coordination based on softer frequency reuse in ofdma cellular systems," in *IEEE International Conference on Neural Networks and Signal Processing*, June 2008.

[15] Y. F. Liu, Y. H. Dai, and Z. Q. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," in *IEEE Transactions on Signal Processing*, vol. 59, no. 3, 2011, pp. 1142–1157.

[16] E. Bjornson, E. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," in *IEEE Communications Magazine*, February 2016.

[17] E. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," in *IEEE Communications Magazine*, vol. 52, no. 2, 2014, pp. 186–195.

[18] 3GPP, "Technical specification group radio access network; study on elevation beamforming / full-dimension (FD) multiple input multiple output (MIMO) for Release 13," TR 36.897 R13, Tech. Rep., 2015.

[19] M. R. I. Azad, "Multiple antenna technique (MIMO)," Ph.D. dissertation, Helsinki Metropolia University of Applied Sciences, Helsinki, May 2012.

[20] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*, 1st ed. Cambridge University Press, 2016, ch. Introduction, pp. 15–18.

[21] X. Rao, V. K. Lau, and X. Kong, "CSIT estimation and feedback for FDD multiuser massive MIMO systems," in *Proc. of the IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[22] A. H. Mehana, "Diversity of MIMO linear precoding," vol. 60, no. 2, 2014, pp. 1019–1038.

[23] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system umder imperfect csit," vol. 14, no. 5, 2015, pp. 2532–2547.

[24] J. Litva and T. K.-Y. Lo, *Digital beamforming in wireless communications*, ser. Wiley series in telecommunications. New York: Artech House,Boston, 1996.

[25] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," in *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, 2010, pp. 3590–3600.

[26] F. Rusek, D. Persson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," in *IEEE Signal Prossesing Magazine*, vol. 30, no. 1, 2013, pp. 40–60.

[27] A. Papazafeiropoulos and T. Ratnarajah, "Linear precoding for downlink massive MIMO with deleyed CSIT and channel prediction," in *Proc. of the IEEE Wireless Communications and Networking Conference*, April 2014.

[28] T. Obara, S. Suyana, J. Shen, and Y. Okumura, "Joint fixed beamforming and eigenmode precoding for super high bit rate massive MIMO systems using higher frequency bands," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2014.

[29] W. Zirwas, M. Sternad, and R. Apelfrojd, "Key solutions for a massive MIMO FDD system," in *Proc. of the 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, October 2017.

[30] R. SivasivaGanesan, W. Zirwas, B. Panzner, K. Pederson, and K. Valkealahti, "Integrating 3d channel model and grid of beams for 5G mMIMO system level simulations," in *Proc. of the IEEE International Vehicular Technology Conference*, September 2016.

[31] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: concepts, performance, and field trial results," in *IEEE Communications Magazine*, February 2011.

[32] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: signal detection with enhanced channel estimation," in *IEEE Journal on Selected Areas on Communications*, vol. 52, no. 9, 2002, pp. 1471–1477.

[33] 5GPPP, "Fantastic5G public deliverable D4.2, EU H2020-ICT-2014-2," Tech. Rep., 2017.

[34] H. Taoka, S. Nagata, K. Takeda, Y. Kakishima, X. She, and K. Kusume, "MIMO and CoMP in LTE-Advanced," NTT DOCOMO Communications Laboratories, Tech. Rep. 2, 2014.

[35] S. Bassoy, H. Farooq, and M. A. Imran, "Coordinated multi-point clustering schemes: a survey," in *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, 2017.

[36] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects," TR 36.819 R11, Tech. Rep., September 2013.

[37] W. Zirwas, L. Thiele, M. Kurras, and G. Wunder, "Flexible 5G below 6GHz mobile broadband radio air interface," in *Proc. of the 83rd IEEE Vehiculat Technology Conference*, May 2016.

[38] IST, "D1.1.2- WINNER II channel models," Tech. Rep., 2008.

[39] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," Tech. Rep., 2009.

[40] Ericsson, "R1-133273 height dependent LOS probability for 3D channel model," Tech. Rep., 2013.

[41] Alcatel-Lucent, "Proposals for Fast Fading channel modeling for 3D-UMa," Tech. Rep., October 2013.

[42] 3GPP, "Technical specification group radio access network; study on physical layer procedures," TR 36.213, Tech. Rep., 2010.

[43] X. Gao, O. Edfors, F. Rusek, and F. Tuversson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. of the 74th IEEE International Vehicular Technology Conference*, June 2011.

[44] C. Park, Y. Byun, A. M. Bokiye, and Y. Lee, "Complexity reduced zero-forcing beamforming in massive MIMO systems," in *Information Theory and Applications Workshop (ITA)*, February 2014.

[45] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," in *IEEE Wireless Communications Letters*, vol. 3, no. 6, 2014, pp. 653–656.

[46] K. S. Ahu, "Performance analysis of MIMO-MRC system in the presence of multiple interferers and noise over rayleigh fading channels," *IEEE Transactions on Wireless Coomunications*, vol. 8, pp. 4409–4418, 2009.

[47] Z. Zhou, S. He, Z. Lou, Y. Huang, H. Wang, and L. Lang, "Non-conservative robust joint transmission beamforming for multicell multiuser systems," in *Proc. of the IEEE International Conference on Wireless Communications and Signal Processing*, 2013.

[48] H. Pennanen, A. Tölli, and M. Latva-aho, "Multi-cell beamforming with decentralized coordination in cognitive and cellular networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 295–308, 2014.

[49] R. M. Radaydeh, "Impact of delayed arbitrary transmit antenna selection on the performance of rectangular QAM with receive MRC in fading channels," in *IEEE Wireless Communications Letters*, vol. 13, no. 9, 2009, pp. 390–392.

[50] L. Yang, "Performance analysis of transmit antenna selection with MRC over correlated fast-fading channels," in *Proc. of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, 2008.

[51] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461–471, 2004.

[52] S. Kaviani and W. A. Krzymien, "Optimal multiuser zero-forcing with pre-antenna power constraints for network MIMO coordination," in *EURASIP Journal on Communications and Networking*, vol. 31, no. 2011, 2011, pp. 1–12.

[53] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Transactions on Signal Processing*, vol. 56, pp. 4409–4418, 2008.

[54] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *Proc. of the IEEE Wireless Communications and Networking Conference*, April 2013.

[55] H. Prabhu, O. Edfors, J. Rodrigues, L. Liu, and F. Rusek, "Hardware efficient approximative matrix inversion for linear pre-coding in massive MIMO," in *Proc. of the IEEE Symposium on Circuits and Systems*, 2014.

[56] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on neumann series in massive MIMO systems," in *Proc. of the IEEE International Conference on Communications*, 2015.

[57] P. Courrieu, "Fast computation of moore-penrose inverse matrices," in *Neural Information Processing Letters and Reviews*, vol. 8, no. 2, 2005, pp. 25–29.

[58] T. Greville, "Some applications of pseudoinverse of a matrix," in *SIAM Review*, vol. 2, no. 1, 1960, pp. 15–22.

[59] D. Pappas, A. Petralias, and V. Katsikis, "An improved method for the computation of the moore–penrose inverse matrix," in *Applied Mathematics and Computation*, 2011.

[60] P. S. Stanimirović and M. B. Tasić, "Computing generalized inverses using LU factorization of matrix product," in *Int. J. Comput. Math.*, vol. 85, 2008, pp. 1865–1878.

[61] G. Strang, *Introduction to linear algebra.* Wellesley-Cambridge Press, May 2016.

[62] G. Golub and W. Kahan, "Calculating the singular value and pseudo-inverse of a matrix," in *SIAM Review*, vol. 2, no. 2, 1965.

[63] A. Bjorck, "Solving linear least squares problems by gram-schmidt orthogonalization," in *BIT Review*, vol. 7, 1967, pp. 1–21.

[64] A. Bjorck and C. Paige, "Loss and recapture of orthogonality in the modified gram-schmidt algorithm," in *SIAM Review on Matrix Analytics and Applications*, vol. 1, 1992, pp. 176–190.

[65] C. Zhang, Y. Jing, Y. Huang, and Y. Luxi, "Performance analysis for massive MIMO downlink with low complexity approximate zero-forcing precoding," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 3848 – 3864, September 2018.

[66] H. Crane, N. Gibbs, W. Poole, and P. Stockmeyer, "Algorithm 508: matrix bandwidth and profile reduction," *ACM Transactions on Mathematical Software*, vol. 2, no. 4, pp. 375–377, 1976.

[67] N. Gibbs, W. Poole, and P. Stockmeyer, "An algorithm for reducing the bandwidth and profile of a sparse matrix," *SIAM Journal on Numerical Analysis*, vol. 13, no. 2, pp. 236–250, 1976.

[68] T. Minka. The lightspeed matlab toolbox. [Online]. Available: https://tminka.github.io/software/lightspeed/

[69] D. Cescato and H. Bolcskei, "Algorithms for interpolation-based QR decomposition in MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1719–1733, 2011.

[70] D. Cescato, M. Borgmann, H. Bolcskei, J. Hansen, and A. Burg, "Interpolation-based QR decomposition in MIMO-OFDM systems," in *Proc. of the IEEE International Workshop on Signal Processing amd Advances on Wireless Communications*, 2005.

[71] J. Choi, B. Mondal, and R. W. Health, "Interpolation-based unitary precoding for spatial multiplexing MIMO-OFDM with limited feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4730–4740, 2006.

[72] Z. Li, C. Jeon, and C. Studer, "Approximate gram-matrix interpolation for wideband massive MU-MIMO," 2016.

[73] P. L. Chiu, L. Huang, L. Chai, and Y. Huang, "Interpolation-based QR decomposition and channel estimation processor for MIMO-OFDM system," in *IEEE Transactions On Circuits And Systems*, vol. 58, no. 5, 2011.