Block Diagonalization Based Beamforming

DARSHAN ARVIND PATIL
Block Diagonalization Based Beamforming

Darshan Arvind Patil

Advisor: Prof. Mats Bengtsson
Abstract

With increasing mobile penetration multi-user multi-antenna wireless communication systems are needed. This is to ensure higher per-user data rates along with higher system capacities by exploiting the excess degree of freedom due to additional antennas at the receiver with spatial multiplexing. The rising popularity of "Gigabit-LTE" and "Massive-MIMO" or "FD-MIMO" is an illustration of this demand for high data rates, especially in the forward link.

In this thesis we study the MU-MIMO communication setup and attempt to solve the problem of system sumrate maximization in the downlink data transmission (also known as forward link) under a limited availability of transmit power at the base station.

Contrast to uplink, in the downlink, every user in the system is required to perform interference cancellation due to signals intended to other co-users. As the mobile terminals have strict restrictions on power availability and physical dimensions, processing capabilities are extremely narrow (relative to the base station). Therefore, we study the solutions from literature in which most of the interference cancellation can also be performed by the base station (precoding). While doing so we maximize the sumrate and also consider the restrictions on the total transmit power available at the base station.

In this thesis, we also study and evaluate different conventional linear precoding schemes and how they relate to the optimal structure of the solution which maximize the effective Signal to Noise Ratio (SINR) at every receiver output. We also study one of the suboptimal precoding solutions known as Block-diagonalization (BD) applicable in the case where a receiver has multiple receive antennas and compare their performance.

Finally, we notice that in spite of the promising results in terms of system sumrate performance, they are not deployed in practice. The reason for this is that classic BD schemes are computationally heavy. In this thesis we attempt to reduce the complexity of the BD schemes by exploiting the principle of coherence and using perturbation theory. We make use of OFDM technology and efficient linear algebra methods to update the beamforming weights in a smart way rather than entirely computing them again such that the overall complexity of the BD technique is reduced by at least an order of magnitude.

The results are simulated using the exponential correlation channel model and the LTE 3D spatial channel model which is standardized by 3GPP. The simulated environment consists of single cell MU-MIMO in a standardized urban macro environment with up to 100 transmit antennas at the BS and 2 receive antennas per user. We observe that with the increase in spatial correlations and in high SNR regions, BD outperforms other precoding schemes discussed in this thesis and the developed low complex BD precoding solution can be considered as an alternative in a more general framework with multiple antennas at the receiver.
Sammanfattning


I denna avhandling studerar vi MU-MIMO-kommunikation och försöker lösa problemet att maximera summadatatransfer i nedlänkskommunikation (även kallad framtåfflan), med begränsad tillgänglig sändeffekt hos basstationen.

I nedlänken, till skillnad från upplänken, så måste varje användare hantera interferens från signaler som är avsedda för andra mottagare. Eftersom mobilterminaler är begränsade i storlek och batteristyrka, så har de små möjligheter att utföra sådan signalbehandling (jämfört med basstationen). Därför studerar vi lösningar från litteraturen där det mesta av interferensundertryckningen också kan utföras vid basstationen (förkodning). Detta görs för att maximera summadatatransfer och även ta hänsyn till begränsningar i basstationens totala sändeffekt.

I denna avhandling studerar vi även olika konventionella linjära förkodningsmetoder och utvärderar hur de relaterar till den optimala strukturen hos lösningen som maximiserar signal till brusförhållande (SINR) hos varje mottagare. Vi studerar även en suboptimal förkodningslösning kallad blockdiagonalisering (BD) som är användbar när en mottagare har multipla mottagarantenner, och jämför dess prestanda.


Resultaten simuleras med både en kanalmodell baserad på exponentiell korrelation och med den spatiella LTE 3D-kanalmodellen som är standardiserad av 3GPP. Simuleringssmiljön består av en ensam makrocell i en standardiserad stadsmiljö med MU-MIMO med upp till 100 sändantenner vid basstationen och 2 mottagarantenner per användare. Vi observerar att BD utklassar övriga förkodningsmetoder som diskuteras i avhandlingen när spatiella korrelationen ökar och för höga SNR, och att den föreslagna lägkomplexa BD-förkodaren kan vara ett alternativ i ett mer generellt scenario med multipla antenner hos mottagarna.
Acknowledgements

I would first like to thank my thesis advisor Prof. Mats Bengtsson of the Electrical School at Royal Institute of Technology (KTH), Sweden. Mats consistently allowed this thesis to be my own work but steered me in the right direction whenever he thought I needed it. I feel privileged to have received guidance from one of the pioneers in this field.

I would like to thank Zaheer Ahmed and Ericsson for providing me this opportunity. I extend my sincere thanks to Margaretha Forsgren and Nameen Abeyratne for the continuous encouragement. I would like to thank the experts who helped in the implementation and validation survey for this research project: Beamforming experts Mats Ahlander and Stéphane Tessier, Researchers Vídit Saxena, David Astely and Miguel Berg. Without their passionate participation and input, the validation could not have been successfully conducted.

I would also like to acknowledge Uri Ramdhani of the Electrical School at KTH as the second reader of this thesis, and I am grateful indebted to her for her very valuable comments on this thesis. I am grateful to Fredrik Lindqvist for the detailed review of the report. I feel obliged to Krister Sundberg and team for their patience and for allowing me to work simultaneously on this report with other assignments.

Finally, I must express my very profound gratitude to my parents Arvind and Jayashree Patil, my sisters (Preeti, Arti, and Deepti) and to all my dear friends for providing me with unfailing support and continuous encouragement throughout my duration of study, the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.
## Nomenclature

### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BD</td>
<td>Block-diagonalization</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>BF</td>
<td>Beamforming</td>
</tr>
<tr>
<td>BS</td>
<td>Base-Station</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Estimation</td>
</tr>
<tr>
<td>DoF</td>
<td>Degree of Freedom</td>
</tr>
<tr>
<td>DPC</td>
<td>Dirty Paper Coding</td>
</tr>
<tr>
<td>enodeB</td>
<td>Evolved node B</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplex</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-Symbol Interference</td>
</tr>
<tr>
<td>KSQR</td>
<td>Kernel Stacked QR decomposition</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-of-Sight</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MBB</td>
<td>Mobile Broadband</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input Multiple Output</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple Input Single Output</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Squared Error</td>
</tr>
<tr>
<td>MPC</td>
<td>Multi Path Component</td>
</tr>
<tr>
<td>MRC</td>
<td>Maximum Ratio Combining</td>
</tr>
<tr>
<td>MU-MIMO</td>
<td>Multi-User MIMO</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>QRD</td>
<td>QR Decomposition</td>
</tr>
<tr>
<td>RB</td>
<td>Resource Block</td>
</tr>
<tr>
<td>SDMA</td>
<td>Space Division Multiple Access</td>
</tr>
<tr>
<td>SIC</td>
<td>Successive Interference Cancellation</td>
</tr>
<tr>
<td>SIMO</td>
<td>Single Input Multiple Output</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference and Noise Ratio</td>
</tr>
<tr>
<td>SM</td>
<td>Spatial multiplexing</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise ratio</td>
</tr>
<tr>
<td>SU-MIMO</td>
<td>Single-User MIMO</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplex</td>
</tr>
</tbody>
</table>
BD Based Beamforming

UE User Equipment
ULA Uniform Linear Array
UMa Urban Macro
UPA Uniform Planar Array
ZF Zero-Forcing
ZFBBF Zero-Forcing Beamforming
ZMCSGRV Zero Mean Circular Symmetric Gaussian Random Variable

Notations

$(\cdot)^H$ Conjugate and Transpose (Hermitian) Operator
$\lambda_c$ Carrier Wavelength
$\mathbb{C}^{m \times 1}, \mathbb{C}^{1 \times n}$ Set of complex vectors
$\mathbb{C}^{m \times n}$ Set of complex matrices
$\mathbb{H}$ Householder transformation
$\mathbb{R}, \mathbb{R}^{m \times n}$ Set of real numbers, set of real complex matrices
$A^\dagger$ Pseudo Inverse of matrix
$H$ Channel matrix
$h$ Channel vector
$x^\perp$ Orthogonal Subspace to vector
$\mathcal{CN}(0, \sigma^2)$ Complex Gaussian with Zero mean and Variance equal to $\sigma^2$
$K$ Numerical Kernel of a matrix
$\perp$ Orthogonal
null$(A)$ Nullity of a matrix
rank$(A)$ Rank of a matrix
rank$^\theta(A)$ Numerical rank of a matrix within $\theta$
span$\{v_1, v_2, \ldots, v_n\}$ Span defined by the vectors
$B_c$ Coherence Bandwidth
$f_c$ Carrier Frequency
$T_c$ Coherence Time
t-f time-frequency

Other Symbols

$\mu$s micro Seconds (unit for time)
bits/s/Hz bits per second per hertz (unit for spectral efficiency)
dB decibel-milliwatts (unit for power)
GHz giga hertz (unit for frequency)
KHz kilo hertz (unit for frequency)
km/h kilometer per hour (unit for velocity)
m meters
ms milliseconds (unit for the time)
# Contents

1 Introduction  
1.1 Background ....................................................... 2  
1.2 Motivation ....................................................... 4  
1.3 Thesis Contributions ............................................. 5  
1.4 Thesis Organization ............................................... 6  

2 Theory  
2.1 Wireless Channel ..................................................... 7  
2.1.1 Physical modelling ............................................ 7  
2.1.2 Channel Coherence ............................................ 9  
2.1.3 Statistical Channel Modelling ................................ 12  
2.1.4 Diversity ....................................................... 13  
2.2 MIMO and its modelling ............................................ 14  
2.2.1 Spatial Diversity and Multiplexing using MIMO ............ 14  
2.2.2 MIMO Channel modelling ..................................... 17  
2.3 MU-MIMO and Its Modelling ....................................... 24  
2.3.1 Channel Estimation and Channel Reciprocity ............... 26  
2.3.2 Exponential Correlation Model ............................... 27  
2.3.3 Uplink data transmission ...................................... 28  
2.3.4 Downlink data transmission .................................... 30  

3 5G and Digital BF  
3.1 NR, Massive-MIMO .................................................. 32  
3.2 Beamforming ....................................................... 34  
3.2.1 Precoding/Transmit beamforming .............................. 35  
3.2.2 Optimal MU-MIMO Precoding Structure ..................... 36  
3.2.3 Traditional linear-precoding methods and their relation to optimal structure ........................................... 38  
3.3 Sumrate Maximization .............................................. 41  

4 Sumrate maximization  
4.1 Motivation ......................................................... 42  
4.2 Using Classic BD ................................................... 43  
4.2.1 Description ..................................................... 43
CONTENTS

4.2.2 Algorithm ............................................. 45
4.2.3 Limitations of BD ................................... 46
4.3 Using QRD-BD .......................................... 47
  4.3.1 Iterative precoder design for BD using QR decomposition 47
4.4 Using Proposed BD ....................................... 50
  4.4.1 Idea .................................................. 50
  4.4.2 Alternatives to SVD .................................. 50
  4.4.3 Gains obtained by using proposed scheme .................. 51
  4.4.4 Preliminaries- KSQR based algorithms .................... 51
  4.4.5 Proposed Algorithm ................................ 54
4.5 KSQR-BD in LTE ......................................... 57
4.6 Complexity Analysis ..................................... 58

5 Simulations and Results ................................. 66
  5.1 Setup .................................................. 66
    5.1.1 Theoretical Channel Model .......................... 67
    5.1.2 3D channel model for LTE .......................... 67
  5.2 Results ............................................... 68
    5.2.1 Theoretical channel model .......................... 69
    5.2.2 3D channel model for LTE .......................... 71

6 Conclusion and future work .............................. 89
  6.1 Conclusion ............................................ 89
  6.2 Future Work .......................................... 92

7 Appendix .................................................. 94
  7.1 Complexity Derivations ............................... 94
  7.2 Additional Results .................................... 98
Chapter 1

Introduction

With the advent of 5G in the telecommunication industry, the need for devices capable of processing high data rates and low complexity is inevitable \[1\]. In spite of the fact that this generation of communication technology will be more focused on connecting things, it is also expected to improve existing Mobile Broadband (MBB) services. Technologies like Multiple Input Multiple Output (MIMO) systems and Orthogonal Frequency Division Multiplexing (OFDM) which were introduced in Long Term Evolution (LTE) and advanced 4G solutions will be playing a fundamental role in 5G technologies. As per the recent trends in data consumption by the consumers, video, and music streaming on smart-phones has shown tremendous growth over past few years. It is worth to note that number of users are engaged in both mobility as well as data consumption which emphasizes the need to evolve the existing telecommunication infrastructure to accommodate such type of usage.

Factors in the radio interface such as noise, propagation environment, and interference limit the communication capacity and reliability severely. In order to obtain an effective communication, radio engineers need to carefully design algorithms which can cancel the effects of such factors. This could be performed at various levels either at analog RF (radio frequency) level or through digital signal processing or both. It is not uncommon to use good filtering mechanisms in order to curtail the noise effects. Signal to Noise ratio (SNR) parameter suggests the impact of noise on the transmitted data signal. On the other hand, Signal to Interference and Noise Ratio (SINR) measure provides the impact of both the noise as well as interference on the transmitted signal. In a non-stationary wireless environment where transmitters and receivers are relatively mobile, the propagation environment is crucial to the design of communication system as it impairs the reception drastically \[2\]. Finally, modelling multi-user communication has recently become extremely important, especially for the wideband systems, as the need to utilize the spectrum resources effectively is vital for inclusive communication environment. Apart from these challenges, the power required to transmit the data has always been a point of concern. Therefore, techniques which help reduce the transmission power have significant
importance. Ideally, a designer would want to fix all the above impairments and concerns, although, often it is more of a trade-off problem. Also, fixing these problems might involve the use of complex algorithms using digital processing or high usage of expensive analog radio devices, which is another trade-off problem.

In this thesis, we emphasize on techniques which will realize multi-user systems in order to cater the increasing demand for network densification. We also emphasize on the optimization problem of maximizing the rate of the data transmitted subject to a power constraint.

1.1 Background

In a very fundamental point to point communication system, a multi-path fading environment is regarded as a limitation. The performance of the channel becomes extremely poor when the channel is in a deep fade as compared to Additive White Gaussian Noise (AWGN) channel. Diversity techniques mitigate these limitations by exploiting such a phenomenon where the same information is sent across possible independent faded paths so that the probability of successful transmission is higher \(^3\). Thus, the reliability of the data is increased and the gain is proportional to the number of independent paths, which also corresponds to the *diversity order* of the system. Some of the diversity techniques include time diversity; where interleaving of coded symbols over time is performed, frequency diversity; where signal is spread over a wide spectrum exploiting frequency-selective fading, spatial diversity; using space-time coding in a multiple transmit and/or receive antennas.

Another way of exploiting the multi-path components is to send different information over them to effectively increase the amount of the transmitted information. This concept is known as *multiplexing*. The gain achieved through such a process is known as *multiplexing gain* and is proportional to the number of independent paths available at a given time instance. Extending the principle followed in spatial diversity to the multiplexing scenario, a technique known as *Spatial multiplexing* (SM) was introduced \(^3\). This technique is one of the foundations on which the multi-antenna technologies like MIMO are based. In MIMO, the transmit and receive antenna arrays are placed in a way that several independent streams of information can simultaneously be communicated. However, the performance depends heavily on the richness of scatterers in the environment which allows the receive antennas to separate out the signals from the different transmit antennas. In this way, the multiple antennas effectively increase the number of *degree of freedom* (DoF) in the system to allow spatial separation \(^3\) and multiplexing. Therefore to summarize, spatial diversity in MIMO leads to reliability whereas, spatial multiplexing leads to increased capacity/data rate and thus improved efficiency of the spectrum usage. Furthermore, techniques like interference cancellation can also be adopted in order to improve the capacity. The choice of the technique depends on the specific requirements and other background factors.

In a wireless access network, the base-station (BS, also called enodeB in 4G)
can serve different geographically located User Equipment (UE) without separate time-frequency resources using Space Division Multiple Access (SDMA). This phenomenon can be seen as a natural extension of SM for Multi-User MIMO (MU-MIMO) environment. In this case, the spectral efficiency can be related to the excess degree of freedom harnessed at the BS in order to serve the multitude of users under normal channel conditions [4].

For Single-User MIMO (SU-MIMO) model, traditional receiver solutions like Zero-Forcing (ZF), Minimum Mean Squared Error (MMSE) and Successive Interference Cancellation (SIC) already exists. ZF solution approximately inverses the channel at the receiver, to estimate the transmitted data. This least squares solution, however, results in noise amplification, which is a major disadvantage. The MMSE based solution is a Bayesian approach, which is formulated by minimizing the mean error in between transmitted data and received estimate. SIC can further enhance the performance by successively canceling the streams as they are decoded and can achieve the capacity of fast fading MIMO channel [3]. However, the computational complexity in SIC solution is relatively higher which limits the use of it in large-scale systems.

Another way to model and evaluate capability of MIMO systems is through Singular Value Decomposition (SVD). Aligning the transmit signal at the transmit antenna array, such that the signals from these antennas align in phase at the receiver is called **transmit beamforming**. Whereas at the receiver side, aligning the received signal in the direction of the transmitted signal, i.e., in the direction of the transmitter is called as **receive beamforming**. **Precoding** is a technique in which appropriate multi-stream beamforming weights are applied at the transmitter such that the received signal at the receiver output is maximized. SVD can help to quantify the number of spatial dimensions for a given channel [3]. The complex channel gains are assumed to be known to both the transmitter (via Channel State Estimation (CSI) fed back by receiver) and the receiver. The channel matrix is formed which consists of the channel gains corresponding to the pairwise transmitter and receiver antennas sampled from the array at both the transmitter and the receiver end. SVD is a mathematical step in matrix computations to decompose the Gaussian vector channel into a set of parallel, independent scalar sub-channels. This degree of freedom can be quantified by a minimum number of transmitter or receiver antennas which is also called as spatial dimensions of the MIMO channel. Further, the channel is decoupled using transmit and receive beamforming and the maximum capacity of the system can be easily computed by addition of the capacities from individual sub-channel links. Usually, one of the requirements for an algorithm designer is to provide a solution which would consume less power. In order to maximize the capacity of a decoupled MIMO system under a given power constraint power can be allocated to only good sub-channels. This can be achieved by water-filling power allocation techniques.

Beamforming (BF) technique (which is application of complex weights to the transmit/receive antennas) could be achieved in different ways like Baseband-BF: precoding and combining is performed in the baseband, RF-BF: precoding and combining is performed in RF and Hybrid-BF: precoding and combining is
performed at both baseband and RF [5].

1.2 Motivation

As discussed earlier, the increasing demand for high data rates by users in a mobile environment motivates researchers to constantly improvise the solutions. However, often one has to also consider the computational resources like memory, power, and complexities needed to implement solutions with high performance. Thus, a trade-off is required in order to combat such a scenario.

The advantage of MIMO systems in terms of capacity using SM techniques is substantially high. Employing such a scheme in the multi-user scenario would result in higher throughput. However, this gain in throughput is limited by the interference caused due to the unintended users, and to cancel such an effect complex signal processing techniques need to be employed. In the uplink (reverse link, from UE to BS), the computational burden to employ these techniques lies at the receiver, i.e., the BS. On the contrary, in the downlink (forward link, from BS to UE), employing such a reception scheme in the user terminal would result in the increase in the power consumption, effectively increasing its physical dimension. Thus, alternative methods are considered in the downlink where the BS employs pre-interference cancellation in order to overcome this challenge.

A non-linear technique known as "Dirty Paper Coding" (DPC), by Costa [6] has gained much attention, since use of it in designing precoder has proven to be optimal capacity-achieving [7–9]. The fundamental idea on which this technique is based is, when the transmitter has the knowledge about the inter-user interference in the channel in advance, it can design a code to compensate for it. Furthermore, there is no penalty on transmit power while canceling the inter-user interference as contrary to the linear techniques. However, the unconventional coding gives rise to increased complexity at both the transmitter and receiver [10]. Another algorithm iteratively cancels the inter-user interference, but again the limitation being high computational cost [10].

Alternatively, linear non-iterative BF algorithms have been proposed which are based on block-diagonalization [10]. Although sub-optimal, it is computationally efficient and convenient to implement. However, this scheme is limited by the number of simultaneously supported users which corresponds to the limitation of number of transmit antennas [11]. The solution to this problem is the selection of an optimal subset of users such that capacity is maximized by brute-force search method which is computationally prohibitive [11]. Other methods [11] involve greedy identification of user such that the sum capacity or of the system with current user and already selected users is maximized. This step is performed iteratively until the maximum number of simultaneously supportable users are reached. Even though this algorithm nearly achieves the sum throughput goal for optimal user set, computation of SVD and water-filling process each time requires more computation.

Recently, using QR decomposition another low-complexity user selection algorithm for sum rate maximization is proposed [12]. As the classical BD method
is already a sub-optimal design, the aim for this thesis would be to design an algorithm which meets the performance results of this original scheme with a much lower complexity than even the one proposed in [12].

1.3 Thesis Contributions

The problem of downlink data transmission in a multi-user MIMO scenario is considered in this thesis. Spatial correlations are one of the key factors on which the system performance is depended along with noise level at the receiver. Thus it is important to evaluate the effect of both interference from co-user channels and the receiver noise on the overall system performance. Considering an urban scenario with multiple users with high data rate requirements we attempt to solve a system capacity maximization problem under limited power availability at the base station.

In order to come to a reasonable solution, from the literature, we understand that the optimal technique is highly complex and non-linear which makes it difficult to be implemented in a realistic scenario. Therefore, we first study the optimal linear structure of the solution to the problem from the literature. Further, we note that the techniques to implement this optimal solution are also relatively complex in terms of computational power requirement and implementation in the product. Thus, we study various alternative beamforming solutions including zero-forcing and heuristic solutions using MMSE in this thesis. These solutions are linear and less complex.

However, in the literature, the above solutions are close to optimal when the receiver has a single receive antenna. In this thesis we attempt to find algorithms which are applicable to a more general scenario. We note from the literature about suboptimal beamforming solutions like block-diagonalization which are more general zero-forcing solutions applicable for the scenario when the receiver has multiple antennas. We study the classic block-diagonalization technique and compare it with traditional precoding solutions by simulating it in a multi-user MIMO urban environment.

In this thesis, primarily, we simulate a theoretical exponential correlation channel model to study the effect of spatial correlations on the system sumrate in a multi-user MIMO setup. Next, we implement different precoding techniques including Channel Inversion (complete zero-forcing, ZF) which cancels the co-channel interference, MMSE transmit which cancels interference as well as maximizes SNR and BD which is generalized ZF solution as discussed before.

Finally, we note that even though BD is promising precoding technique, it is computationally complex. Thus, one of the contributions in this thesis is the proposal of a low complex BD solution based on perturbation theory and exploiting the coherence principle. Additionally, we also develop this solution and compare it with the classic BD solutions by simulations using 3D beamforming spatial channel model which is standardized by 3GPP. We observe the effects of change in the dimensions of the MU-MIMO setup (considering potential application of BD precoding to Massive MIMO), SNRs and spatial correlation on
the system sumrate and bit error rates (BER).

1.4 Thesis Organization

In Chapter 2 we summarize the theory of wireless communications including various types of channels and their physical and statistical modelling. We also discuss the key concepts in multi-antenna systems including spatial multiplexing and channel modelling for MIMO systems which are foundations of multi-user MIMO systems. Finally, we discuss the modelling of multi-user MIMO along with how data transmission is performed using advanced signal processing methods at the BS. This is essential introduction to understand the problem that we attempt to solve.

Chapter 3 elaborately explains beamforming concept and how it is implemented at the base station. It also explains the relevant concepts and assumptions used to solve the problem. In the same chapter, we emphasize on learning optimal structure for linear transmit beamforming and how the conventional linear BF solutions relate to it.

In Chapter 4 we elaborately explain the BD technique as a solution to the problem we are attempting to solve. In the same chapter we discuss the classical BD algorithm and alternative method to implement the same with lower complexity. In the concluding sections of the same chapter we propose a novel alternative algorithm to implement the BD technique with much lower complexity than the classical BD. Finally, we compare the complexity of different solutions and show the benefit of using proposed solution along with the allied assumptions.

Chapter 5 in this report presents the numerical results obtained to evaluate different precoding methods which were discussed in the previous chapter. In the same chapter we describe the simulation environment used in order to obtain these results. Further, we also discuss the results obtained and associate it with the theoretical explanation given in the earlier chapters.

Finally, in the concluding Chapter 6 we highlight the important results and mention the key takeaways from this thesis work. At the end of the report we note some comments which can improvise the solution to the problem which we attempted to solve in this thesis and can also be seen as future work for the project.
Chapter 2

Theory

In this chapter we describe the relevant theory from the literature (prominently from [3]) which is required to get better understanding of the problem which we attempt to solve in this project. In Section 2.1 we describe the fundamental concepts in wireless communication including the fading behaviour in a channel and its modelling. We also explain the Diversity concepts which will help us set a background for the introduction to MIMO in Section 2.2 In Section 2.2.2 we advance a little bit further and describe the concepts which characterize the spatial domain in MIMO channels. Finally, we explain the MU-MIMO and its modelling in Section 2.3 and also explain how uplink and downlink data transmissions are performed in MU-MIMO.

2.1 Wireless Channel

2.1.1 Physical modelling

Modelling wireless channels is a key aspect in designing any wireless communication system. For any moving receiver antenna in free space, the electric far field is given by

\[ E_r(f, t, (r_0 + vt, \theta, \psi)) = \alpha(\theta, \psi, f) \cos 2\pi f [(1 - v/c)t - r_0/c] \]

\[ r_0 + vt \] (2.1)

The electric far field is a function of \((\theta, \psi)\), which represents vertical and horizontal angles in the direction of signal arrived at the receiver antenna. Since the receiver is moving with a velocity \(v\), the corresponding distance from the transmitter at any given time instance, i.e., \(r(t)\) is given by \(r_0 + vt\) where \(r_0\) is an initial distance to the transmitter. Note that the transmitted sinusoid at frequency \(f\), i.e., \(\cos 2\pi ft\) has been converted to a sinusoid of frequency \(f(1 - v/c)\); there has been a Doppler shift of \(-fv/c\) due to the motion of the receiver which is a function of \(f\) and \(v, c\) being velocity of light. \(\alpha(\theta, \psi, f)\) denotes the product of radiation patterns of both the transmit as well as receive antennas in the given direction.
From the Equation (2.1) we can see that the far field is inversely proportional to distance \( r \) and thus the power density decreases as \( r^{-2} \) which is known as *Inverse Square Law*. In such a type of free space environment, the attenuation of the transmitted signal is solely on account of the expansion of the signal wavefront, and this phenomenon is known as *free space path loss*. Typically, there are environments in which the distance in between antenna pairs is relatively larger than the heights of individual antennas from the ground. Such a scenario causes part of the signal to get reflected from the earth’s surface and interfere with the primary wavefront which causes the power density to decrease by a *path loss exponent* of 4. Empirical evidence from field experiments suggests that power decay varies by path loss exponent of 2.5 near transmitter to 6 at farther distances [13].

Distinctly from the wired communication system, wireless communication is a linear time-varying system in which the quality of communication is highly influenced by the characteristics of the channel present in between the transmitter and the receiver. Thus, the study of such varying channel characteristics is of prime concern and one of the research areas in the wireless communication system.

*Fading* is a phenomenon in which the signal is attenuated over time or over frequency. Figure 2.1 shows such a phenomenon. *Large scale fading* is a loss of signal caused by shadowing of the mobile user because of large obstacles like buildings and hills. Such fades are typically frequency independent and are a function of distance and size of the shadowing objects. Practical measurements have shown that logarithm of such type of attenuation from several scatterers typically exhibits normal distribution, so it is also known as *log-normal fading*. *Shadowing* is considered to be a greater trouble for system designers compared to the small-scale fading scenario and is resolved by a better cell-site-planning. Even though the primary wavefront is obstructed due to scatterers, fractions of
signal power reach the receiver through diffraction at the edges of the scatterers.

Another important type of fading is small scale fading which occurs due to a constructive or destructive interference of a large number of multiple individual paths (MPC: Multi-Path Component), caused from the reflections of the transmitted signal wavefront from atmospheric scatterers or very rough objects in the environment. Each MPC has a unique attenuation and phase shift. In the urban environment with buildings and trees, there is less possibility of reception of the dominant signal wavefront at the receiver. In fact, the effective received signal is the superposition of the MPC which can be assumed to be generated from uncorrelated scatterers. If the number of in-phase and quadrature components of the received signal with small-scale fading is large enough to satisfy the central limit theorem, the resultant impulse response could be modeled as a Gaussian process and the envelope follows Rayleigh distribution given by density Equation (2.2), irrespective of the distributions of the individual components. \( \sigma_l^2 \) is the variance of the impulse response of \( l \)th tap, i.e., \( h_l[m] \) at time instance \( m \). However, if there is also a presence of dominant line of sight (LOS) signal wavefront, the effective distribution follows the Rician distribution. Additionally, usually the receiver is in motion with respect to the transmitter which impacts fading over time as well as frequency (owing to the Doppler shifts). In the next section we discuss the impact of variation of time and frequency on the resultant signal strength at the receiver.

\[
x \frac{1}{\sigma_l^2} \exp \left\{ -\frac{x^2}{2\sigma_l^2} \right\}, \quad x \geq 0
\]  

(2.2)

To summarize fading, the key difference in between large scale and small scale fading is the variation of the signal strength over distance measure. In small scale this measure is of the order of few carrier wavelengths whereas, in large scale it is relatively larger. Also, small-scale fading is typically frequency and time-dependent. Modelling such a complex type of faded environment and estimating the channel as accurately as possible and more real-like is a key challenge. However, many models have been developed and proposed in an attempt to mitigate this challenge, some of them will be explored in the later section.

2.1.2 Channel Coherence

Frequency Selective Fading

In a small scale fading scenario including a Line-of-sight (LOS) signal, each received \( l \)th MPC has a phase shift with respect to the directly received signal and is given by

\[
\Delta \theta_l = \frac{2\pi f \Delta_l}{c},
\]

(2.3)

where \( \Delta_l \) is a difference in path lengths travelled by the dominant signal and the \( l \)th MPC.
In other words, if there are many MPC’s, the received signal would consist of impulse responses with different time delays ($\tau_l$) and attenuation factors ($\alpha_l$) if a signal $s(t)$ was transmitted, as can be seen in the following equation.

$$y(t) = \sum_l \alpha_l(t)s[t - \tau_l(t)]$$  \hspace{1cm} (2.4)

The impulse response $h(\tau, t)$ for such type of a fading multipath channel can be given by

$$h(\tau, t) = \sum_l \alpha_l(t)\delta[t - \tau_l(t)]$$  \hspace{1cm} (2.5)

where $\delta(t)$ is a narrow impulse signal at the transmitter. Thus, in order to measure this time dispersion we define delay spread ($T_d$) which is the difference in propagation time between the longest and the shortest path, counting only the paths with significant energy. Typically, it is the coherence bandwidth ($B_c$) which is more commonly described as a measure of the impact of multipath fading and is approximately given by

$$B_c = \frac{1}{2T_d}$$  \hspace{1cm} (2.6)

where $T_d$ is the delay. A better way to calculate coherence bandwidth is using RMS (root mean square) delay Spread, as the different channel will experience different signal intensity over different delay span with same delay spread.

When the bandwidth of the input signal ($B_s$) is less than the $B_c$ (i.e., narrow-band), then the channel is said to be non-frequency selective or flat. Whereas, when $B_s$ is much greater than $B_c$ (i.e., broadband) the channel is referred to as frequency selective.

**Time Selective Fading**

As discussed in the earlier section, the motion of the receiver or the scatterers with respect to the transmitter or vice-versa or both causes a Doppler shift in the frequency of the received signal and is proportional to the velocity of the motion and carrier frequency. Frequency dispersion is a phenomenon in which there is a broadening of the frequency spectrum of the received signal because of the superposition of the MPC with independent Doppler shifts. This causes the channel response to vary with time and hence is known as time-selective fading. Coherence time ($T_c$) is defined as the duration of time for which the channel response of the MPC remains constant over such type of a fading environment. If the transmission of the symbol duration ($T_s$) is greater than the coherence time, the channel response is varying and changes significantly, hence the channel is called fast fading. On the contrary, if $T_s$ is very small in relation to $T_c$, then the channel is said to be slowly fading. The relation (see Equation (2.7)) of coherence time with respect to Doppler shift and thus the velocity of the relative motion is important during the design of any mobile communication system.

$$T_c \approx \frac{c}{4f_c v}$$  \hspace{1cm} (2.7)
Coherence Interval

Typically, in urban environments with mobile users and multiple paths caused by scatterers like buildings and trees etc., it becomes extremely crucial for communication system designers to consider fading caused by both time as well as frequency dispersion. If such a multi-fading channel is converted into a time-invariant system, it would be easier for them to design and develop efficient transmission and reception techniques and reduce BER. Coherence Interval can be regarded as a measure which places limits on the duration of time and bandwidth over which the communication channel can be assumed to be constant.

Orthogonal Frequency Division Multiplexing system (OFDM) can be considered to illustrate this phenomenon. In OFDM technology, the time-frequency (t-f) resources are divided in such a way that they can maintain the ortho-
nality in between two consecutive resource elements and avoid interference. A block of t-f resource elements where the channel is assumed to be constant is dependent on the coherence bandwidth and coherence time, which in turn depends on the worst case delay spread, operating frequency and the motion of the user equipment as shown in Equation (2.6) and (2.7).

Figure 2.2 illustrates downlink resource grid in LTE system. In a typical Urban Macro environment (UMa), the coherence interval in LTE system is defined as a Resource Block (RB) pair which consists of 12 sub-carriers representing a bandwidth ($B_c$) of 180 KHz on frequency axis and 14 symbols on time axis. A RB constitutes of 7 OFDM symbols with each symbol representing 71.4 µs on time axis. It could be noted that this numbers can be seen as coherence interval calculated considering a worst case user velocity of 250km/h and operating frequency of 2 GHz.

2.1.3 Statistical Channel Modelling

As discussed in the previous section, we need to determine a way in order to characterize the channel which is valid over a range of conditions like multipath spread and Doppler spread. It is difficult to identify this characterization of the channel by simply experimenting or through user experience. Thus, one methodology to perform characterization is through statistical modelling. Even though this approach is highly oversimplified and far less believable for such a complex t-f spread channel characterization as compared to that of modelling noise, designers get to compare results from different methodologies. One of the reasons probabilistic modelling is not so accurate is the assumption of inter

iden
ticality of the characteristic of different channels as the channels are very different from each other. However, it has been shown that probabilistic modelling can provide good insights and order-of-magnitude guides about wireless systems in order to design and analyze their performance [3].

The aim of modelling a t-f spread channel is to identify the number of filter taps $\{h_l[m]\}$ and the range of intensities in which they vary, through statistical measurements of the channel. The discreet impulse response of the linearly time-varying channel can be represented in terms of $l$ filter taps. Each filter tap consists of an aggregation of multiple paths of the transmitted signal due to reflectors, scatterers etc. At a given time instance $m$, the filter tap can be modeled by

$$h_l[m] = \sum_i a_i(m/W) \exp(-j2\pi f_c \tau_i(m/W)) \text{sinc}[l - \tau_i(m/W)W],$$  \hspace{1cm} (2.8)

where $\tau_i(m/W)$ and $a_i(m/W)$ are the propagation delay and overall attenuation factor for $i^{th}$ path at $m^{th}$ time instance. The overall attenuation factor corresponds to product of attenuation factors due to transmitter and receiver antenna radiation patterns, reflector properties and as a factor that is a function of the distance from the transmitting antenna to the reflector and from reflector to the receive antenna. $W$ is the bandwidth of the input signal, $f_c$ is the carrier frequency and $\text{sinc}[l - \tau_i(m/W)W]$ is the orthogonal basis.
CHAPTER 2. THEORY
2.1. WIRELESS CHANNEL

Our assumption as mentioned earlier is, phase of each multipath component is independent, identical in nature and is uniformly distributed between 0 and $2\pi$. It can be modelled as a circular symmetric random variable considering that the reflectors and scatterers are far away relative to the carrier wavelength. Further, using Central Limit Theorem we can conclude that sum of many such paths can be reasonably modelled as Zero Mean Circular Symmetric Gaussian Random Variable (ZMCSGRV) with variance $\sigma^2_l$, i.e., $CN(0, \sigma^2_l)$, including the phase component $\phi = -j2\pi f_c \tau_i (m/W)$ of the exponential function in Equation (2.8). Finally, the magnitude function $|h_i(m)|$ is defined by Equation (2.2).

However, as discussed earlier, typically there is also a LOS component in addition to MPC in some environments which results to Rician distribution of the channel response given by equation

$$h_i[m] = \sqrt{\frac{K}{K+1}} \sigma_l e^{j\theta} + \sqrt{\frac{1}{K+1}} CN(0, \sigma^2_l),$$

where $\theta$ is phase of the dominant arriving signal at the receiver and $K$ (K-factor) is the ratio of energy in the dominant path to the energy in the scattered paths.

2.1.4 Diversity

The multipath environment at the first place would appear to be a key challenge for the communication. However, it is not. It could also be seen as paths having independent fading gains, making sure that the communication exists as long as there is one strong path. This phenomenon could be further exploited by sending the same information over different paths having independent fading gain such that it is successfully transmitted across at least through one of the strong paths. This technique is known as diversity. Quite likely the channel at one instance of time would not be same as at the other instance; diversity technique could thus be employed in the time domain which is called as time diversity. Coding and interleaving provide time diversity where information is coded and dispersed across the channel at different time instances. Repetition coding and rotation codes are usually followed in order to employ time diversity. At the receiver, a maximal ratio combiner could be employed which weighs the received signals across different time branches in proportion to the signal strength and also aligns the phases of the signals in the summation to maximize the output SNR [3].

Analogously, frequency diversity is a technique in which the variation of channel response (frequency selectivity) can be exploited over a wideband frequency range. Thus, same information signal with bandwidth less than coherence bandwidth can be sent across different smaller flat-faded frequency bands. At the receiver, since the same information symbol is repeated more frequently, there is a possibility of occurrence of Inter-Symbol Interference (ISI). In order to resolve this challenge, either the information sequence can be modified in a way that inter-symbol signal is seen as a pseudo-noise or the wideband frequency
spectrum can be divided into smaller narrow-band frequencies in such a way that the sub-carriers are orthogonal to each other. A popular method which is based on the prior principle is Direct-sequence spread spectrum (DSSS) and OFDM method is based on the later principle (see also Figure 2.2). A detailed description of transmission and detection mechanisms in time and frequency diversity systems are out of the scope of this thesis, however, readers can refer to [3] for more details.

2.2 MIMO and its modelling

2.2.1 Spatial Diversity and Multiplexing using MIMO

The degree of challenge in obtaining frequency and time diversity gain depends on the duration of coherence bandwidth and coherence periods respectively. For instance, when the coherence period is large, the channel response over this duration would be constant and thus diversity cannot exist, in a constrained delay requirement. Antenna diversity also known as Spatial diversity exploits the spatial dimension in order to obtain reliable communication. In this technique antennas are spaced sufficiently far apart such that there are multiple independent fading channel coefficients for the same t-f resource. As mentioned earlier, the spatial diversity phenomenon depends on the richness of the scattering environment as well as the carrier frequency. We discuss this relationship in more detail when we look into channel modelling for MIMO. Antenna diversity can be further classified into transmit diversity and receive diversity. In receive diversity (also known as Single Input Multiple Output or SIMO channels), multiple receive antennas are separated far apart such that the channel response for the individual path is uncorrelated with each other. Considering a flat fading channel with L receive antennas and a single transmit antenna the channel model is given by

\[ y_l[m] = h_l[m]x[m] + w_l[m] \quad l = 1, \ldots, L \]

(2.10)

It can be noted that the received signal \( y_l[m] \) across the L receive antennas, is a linear combination of the same transmitted signal \( x[m] \), however with different channel coefficients \( h_l[m] \) and the independent complex Gaussian noise \( w_l[m] \) at the same time instance. If the receive antennas are spaced sufficiently far apart, the detection principle followed at the receiver is similar to the detection problem in time diversity. Thus, it can also be noted that with the increase in the number of receive antennas the diversity gain increases up to a certain degree, until the fading paths are sufficiently independent to each other. If the number of receive antennas is increased further under the constraint of a limited area or if all the incoming rays arrive from a limited cone of directions, the gain obtained from independent path diminishes, which is one of the limitations of such techniques.

Generally, in the downlink, it is more common to have multiple antennas at the base station side as it would be cheaper to add antennas to a single unit.
of a network than every user terminal. Transmit diversity or *Multiple Input Single Output* (MISO) channels use sufficiently spaced transmit antennas in a timely fashion to send the information symbols to the receiver which is then detected by the receiver. These techniques are also called as space-time codes. *Alamouti scheme* is one such space-time coding techniques in which the number of transmit antennas is set to two. The central idea of Alamouti scheme is to use the additional degree of freedom obtained by having multiple antennas at the transmitter. Instead of sending the same symbols across multiple antennas in two different time instances as in the case of repetition coding, symbols equal to number of transmit antennas are sent every instance. Thus, the utilized energy per symbol is increased by a factor of 2 than repetition coding. Interestingly, the symbols are sent in such a way that the detection problem for different symbols decomposes into separate, orthogonal, deterministic problem. Assuming that the channel remains constant over the two symbol duration and considering flat fading with two transmit antennas and a single receive antenna, received signals at two different time instances are given by

\[
\begin{align*}
\end{align*}
\]

(2.11)

where \( h_1 \) and \( h_2 \) represent channel gain in between transmit antennas 1, 2 and receive antenna respectively. Two separate coded information symbols are sent this time as against earlier diversity schemes, which are \( x_1 \) and \( x_2 \). The trick in this technique is to send the two complex symbols \( u_1 \) and \( u_2 \) over two symbol times, i.e., \( x_1[1] = u_1 \) and \( x_2[1] = u_2 \) at time 1 and \( x_1[2] = -u_2^* \) and \( x_2[2] = u_1^* \) at time 2. Rewriting the equation in a desirable form which is convenient for detection, we obtain

\[
\begin{pmatrix} y[1] \\ y[2]^* \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} w[1] \\ w[2]^* \end{pmatrix}
\]

(2.12)
Multiple Input Multiple Output (MIMO) channels (Figure 2.3) where both transmitter as well as receiver have multiple numbers of antennas. Such a channel can potentially provide higher diversity gain as compared to MISO as long as the fading of the different channel coefficients in between transmit and receive antenna pairs are independent of each other. Thus, the diversity order for full spatially diverse MIMO channel is the product of the number of receive antennas and the number of transmit antennas when the fading of the different channel coefficients are independent. In all the diversity schemes, the error probability is given by

\[ p_e \approx c \cdot SNR^{-L}, \]  

where \( L \) is the number of i.i.d diversity branches and \( c \) is the coding gain for the diversity scheme used.

As mentioned earlier, space-time coding using MISO channel provides a degree of freedom equal to the number of transmit antenna divided by the time units required to send the symbols with an assumption of constant and independent channel gains over number of symbol times. Under the same assumption, space-time codes in MIMO provide additional gain in the degree of freedom as a result of an increase in dimension at receiver caused due to increase in number of receive antennas.
Figure 2.4: Degree of Freedom

Figure 2.4 provides a comparison of signal space spanned by a single receiver antenna in the case of repetition coding enabled in MISO channel against multiple (two in this example) receiver antennas in MIMO channel. Now, if we consider harnessing the entire degree of freedom in MIMO channel, the time dimension of the space-time block coding could be waived off. Therefore, the resulting "space only" scheme provides the degree of freedom equal to the complete degree provided by the MIMO channel. This scheme is known as spatial multiplexing (SM) in which independent information symbols are multiplexed in space over the different antennas as well as over the different symbol times. The obtained multiplexing gain increases further if the independent fading paths have identical coefficients and higher gains. It can also be noted that the additional degree of freedom exploited in MIMO channel improves the efficiency of packing of more information bits resulting in a better coding gain [3].

2.2.2 MIMO Channel modelling

As discussed in the previous section, MIMO channels are very important as they provide an additional degree of freedom which could be used to increase the data rate of the system using spatial multiplexing. We mentioned a key assumption to obtain higher spatial multiplexing gain which is the fading of the different channel coefficients be independent and equal. This depends on the richness of the scatterers present in the environment. However, in a realistic environment, fading conditions do not strictly allow the channels to be independent in spite of sufficient spacing of the transmitter and receiver antennas. Thus, there is a need to model MIMO channels in such an environment to analyze the effects of the inconsistencies caused and study its effects on the spatial multiplexing capabilities of the system. Angular separation essentially has a major role in MIMO modelling. We further attempt to discuss this in detail.

We consider a narrowband model with time-invariant and multipath wireless
In this model user data stream is split into multiple vectors which are then independently transmitted via different transmit antennas of an antenna array, at the transmitter side. At the receiver end, the transmitted data streams are received by an antenna array. The above described model can be represented by

$$x = Hs + n,$$

(2.14)

where $s \in \mathbb{C}^{n_T \times 1}$ is the transmitted signal, $x \in \mathbb{C}^{n_R \times 1}$ is the received signal and $n \sim \mathcal{CN}(0, N_0 I_{n_R})$ is Gaussian noise at a symbol time (for simplicity, the time index is dropped). And the channel can be represented in terms of matrix $H \in \mathbb{C}^{n_R \times n_T}$, where the number of rows and columns of the matrix indicate the number of receive and transmit antennas. Every entry of the matrix $H$ corresponds to equivalent channel gain $h_{ij}$ in between corresponding $i$ receive-$j$ transmit antenna pair. Before proceeding ahead, we note down some key definitions and parameters which will help in order to understand the performance of this channel model.

**Rank**

The rank of a matrix is the number of non-zero singular values ($\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_{\min}^2$). In our case, if the channel matrix $H$ is decomposed into set of unitary rotation matrices (represented as $U$ and $V^*$) and scaling matrix ($\Sigma$) using singular value decomposition, we can obtain the rank of the channel which will also indicate the number of spatial degree of freedom per second per hertz. Due to fading conditions, the transmitted signal is modified by the MIMO channel and the rank provides the measure of this change. As per our earlier assumption, if the MIMO channels are independently fading, then the channel matrix might be full rank and is equal to the minimum number of transmit or receive antennas $n_{\min} = \min(n_R, n_T)$. However, strictness of this assertion depends on the condition of angular separability.

**Condition number**

The ratio in between $\max_i \lambda_i$ and $\min_i \lambda_i$ is defined as condition number [3] of the matrix $H$. This figure suggests the spread in between the channels.

**Antenna separation and Angular separation**

The Antenna Separation $\Delta_t \lambda_c$ or $\Delta_r \lambda_c$ in between two transmit antennas or receive antennas respectively and is proportional to the carrier wavelength $\lambda_c$. $\Delta_t$ and $\Delta_r$ are normalized transmit and receive antenna separations respectively, with respect to the carrier wavelength. Generally, the distance between the transmit antenna and the receive antenna $d$ is much larger than the antenna separation. This is the reason why the paths from the transmit antenna to receive antenna are to a first order parallel. Thus, it is possible to represent the channel gain of any transmit-receive antenna pair with respect to the channel
gain of the first transmit-receive antenna pair using *directional cosine* given by \( \Omega := \cos \phi \), where \( \phi \) is *angle of departure* (AoD) or *angle of arrival* (AoA) with reference to transmit antenna array or receive antenna array respectively. This form of representation is known as *spatial signature*. To illustrate this, for a Rayleigh fading channel (as in Equation (2.8)), in a *Single Input Multiple Output* (SIMO) setting, the channel gain can be represented by,

\[
h = a \exp\left(-\frac{j2\pi d}{\lambda_c}\right) \begin{pmatrix}
1 \\
\exp(-j2\pi \Delta r, \Omega) \\
\exp(-j2\pi 2\Delta r, \Omega) \\
. \\
\exp(-j2\pi (n_R - 1)\Delta r, \Omega)
\end{pmatrix}
\]  
(2.15)

Above equation can also be represented in a simpler form as a unit spatial signature in the directional cosine \( \Omega \), given by

\[
e_r(\Omega) := \frac{1}{\sqrt{n_R}} \begin{pmatrix}
1 \\
\exp(-j2\pi \Delta r, \Omega) \\
\exp(-j2\pi 2\Delta r, \Omega) \\
. \\
\exp(-j2\pi (n_R - 1)\Delta r, \Omega)
\end{pmatrix}
\]  
(2.16)

In Figure 2.5 we illustrate the line of sight reception when there are multiple receive antennas. It can be seen that the signal received by the \( i^{th} \) Rx antenna travels additional small distance of "\( (i-1)\Delta r, \lambda_c \cos \phi \)" as compared to the distance \( d_1 = d \), which the signal received by the first Rx antenna travels. This additional distance is proportional to the antenna separation in between the first receive antenna and the \( i^{th} \) receive antenna, which is the only variable term which affects the channel gains in between the respective transmit-receive antenna pair as compared to the channel gain for the first pair.

It can be noted that there is no gain in the degree of freedom achieved in the above line-of-sight based channel when the antenna separation distance is much lower than the source and destination distance. It just harnesses the power gain by receiving in the direction of the respective directional cosine. The possibility to harness degree of freedom is not possible in this case mainly due to the fact that there is no change in the phase of the received or transmitted signals as there is no change in the angle of arrival or departure.

Further, in order to obtain the degree of freedom (can be also seen as rank), the antenna elements must be placed far away in distance (of the order of the distance between the transmitter and receiver). One condition to obtain distinct channel gains in the case of line of sight channels is related to the separation of the directional cosines obtained from AoA of signal from one antenna to another as given by \( \Omega_r := \Omega_r - \Omega_r \neq 0 \mod \frac{1}{\Delta r} \), where \( \Omega_r \) and \( \Omega_r' \) are two
directional cosines obtained because of two different AoA’s. Also, despite having a good rank, the channel matrix can be ill-conditioned whenever $|\cos \theta| \approx 1$ and well-conditioned otherwise \[3\], where $|\cos \theta|$ is given by

$$
|\cos \theta| = \left| \frac{\sin(\pi L_r \Omega_r)}{\sin(\pi L_r \Omega_r/n_R)} \right| \tag{2.17}
$$

Thus, condition number of channel matrix depends on length of antenna array given by equation

$$
L_r := n_R \Delta_r \tag{2.18}
$$

Angular separation $|\Omega_r|$ at the receive array if larger than the measure of resolvability in angular domain $\frac{1}{L_r}$, then the signals received by the same receive antenna can be resolved and thus gain in the degree of freedom could be harnessed. Similar conditions are valid at the transmitter side respectively in order to exploit the degree of freedom.

All the above conditions can be met even if the antennas are co-located by exploiting the richness of the scatterers.
To illustrate this phenomenon in a MIMO fading environment we take help of Figure 2.6. There are two paths followed by the signal from the same transmit antenna 1 of the transmit array. First is a clear line-of-sight path while second is a reflection of the same signal caused by an object in the environment. Both the direct and the reflected path signals are received by receive antenna 1 of the receive array. $\phi_{t1}$ and $\phi_{t2}$ are AoD’s of the direct and reflected path respectively whereas $\phi_{r1}, \phi_{r2}$ are the corresponding AoA’s. The effective channel is a superposition of channel gains obtained by both the paths and given by

$$H = a_1^b e_r(\Omega_{r1}) e_t(\Omega_{t1})^* + a_2^b e_r(\Omega_{r2}) e_t(\Omega_{t2})^*, \quad (2.19)$$

where it can also be noted from the spatial signature that the signal follows in the direction of both transmit as well as receive antenna array. Transmit spatial signature $e_t(\Omega)$ is given by

$$e_t(\Omega) := \frac{1}{\sqrt{nT}} \begin{pmatrix} 1 \\ \exp(-j2\pi\Delta t\Omega) \\ \exp(-j2\pi2\Delta t\Omega) \\ \vdots \\ \exp(-j2\pi(nT-1)\Delta t\Omega) \end{pmatrix} \quad (2.20)$$

Thus, in order to exploit the maximum degree of freedom in this scenario which is equal to number of effective paths in between the transmitter and receiver, following conditions must be met:

- Unequal directional cosines $\Omega_{t1}$ and $\Omega_{t2}$ for both the transmit signals.
- Unequal directional cosines $\Omega_{r1}$ and $\Omega_{r2}$ for both the receive signals.
• Angular separation $|\Omega_t|$ of the two paths at the transmit array should be of the same order or larger than angular resolvability $\frac{1}{L_t}$ for the transmit array. $L_t$ is given by equation analogous to Equation (2.18).

• Angular separation $|\Omega_r|$ of the two paths at the receive array should be of the same order or larger than angular resolvability $\frac{1}{L_r}$ for the receive array. $L_r$ is given by Equation (2.18).

It can be further noted for this example that the rank of the channel matrix $H$ in Equation (2.19) is equal to 2.

Finally, for the multipath channel model described in Equation (2.14), the channel matrix can be represented in terms of spatial signatures by

$$H = \sum_i a_i \sqrt{n_T n_R} \exp\left(-j\frac{2\pi d^{(i)}}{\lambda_c}\right) e_r(\Omega_r_i) e_t(\Omega_t_i)^*, \quad (2.21)$$

where $e_r(\Omega_r_i), e_t(\Omega_t_i)$ are receive and transmit spatial signatures respectively and $a_i$ is the attenuation factor for $i^{th}$ multipath. $d^{(i)}$ is the distance between transmit antenna 1 and receive antenna 1 along path $i$.

To conclude, it is evident that the degree of freedom can be harnessed with the help of multipath components in a MIMO environment. The quantity (in terms of rank) and quality (in terms of condition number) can be achieved quite easily through the richness of the scattering environment and strategically placing the multiple antennas. Spatial multiplexing can then be accomplished by taking advantage of the degree of freedom gain in order to increase the system capacity.

**Spatial correlation**

In a slow and small scale fading MIMO environment with $n_T$ transmit and $n_R$ receive antennas, the channel matrix $H$ can be split into $n_R \times 1$ dimensional channel vectors $h_k$ of $1 \times n_T$ dimension each as shown in Equation (2.22). Every channel vector represents complex values of channel gains in between $n_T$ transmit antennas and $k^{th}$ receive antenna.

$$H = [h_1, h_2, ..., h_{n_R}]^T,$$

where $$h_k = [h_{1,k}, h_{2,k}, ..., h_{n_T,k}]$$

Ideally, the MIMO channel in a small scale fading environment is *spatially white*, which means that the channel gain for every pair of transmit-receive antenna is modelled as independent and identically distributed (i.i.d) Rayleigh fading. And the channel vector $h_k$ can be modelled as Zero Mean, Circularly Symmetric Gaussian (ZMCSG) random variable. Equation (2.23) is a mathematical representation of uncorrelated $n^{th}$ and $n^{th}$ antenna channel gain out of $n_T$ transmit antenna array.

$$E\{h_{m,k}^* h_{n,k}\} = 0 \quad \text{if} \quad m \neq n,$$

$$h_k = h_w \quad (2.23)$$
As seen in the previous section, significant angular separation of the MPC’s at both the transmit and receive antenna arrays and the antenna separation is crucial to have well-conditionedness of the channel matrix $H$. However, in practical cellular systems, the base station is located at higher altitudes than the mobile handsets which have most of the scatterers and reflectors locally around it. Antenna spacing in the GSM BS is usually of the order of several tens of wavelengths as against half the wavelength in case of mobile terminals to be able to maintain the angular resolvability. However, in the case of MIMO setting used in LTE, much smaller antenna separations might be needed in order to maintain the compactness of the antenna system at the BS, especially in small-cell systems. Use of dual polarization is considered to be beneficial in order to increase the compactness and improve the spectral efficiency due to low fading correlations.

It is highly probable that the channel vectors are correlated due to the limitations in the propagation environment and interdependent antenna radiation pattern (as seen in spatial signatures). This phenomenon is called as spatial correlation and is given by co-variance matrix $R_k$ as in Equation (2.24). Here $(\cdot)^H$ represents conjugate and transpose (Hermitian) operator. The off-diagonal elements of $R_k$ model the correlation between the antenna elements while the diagonal elements model the path-loss. The co-variance matrix varies slowly as compared to small scale fading and can be approximated from the channel estimates [4].

$$h_k = R_k^{1/2} h_w$$

where $R_k = E\{h_k^H h_k\}$

If the spatial correlation is high, the degree of freedom gets affected and also the rank of the channel matrix.
2.3 MU-MIMO and Its Modelling

Multi-user MIMO (MU-MIMO) is a very practical situation in commercial mobile communications. By obtaining more degree of freedom through the spatial domain, the excess utilization of the time and frequency resources is effectively reduced. Thus, the use of the spatial domain is considered to be beneficial to the existing communication technology at the cost of additional radio frequency circuits including antennas at the transceivers. Even though currently it is challenging to fit multiple antennas in a mobile handset (due to its small size and constraint on total power availability required for complex signal processing), from the trend over past few decades and Moore’s law, it can be predicted that the size of electronic circuits might reduce. Therefore, the possibility of the addition of multiple antennas and processors to mobile terminals in near future seems to be high.

On the other hand, adding multiple antennas and complex signal processing capabilities at the base station is much easier due to relaxed constraints on size and power consumption. Therefore, currently, a common setup in cellular wireless systems is a combination of multiple single-antenna mobile handsets and multiple-antenna base stations. However, in this thesis we attempt to solve a problem in MU-MIMO setup considering multiple antennas at the mobile terminal.
Considering a top view of a radio access network as illustrated in Figure 2.7, typically a region with several users is served by multiple base station sites. Each station is further divided into sectors in order to reuse the frequency spectrum and reduce the interference caused due to the allocation of the same carrier frequency to all the base stations. While modelling, often a hexagonal cell structure is considered with three sectors/cell-sites per base station site. Further, every cell-site has its own set of antenna array covering an angle of 120° and serving several indoor and outdoor users. The mobile terminals can have one or multiple antennas.

The structure between a cell-site BS with multiple antenna array and several users each having multiple antennas can be represented as per the block diagram shown in Figure 2.8.

The data communication in MU-MIMO scenario depends on the link direction in which the data has to be transmitted. Accordingly, there are two commonly known link transmissions, forward-link data transmission (Downlink): when the data is transmitted from BS to UE and reverse-link data transmission (uplink): when the data transmission is from UE to BS. The reason for having two separate transmission methods is because of the propagation environment surrounded by the network element. UE is often surrounded by the good amount of scatterers and reflectors, whereas the BS is usually located in an elevated position. Thus, as discussed earlier, there is a corresponding impact on the spatial correlations which leads to an impact on the degree of freedom.

With the use of multiple antennas at the transmitter and receiver there is a need to exploit the spatial dimension effectively. In order to perform this, the MU-MIMO channel needs to be parallelized, and hence the channel is split.
into multiple SU-MIMO channels by applying optimal beamforming weights to the antennas. Depending on the position where the weights are applied, the beamforming types include transmit beamforming and receive beamforming.

Precoding is a transmit beamforming method used when the user has multiple streams, where the multi-antenna transmitter can precode the signal such that the radiated energy from each antenna adds constructively or destructively in desired directions. On the same lines, in receive beamforming, complex optimal weights are applied to the receiver antennas so that the noisy received signal is projected onto the direction of the intended signal and the total signal power is maximized. It is important to note that the knowledge of channel is crucial in order to determine the optimal beamforming weights. We briefly discuss this in the next subsection. However, sharing this knowledge in between the transmitter and receiver might increase the overhead transmission cost.

2.3.1 Channel Estimation and Channel Reciprocity

This is a wide area of a research study in itself, however, we will rather have a brief discussion considering a systemic view of practical systems. Since the central topic of this thesis is applicable in both LTE MU-MIMO environments as well as futuristic massive-MIMO environments, we can base our discussions considering LTE for now.

Channel estimation is a process in which a known sequence of symbols (typically known as pilots or training sequence) is transmitted through the channel and received at the receiver. In LTE, these symbols are called as reference symbols. In this way, the transmission coefficients which are complex valued numbers representing the gain imposed by the channel (channel effect) are evaluated. There is a possibility to estimate the channel both in downlink as well as uplink. In the downlink, BS periodically sends cell-specific reference signals (CRS) used by the user terminals for initial acquisition, channel quality indicator (CQI) measurement and channel estimation for coherent detection. Further, the measured CQI and channel estimates can be fed-back to the BS in order to perform power control, rate adaptation, and scheduling [15]. Since BS can also be seen as a decoder in the uplink, user terminals can also transmit sounding reference symbols (SRS) to the BS which can be used for coherent demodulation in the uplink.

For precise demodulation or precoding, it is a key requirement that the channel estimates must closely resemble the true channel. However, in LTE, it is difficult to obtain such estimates. The reason for the difference in true and estimated channel information is a delay in between transmission of the pilots to reception of the channel estimates (latency) and the interference in between two pilots. However, in practical scenarios in LTE systems, under the assumption of coherence interval, the channel can be estimated with a minor difference as compared to the true channel.

Further, the quality of channel estimation also depends on the modes of the duplexing used for the uplink-downlink data transmissions. In Frequency Division Duplex (FDD) mode, uplink and downlink are separated over frequency
bands which allows simultaneous transmission and reception at the BS (effectively increasing estimation quality). However, due to the separation in frequency bands, the channel fading conditions would be independent of uplink and downlink which results in a setback for CSI estimation. Thus, this comes with an "overhead cost" of obtaining DL link estimates separately. FDD mode is also unsuitable in the case of massive-MIMO as the latency increases with the number of transmit antennas. In the case of TDD, since both the uplink and downlink transmissions are duplex over the time dimension, with some amount of pre-calibrations and under the assumption of coherence interval, the estimates are close to the true channel and can be utilized for precoding or decoding by the BS in MIMO scenario \[16,18\]. This is also known as channel reciprocity.

Channel estimation can be performed using linear low complex solutions like least squares: which minimizes the squared error in between the received pilot sequence and the interference and noise free version of it \[19\], Bayesian based MMSE approach \[19\], in which spatial direction of the intended user terminal is maximized and at the same time attenuating interference with the help of prior knowledge of the channel covariance matrix (which is described in earlier section, and Equation (2.24)) or methods like Normalized Mean Squared Error (NMSE) when size of antenna array is large \[4\]. Under some conditions, MMSE based approach is shown to have eliminated the interference completely \[20\].

In the context of this thesis, please note that we assume the channel estimation to be perfect and available to the BS.

2.3.2 Exponential Correlation Model

In Section 2.2.2 we studied that Spatial Correlations are one of the key factors which affect the performance of the MIMO channels. Furthermore, to gain better insights on the order of impact of it on the system performance, we utilize a simple theoretical model based on \[21\], known as exponential correlation model.

For this model, the components of correlation matrix \( \mathbf{R} \) can be given by

\[
[R(\rho, \theta)]_{ij} = \begin{cases} 
(\rho e^{i\theta})^{j-i}, & i \leq j \\
(\rho e^{-i\theta})^{i-j}, & i > j 
\end{cases}, \quad |\rho| \leq 1 \quad \text{and} \quad \theta \sim U[0, 2\pi). \tag{2.25}
\]

where \( \iota = \sqrt{-1} \), \( \rho \) is the correlation factor between adjacent antennas, \( \theta \) is related to the AoA or AoD and assumed to be distributed uniformly. Maximum value of \( \rho \), i.e., \((-1)\) corresponds to highest correlation level and \( \rho = 0 \) corresponds to no spatial correlations. Another assumption for the modelling in case of MU-MIMO is \( \rho \) is same for all the users while \( \theta \) is different. This model will also be utilized for simulation. Even though this model might not be applicable for certain scenarios, it is based on the basic principle, i.e., the correlation decreases with increasing distance between receive antennas. In highly spatially correlated channel with typical angular spread of \( 10-20^\circ \), the correlation factor is \( \rho \approx 0.9 \). Please refer to \[21\] and \[22\] for further details of the model.
2.3.3 Uplink data transmission

In MU-MIMO uplink, base station with multiple antennas act as the receiver and multiple users with one or several antennas act as transmitters. A simple block diagram representation for MU-MIMO uplink data transmission and reception using linear beamforming is shown in Figure 2.9. With such a setup and for the same t-f resource, received signal at the base station can be discriminated in spatial domain using SDMA: in which the user information is classified based on their individual spatial signatures as seen in Equation (2.20), however now, there would be $K$ such signatures for $K$ users. Such a setup can be mathematically denoted by

$$ y = \sum_{k=1}^{K} H_k x_k + w $$  \hspace{1cm} (2.26)

Here, $H_k \in \mathbb{C}^{n_R \times n_{T_k}}$ is $k^{th}$ user’s channel gain with $n_{T_k}$ dimensions corresponding to number of transmit antennas per user and $n_{R}$ dimensions corresponding to number of receive antennas at the BS. $y$ is the received signal vector corresponding to the $k^{th}$ user transmit signal $x_k$, which is a vector of dimensions $n_{T_k} \times 1$ and $w$ is $\mathcal{CN}(0, N_0 I_{n_R})$ noise. We should also note that there would be additional spatial signatures from other cell-site users creating more interference to the desired user’s information, although, in above equation, we ignore this effect for simplicity. We also assume the channel to be slowly fading.

As the BS has relatively lower constraints on power consumption and complexity than mobile terminal, there is a possibility to employ different reception methods based on different conditions.

At high SNR, interference from other users can be cancelled effectively using technique like zero-forcing (ZF) also known as decorrelator or interference nulling. This technique includes projecting the received signal onto a subspace $V_k$, where $V_k$ is a subspace which is orthogonal to the one spanned by the matrix $H_k$ ($k = 1 : K$, $k \neq k$) and demodulating the intended $k^{th}$ user’s data.
stream by projecting the obtained resultant signal on to the spatial direction of the \( k^{th} \) user. The entire technique can be seen as channel inversion mechanism given by

\[
\hat{x}_k = C_{ZF,k}^H y \quad \text{where} \\
C_{ZF}^H = H^\dagger = (H^*H)^{-1}H^*
\]

(2.27)

where \( H \) is the channel matrix formed by stacking all the \( K \) users channel gain, i.e., \( H \triangleq [H_1, H_2, \ldots, H_K] \). \( H^\dagger \) is pseudo inverse of \( H \) and \( C_{ZF}^H \in \mathbb{C}^{n_T \times n_T} \) is beamforming matrix, where \( n_T = \sum_k n_{Tk} \).

We note that the ZF solution is limited at low SNR as channel inversion also "amplifies" noise and thus is not optimal. Rather, a simple matched filter operation like Maximum Ratio Combining (MRC) without any interference cancellation is well suited for low SNR applications. In MRC, the BS maximizes the received SNR of each \( l^{th} \) stream, ignoring interference caused by other streams. This can be shown by the following equation,

\[
c_{MRC,k,l}^H = \arg \max_{\|c_{k,l}\|^2=1, c_{k,l} \in \mathbb{C}^{n_R \times 1}} \frac{\text{desired signal power}}{\text{noise power}} = \arg \max_{\|c_{k,l}\|^2=1, c_{k,l} \in \mathbb{C}^{n_R \times 1}} p_t |c_{k,l}^H h_{k,l}|^2
\]

(2.28)

where \( p_t \) is transmit power for \( l^{th} \) stream of \( k^{th} \) user. The beamforming weights are given by,

\[
c_{k,l}^H = \frac{h_{k,l}^H}{\|h_{k,l}\|}
\]

(2.29)

Another linear solution is minimum mean squared error (MMSE) solution, which achieves sum rate results of both matched filter at low SNR and ZF at high SNR and is given by

\[
\hat{x}_k = C_{MMSE,k}^H y \quad \text{where} \\
C_{MMSE}^H = PH(H^*H + N_0 I_{n_T})^{-1}
\]

(2.30)

In this equation, \( C_{MMSE}^H \in \mathbb{C}^{n_R \times n_T} \), where \( n_T = \sum_k n_{Tk} \), is the analytic solution, which minimizes the mean squared error, \( P \) is the power allocation matrix, \( N_0 \) is noise power and \( H \) is the channel matrix.

We also note that the channel gain of every user can be single dimensional or multi-dimensional based on the number of antennas they possess and the channel conditionedness. Thus, the maximum number of data streams (also commonly known as "layers" in LTE) on which information can be transmitted equals the number of antennas.

Additionally, there are also non-linear mechanisms which have proven to be optimal (e.g MMSE based Successive Interference Cancellation), however, factors like the degree of complexity and power consumption set limitations on their usage as compared to the linear methods.
2.3.4 Downlink data transmission

Analogous to uplink data transmission, the multi user MIMO downlink data transmission setup could be summarized by

\[ x_k = \sum_{i=1}^{K} H_k s_i + n_k \]  \hspace{1cm} (2.31)

We explicitly state notations here in order to avoid confusion with uplink data transmission and also to maintain consistency throughout the explanation of precoding in the later section. Here, \( n_T \) is the number of transmit antennas whereas \( n_R = \sum_{j=1}^{K} n_{R_j} \) is the total number of receive antennas in the system with \( K \) users. \( x_k \) is the received signal vector with \( n_R \) dimension corresponding to the number of receive antennas of the \( k \)th user, \( H_k \in \mathbb{C}^{n_{R_k} \times n_T} \) is the channel matrix for \( k \)th user seen from BS. \( s_i \) denotes transmitted signal vector for \( i \)th user among \( K \) users having dimension of \( n_T \times n_{R_i} \) and \( n_k \) is Gaussian noise \( \mathcal{CN}(0, N_0 I_{n_{R_k}}) \) at receiver \( k \). Equation (2.31) can be further written as

\[ x_k = H_k s_k + \sum_{i=1,i\neq k}^{K} H_k s_i + n_k, \]  \hspace{1cm} (2.32)

where the first term consists of the desired signal in the direction of user \( k \) and second term represents interference signal caused due to other users from the same cell in the direction of the user \( k \). Please note that we have omitted interference caused due to users from nearby cells for simplicity.

Now, in order to cancel the inter-user interference at the receiver, the user terminal has to employ some kind of filtering mechanism. The simplest approach is to use ZF solution to null out the interfering signals as discussed in the uplink transmission scenario. However, this channel inversion solution can cause noise amplification, especially in low SNR regime, thus not very much useful. MRC linear solution can be used in low SNR. Thus, the use of receiver structure depends on the applications based on SNR. Ideally, the filtering should also maximize the received SNR in the direction of the desired signal along with canceling the interference due to other users. As discussed earlier, if we somehow diagonalize the effective MU-MIMO channel matrix by applying complex beamforming weights, we will be able to get rid of the interference signals. Again, analogous to uplink scenario, another linear option is to use MMSE based beamforming solution which maximizes SINR, given by

\[ \tilde{W}_k = H_k (I_{n_T} + \sum_{i=1}^{K} \frac{P_K}{N_0} H_i^H H_i)^{-1} \]  \hspace{1cm} (2.33)

where \( \tilde{W}_k \in \mathbb{C}^{n_{R_k} \times n_T} \) is the beamforming weight matrix to be applied at the \( K \) receiver.

However, it is also important to take into consideration, the complexity of the solution and transmit power requirements. In a user terminal, both...
transmit power and processing capabilities are limited, which limits the scope to implement complex beamforming solutions. Therefore, alternative techniques are being considered which enforce BS rather than user terminals to take care of the potential interference to the desired received signals. Precoding is one such technique and will be discussed in the next chapter.
Chapter 3

5G, Massive-MIMO and Digital Beamforming

In this chapter we discuss aspects which are or will be part of the emerging 5G technology. In Section 3.1 we briefly familiarize with the terminologies and advances in 5G technology. In the same section we discuss the potential ways to leverage the advancement of the technology by overcoming key challenges. In Section 3.2 we explain the Beamforming concept at length, especially transmit digital beamforming. Understanding of the optimal structure of transmit beamforming presented in this section proves to be a guideline in solving the problem of this thesis work. In the same section we describe three different linear transmit beamforming solutions which are commonly used in the commercial setup. Finally, we conclude this chapter by reformulating the optimization problem for this thesis.

3.1 New radio, Massive-MIMO

In the traditional 4G communication technology use of MU-MIMO along with t-f resources has already made a reasonable impact on the data rate which the users can achieve. However, with the advent of 5G communication technology the overall system data rates are expected to increase by as much as a thousand times. Therefore, it has become more important than ever to explore the possibilities to enhance the existing resource utilization. As compared to LTE release 8, where along with traditional t-f resources, spatial domain is also exploited by having up to 8 antennas at the enodeB (BS in 4G), 5G technology revolves around the buzz words like "New Radio (NR)", "Massive MIMO" and "Beam management". This can be confirmed by looking at the drafts of 3rd Generation Partnership Project (3GPP) specification 38.000 series and most of the industry white papers (for instance, refer [23]).
With the huge bandwidth availability in the range from 30 GHz up to 300 GHz, millimeter-wave (mm-Wave) based communication systems are proposed to be an important part of 5G. Theoretical results published by various research groups including NYU wireless (http://wireless.engineering.nyu.edu/), WNCG (http://www.wnnc.org/), etc., have shown a tremendous increase in the spectral efficiency by using mm-Wave RF in spite of many challenges. Researchers in various industry associations including Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute HHI (e.g., mm-Wave group), Ericsson Research etc. are focusing on developing this technology in a more realistic environment. The use of high-frequency carriers will allow exploiting the unused spectrum of frequencies which provide the possibility to increase the system bandwidth. However, in the higher frequency ranges, use of FDD modes would be limited and instead TDD would be a preferred choice for dense deployments.

On the other hand, in order to boost the system performance further, Multi-antenna Transmission schemes seem to be complementary. Increasing the number of transmit antennas at the BS in the order of tens to several hundred will lead to enhanced t-f resource utilization. Such a setup together with an increased number of user terminals in a cell is called as Massive MIMO \cite{24}. The asymptotic increase in the number of antennas at the BS is highly beneficial in terms of reducing the overhead cost required to obtain the channel estimates in a cell (as discussed in \cite{4}) and to obtain optimal performances with simple linear processing like Maximal Ratio Transmission (MRT) or MRC. This is possible because the channel vectors for the terminals become asymptotically orthogonal \cite{4}. Thus, at the cost of increasing number of antennas at the BS spectacular gains in spectral and radiated energy efficiency is obtained as compared to traditional point-to-point MIMO systems.

However, in practical implementations, the number of antennas at the BS is limited to a finite amount (several hundred) with several tens of user terminals which brings another set of challenges. The challenge is to gain the channel information especially in FDD mode of operation, which is primarily focused in LTE, however intractable when used in massive MIMO environment. Even though TDD mode is a preferred choice, the UL-DL channel reciprocity assumption in it might not hold true in practice. However, with better measurements and calibration methods TDD is popular and looked upon as a key contender in 5G NR. One of the sub-problems in gaining channel information is mitigating the pilot contamination. As we discussed earlier, coherence interval is the fundamental limit of t-f resource to obtain channel estimates. Within these limited resources, both data symbols, as well as pilots, need to be sent. In order to reduce the overhead caused due to the transmission of new pilots, some pilots are reused in between multiple cells. However, this leads to the introduction of interference in the channel estimates of the desired cell and the effect is known as "pilot contamination". This effect impacts the quality of the beamforming weights adversely. Articles \cite{4} and \cite{25} discuss this impact in great detail.

Finally, as discussed earlier in this chapter and various sections of Chapter 2, spatial multiplexing technique using advanced signal processing capabilities is of prime importance.
Phase difference between 2 adjacent beams is \( \Psi = (2\pi d/\lambda) \cdot \sin(\theta) \)

However, user terminals might be incapable to perform such procedures due to their size and power dimensioning constraints. Furthermore, as the cost to add antenna elements at the BS is less (owing to the low-cost linear amplifiers operating at lower transmit powers), the burden of performing efficient signal processing lies on it. Thus, massive MIMO shares the common challenge of developing practical low complex linear filtering solutions, as in the MU-MIMO scenario.

### 3.2 Beamforming

As discussed in the introduction and some of the earlier sections, beamforming (BF) is a technique to steer the radiation power of the signal in the desired direction by changing the amplitude and or phase difference in between adjacent elements of an antenna array. Figure 3.1 shows an antenna array with six isotropic elements spaced at a distance \( d \). With a spacing of \( d = \lambda/2 \) (i.e., half the wavelength) and an equal amount of power phase and gain, the beamforming direction obtained is \( \theta = 0^\circ \), i.e., along the direction of the z-axis. The control over individual element’s amplitude and phase results in the formation of a pattern with constructive and destructive interference in desired and undesired directions respectively.

In analog transmit beamforming, a modulated data stream generated in the baseband unit is passed through a single digital to analog converter (DAC), a transceiver unit followed by separating and passing through a branch of analog phase shifters and amplifiers before aligning to the respective antenna elements of an array. Whereas, in digital beamforming, each modulated data stream is generated, phase shifted and scaled digitally in different branches (corresponding to the number of antenna elements) and then converted into analog by DAC in order to align with the respective antenna elements.
3.2.1 Precoding / Transmit beamforming

In Sections 2.3.3 and 2.3.4, we have seen how linear signal processing techniques are employed in order to perform uplink and downlink data transmission respectively. In the uplink, since the BS is at the receiver side, receive beamforming could be effectively applied in order to get optimal performance. In general, BS can afford to provide better power amplification (but, under the constraint of total transmit power available) as well as computation power as compared to user terminals (in which constraints are caused due to the limitation on physical dimensions). Also, as in the case of massive-MIMO where the number of BS antennas is large or if there are a large number of scatterers causing an increase in multipath components, the user terminal needs to employ more complex receive beamforming techniques which is practically difficult. Thus, somehow if the computational load is shifted to BS so that there is effectively less interference at the desired user, the performance in the downlink could be improved without the need for the user terminal to employ advanced receive beamforming techniques. This is performed by transmit beamforming. Transmit beamforming is also known as precoding in the scenario when there are multiple streams per user.

In the context of digital beamforming, precoding in short is applying complex weight vectors to BS antennas which steer the resultant beam in the direction of intended user such that either the received signal power is maximized or interference is cancelled or both, at the user terminal. The later method can also be seen as maximization of signal-to-noise-plus-interference (SINR). The weights to be applied can be determined with the help of channel knowledge as they provide information about both constructive as well as destructive interference in the form of complex gains for every pair of transmit-receive antennas in between BS and user terminals. We denote such channel information in the form of a matrix as defined in Equation (2.22). Thus precoding process can be mathematically written as,

\[
\hat{x}_k = \sum_{i=1}^{K} H_k M_i d_i + n_k = H_k M_k d_k + \sum_{i=1, i \neq k}^{K} H_k M_i d_i + n_k
\]

(3.1)

where

\[
M = [M_1 \ldots M_k \ldots M_K]
\]

and

\[
d^T = [d_1^T \ldots d_k^T \ldots d_K^T]
\]

where the first term denotes beamforming in the intended direction (desired user \( d_k \)) and the second term denotes interference caused due to undesired users (\( d_i \)) and \( n_k \) is thermal noise at the receiver with Gaussian property, i.e., \( \mathcal{C}N(0, I_{n_k} N_0) \). The dimensions of MU-MIMO setup resembles the one discussed in section of downlink data transmission. In the proceeding subsections,
we provide detailed explanation of several linear precoding algorithms which are used to determine the beamforming weights.

The post-processing operation at the user terminal is given by

\[ \hat{d}_k = D(\hat{x}_k + n_k), \]  

(3.2)

where \( D \in \mathbb{C}^{m \times n_k} \) is the demodulation matrix which converts received estimate signal \( \hat{x}_k \) into estimate of transmitted data signal \( \hat{d}_k \). Here \( m \) is the the number of modulation symbols.

Figure 3.2 shows the block diagram of downlink data transmission using precoding, which is on similar lines as receive beamforming in the case of uplink data transmission in MU-MIMO environment.

### 3.2.2 Optimal MU-MIMO Precoding Structure

Despite the fact that BS is comparatively more capable of handling the power, processing, computational and memory requirements of precoding than user terminal, there are obvious practical limitations. All the above-mentioned resources have an upper limit or threshold, utilization above which might impact the system adversely. Thus, there needs to be some kind of trade-off between maximization of performance and utilization of the resources at the BS (especially power). Therefore, finding an optimal structure for precoding deals with two major issues. On the one hand, there is a problem to maximize the performance of the system with a constraint on total transmitted power at the BS. On the other hand, the power control problem deals with minimizing the total transmitted power while achieving minimal Quality of Service (QoS) requirements for each user in the network [10]. Article [18] was among the very first, which intrigued to formulate such optimal structures. In this report, even though we limit our discussion to the maximization of performance (system throughput or capacity to be specific), it has been shown in [26] that the optimal solution to this problem also solves the later problem of power control.
CHAPTER 3. 5G AND DIGITAL BF

3.2. BEAMFORMING

Zero inter-user interference

(ZFBF) Optimal beamforming

channel direction

\[ \mathbf{H}_k \]

Maximize signal power (MRT)

subspace of co-user channels \( \mathbf{H}_1, \ldots, \mathbf{H}_{k-1}, \mathbf{H}_{k+1}, \ldots, \mathbf{H}_K \)

Figure 3.3: Optimal beamforming- a balance between MRT and ZFBF

We know from our earlier discussions that capacity achieving techniques like Dirty Paper Coding (DPC) perform optimally in all the SNR regions. Ideally, when the noise power is high as compared to signal power, a technique which maximizes the received power over noise is a preferred choice whereas in the high SNR region, the BS can prefer to cancel the interference caused due to unintended users. In mid-SNR regions, an optimal scheme should balance in between the solutions of extreme SNRs.

Geometrically, optimal beamforming can be represented by a vector based on the channel direction which can be rotated to balance between high signal power (MRT solution) and being orthogonal to the co-user channels (zero-forcing solution). Figure 3.3 shows this phenomenon diagrammatically.

Thus, it can be understood that the general optimization problem for a MU-MIMO system including a BS with \( n_T \) transmit antennas and \( K \) users deals with maximization of the performance utility metric, which is generally a function of the signal-to-noise-plus-interference ratio (SINRs) of the \( K \) users and can be formulated as shown in \([27]\) by,

\[
\begin{align*}
\text{maximize} & \quad f(\text{SINR}_1, \ldots, \text{SINR}_K) \\
\text{subject to} & \quad \sum_{k=1}^{K} ||\mathbf{M}_k||^2 \leq P,
\end{align*}
\]

(3.3)

where \( \mathbf{M}_k \in \mathbb{C}^{n_T \times n_{R_k}} \) is the optimal beamforming weight for \( k^{th} \) user with \( n_{R_k} \) receiver antennas, \( P \) is the limit on the total transmitted power at the BS. This problem is non-deterministic polynomial-time hard, especially for utility metric like sum rate \([27]\). However, the structure of the solution (obtained using Lagrangian duality and Karush-Kuhn-Tucker (KKT) conditions) to this problem provides optimal beamforming weights.

The following solution is obtained for an optimization problem similar to Equation (3.3), but considering single antenna per user.

\[
\mathbf{M}_k = \sqrt{p_k} \left( \mathbf{I}_{n_T} + \sum_{i=1}^{K} \frac{\mathbf{h}_i^H \mathbf{h}_i}{N_0} \right)^{-1} \mathbf{h}_k^H \\
\frac{|| \left( \mathbf{I}_{n_T} + \sum_{i=1}^{K} \frac{\mathbf{h}_i^H \mathbf{h}_i}{N_0} \right)^{-1} \mathbf{h}_k^H ||}{|| (\mathbf{I}_{n_T} + \sum_{i=1}^{K} \frac{\mathbf{h}_i^H \mathbf{h}_i}{N_0} )^{-1} \mathbf{h}_k^H ||}
\text{ for } k = 1, \ldots, K
\]

(3.4)
where $\lambda_i \geq 0$ is the Lagrange multiplier associated with $i^{th}$ SINR constraint, also called as a regularization term. $h_k \in \mathbb{C}^{1 \times n_T}$ is channel estimate of the intended user and $\sqrt{p_k}$ is the beamforming power. Note that here $M_k \in \mathbb{C}^{n_T \times 1}$.

In a compact form, Equation (3.4) can be re-written as

$$M = (I_{n_T} + \frac{1}{N_0}H^H\Lambda H)^{-1}H^H P^{\frac{1}{2}},$$  \hspace{1cm} (3.5)$$

where $H = [h_1^T \ldots h_K^T]^T \in \mathbb{C}^{n_R \times n_T}$, $M = [M_1 \ldots M_K] \in \mathbb{C}^{n_T \times n_R}$, power allocation matrix $P$ is,

$$\text{diag}(p_1/(||(I_{n_T} + \frac{1}{N_0}H^H\Lambda H)^{-1})h_1^H||, \ldots, p_K/(||(I_{n_T} + \frac{1}{N_0}H^H\Lambda H)^{-1})h_K^H||),$$

and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ is a diagonal matrix with $\lambda$-parameters, $n_R = \sum_{k=1}^K n_{R_k}$ is the sum of all the receive antennas of all the users in the system and with $(n_T \geq n_R)$.

The resulting solution for problem in Equation (3.3) can be seen as a natural extension to the solution mentioned in Equation (3.4), as, for any fixed receive beamformer, the MIMO channel turns in a MISO channel [18].

### 3.2.3 Traditional linear-precoding methods and their relation to optimal structure

After looking into the optimal structure of the precoder in the prior section, we briefly summarize the existing linear precoding techniques and how they relate to it. Following explanation is considering single antenna users without any assumptions on receive beamforming. Some notations are followed as in Equations (3.4) and (3.5) Please note that as mentioned earlier, non-linear precoding technique using Dirty Paper Coding (DPC) by Costa [6] has been shown to be optimal in terms of sum-rate maximization. However, because of the complexity involved in the implementations and processing capabilities, the scope of this study is being limited to linear techniques.

**MRT**

Maximum ratio transmission (MRT) or matched filtering method of precoding involves the determination of complex BF weights such that the beamforming direction is in the same direction as the desired users channel, i.e., the received signal power of the intended user is maximized. Mathematically, a closed expression can be formulated for this solution as

$$\arg \max_{M_{\text{MRT}}} \{||H^HM_{\text{MRT}}||^2 = \frac{H^H}{||H||}, \}$$

where the $k^{th}$ column of the matrix $M_{\text{MRT}}$ corresponds to beamforming weights for $k^{th}$ user. Intuitively, it is obvious that MRT solution will perform optimally.
only when the number of users is restricted to one \((K = k = 1)\), as otherwise the interference in the received signal due to unintended users is left unaccounted. From Equation (3.5) and (3.6) we note that optimal BF solution is reduced to MRT solution in the scenario where the thermal noise at the receiver is large as compared to the received signal power. In such situations, the system is rather noise-limited than interference and thus it becomes more important to maximize the received signal power.

ZF or Channel Inversion

As shown in Figure 3.3, zero-forcing beamforming solution cancels the interference caused due to unintended co-channel users by directing the beam in the "direction orthogonal to the subspace spanned by all the co-user channels" \(^7\). This is same as creating nulls in the direction of co-user channels and hence this subspace is also known as nullspace of the unintended user. Intuitively, this process is similar to inverting the channel characteristics and thus is also popularly known as "channel inversion" \(^28\). Mathematically, in the matrix form, computing \(k^{th}\) column of the matrix \(M_{ZF}\) corresponds to beamforming weights for \(k^{th}\) user. \(\bar{M}_{ZF}\) corresponding to BF weights for multiple users can be determined by calculating pseudo-inverse given by

\[
M_{ZF} = H^\dagger = (H^H H)^{-1} H^H P^\frac{1}{2}
\]

which is generalized inverse of a matrix where Power allocation matrix \(P\) is given by,

\[
\text{diag}(p_1/||H^H H||^{-1} h_i^H||^2, ..., p_K/||H^H H||^{-1} h_K^H||^2)
\]

In the case when matrix is invertible, pseudo-inverse is same as matrix inverse (i.e., \(H^\dagger = H^{-1}\)). Further, when the BF weights are precoded during the transmission (i.e., mapped to BS antenna elements), the resulting signal at the receive antennas of the intended user is free from inter-user interference. In the matrix form for MU-MIMO setup, the effective received signal for all the users can be denoted as a diagonal matrix (although, practically it is close to diagonal), as off-diagonal elements representing interference are forced to zero \((H^H M_k = 0\) for \(i \neq k\)). Therefore, ZFBF solution can also be seen as channel diagonalization at high SNR.

The optimal structure of BF solution in earlier section confirms the principle of zero forcing asymptotically. In contrast with asymptotic analysis for MRT solution, it can be seen at high SNR (when \(N_0 \to 0\)), \(M\) in Equation (3.3) reduces to ZF solution given by

\[
\tilde{M}_{N_0 \to 0} = (0 \cdot I_n T + H^H \Lambda H)^{-1} H^H P_{N_0 \to 0}
\]

where \(\tilde{P}_{N_0 \to 0}\) denotes correspondingly re-written asymptotic power allocation matrix \(^27\).
MMSE or Regularized channel Inversion

We have seen that the MRT and ZFBF precoding solutions perform well (rather asymptotically optimally) at their respective extremes of SNRs. However, for moderate SNR region, both of them suffer from respective interference-noise limitations. To resolve this issue, one can make use of the optimal structure of the BF solution seen in the earlier section. Instead of prioritizing over either maximizing SNR or cancelling interference, the optimal BF solution maximizes signal-to-noise-plus-interference ratio (SINR). Therefore, one heuristic approach to determine the BF weights (which are close to optimal) could be to select the Lagrange parameters in Equation (3.4) heuristically, as finding the true values of $\lambda$-parameters is generally difficult.

Regularized channel inversion solution is one such heuristic method where a regularization term $I_n$ is added to the ZF solution in Equation (3.8) in order to make the ZF solution more robust for low and moderate SNR regions. This solution corresponds to the problem of maximization of signal power to generated interference. Interestingly, the structure of this solution is same as the optimal structure in Equation (3.4) when the $\lambda$-parameters are set homogeneously as average transmit power ($\lambda = P/K$) [27].

$$M = (I_{nt} + \frac{1}{N_0}HH^H)^{-1}HP^{\frac{1}{2}} \quad (3.9)$$

It can be noted that the same solution has been referred differently in the literature. For example, transmit Wiener filter, signal-to-leakage-and-noise ratio beamforming, transmit MMSE beamforming are popular names of the same solution.

It has been discussed in [18] how the same optimal beamforming solution employed for receive beamforming at receiver would not be optimal when employed for transmit beamforming at the transmitter. The same principle applies to MMSE based beamforming where it is optimal for receive BF but not for transmit BF in general. However, in special cases where the co-user channels are equally strong with superior spatial properties, the optimal structure of BF solution automatically reduces to the MMSE based solution as the $\lambda$-parameters in Equation (3.4) have to be symmetric for maintaining optimality. Thus, in symmetric scenario with equally strong channel having well separated directivity, transmit MMSE beamforming is the optimal solution [27].
3.3 Capacity maximization formulation

The practical communication systems can have various utility metrics mentioned in the optimal beamforming problem formulation. Most commonly, system capacity is regarded as a popular metric since the increase in consumer demand for services like video streaming needs to be addressed. Thus the SINR function in Equation (3.3) can be related to system capacity (also known as system throughput). Therefore, the SINR maximization problem with constraints on total transmit power can also be simply seen as throughput maximization problem given by

\[
\text{maximize } C = \sum_{k=1}^{K} c_k \\
\text{subject to } \sum_{k=1}^{K} ||M_k||^2 \leq P,
\]

where \( C \) is the system capacity or sum rate, which is a function of SINR and can be determined using Shannon information theory framework \([2,3]\). Even though there is no general formula for capacity in fading channels, we stick to the definition given by

\[
C = \sum_{k=1}^{K} c_k \\
= \sum_{k=1}^{K} \log_2(1 + \text{SINR}_k)
\]

which is most commonly used for MU-MIMO setup with Rayleigh fading \([10,27]\). It can be noted that when the interference from the co-users in MU-MIMO is zero, the SINR in Equation (3.11) is simply reduced to SNR.
Chapter 4

Sumrate Maximization using Block-Diagonalization and Complexity Analysis

We begin this chapter by discussing the motivation required to use the Block-
diagonalization based precoding methods. In Section 4.2 we provide a brief
description about classical BD precoding including the algorithm, as discussed in [10]. Furthermore, in the same section, we also state the key limitations of this BD solution. Section 4.3 reviews a low complex BD implementation from the literature [12], which uses QR decomposition instead of SVD. In Section 4.4, we discuss our proposed low complex solution in detail along with its required computational complexity. Section 4.5 exemplifies the use of proposed BD algorithm in LTE. Finally, in Section 4.6, we provide a comparison of the proposed solution with the other precoding schemes and discuss the benefits of using our solution.

4.1 Motivation

We have considered some fundamental BF techniques in the previous section assuming MU-MIMO setup, where the base station has an array of multiple antennas and users have a single antenna. However, it would be interesting to generalize these solutions, for a scenario when users also have an array of multiple antennas. Therefore, in this chapter, we look into a low complex precoding technique which generalizes the zero-forcing solution for users with multiple antenna arrays. We have already discussed that ZFBF solution is only optimal for asymptotically high SNR, so any extension of this method would also be sub-optimal in achieving the channel capacity.

Fundamental Zero-Forcing Beamforming (ZFBF) solution is applicable for single antenna user, where it diagonalizes the channel completely by enforcing
the interference caused due to unintended users to zero. However, this solution might not achieve the same results when the user has multiple antenna arrays since now there is a possibility for the user to coordinate the processing of its own receiver output. Additionally, when there is a significant spatial correlation between the antennas at the receiver, complete diagonalization will cause reduced throughput. Instead, if the BF solution is block-diagonal, (where block form indicates that the signal received from the intra-streams of the same user are not considered as interference and thus are not enforced to zero, while on the other hand, all the inter-user streams are forced to zero which is as desired) there is improvement in spectral efficiency. Additionally, use of power-allocation methods like water-filling [3] can boost the spectral efficiency in limited power availability, at the BS.

The solution to the problem of precoding in MU-MIMO with multiple antennas user has also been approached in several different ways, we highlight two of them. In the first method, an iterative scheme is proposed to cancel the inter-user interference, allowing multiple data subchannels per user as in classical MIMO [29]. Whereas, in the second method, the single-antenna algorithms discussed in the previous section are generalized by including beamforming at the receiver [30].

4.2 Block-diagonalization precoding

4.2.1 Description

Block-diagonalization is a precoding technique in which the ZFBF solution is generalized when the user terminals have multiple antennas [10]. Let us consider a multiuser downlink data transmission with $K$ users and a single BS. The base station has $n_T$ antennas and the $k^{th}$ receiver has $n_{Rk}$ antennas. The total number of antennas at all receivers is defined to be $n_R = \sum_{k=1}^{K} n_{Rk}$. We denote the dimensions of this channel by the notation $\{n_{R1}, \ldots, n_{Rk}, \ldots, n_{RK}\} \times n_T$. The channel from BS to user $k$ is denoted as $H_k \in \mathbb{C}^{n_{Rk} \times n_T}$. The structure is similar to the one discussed in the earlier section of precoding 3.2.1. Thus the received signal at the $k^{th}$ user is represented by

$$\hat{x}_k = \sum_{i=1}^{K} H_k M_i d_i + n_k$$

$$= H_k M_k d_k + H_k \tilde{M}_k \tilde{d}_k + n_k$$

(4.1)

where

$$M_k = [M_1 \ldots M_{j-1} M_{j+1} \ldots M_K]$$

$$\tilde{d}_k = [d_1^T \ldots d_{j-1}^T d_{j+1}^T \ldots d_K^T]$$

which is a more compact form of Equation (3.1), however, here $\tilde{M}_k$ and $\tilde{d}_k$ are combined modulation matrix and transmit vector for all users other than user $k$. 43
We concatenate the channel matrices of all the users and represent them in a block matrix form as $H_S = [H^T_1 \ H^T_2 \ \cdots \ H^T_K]^T \in \mathbb{C}^{n_R \times n_T}$. We can similarly form a block modulation matrix which constitutes modulation matrices for all the users as $M_S = [M_1 \ M_2 \ \cdots \ M_K] \in \mathbb{C}^{n_T \times n_R}$. Therefore, the generalized zero forcing precoding solution, i.e., $H_S M_S \in \mathbb{C}^{n_R \times n_T}$ would be a block-diagonal, where off-diagonal elements for each block representing $H_k M_i = 0$ for $i \neq k$. In terms of capacity maximization with sum power constraint, the modulation vectors which satisfy Equation (4.2), form the resulting precoding block-diagonal solution which is less than or equal to the sum capacity of the system.

$$C_{BD} = \max_{M_S, H_k M_i = 0, i \neq k} \log_2 |I + \frac{1}{N_0} H_S M_S M_S^H H_S^H|$$  \hspace{1cm} (4.2)

Like channel inversion, one of the key assumptions to use BD precoding is that the total number of transmit antennas $n_T$ must be larger than or equal to the total number of receive antennas $n_R$, which comes from the requirements that data can be transmitted to the intended user if the nullspace of all the unintended users has a dimension greater than 0, which limits $\text{rank}(H_k < n_T)$ [10]. Mathematically, it boils down to $\hat{L}_k = \text{rank}(\tilde{H}_k) \leq n_R - n_{R_k} < n_T$, where $\tilde{H}_k = [H^T_1 \ \cdots \ H^T_{k-1} H^T_{k+1} \ \cdots \ H^T_K]^T \in \mathbb{C}^{(n_R - n_{R_k}) \times n_T}$ is the concatenation of channel matrices of all the users other than intended user. Singular Value Decomposition (SVD) [31] can be used to decompose a matrix into matrices representing rotation and scaling. Thus, with the help of SVD, we can easily determine the nullspace for all the unintended users as shown below

$$\tilde{H}_k = \tilde{U}_k \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{V}^{(1)}_k \\ \tilde{V}^{(0)}_k \end{bmatrix}^H$$  \hspace{1cm} (4.3)

Desired nullspace are held by $\tilde{V}^{(0)}_k$ denoting last $(n_T - \hat{L}_k)$ right singular vectors in the form of column vectors, and are the required orthogonal basis for the nullspace of unintended users $\tilde{H}_k$.

In the second step, intended user’s channel is projected on the nullspace in order to have the transmission under the constraint of zero-interference (i.e., $H_k \tilde{V}^{(0)}_k$). However, another condition which needs to be satisfied for the transmission to take place is that at least one sub-channel of the user (represented as a row in $H_k$) must be linearly independent to the subchannels of the unintended users (represented as rows of $\tilde{H}_k$). Mathematically, the rank of the product $H_k \tilde{V}^{(0)}_k$ should be at least one (denoted by $\hat{L}_k \geq 1$). The condition might be difficult to be satisfied in a scenario where co-user channels are highly spatially correlated. Expressions for bounds of this rank can be studied in [10]. Comparing these conditions of dimensions and channel independence to the solution in channel inversion, we can see that they are stricter in the channel inversion case, where all the rows of the intended users should be strictly linearly independent to the rows of $H_k$. In both channel inversion as well as block diagonalization we can conclude that, with the increase in the number of linearly independent rows, the degree of freedom is increased, resulting in more sumrate.
After the inter-user-stream cancellation, the MU-MIMO setup is decomposed into parallel SU-MIMO setup. Thus, in the third step, by finding the vectors which span the subspace in the direction of the intended user (also known as range of a matrix in linear algebra), the information rate maximization problem (as in Equation (4.2)) is solved and the resulting modulation matrix is the BD-BF solution. In matrix algebra, range of a matrix can be determined by an SVD operation. The SVD of the product $H_k \tilde{V}_k^{(0)}$ is given by

$$H_k \tilde{V}_k^{(0)} = U_k \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_k^{(1)} \\ V_k^{(0)} \end{bmatrix}$$

(4.4)

where $\Sigma_k \in \mathbb{R}^{L_k \times L_k}$ is the singular value matrix denoting the scaling of the corresponding right singular vectors $V_k^{(1)}$ (range of the intended user’s channel, under zero-interference constraint), it can also be seen as effective SINR for all the streams of the user $k$ and $\tilde{V}_k^{(0)}$ is same as in Equation (4.3).

Effectively, the product of $V_k^{(1)}$ and $\tilde{V}_k^{(0)}$ forms the final set of modulation vectors which are orthogonal basis of dimension corresponding to the rank of the intended user’s channel under the constraint of zero inter-user interference. It is represented in the matrix form as

$$M_S = \begin{bmatrix} \tilde{V}_1^{(0)} V_1^{(1)} & \tilde{V}_2^{(0)} V_2^{(1)} & \ldots & \tilde{V}_K^{(0)} V_K^{(1)} \end{bmatrix} \Delta^{\frac{1}{2}}$$

(4.5)

$\Delta$ is the optimal power-loading diagonal matrix with elements $\delta_i = \max(\mu - \frac{1}{\Sigma_i}, 0)$ computed by applying water-filling operation [2,3] on $\Sigma = \begin{bmatrix} \Sigma_1 \\ \Sigma_2 \\ \vdots \\ \Sigma_K \end{bmatrix}$, assuming a total power constraint $P$, as shown in Equation (4.6), where $\mu$ is the waterlevel, selected such that

$$\sum_{k=0}^{K} \sum_{i=1}^{L_k} \left[ \mu - \frac{1}{\Sigma_{k,i}} \right]^+ = P,$$

(4.6)

where $x^+$ denotes $\max(x, 0)$.

Finally, the sumrate capacity of BDBF algorithm in Equation (4.2) can be reformulated using the determined modulation matrix $M_S$, as shown below.

$$C_{BD} = \log_2 \left| I + \frac{\Sigma^2 \Delta}{N_0} \right|$$

(4.7)

4.2.2 Algorithm

Alg. 1 summarizes the Block-diagonalization precoding process in the form of a pseudo code.
CHAPTER 4. SUMRATE MAXIMIZATION 4.2. USING CLASSIC BD

Algorithm 1: Block-Diagonalization algorithm using SVD [10]

1: Inputs: \( H_S, \) SNR, \( K = \) number of users
2: Initialize: Empty matrices \( \Sigma \) and \( M_S \)
3: for every \( k^{th} \) user do
4: \( \tilde{H}_k = \tilde{U}_k \tilde{\Sigma}_k \begin{bmatrix} \tilde{V}_k^{(1)} \\ \tilde{V}_k^{(0)} \end{bmatrix}^T \)
5: \( H_k \tilde{V}_k^{(0)} = U_k \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_k^{(1)} \\ V_k^{(0)} \end{bmatrix}^T \)
6: \( M_k = [\tilde{V}_k^{(0)} \tilde{V}_k^{(1)}] \)
7: Set \( \Sigma = \begin{bmatrix} \Sigma_1 & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \end{bmatrix} \)
8: Compute power loading matrix \( \Delta = \text{waterfill}(\Sigma) \)
9: Set final BF weights as \( M_S = [M_1 \ M_2 \ \ldots \ M_k \ \ldots \ M_K] \Delta \)

4.2.3 Limitations of BD

- Even though the classical BD-BF solution presented above provides far better results than conventional zero-forcing method, this comes at the cost of increased implementation and processing complexity. It can be noted that the BD implementation presented in Alg. 1 requires computation of two SVD operations per intended user. One of the SVD operation is performed on a matrix which has huge dimensions (almost equal to the size of the number of users) upper bounded by the number of the transmit antennas at the base station. Therefore, the high computational power requirement limits the usage of this algorithm in practice. Recently, there has been some work to reduce the complexity of the BD implementation [12,32,33]. We look into one of these methods in the proceeding section and also propose a new method.

- An important limitation of the BD method is the applicability of it in low SNR regions, which comes from zero-forcing BF solution. As compared to MMSE based solution, which has comparable performance with the MRT solution at low SNR regions, the sumrate of the BD method suffers due to lack of regularization term, which compensates the problem of amplification of the noise during interference cancellation. However, beamforming solutions in the downlink are usually preferred for high SNR region, where BD performs reasonably well as MMSE based solution. Also, new techniques such as Regularized BD have been proposed recently to cope up with low SNR situations [32,34,36].

- Another limitation of the BD method comes from the dimension requirements as discussed in the previous section. We note that the total number
of receive antennas and hence the size of the users in the MU-MIMO setup, where BD would be applicable, is limited by the number of the transmit antennas (i.e., $n_R = \sum_{k=1}^{K} n_{R_k} \leq n_T$).

This problem can be mitigated by using user selection procedure, however, it becomes quite challenging to select the optimum set of users by searching exhaustive search, such that the sumrate is maximized, especially, in a highly spatially correlated environment. Article [12] surveys this problem in detail along with proposals which perform at par with solutions presented in [11].

4.3 Low complex BD implementations

In this section, we highlight the recent works which have targeted the problem to reduce the complexity of the original BD method, by providing a workaround to compute the first SVD operation in classic BD solution (which is the major cause for complexity increase). Please note that we interchangeably use the terms original BD or classic BD as a nomenclature for BD implementation using SVD proposed by Spencer et.al. in their seminal work presented in the article [10].

In the article [32], the authors have proposed an alternative for the first SVD operation. Instead of obtaining an SVD decomposition of the channel matrix, a combination of operations is performed to determine the nullspace of the unintended users. The implementation includes obtaining orthonormal basis vectors by QR decomposition [31] of the low complex pseudo channel inverse [28] of the intended user channel matrix. These basis vectors denote nullspace of the unintended users which could be confirmed by projecting them on the unintended user channel matrix to obtain zero vectors. In [33], a novel channel extension approach is used in order to reduce the complexity. Article [12] suggests an iterative approach in which the precoding matrix of each user sequentially cancels the interference caused by other users. QR operation (using Givens rotations or Householder transformation) is suggested instead of SVD for low dimension matrix decomposition. We look into this algorithm in more detail as follows.

4.3.1 Iterative precoder design for BD using QR decomposition

We consider the same MU-MIMO setup as discussed in the earlier section. The received signal at the $k^{th}$ user is shown by the Equation (4.1). We follow the same notation for the symbols as in the earlier section. Thus, the BD solution using QR can be represented as BF weights in Equation (4.5). To obtain $V_k^{(1)}$, which is the vector in the direction of an intended user under the constraint of zero inter-user interference, SVD method is used similarly to the original BD solution. The power allocated in this direction maximizes the sumrate under a total transmit power constraint. However, the orthonormal basis vector
which forms the nullspace of the unintended users (i.e., $\hat{V}_k^{(0)}$) are computed in a different way. For time being and convenience, we denote this vector as $B_k$.

Instead of concatenating all the channels of unintended users together and performing SVD, QR decomposition is applied on small matrices iteratively to compute the intersection of the nullspace of all the unintended users. Following example shows this process for four users with $n_{R_k}$ receive antennas and a BS with $n_T$ antennas.

**Example:** showing determination of orthonormal basis of nullspace using iterative QR approach in the case of 4 users

1. **Inputs:**
   \[ H_S = [H_1^T \ H_2^T \ H_3^T \ H_4^T], \quad K = 4 \quad \text{(number of users)} \]

2. **Initialize:**
   \[ B_1^1 = I_{n_T} \]

3. \[ B_2^2 = B_1^1 \times \text{null}(H_1 B_1^1) \] step computes nullspace orthonormal matrix for user 2 by canceling interference caused due to user 1. (Note, interference caused due to users 3 and 4 will be cancelled iteratively at a later stage.)

4. \[ B_3^3 = B_2^2 \times \text{null}(H_2 B_2^2) \] step computes nullspace orthonormal matrix for user 3 by canceling interference caused due to user 1 and 2. (Note use of $B_2^2$ from the previous step.)

5. \[ B_4^4 = B_3^3 \times \text{null}(H_3 B_3^3) \] step computes nullspace orthonormal matrix for user 4 by canceling interference caused due to user 1, 2 and 3.

6. **Update steps:**
   7. **user 1,** \[ B_1^2 = B_1^1 \times \text{null}(H_2 B_1^1) \] cancels interference caused due to user 2.

8. **user 1,** \[ B_2^2 = B_1^1 \times \text{null}(H_3 B_2^2) \] cancels interference caused due to user 2 and 3. (Note use of $B_1^2$ from the previous step.)

9. **user 2,** \[ B_3^3 = B_3^3 \times \text{null}(H_4 B_3^3) \] cancels interference caused due to user 3. (Note step 3 above.)

10. **user 1,** \[ B_1^2 = B_1^1 \times \text{null}(H_4 B_1^1) \] cancels interference caused due to user 2, 3 and 4. (Note step 7 above.)

11. **user 2,** \[ B_2^2 = B_3^3 \times \text{null}(H_4 B_2^2) \] cancels interference caused due to user 1, 3 and 4. (Note step 8 above.)

12. **Update step for user 3,** \[ B_3^3 = B_3^3 \times \text{null}(H_4 B_3^3) \] cancels interference caused due to user 1, 2 and 4. (Note step 4 above.)

It can be observed that, for the $k^{th}$ user, after the $i^{th}$ iteration, $H_j B_k^{(i)} = 0$, where $k \neq (1 < j \leq i)$ and $B_k^{(i)} \in \mathbb{C}^{n_T \times \left(n_T - \sum_{j=1,j \neq k}^{n_{R_k}} n_{R_j}\right)}$. The null($T$) shown in the example signifies the orthonormal basis vectors which span the nullspace of the matrix ($\mathcal{N}(T)$), determined by QR decomposition based on Givens rotation [31] as shown below

\[
T = \begin{bmatrix} L_{m \times m} & 0_{m \times (n-m)} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_{n_{null}} \end{bmatrix}, \quad (4.8)
\]

where $L \in \mathbb{C}^{m \times m}$ is an upper triangular matrix, $Q_1 \in \mathbb{C}^{m \times n}$ contains the
orthonormal basis of the row space of $T$ and $Q_{\text{null}} \in \mathbb{C}^{n \times m}$ forms an orthonormal basis of $\mathcal{N}(T)$, i.e., $\text{null}(T)$. The orthonormal basis should satisfy $TQ_{\text{null}} = 0$ and $Q_{\text{null}}^H Q_{\text{null}} = I$.

The iterative algorithm discussed above could be looked into further detail along with a summarized pseudo code in the article [12].
4.4 Proposed solution

We have seen that the iterative QR decomposition based BD scheme has much lower computational complexity as compared to the original BD due to the use of alternate decomposition to compute the nullspace. In this novel proposal, we attempt to reduce the complexity further by using approximations in the nullspace determination. Eventually, the proposed algorithm reduces the complexity of the entire BD scheme and makes it viable to be utilized in practical situations.

4.4.1 Idea

We exploit the principle of coherence interval (explained in Section 2.1.2): which suggests that the amplitude and phase of the channel is on an average predictable over a small interval on the t-f resource, and perturbation theory [37], which intuitively states that when there is a minor change in the entries of a matrix, the decomposition of the matrix and the rank does not change heavily. These ideas enable us to avoid the entire computation of the matrix decomposition in order to determine, for example, the nullspace of a matrix.

We apply the ideas in our proposed algorithm in two steps. In the first step, for every intended user in the first subcarrier, we recursively compute the intersection of the nullspace of the unintended users using an approximation algorithm known as KSQR algorithm. In the second step, for every intended user in the subsequent subcarriers, we determine the intersection of the nullspace of the set of unintended users using KSQR update algorithm. Note that to determine the nullspace in the second step we do not compute the nullspace again but rather update the already computed nullspace in the first step. This updating is possible because, when the channel is frequency non-selective or flat, there is a very slight variation in the channel response over subsequent subcarriers as compared to the first subcarrier.

4.4.2 Alternatives to SVD

To implement the above scheme, we make use of a low complex linear algebra technique called kernel stacked QR factorization (KSQR) [38], which replaces the SVD operation which was computationally heavy. Further, while implementing BD scheme in an OFDM based system (e.g., in LTE. See Figure 2.2), instead of computing the nullspace of the unintended user channel matrix for every subcarrier, we compute it only for the first and approximate for all the subsequent subcarriers. The approximations are determined by updating the computed nullspace for the first subcarrier. In linear algebra, this process is known as subspace tracking. To update the nullspace, algorithms like rank-revealing ULV Decomposition [37], QR (RRQR) decomposition [39] and KSQR algorithm [38], which are based on the perturbation theory [40], could be employed. As the proposed algorithm uses a KSQR method instead of traditional SVD based method to compute the nullspace for the first subcarrier as discussed above,
we prefer to use the KSQR based updating algorithm to update the nullspace for the subsequent subcarriers. In this thesis, terminology "KSQR" symbolizes that the nullspace is computed using kernel stacked QR decomposition algorithm whereas "KSQR update" symbolizes that the nullspace is updated using kernel stacked QR decomposition update algorithm.

4.4.3 Gains obtained by using proposed scheme

For the same sumrate performance, when the proposed method is used, the complexity of BD reduces by at least an order and further keeps reducing exponentially with the increase of transmit antennas and number of subcarriers up until the condition of the channel being frequency flat is maintained.

4.4.4 Preliminaries- KSQR based algorithms

In order to understand the use of KSQR based methods in the proposed algorithm and for completeness, we briefly describe them in this subsection, however, article [38] can be referred for detailed explanation.

KSQR

KSQR algorithm is an alternate way to determine the singular values $\sigma_n \leq \sigma_{n-1} \ldots \leq \sigma_k$ along with their associated right singular vectors $w_n, w_{n-1}, \ldots, w_k$ which form the orthonormal basis for the approximate nullspace of the overdetermined matrix $T \in \mathbb{C}^{m \times n}, m \geq n$. Here $\sigma_n = \sigma_{\text{min}}$ is the smallest singular value of the matrix $T$ and $\sigma_{r+1}$ denotes a singular value which is lower than a prescribed approximation threshold $\theta > 0$. $\theta$ ensures the determination of numerical rank $r$, $w_k$ is the associated singular vector for $\sigma_k$ and $k = n - r$ where $k$ is numerical nullity of the matrix $T$.

The problem of finding

$$\sigma_{\text{min}} = \min_{||x||_2=1} ||Tx||_2$$

is converted to solving the overdetermined system

$$\begin{bmatrix} \tau x^H \\ T \end{bmatrix} x = \begin{bmatrix} \tau \\ 0 \end{bmatrix} \quad \text{where scaling factor } \tau > \sigma_n$$

for its least squares solution $x$, using Gauss-Newton iteration given by

$$x_{j+1} = x_j - \left[ 2\tau x_j^H T \right] ^{\dagger} \left[ \tau x_j^H x_j - \tau \right] T x_j$$

$$\varsigma_{j+1} = \frac{||T x_{j+1}||_2}{||x_{j+1}||_2} \quad j = 0, 1, \ldots$$

or computationally low complex inverse iterations. In the article, it is suggested to begin with a random vector $x_j$ with its associated singular value $\varsigma_j$ and
continue applying the Gauss-Newton iteration as given in Equation (4.11) until a pair of lowest singular value and its associated singular vector \((\sigma_n, w_n)\) is obtained for a matrix \(T\). Furthermore, in order to obtain the subsequent singular values \(\sigma_{n-1} \leq \ldots \leq \sigma_k\) and their associated vectors, i.e., \(w_{n-1}...w_k\), following deflation procedure is suggested.

A kernel stacked matrix is formed given by,

\[
T_\varrho = \begin{bmatrix} \varrho w_n^H \\ T \end{bmatrix}
\]

where \(\varrho \in \mathbb{R}\)

Now, by applying Equation (4.11) to this matrix, a new set of singular values and their associated singular vectors are obtained. It can be noted that the smallest singular value of this kernel stacked matrix \(\sqrt{\varrho^2 + \sigma_n^2}\) replaces the smallest singular value \(\sigma_n\) of the matrix \(T\), however, maintaining rest of the singular values same as that of \(T\). Now, by assigning \(\varrho = ||T||_F\), the smallest singular value \(\sqrt{\varrho^2 + \sigma_n^2}\) becomes the largest singular value of \(T_\varrho\) and a new smallest singular value is obtained which is same as \(\sigma_{n-1}\) of \(T\). Same approach could be followed to obtain the subsequent singular values \(\sigma_{n-1} \leq \ldots \leq \sigma_k\) and their respective singular vectors \(w_{n-1}...w_k\), until they cross the threshold \(\theta\). The computed \(w_k,...,w_n\) form an orthonormal basis vectors for the approximate nullspace of \(T\) denoted by \(W = [w_k \ldots w_n]\).

Further, it is also shown in the article [38] that performing the iterations in Equation (4.11) to find singular vector \(w_n\) is equivalent to performing inverse iterations on \(T^H T\). Considering \(T = Q \begin{bmatrix} R \\ 0 \end{bmatrix}\) and since \(T^H T = R^H R\) (and both \(T\) and \(R\) share the same numerical kernel \(\mathcal{K}_\varrho(T)\)), \(R\) can be used instead of \(T\) in Equation (4.11) to find the singular vectors.

The above algorithm is discussed step wise as following. In the first step, by taking QR factorization (using Givens rotation or Householder transformation) [31] of the matrix \(T\) followed by using inverse iteration [31] operation on \(R^H R\) to find a unit vector \(w_n \in \mathcal{K}_{\varrho}(T)\). In the second step, to obtain a second unit vector \(w_{n-1} \in \mathcal{K}_{\varrho}(T)\), a matrix \(\tilde{T} \in \mathbb{C}^{(m+1) \times n}\) is formed by stacking a row \((\tau w_n^H)\) on top of the matrix \(T\), where \(\tau = \sqrt{n} \times ||T||_\infty\). By performing this operation, the dimension of the numerical kernel of the new matrix \(\tilde{T}\) decreases by one. Note that \(w_{n-1}\) vector also belongs to \(\mathcal{K}_{\varrho}(\tilde{T})\). Further the QR factorization of the new matrix is updated as \(\tilde{T} = \tilde{Q} \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix}\) and followed by inverse iteration operation on \(\tilde{R}^H \tilde{R}\) to find a unit vector \(w_{n-1} \perp w_n\). The process is continued until the nullity of the numerical kernel tends to zero and we obtain all the orthonormal vectors of matrix \(W\). Pseudo code for KSQR algorithm can be found in [38].

To summarize, the input to this algorithm is a matrix \(T\) and a prescribed approximation threshold \(\theta\). The output of this algorithm is the corresponding approximate nullspace of the input matrix, denoted by \(W\). It is important to note that the choice of \(\theta\) is application specific. We motivate its choice in our application in the next subsection when we discuss the proposed algorithms.
in detail. Further, the above procedure can be summarized in the form of a function call which we have used for the KSQR algorithm in the pseudo-code of our proposed algorithms.

\[ W = \text{KSQR}(T, \theta) \]

Note that the nomenclature of input and output variables in the function call is different than mentioned here, however, all the notations are explained when we discuss the pseudo-codes.

**KSQR update**

We make use of a trick from [38] to update the already computed nullspace \( W \) (see previous item in this subsection) of a matrix \( T \) when a new row \( b^H \) is inserted into the matrix \( T \). Mathematically, given a matrix \( W \) which corresponds to the nullspace of \( T \), a new row vector \( b^H \) and rank approximation threshold \( \theta \), the algorithm finds a corresponding \( \hat{W} \) which represents the updated nullspace matrix for the row updated matrix \( \hat{T} \). A quick note, we do not use the complete KSQR update algorithm in [38] as we do not need to obtain the rank of the matrix neither the QR factorization. Instead, we only use the byproduct of the algorithm (i.e., Kernel matrix).

As per [38], the approximate rank \( r \) of a matrix \( \hat{T} \) will remain the same as of matrix \( T \), even after the insertion of a row \( b^H \) into it, unless the approximate rank of matrix \( T \) is less than \( n \) where \( n \) represents number of columns in matrix \( T \). Further, approximate nullspace \( \hat{W} \) of matrix \( \hat{T} \) is a subset of the approximate nullspace \( W \) of original \( T \) and they are equal if \( b \) is approximately orthogonal to \( W \). Mathematically, if \(||W^Hb||_2 \leq \theta \) then \( W = \hat{W} \). It can also be said that when this condition is met, the "approximate rank of the resulting matrix remains constant".

However, when \(||W^Hb||_2 > \theta \), the approximate rank of the new matrix \( \hat{T} \) becomes \( r + 1 \), and we name this scenario as "variable rank condition". In order to determine the orthonormal basis vectors of the matrix \( \hat{W} \) Householder transformation \( H \in \mathbb{C}^{k \times k} \) is performed on \( y = W^Hb \), i.e., \( H(y) = ||y||_2, 0, ..., 0 \)\(^T \). We obtain \( H = [y_k, ..., y_n] \) and \( \{y\}^\perp = \text{span}\{y_{k-1}, ..., y_n\} \). For \( E = [y_{k-1}, ..., y_n] \), let \( WE = [\hat{w}_{k-1}, ..., \hat{w}_n] \in \mathbb{C}^{n \times (k-1)} \). Thus, it can be shown that the columns \( \{\hat{w}_{k-1}, ..., \hat{w}_n\} \) form orthonormal basis for \( \hat{W} \) which is the desired updated approximate nullspace.

We summarize the two cases which we discussed above having an effect on the rank of the new matrix as follows

\[
\text{If}, \quad ||W^Hb||_2 \leq \theta \quad \text{then no effect on the approxi-rank} \\
\text{If}, \quad ||W^Hb||_2 > \theta \quad \text{then effect on approxi-rank} \quad (4.12)
\]

Finally, the Pseudo code for KSQR update algorithm used in our application is shown in Alg. 3.
Algorithm 3: KSQR update algorithm

\[ \hat{W} = \text{KSQRUpdate}\left(b^H, W, \theta \right) \]

1: **Inputs:**
- \( W \) obtained from KSQR algorithm, new row to be inserted
- \( b^H \), prescribed approx threshold obtained from KSQR algorithm

2: **Outputs:**
- Updated approximate nullspace \( \hat{W} \)

3: if \( \|W^H b\|_2 \leq \theta \) then (for constant rank)
4: Output \( \hat{W} = W \)
5: else (for variable rank)
6: Construct the Householder transformation \( H \) such that \( H(W^H b) = (\ast, 0, ..., 0)^H \)
7: Obtain \( \hat{W} = W E \), where \( E \) is computed as mentioned in Section 4.4.4

4.4.5 Proposed Algorithm

Assuming a low Doppler channel model with relatively flat or frequency non-selective fading in MU-MIMO setup with constant user set \( S = \{1, 2, ..., K\} \), in this section, we propose low complex algorithms to compute the BD BF weights. In conjunction to this assumption, we expect the channel estimates represented in the form of a matrix obtained for every subcarrier in a given coherence bandwidth, to vary (or perturb) in small amount. The effect of this variation in the channel estimates over the subsequent subcarriers does not much affect the nullspace and rank of the system. From numerical linear algebra, when \( \text{rank}_2(H_{1,k}) = r \), we call the nullspace of \( H_{l,k} \) the approximate nullspace of \( H_{1,k} \) within 2-norm distance \( \theta \) (i.e., mathematically, \( \|H_{1,k} - H_{l,k}\|_2 \leq \theta \)), since \( H_{l,k} \) is the nearest matrix to \( H_{1,k} \) with rank \( r \). Here, \( H_{1,k} \) denote the channel matrix of unintended users for \( k = 1 : K \) intended user from set \( S \) in 1st subcarrier and \( H_{l,k} \) denote the channel matrix of unintended users for the same \( k = 1 : K \) user from set \( S \) in \( l^{th} \) subcarrier where \( l = 2 : N_{sc} \) and \( N_{sc} \) is the total number of subcarriers for a given t-f resource block.

Using this definition, we propose two algorithms in order to determine the nullspace and thus BD beamforming weights for all the users in set \( S \) and subcarriers iteratively. In Alg. 4, nullspace and BF weights are computed for the first subcarrier. In Alg. 5 we reuse the nullspace computed from Alg. 4 and update it for the subsequent subcarriers using KSQR update as shown in Alg. 3.

Following are the general notations used in the algorithms. \( H_{SC_1} \) and \( H_{SC_2} \) are the channel estimates for all the \( K \) active users from set \( S \) in the first subcarrier and subsequent (i.e., \( l = 2 : N_{sc} \)) subcarriers respectively, where \( N_{sc} \) is the total number of the subcarriers for a given t-f resource block. \( \Sigma_{SC_1} \) and \( \Sigma_{SC_2} \) having the same dimensions as in the classic BD method, hold the effective SINR values which are utilized to compute the optimal power loading.
matrix $\Delta_{SC_1}$, $\Delta_{SC_l}$ and system sumrate $C_{BD_{SC_1}}, C_{BD_{SC_l}}$ given by Equation (4.7). Similarly, matrices $M_{SC_1}$ and $M_{SC_l}$ hold the final beamforming weights for respective subcarriers.

Algorithm 4: Proposed precoder design for BD - first subcarrier in a resource block

1: Inputs: $H_{SC_1}$, SNR and $K$
2: Initialize: Empty matrices $\Sigma_{SC_1}$, $M_{SC_1}$ and $W_{SC_1}$
3: Outputs: $\theta_k$, $M_{SC_1}$ and $W_{SC_1}$
4: for user $k = 1 : K$ from set $S$ do
5: Form matrix of unintended users $\tilde{H}_{1,k}$ from the block matrix $H_{SC_1}$.
6: $\hat{H}_{1,k} = H_{1,k}^H \tilde{H}_{1,k}^H$
7: $\theta_k = \text{Smallest singular value of } (\hat{H}_{1,k})$
8: $\tilde{V}^{(0)}_{1,k} = \text{KSQR}(\hat{H}_{1,k}, \theta_k)$
9: $H_{1,k} \tilde{V}^{(0)}_{1,k} = U_{1,k} \begin{bmatrix} \Sigma_{1,k} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{V}^{(1)}_{1,k} \\ V_{1,k}^{(0)} \end{bmatrix}^T$
10: $M_{1,k} = [\tilde{V}^{(0)}_{1,k} \tilde{V}^{(1)}_{1,k}]
11: Set $\Sigma_{SC_1} = \begin{bmatrix} \Sigma_{11} & \cdots \\ \vdots & \Sigma_{1,k} \end{bmatrix}$
12: Compute power loading matrix $\Delta_{1} = \text{waterfill} (\Sigma_{SC_1}, \text{SNR})$
13: Set final BF weights as $M_{SC_1} = [\tilde{V}^{(0)}_{11} \tilde{V}^{(1)}_{11} \tilde{V}^{(0)}_{12} \tilde{V}^{(1)}_{12} \cdots \tilde{V}^{(0)}_{1,k} \tilde{V}^{(1)}_{1,k} \Delta_{1}^{-1/2}]
14: $W_{SC_1} = [\tilde{V}^{(0)}_{11} \tilde{V}^{(0)}_{12} \cdots \tilde{V}^{(0)}_{1,k} \cdots \tilde{V}^{(0)}_{1,k}]$

We discuss the steps in the algorithms as follows:

- Pseudo code for Alg. 4 showcases the determination of BF weights based on BD precoding for the first subcarrier. In this pseudo code, we compute the nullspace $W_{SC_1}$ for $\tilde{H}_{1,k}$ for all $k = 1 : K$ users from set $S$ by using KSQR algorithm, as discussed in Section 4.4.4.

  - In our proposed Alg. 4, Step 8 corresponds to nullspace computation. The input to the KSQR algorithm is the matrix $\tilde{H}_{1,k}$ which corresponds to matrix $T$ and $\theta_k$ corresponds to $\theta$ and the output $\tilde{V}^0_{1,k}$ corresponds to $W$ in Section 4.4.4.

  - As the KSQR algorithm is formulated for overdetermined systems, we have performed an additional operation in Step 6 of Alg. 4 to maintain this compatibility.

  - Several different values were tried for $\theta_k$ and the chosen value in Step 7 provided good subspace tracking as compared to SVD. Please note
Algorithm 5: Proposed precoder design for BD- subsequent subcarriers of a resource block

1: **Inputs:**
   \( H_{SC_l} \), SNR, \( K \), \( \theta_k \) and \( W_{SC_1} \), where \( l = 2 : N_{sc} \)

2: **Initialize:**
   Empty matrices \( \Sigma_{SC_l} \) and \( M_{SC_l} \), where \( l = 2 : N_{sc} \)

3: **for** subcarrier \( l = 2 : N_{sc} \) **do**
   
4:   **for** user \( k = 1 : K \) from set \( \mathcal{S} \) **do**
   
5:     Form matrix of unintended users \( \tilde{H}_{l,k} \) from the block matrix \( H_{SC_l} \).

6:     **for** 1st row of \( \tilde{H}_{l,k} \) **do**
   
7:       \( \tilde{V}_{l,k,1}^{(0)} = \text{KSQRUpdate}(\tilde{H}_{l,k}(1,:), \tilde{V}_{l,k,1}^{(0)}, \theta_k) \)

8:     **for** row \( i = 2 : \text{last} \) of \( \tilde{H}_{l,k} \) **do**
   
9:       \( \tilde{V}_{l,k,i}^{(0)} = \text{KSQRUpdate}(\tilde{H}_{l,k}(i,:), \tilde{V}_{l,k,i-1}^{(0)}, \theta_k) \)

10:      \( \tilde{V}_{l,k}^{(0)} = \tilde{V}_{l,k,i}^{(0)} \) representing final updated nullspace matrix, after appending all the rows in Step 8

11:     \( H_{l,k} \tilde{V}_{l,k}^{(0)} = U_{l,k} \begin{bmatrix} \Sigma_{l,k} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{l,k}^{(1)} \\ V_{l,k}^{(0)} \end{bmatrix}^T \)

12:     \( M_{l,k} = [\tilde{V}_{l,k}^{(0)} \ V_{l,k}^{(1)}] \)

13:     \( W_{SC_l} = [\tilde{V}_{l,1}^{(0)} \ V_{l,2}^{(0)} \ldots \tilde{V}_{l,k}^{(0)} \ldots \tilde{V}_{l,k}^{(0)}] \)

14:     Set \( \Sigma_{SC_l} = \left[ \begin{array}{c} \Sigma_{l,1} \\ \vdots \\ \Sigma_{l,k} \end{array} \right] \)

15:     Compute power loading matrix \( \Delta_{SC_l} = \text{waterfill}(\Sigma_{SC_l}, \text{SNR}) \)

16:     Set BF weights as \( M_{SC_l} = [\tilde{V}_{l,1}^{(0)} \tilde{V}_{l,1}^{(1)} \ V_{l,2}^{(0)} \ V_{l,2}^{(1)} \ldots \tilde{V}_{l,k}^{(0)} \tilde{V}_{l,k}^{(1)}] \Delta_{SC_l} \)

that the same threshold value \( \theta_k \) will be used in the KSQR update Alg. 5

Finally, in Step 14, we stack the resulting nullspaces for all the \( K \) users from set \( \mathcal{S} \) for the first subcarrier in a matrix \( W_{SC_1} \) which will be utilized in determination of nullspace for the subsequent subcarriers in KSQR update Alg. 5

Please note that in the same loop of computation of nullspace for every user from set \( \mathcal{S} \) in the first subcarrier, we also determine the corresponding beamforming weights for the user as shown in Steps 10 to 14, which follow straightforwardly from classical BD method (see Alg. 1). The resulting BF weights for all the users are stored in a matrix \( M_{SC_1} \).

- In the pseudo code represented by Alg. 5, the computed nullspace \( W_{SC_1} \) is updated using KSQR update Alg. 3 for every subsequent subcarrier as
discussed in Section 4.4.1. Further, we also determine the BF weights for the users in this subcarrier.

- In order to perform this, as shown in Step 5, unintended users channel matrix is formed by concatenating channel estimates from all users but the intended user \( k \) from set \( S \).

- In the Step 7, \( \hat{H}_{l,k}(1,:) \) corresponds to the first row vector of the matrix \( \hat{H}_{l,k} \) which also corresponds to \( b^H \) in Section 4.4.4. Instead of performing SVD operation on the entire concatenated matrix \( \hat{H}_{l,k} \) of dimension \( n(l-R_k) \times n_T \), a simple operation is performed on a single row vector (i.e., \( 1 \times n_T \)). In the same step, we use KSQR update Alg. 3 in order to update the already computed nullspace \( \tilde{V}^0_{l,1,k} \) which corresponds to \( W \) in Section 4.4.4 and was obtained from Alg. 4. The updated nullspace represented by \( \tilde{V}^0_{l,k,1} \) is the approximate nullspace corresponding to \( \tilde{W} \) in Section 4.4.4 for the modified channel matrix \( \hat{H}_{l,k} \) row vector in the matrix \( \hat{H}_{l,k} \). \( \theta_k \) corresponds to \( \theta \) in Alg. 3.

- This entire update process is performed iteratively over all the rows of channel information for a given subcarrier, as shown in Step 9. Note that the reason to split the process in Step 6-7 and 8-9 is that for updating the nullspace due to insertion of first row, we use the computed nullspace from Alg. 4 while for updating due to the insertion of the subsequent rows \( i \), we use the updated nullspace determined from the same Alg. 5.

- The nullspace matrix for the entire subcarrier is denoted by \( W_{SC_l} = [\tilde{V}^{(0)}_{l,1} \ldots \tilde{V}^{(0)}_{l,k} \ldots \tilde{V}^{(0)}_{l,k}] \), with \( k^{th} \) column indicating nullspace for \( k^{th} \) user.

- The rest of the Steps, i.e., from 11-16 except 12, are straightforwardly followed from classical BD technique (see Alg. 1) to obtain the corresponding BF weights.

### 4.5 Using proposed scheme in LTE system

To exemplify the proposed solution, let us consider MU-MIMO setup in an LTE communication system with multiple receiver antennas. Figure 2.2 shows the t-f resource mapping in the LTE system.

The resource block (RB) is the basic scheduling t-f resource which is allocated to different users based on different physical channel and control signal functions. We are interested in the channel estimates which can be fetched from different Sounding Reference symbols (SRS) in the uplink as defined in the LTE standard. Considering a standard 20 MHz bandwidth and 15 KHz of subcarrier spacing on the frequency axis (vertical axis in the figure), for a single SRS, we have
transmission bandwidth configuration of 100 RB’s and every RB retains 12 subcarriers. Every subcarrier contains uplink channel information of the active users which might cause interference.

The proposed method initially computes the nullspace and the BF weights for the first subcarrier using proposed KSQR Alg. 4. Next, for every subsequent subcarrier, nullspace of the first subcarrier acts as an input and the nullspace is updated for respective subcarrier using proposed KSQR update Alg. 5. Thus, many computations can be potentially saved (by at least one order of magnitude, for a number of antennas greater than 64).

### 4.6 Complexity comparison

The notations in this discussion follow from the MU-MIMO precoding setup discussed in Section 3.2.1. For simplicity, we have considered two receive antennas per user for all the users, i.e., $n_{R_k} = 2$. $K_{\text{max}} = n_T/n_{R_k}$ denote the maximum number of users that the BS can support simultaneously. We consider the dimensionality constraint in BD and thus have $n_T \geq n_{R_k}$. Additionally, $\text{nul}$ indicates the nullity of the unintended users channel matrix $\hat{H}_k$. It can also be seen as a number of orthonormal basis vectors corresponding to the columns in the nullspace matrix $W$ (see Section 4.4.4 for the dimensions of the matrix).

For the classic SVD based precoding and the iterative QRD based precoding algorithms we use the complexity derivations from [12] as it is.

We tabulate the total number of floating point operations (Flops) required to implement the various nullspace determination algorithms for BD based precoding in Table 4.1. Note that in this case, we do not account for the number of flops required to compute the nullspace using KSQR which is fed to the update algorithm. This would be a valid scenario for nullspace determination of a system with a large number of subcarriers as the complexity required to compute the nullspace for the first subcarrier can be ignored.
CHAPTER 4. SUMRATE MAXIMIZATION.6. COMPLEXITY ANALYSIS

<table>
<thead>
<tr>
<th>Method</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD</td>
<td>$F_{SVD} = K_{\max}(24n_R(K_{\max} - 1)n_T^2 + 48n_R^2(K_{\max} - 1)^2n_T + 54n_R^3(K_{\max} - 1)^3$</td>
</tr>
<tr>
<td>Iterative QRD</td>
<td>$F_{QRD} = 8n_R^2(3n_T - n_{R_k})$</td>
</tr>
<tr>
<td></td>
<td>$+ \sum_{i=2}^{K_{\max}-1} (i + 1){8n_R^2n_T[n_T - (i - 1)n_{R_k}]$</td>
</tr>
<tr>
<td></td>
<td>$+ 4n_R^2[3(n_T - (i - 1)n_{R_k}) - n_{R_k}]$</td>
</tr>
<tr>
<td></td>
<td>$+ 8n_T[n_T - (i - 1)n_{R_k}][(n_T - n_{R_k})]$</td>
</tr>
<tr>
<td>KSQR update</td>
<td>Two variants as below:</td>
</tr>
<tr>
<td>- for constant rank</td>
<td>$F_{\text{proposed,,const}} = K_{\max}(n_T - n_{R_k})[6(n_T^2 + n_T)]$</td>
</tr>
<tr>
<td>- for variable rank</td>
<td>$F_{\text{proposed,,var}} = K_{\max}\left[ \sum_{\text{col}=(n_T-1)}^{\text{nul}} 6(n_T + 3n_T \text{col}$</td>
</tr>
<tr>
<td></td>
<td>$+ 7\text{col})\right]$</td>
</tr>
</tbody>
</table>

Table 4.1: Number of FLOPs required per subcarrier and $K_{\max} = K = n_T/n_R$ users for nullspace determination

Here, "col" corresponds to the number of columns in matrix $W$ (see Alg. 5). Appendix 7.1 can be referred for detailed derivations.

From the complexity derivations for KSQR update algorithms, we observe that the complexity is heavily dependent on the rank of the matrix $\hat{H}_k$. Following the dimensions and notations of the matrix $\hat{H}_k$ from earlier sections and assumptions on $K_{\max}$, $\text{nul}$ can also be written as follows

$$\text{nul} = n_T - \text{rank}_\theta(\hat{H}_k)$$
$$= n_T - (n_R - n_{R_k})$$
$$= n_T - (K_{\max}n_{R_k} - n_{R_k})$$
$$= n_T - n_{R_k}(K_{\max} - 1)$$

(4.13)

We also observe that in the case of constant rank KSQR update, col remains constant throughout the entire KSQR update process.

Note that even though we have shown two separate cases of KSQR update algorithm, the proposed solution adapts to either of this cases autonomously. Typically, it has been observed that when we perform the subspace tracking over multiple subcarriers, due to relatively higher variations of the channel response of the subcarriers with respect to the channel response of the first subcarrier, variable rank KSQR update is more often being utilized.

Further, we also perform complexity analysis comparison of MMSE transmit precoding technique with all the BD based precoding variants in Table 4.2. The derivations of complexity for MMSE transmit precoding method can be referred.
in Appendix 7.1. The complexity derivations performed for all the precoding methods are per subcarrier.

<table>
<thead>
<tr>
<th>Method</th>
<th>Complexity</th>
</tr>
</thead>
</table>
| MMSE Transmit        | \[6\frac{7n_k^2}{3} + n_T^2 K_{max}n_{R_k} \]
|                      | \[+ 2n_T^2 + 6n_T + K_{max}^2 n_T^2 n_{R_k} \]                           |
| SVD                  | \[F_{\text{SVD}}\text{ (Table 4.1)}\] + \[F_{\text{op}}\text{ (Equation (4.15))}\] |
| Iterative QRD        | \[F_{\text{QRD}}\text{ (Table 4.1)}\] + \[F_{\text{op}}\text{ (Equation (4.15))}\] |
| KSQR update          | Two variants as below:                                                      |
| -for constant rank   | \[F_{\text{proposed}_\text{const}}\text{ (Table 4.1)}\] + \[F_{\text{op}}\text{KSQR}_\text{const}\text{ (Eq. (4.14))}\] + \[F_{\text{KSQR compute}}\text{ (Eq. (4.16))}\] |
| -for variable rank   | \[+F_{\text{opSVD,QR,var-KSQR}}\text{ (Eq. (4.15))}\] + \[F_{\text{KSQR compute}}\text{ (Eq. (4.16))}\] |

Table 4.2: Number of FLOPs required per subcarrier and \(K_{max} = K = n_T/n_R\) users to obtain precoding weights

Note that, in the BD based precoding variants, apart from the flops required for nullspace determination additional operations like the projection of user matrix on the nullspace, range space determination using SVD, and water-filling are also considered. Following is the resulting complexity for such operations. For variable KSQR update algorithm, we assume that the resulting col = null and thus complexity required to compute these operations is equal to the BD techniques from literature. Appendix 7.1 can be referred for the derivations of individual operations.

\[
F_{\text{opKSQR const}} = K_{max}(8n_{R_k}n_T^2 + 24n_Tn_{R_k}^2 + 48n_{T}^2n_{R_k} + 54n_T^3 + n_T^2n_{R_k}) + 2n_T^2 + 6n_T; \quad (4.14)
\]

\[
F_{\text{opSVD,QR,var-KSQR}} = K_{max}(8n_{R_k}n_T(n_T - (K_{max} - 1)n_{R_k}) + 24(n_T - (K_{max} - 1)n_{R_k})n_{R_k}^2 + 48(n_T - (K_{max} - 1)n_{R_k})^2n_{R_k} + 54(n_T - (K_{max} - 1)n_{R_k})^3 + n_T(n_T - (K_{max} - 1)n_{R_k})n_{R_k} + 2n_T^2 + 6n_T; \quad (4.15)
\]
Additionally, maximum flops required to compute the nullspace for the first subcarrier using proposed KSQR Alg. is given by

\[ F_{KSQR_{compute}} = K_{max} (106n_T^3 + 54n_T^2 + 12n_T), \]

with assumption \( n_R = n_T \) \quad (4.16)

Please refer to Appendix 7.1 for detailed derivation of the above equation. Moreover, the complexity to compute the matrix-matrix multiplication \((\hat{H}^H\hat{H})\) in Step 6 and determination of the threshold \(\theta_k\) in Step 7 is also taken into consideration.

Finally, we observe that the proposed BD algorithm is robust to small variations in the rank corresponding to change in the channel matrix.
Figure 4.1: Complexity comparison to determine nullspace in Block Diagonalization (BD) based precoding using SVD-based method, iterative QRD based method and proposed algorithm, for increase in number of transmit antennas \( n_T \). Here \( n_R \) is number of receive antenna for the user and \( K \) is maximum number of users in the system.

In Figure 4.1 we compare the complexity (number of floating point operations or FLOPs) required to determine the nullspace per subcarrier for the proposed BD algorithm with the classic BD and iterative QR based precoder design which is discussed in the chapter so far. However, this comparison does not account for the number of flops required to compute the nullspace using KSQR, which is fed to the update algorithm. In short, this graph would be valid for nullspace determination for all the subcarriers except the first one. Clearly, as the number of transmit antennas increases (notably after 20), the gap in between complexity of the proposed algorithms and rest of the methods increases. We have considered both the cases in Equation (4.12) for the proposed algorithm.
In Figure 4.2, the complexity in terms of the number of flops is compared for different precoding schemes discussed in this report. Block-diagonalization scheme using three variants, namely classic, iterative QRD and proposed algorithm are compared with precoding scheme using MMSE based solution for increasing number of transmit antennas. It can be seen that for a fixed subcarrier size of 12 (corresponding to 1 Resource Block), the complexity of MMSE transmit linear precoding scheme is much lower compared to all the BD variants. Results in this figure has also taken into account the complexity required for additional operations than determination of nullspace (see Equations (4.14) and (4.15)) and complexity required to compute the first QR factorization.
CHAPTER 4. SUMRATE MAXIMIZATION

6. COMPLEXITY ANALYSIS

Figure 4.3: Complexity comparison in between three variants of Block Diagonalization (BD) based precoding including proposed method and transmit MMSE precoding, for increase in total number of subcarriers (scsize). Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 100$.

In Figure 4.3, the complexity (in terms of number of flops) for the same precoding schemes as in the previous figure are compared for increasing subcarrier sizes (total number of subcarriers) in the system. The maximum subcarrier size considered is chosen considering a transmission bandwidth configuration of 100 RB’s and every RB retaining 12 subcarriers (thus, approx scsize = 1200). It is evident that transmit MMSE based precoding technique has low complexity. However, as shown in Figure 4.4, the proposed algorithms performs better than MMSE transmit precoding correspondingly when we consider the parallel computation of beamforming weights for all the users in the user set. Additionally, using the proposed algorithm, the BF weights for every subcarrier could also be determined in parallel.
Figure 4.4: Complexity comparison in between Block Diagonalization (BD) based precoding using proposed methods and transmit MMSE precoding, for increase in total number of subcarriers (scsize). Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system.
Chapter 5

Simulation setup and Numerical results

In this chapter, we evaluate the performance of various linear precoding schemes discussed in this thesis so far including the proposed one. In Section 5.1, we describe the simulation environment utilized for the numerical evaluations which are performed in Section 5.2. Further, we look into two performance characteristics, namely system sumrate and bit error rate (BER) in downlink data transmission using these schemes.

The numerical results are then evaluated for different SNR regions. Additionally, we also evaluate the performance by varying number of transmit antennas, the number of users in a cell and spatial correlations in between the transmitter and the receiver antennas. Further, we evaluate the performance using two different channel models. The first model is theoretical exponential correlation model and is generated in MATLAB, whereas the second model is standardized, and is preferred in commercial environments. Please note that the proposed BD based scheme is implemented in two different ways while using each of these channel models. Specifically, while using theoretical model, we consider the system as a single-carrier; without splitting it further into several subcarriers. Due to this, we cannot feed computed nullspace for all the unintended users, as we do in the case of the multi-subcarrier system. Instead, we compute nullspace for a single unintended user at the beginning and feed it to the algorithm in order to update the nullspace for rest of the un-intended users. On the other hand, while using the ITU based standardized channel model, we use algorithms 4 and 5 strictly as discussed in Section 4.4.5 in order to determine the beamforming weights.

5.1 Simulation setup

For simplicity of analysis, we consider single cell environment in our simulations and restrict the user terminals to two receiver antennas. However, our results
may be extended to a larger number of antennas in a straightforward manner. Also, we assume that the user set always remains same and the number of users in the set is always less than or equal to half of the number of transmit antennas (i.e., $K \leq \frac{n_T}{2}$). This assumption follows from the requirements of block-diagonalization precoding scheme. In all the simulations we assume perfect CSI estimation.

5.1.1 Theoretical Channel Model

We begin with a simple exponential correlation model from [21] which is discussed in Section 2.3.2. The motivation to use the theoretical channel model is to have a better control on the variables like spatial correlations, which are generally difficult to track in a realistic model. This model follows the MU-MIMO setup as discussed in section 2.3. The inputs to the model are, randomly generated complex matrix without any correlations, angle of arrival and departures, and degree of spatial correlations required at the output. The uncorrelated matrix has dimensions considering a multiple of antennas at transmitter and receivers and multiple users in a cell. This model is generated using MATLAB as a tool and the program used is from a reproducible academic research article [22]. The generated channel matrix has spatial correlations which can be changed to various degrees using a parameter correlation factor.

5.1.2 3D channel model for LTE

One of the aspects of our work is also to evaluate how the proposed low complex algorithm performs against the classical BD solution. As the algorithms are based on coherence principle, we make use of existing OFDM based systems where the t-f resources are represented in the form of a grid of resource blocks, as shown in Figure 2.2. We assume that during a coherence interval (represented by a block of subcarriers and OFDM symbols) the channel remains constant. From equations of coherence bandwidth (i.e., Equation 2.6), coherence time (i.e., Equation 2.7), communication environments like usage in indoor-outdoor, carrier frequency, and relative velocity of the user-terminal, we determine a minimum block of subcarrier-OFDM time symbols where the channel remains constant in our setup. Considering a conservative RMS maximum path length difference between the multi-path components (MPC’s) as 1500 m (order of inter-site distance), we obtain $B_c = 180$ KHz which can be further divided into 12 subcarriers of 15 KHz and corresponds to an LTE system. Similarly, considering a conservative value of 250 km/hour as relative velocity of the user terminal with respect to BS, we obtain coherence time as 1 ms (for the carrier frequency of 2 GHz), which corresponds to 14 OFDM symbols in LTE. Thus, the resulting coherence block consists of 12 subcarriers and 14 OFDM symbols representing a resource block (RB) pair [15]. Additionally, we analyze a scenario where low complex BD scheme using proposed method performs at par with the classical BD. In order to do this, we assume a system with high bandwidth split.
into several subcarriers and having large coherence bandwidth (equal to the number of subcarriers).

We specifically consider downlink data transmission, i.e., data is transmitted from the BS to user terminals and the setup is followed from Section 2.3.4. We consider a TDD duplexing mode in order to exploit the channel reciprocity principle. Thus, we use the same channel from the uplink in order to compute the beamforming weights in the downlink. The channel estimate matrix is generated using the drop methodology, where random properties like the location of the terminals, slow fading, angle of arrival, etc., do not change during the simulation.

The channels are generated in a UMa environment with hexagonal coverage around the cell. The area is covered with 3 sites having 3 cells each. The user terminals are randomly distributed in the area with 9:1 ratio of indoor-to-outdoor usage.

The antenna setting at the BS follows a cross polarization with a possibility of both linear as well as planar uniform array geometry, as the channel model considers elevated beamforming methodology in order to capture the scenario where users are situated at an elevation from the ground surface. The antenna element separation distance follows from the specifications. We assume full buffer operation, i.e., all the terminals have content to transmit/receive at all times. The BS antenna elements are modeled according to specifications in [41], which may be deployed in practice using Kathrein antennas.

Considering a more realistic situation in an urban macro LTE environment, where users are also positioned at a certain height from the ground and a need for beamforming due to multi-antenna multi-user setup, 3rd Generation Partnership Project (3GPP) has undertaken a study to model such a scenario. Recently, they have standardized such a channel model [41] which is based on ITU Urban Macro model [42]. We use this model and parameters in Table 5.1 to perform the simulations. The model is programmed in C/C++, however, MATLAB wrappers (MEX) are used to perform the simulations.

It is important to note that when we evaluate the sumrate performance with respect to SNR, we consider an average of SNR values for the selected users. These SNR values are computed based on path loss in between the BS and the respective users and considering standard constants and parameters used to determine the thermal noise at the user terminals.

### 5.2 Numerical results and discussion

In following figures, $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas per user and $K$ is number of users in the system. All the comparisons includes three variants of Block Diagonalization (BD) based precoding schemes including the method proposed in Section 4.4 and ZF channel inversion and transmit MMSE precoding.

Sumrate or spectral efficiency (unit bits/s/Hz) wherever computed is based on Equation (4.7) for BD based schemes with water-filling power allocation. In
### Table 5.1: Simulation Parameters

<table>
<thead>
<tr>
<th>Scenario</th>
<th>ITU Urban Macro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifications</td>
<td>3D channel model for LTE</td>
</tr>
<tr>
<td>Network Deployment</td>
<td>Single cell</td>
</tr>
<tr>
<td>Maximum Indoor terminals per cell</td>
<td>90%</td>
</tr>
<tr>
<td>Maximum Outdoor terminals percentage</td>
<td>10%</td>
</tr>
<tr>
<td>Carrier Frequency ($f_c$)</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 KHz</td>
</tr>
<tr>
<td>Horizontal antenna separation</td>
<td>$0.5\lambda_c$</td>
</tr>
<tr>
<td>Vertical antenna separation</td>
<td>$0.5\lambda_c, 0.8\lambda_c$</td>
</tr>
<tr>
<td>Wrapping link gain derivation methodology</td>
<td>Slow fading gain</td>
</tr>
<tr>
<td>Wrapping methodology</td>
<td>Wrap to strongest point 1 lap’</td>
</tr>
<tr>
<td>Antenna Array Subelements</td>
<td>’3GPP subelement model’</td>
</tr>
<tr>
<td>Antenna xpol name</td>
<td>3GPP X</td>
</tr>
<tr>
<td>Link gain method</td>
<td>Spatial Channel correlation</td>
</tr>
</tbody>
</table>

case of computation of sumrate in transmit MMSE and ZF precoding schemes, we have used code from the article [22], which also uses water-filling power allocation.

#### 5.2.1 Theoretical channel model

The results presented in this discussion are based on setup mentioned in Section 5.1.1. We plot sumrate performance against correlation factor which denotes the degree of spatial correlation. The results make a note on the basic gain obtained from the use of BD precoding method over other solutions.
Figure 5.1: Sumrate performance for various precoding schemes including BD is mapped against different degree of spatial correlations at both transmitter (BS) and receivers (users) at SNR=46dB. Channels are generated using exponential correlation model \[22\]. Here \(n_T\) is number of transmit antennas, \(n_{R_k}\) is number of receive antennas for the user and \(K\) is maximum number of users in the system. Maximum number of layers (streams) is \(K \times n_{R_k} = 16\).

In Figure 5.1, sumrate performance characteristics is plotted against both transmit and receive spatial correlations for SNR value of 46dB (assuming the BS is capable of maximum transmit power of 40W). The chosen SNR values represent high SNR region. We do not consider low SNR region as from our understanding of optimality conditions in Section 3.2.2, BD is another form of ZF solution which is not suitable for low SNR regions. However, for completeness, we have simulated curves for this region using realistic channels, as shown in the later results. Results of simulation for transmit antenna array of 64 elements at the BS with maximum 16 layers (8 users \(\times\) 2 receive antennas per user) is shown in the figure. We observe that the margin of difference in sumrate between BD and other methods (both MMSE transmit and ZF) increase with the increase in the degree of spatial correlations. Specifically, for high spatial correlations (degree > 0.5), the gain in sumrate increases significantly. Thus, the results re-confirm the findings in the article \[10\] for a higher number of transmit antennas and layers in MU-MIMO setup. We have plotted additional result (Figure 7.1) with number of layers equal to 8 for 4 user MIMO (please see appendix).
5.2. RESULTS

BD using classical SVD
ZF
MMSE Transmit
BD using iterative QRD
BD using iterative KSQR update

Figure 5.2: Sumrate performance for various precoding schemes including BD is mapped against different degree of spatial correlations at both transmitter (BS) and receivers (users) at SNR=46dB. Channels are generated using exponential correlation model from literature [22]. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$.

We performed another similar simulation of sumrate against spatial correlations, however now with a lower number of BS antennas (16) and an equal number of layers (corresponding to 8 users $\times$ 2 receive antennas) and have plotted results in Figure 5.2. For such a setup, we observe that even though BD performs better than ZF, it is outperformed by transmit MMSE solution for all the degree of correlation. The reason for this lack of performance could be conditions of dimensional requirements in BD, which are one of the limitations for employing this method as discussed in 4.2.3. We note that, surprisingly, the sumrate performance gets degraded in spite of the fact that the number of transmit antennas and layers are equal (which is the dimensional condition, i.e., $n_R = \sum_{k=1}^{K} n_{R_k} \leq n_T$).

5.2.2 3D channel model for LTE

Following are the results when simulations are performed using realistic channel models defined in Section 5.1.2. In these simulations, we also observe the effect of using two different antenna array layouts. Prominently, we discuss ULA (uniform linear array) and study the implications of different variables on the system capacity. We also present results for UPA (uniform planar array), where the BS
transmit antenna elements are placed in a rectangular plane forming a grid-like structure with rows and columns. Due to this, the antenna separation distance varies which causes transmit spatial correlations to vary. Cross polarization is used with antennas separation of $0.5\lambda_c$ unless mentioned explicitly.

**Sumrate tracking subcarrier channels**

We analyze the subcarrier tracking and how sumrate for each subcarrier varies.

![Sumrate Comparison](image)

Figure 5.3: Tracking of system sumrate for different subcarriers using proposed algorithm at SNR=46dB showcases precision of nullspace approximation. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 8$.

In Figure 5.3, we plot the system sumrate for all the precoding methods for 48 subcarriers representing frequency dimension of 4 resource blocks in LTE OFDM system assuming minor changes or updates in the channel estimates in between these subcarriers. For the first subcarrier, the beamforming weights and the channel capacity is based on computed nullspace for all the BD precoding schemes. However, for second subcarrier onward, proposed iterative KSQR update algorithm determines the nullspace and hence BF weights approximately as against all other BD methods. Transmit MMSE and ZF solutions are shown for reference and uniformity. The simulations are performed for setup with 16 transmit antennas at the BS with ULA structure and 8 spatial layers (4 users...
The simulations are performed in a spatially correlated environment with a magnitude of correlation having a random distribution. The link gains are derived considering small-scale fading scenario and zero Doppler.

Primarily, we observe that the sumrate for proposed BD algorithm closely follows the sumrate of other BD schemes, suggesting a good numerical precision while approximating the nullspace of the unintended users. The curve representing proposed algorithm tracks the channel well in this case, which suggests that it is adaptable to small variations in the channel due to the change in the frequency. Finally, we also observe that all the BD schemes outperform the transmit MMSE and ZF solutions. The relative sumrate difference in between BD and other methods is approximately 1.5 bits/s/Hz on an average which is minor (however, note that in these simulations, the amount of spatial correlations is unknown and could be low).

In Figure 5.4, we plot similar results for UPA alignment at the BS. The observations in terms of nullspace tracking and precoding scheme sumrate comparisons are same as in the above case. The only difference is the reduction in system throughput for all the schemes as compared to the ULA model and the increase in the relative sumrate difference in between BD based and non-
BD based schemes. Further, we plot similar graphs (Figures 7.2 and 7.3 in appendix) for larger antenna array at BS (64 elements) along with 16 layers for both ULA and UPA and deduce same results.

Figure 5.5: Tracking of system sumrate for different subcarriers using proposed algorithm at SNR=46dB showcases precision of nullspace approximation. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here \( n_T \) is number of transmit antennas, \( n_{R_k} \) is number of receive antennas for the user and \( K \) is maximum number of users in the system. Maximum number of layers (streams) is \( K \times n_{R_k} = 16 \).

In Figures 5.5 and 5.6 we plot similar graphs for BS antenna array size of 16 along with an increased number of layers, i.e., 16 for ULA and UPA, respectively. We have used \( 0.5 \lambda_c \) as horizontal antenna separation distance and both \( 0.5 \lambda_c \) and \( 0.8 \lambda_c \) vertical antenna separation. In the case of sumrate comparison for ULA, proposed algorithm tracks the sumrate implying success of subspace tracking algorithm with a minor relative difference, and sumrate of all the BD schemes outperform transmit MMSE and ZF solutions. This is in contrast to the result observed in a theoretical model, where all the BD schemes failed. However, considering that the two setups differ in terms of how the algorithms are applied as discussed at the beginning of the current chapter, we assume the results are uniform. On the other hand, the results for UPA confirm results from the theoretical model. In this case, all the BD schemes fail as compared to transmit MMSE solutions. Also, the proposed algorithm fails to track the sumrate with the change in subcarrier frequency.
Figure 5.6: Tracking of system sumrate for different subcarriers using proposed algorithm at SNR=46dB showcases precision of nullspace approximation. Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 
CHAPTER 5. SIMULATIONS AND RESULTS

5.2. RESULTS

Figure 5.7: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here \( n_T \) is the number of transmit antennas, \( n_{R_k} \) is the number of receive antennas for the user and \( K \) is the maximum number of users in the system. Maximum number of layers (streams) is \( K \times n_{R_k} = 8 \).

**Effect of SNR on Sumrate performance**

In this setup, we plot sumrate performance curves for different SNR values.

In Figure 5.7, we plot sumrate as a function of SNR for ULA setting at the BS having 16 antenna elements with antenna separation of \( 0.5\lambda_c \). We vary SNR from -14dB to 46dB to ensure all the SNR regimes are covered. Different curves indicate sumrate for all the precoding methods discussed in this report. We observe that with the increase in SNR, sumrate for all the methods increases. We observe that in low SNR region until 0dB in the figure, MMSE transmit precoding solution performs better than BD and until 23dB, better than zero-forcing. MMSE transmit BF as compared to ZF solutions performs better in low SNR region because it allows for a certain amount of interference in order to avoid noise amplification caused due to complete channel inversion. The results are in congruence with the phenomenon discussed in Section 3.2.2.

However, in the high SNR region in the figure, we observe that BD methods perform close to MMSE and ZF solution or rather relatively better. The minor improvement in the performance could be due to spatial correlations as seen in the theoretical results earlier. Thus, these results justify the general optimality
CHAPTER 5. SIMULATIONS AND RESULTS

5.2. RESULTS

Figure 5.8: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here \( n_T \) is number of transmit antennas, \( n_{R_k} \) is number of receive antennas for the user and \( K \) is maximum number of users in the system. Maximum number of layers (streams) is \( K \times n_{R_k} = 8 \).

Figure 5.8 shows sumrate as a function of SNR for UPA setting at the BS having 16 elements arranged in rectangular plane with separation distance of 0.5\( \lambda_c \). Similar to the previous case, we vary SNR from -14dB to 46dB and compare sumrate characteristics of different precoding schemes. We observe that the MMSE transmit precoding scheme outperforms all the solutions significantly for low SNR regions. With the rest of the setup being constant as relative to the previous setup, the cause in this performance boost can be regarded to lack of adaptability of ZF solutions to increase in noise. As ZF solutions completely nullify the correlation terms, while doing so, they boost the effect of noise giving rise to a reduction in sumrate.

Similar observations could be made for both ULA and UPA BS setting with 64 antenna elements at BS and 16 layers from the figures 7.4 and 7.5 respectively. We also observe that the overall system capacity increases in these figures owing to increase in the number of BS antennas and number of layers as compared to the previous case.
BD using classical SVD
ZF
MMSE Transmit
BD using iterative QRD
BD using iterative KSQR update

Figure 5.9: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is the number of transmit antennas, $n_{R_k}$ is the number of receive antennas for the user and $K$ is the maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$.

In Figure 5.9 we plot the comparison of sumrate as a function of SNR for various precoding schemes when BS antenna setting is ULA with 16 elements along with an equal number of layers. The entire graph could be visualized in three parts based on different SNR regions. We observe that even though BD based solutions outperform ZF solutions at all the SNRs, for lower and mid-SNR values, sumrate for transmit MMSE method is more as compared to both the ZF solutions. We also observe that sumrate for BD is marginally higher than MMSE transmit at high SNRs. We also note in the case of BD method that the proposed algorithm tracks the sumrate of other BD based schemes very well in spite when the dimensions of the MU-MIMO setup approaches close to dimensionality constraints for BD algorithm.
CHAPTER 5. SIMULATIONS AND RESULTS  5.2. RESULTS

Figure 5.10: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. Note that number of transmit antenna and layers are equal.

We now perform same simulations as above with UPA BS antenna setting and plot the results in Figure 5.10. We observe that when the dimensions of MU-MIMO setup tend to approach close to the dimensionality constraint for BD algorithm, the performance of all the BD methods degrades. Additionally, the proposed BD algorithm is unable to track the nullspace of the unintended users and hence the sumrate. This result is consistent with our discussion for results from Figure 5.6.
Effect of number of transmit antennas on sumrate

In this analysis, we focus our discussion on the effect on sumrate caused due to increase in the number of transmit antennas at the BS and having fixed the number of layers (i.e., a product of the number of users and number of receive antennas per user.)

![Figure 5.11: Sumrate performance for various precoding schemes including BD is mapped as a function of number of transmit antennas at the BS for SNR=46dB. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$.]

In Figure 5.11 we plot results considering constant 16 layers. We vary the number of transmit antennas from 16 to 100 and note the effect of variation on sumrate for all the precoding schemes. Note that we regenerate the channel when the number of antennas is changed. The simulation is performed for a high SNR value of 46dB and the sumrate is evaluated for subcarrier number 2. We observe that the sumrate for all the precoding schemes increases with the increase in the number of transmit antennas showing the gain obtained due to spatial multiplexing. Additionally, we also observe that there is a marginal improvement in the sumrate for all the BD schemes as compared to MMSE transmit and ZF solutions. This result is consistent with the results obtained in the article [10] but now considering the setup with much higher dimensions. The reason for this improvement is the ability of BD algorithm to optimally use
the excess degree of freedom available at the transmitter. We also note that the sumrate for proposed BD algorithm overall tracks the sumrate computed using other BD methods fairly well especially when the number of transmit antennas is equal to the number of layers.

In Figure 5.12, we plot sumrate performance curves for a constant number of layers 16, but varying the number of transmit antennas from 16 to 49 to see the effect on sumrate. We use UPA setting at BS with antenna separation of 0.8λc in both vertical and horizontal directions. We note that the sumrate for all the schemes increase with the increase in the number of antennas. BD based schemes performing better than MMSE transmit solution at all the antenna setups except when the number of transmit antennas is equal to 16 (which is the number of layers). The reason for this could be the limitation of dimensionality requirement for BD based solutions discussed in Section 4.2.3. We also observe that for the number of transmit antennas equal to the number of layers (16 in this case), proposed BD fails to track the sumrate with other BD methods which are consistent with all our previous discussions for UPA settings.
CHAPTER 5. SIMULATIONS AND RESULTS 5.2. RESULTS

BD using classical SVD
ZF
MMSE Transmit
BD using iterative QRD
BD using iterative KSQR update

Figure 5.13: Sumrate performance for various precoding schemes including BD is mapped as a function of number of users at SNR=46dB. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T = 16$ is number of transmit antennas, $n_{R_k} = 2$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Note that maximum number of layers (streams) $K \times n_{R_k}$ is varying.

Effect of number of users on sumrate

Now we study the effect of variation of the number of users on the sumrate for both BD based solutions and conventional linear precoders. We keep the number of transmit antennas fixed.

In Figure 5.13 we plot the sumrate comparison results for various precoding schemes as a function of number of users keeping the number of transmit antennas fixed. We use ULA BS setting to simulate this result. We observe that the sumrate performance for all the precoding schemes shown increases with the increase in the number of users. This can be intuitively understood as a gain due to the exploitation of additional degree of freedom by more number of users. However, it could also be noted that the gain also depends on the effect of spatial correlations on the channels especially due to the location of the users. The closer the location of users to each other, the lower would be the gain in sumrate. Thus, it becomes very important to select users and co-schedule them such that the effective sumrate increases rather than selecting users randomly. Results in Figure 5.14 justify this discussion, where at an instance when the number of users is equal to 5, the sumrate is lower than that of the sumrate at
Figure 5.14: Sumrate performance for various precoding schemes including BD is mapped as a function of number of users at SNR=47dB. Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T = 16$ is number of transmit antennas, $n_{R_k} = 2$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Note that maximum number of layers (streams) $K \times n_{R_k}$ is varying.

an instance when the number of users is equal to 4.

Further, in Figure 5.14 we observe that, when the number of layers becomes equal to the number of transmit antennas, the proposed BD algorithm fails to track the sumrate and overall BD precoding schemes perform lower than MMSE transmit owing to the dimensionality constraints analogous to previous UPA discussions.
CHAPTER 5. SIMULATIONS AND RESULTS

5.2. RESULTS

BD using classical SVD

ZF

MMSE Transmit

BD using iterative QRD

BD using iterative KSQR update

Figure 5.15: Sumrate performance for various precoding schemes including BD is mapped as a function of number of users at SNR=46dB. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T = 32$ is number of transmit antennas, $n_{R_k} = 2$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Note that maximum number of layers (streams) $K \times n_{R_k}$ is varying.

In figures 5.15 and 5.16 we plot the sumrate as a function of number of users using fixed number of transmit antennas as 32 and 64 respectively. From these results, we observe that with the fixed increase in number of transmit antennas at the BS, the possibility of co-scheduling maximum number of users in proposed BD scheme decreases. Our analysis follows from the three plots for three different fixed number of transmit antennas at the BS using ULA setting. When the number of antennas at the BS is fixed to 16 as shown in Figure 5.13, nullspace and thus the sumrate is tracked for all the 8 users. Whereas, in Figure 5.15 nullspace is accurately tracked when only 14 out of 16 users are co scheduled as a set. Further, same effect can be observed in Figure 5.16 where nullspace is accurately tracked for 23 co scheduled users out of maximum 32 users. The ideal figure for maximum number of co-scheduled users in a BD scheme depends on the total number of transmit antennas and number of receive antennas each user has, as discussed in description of BDZF in Section 4.2.1.

84
Figure 5.16: Sumrate performance for various precoding schemes including BD is mapped as a function of the number of users at SNR=47dB. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T = 64$ is the number of transmit antennas, $n_{R_k} = 2$ is the number of receive antennas for the user and $K$ is the maximum number of users in the system. Note that maximum number of layers (streams) $K \times n_{R_k}$ is varying.
Figure 5.17: BER performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 8$.

**Bit error rate comparison**

For completeness of the overall study, we base our discussion on another performance characteristic. Bit error rate is the measure of knowing the number of erroneous bits received at the receiver as compared to the total transmitted bits at the transmitter. The errors are caused due to various factors such as channel path loss, interference due to other users and noise, etc. In the following setup we have used QPSK modulation scheme to avoid BER increase caused due to other higher order modulation schemes, and base our comparisons more on the effect of precoding schemes. We send 100 packets each having packet size of 1000 QPSK symbols. We use standard modulation and demodulation procedure for QPSK. The same channel is used on which the precoding algorithms are based on. All the vectors generated are normalized and scaled. At the receiver, added noise is modeled as Gaussian with zero mean and variance corresponding to the noise floor. We plot BER plot for various SNRs in the following graphs.

In Figure 5.17 we plot BER curves for various precoding schemes against SNR values from -16dB to 46dB. The channel is based on ULA setting at the BS with 8 layers. We observe that all the precoding schemes have comparable
performance in terms of BER. If we observe very carefully in the low SNR regime, it can be noted that BER for channel inversion ZF method is negligibly lower than MMSE transimt. BD based schemes perform at par with MMSE in this region. For 15dB to 25dB, the relative difference in between BD and other solutions is higher. If we consider the high SNR regime (from 30dB onward in the figure), we can observe that BER curves for all the solutions merge together with BER=0, implying that ZF and BD solutions converge to MMSE transmit performance. This is a confirmation of the asymptotic analysis which we discussed in Section 3.2.2 for ZF based solutions.

In Figure 5.18 we plot BER curves against different SNRs similar to the plot in the previous figure. However, now we consider UPA setting at the BS. It is more visibly clear in this case that ZF based solutions (both channel inversion and BD) lack marginally behind the MMSE transmit solution in low SNR regimes. The performance of all the precoding schemes, however, are comparable in high SNR regime as in the case BER analysis for ULA setting.

Similar conclusions are made for BER performance curves in the case when the system dimensions are increased (i.e., for 64 antenna elements at the BS and 16 layers). The results are shown for both ULA and UPA in figures 7.6 and

![BER Comparison for precoding (n_T, K, n_{R_k}) = (16, 4, 2)](image-url)
BD using classical SVD
ZF
MMSE Transmit
BD using iterative QRD
BD using iterative KSQR update

Figure 5.19: BER performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_R$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_R = 14$.

Finally, in Figure 5.19 we plot results for a special case of UPA when the number of transmit antennas is closely equal to the number of layers, in order to provide completeness. We observe that transmit MMSE precoding adapts to the 3D environment well and outperforms all the ZF based solutions. As discussed earlier, BD schemes are not suitable for UPA setting when the number of transmit antennas is closely equal to the number of layers. Further, proposed BD scheme also fails to track the nullspace precisely in such a setup.
Chapter 6

Conclusion and future work

6.1 Conclusion

In this thesis we investigated different conventional precoding techniques applied to multi-user MIMO communication setup in downlink (forward-link) data transmission for a single cell with up to $n_T = 100$ BS antennas with two different array alignment formats (i.e., Uniform Linear Array and Uniform Planar Array) and up to $K = 8$ users using the 3GPP 3D spatial correlation channel model for slow fading scenario. The optimal beamforming weights are obtained using a constrained power optimization problem. We also investigated the effect of spatial correlation on the system sumrate using the theoretical exponential channel model. We used a LTE OFDM based t-f resource grid assuming coherence interval of 14 OFDM symbols (1 ms) and 12 subcarriers in most of the simulations. In some simulations, we assumed coherence over a larger set of subcarriers. We compared more advanced techniques like classical block-diagonalization (BD) with the conventional precoding schemes. We also proposed and developed a BD algorithm which is low complex and compared it with high complex classical BD solution. The key findings of this thesis are summarized below:

1. Unlike MMSE receive beamforming, MMSE transmit beamforming is suboptimal for both low as well as high SNRs. The optimal structure for transmit BF weight is given by Equation (3.5) as discussed in the article [27]. We confirm this finding through our results, especially at high SNRs.

2. Block-diagonalization is a generalized ZF solution when the receiver has multiple antennas. From the results (in low SNR regions), we confirm that BD is also a suboptimal precoding scheme.

3. Comparison of the developed BD algorithm with classical BD solution:

   (a) For ULA setting at the BS, when the number of transmit antenna elements is equal to 16 along with up to 16 layers, developed BD solution tracks the nullspace for at least 48 subcarriers and hence the
sumrate performance of the classic BD solution precisely. Same is the conclusion for the case of 64 transmit antennas along with up to 16 layers and at least 24 subcarriers.

(b) For UPA setting, except for the case when the number of transmit antennas is closely equal to number layers, the developed BD solution performs at par with the classical BD solution in flat fading Rayleigh channel.

(c) Number of flops required to determine the BF weights in developed BD solution is much less than the other BD algorithms. Thus, the complexity of the developed solution is much lower (by at least an order of magnitude).

(d) Thus, the developed BD solution can be considered as an alternative to the classic BD solution for low Doppler flat fading Rayleigh channel in wideband communication systems.

4. BD based precoding solutions outperform both ZF as well as MMSE transmit based solution in high spatially correlated environments. And the obtained gain in the sumrate increases with the increase in spatial correlation.

5. Summary of analysis of Sumrate for different SNR regions:

(a) Sumrate of all the precoding schemes including BD increases with increase in SNR.

(b) The sumrate performance of BD precoding scheme is marginally better or equal, than that of the ZF solution when the receiver has multiple antennas in all the SNR regions and both the BS antenna settings.

(c) The sumrate of BD scheme is marginally better or equal than MMSE transmit BF at high SNRs for ULA setting of antenna array and number of transmit antennas at BS equal to 16 and number of layers up to 16. Same is the result for the case when transmit antenna array size is of 64 along with the number of layers up to 16.

(d) The sumrate of BD scheme is marginally better or equal than MMSE transmit BF at high SNRs when the number of layers is strictly less than the number of transmit antennas for UPA setting of the antenna array at BS.

(e) MMSE transmit BF solution outperforms all the ZF solutions for UPA setting and number of transmit antennas at the BS is equal to the number of layers at all SNRs.

(f) At low SNRs, MMSE transmit solution typically performs better than both the ZF based solutions in terms of system sumrate for all the dimensions.
(g) For UPA setting with 16 transmit antennas and 16 layers, the performance of the developed BD solution is lower than the classic BD solution.

6. Summary on analysis of the increase in the number of transmit antennas with a fixed number of users.

(a) Sumrate of all the precoding schemes including BD increases with the increase in the number of transmit antennas. The results are confirmed using 100 transmit antennas in ULA setting and 50 antennas in UPA setting along with 16 layers for SNR value of 46dB corresponding to high SNR region.

(b) There is a marginal improvement in the sumrate for all the BD schemes as compared to MMSE transmit and ZF solutions in both UPA as well as ULA settings with the same setup as mentioned in above point with one exception for UPA setting. When transmit antenna array size is equal to the number of layers in UPA setting, MMSE transmit solution performs better than all other schemes.

7. Summary on analysis of the increase in the number of users with a fixed number of transmit antennas:

(a) The sumrate of all the precoding schemes increases with the increase in the number of co-scheduled users for a fixed number of transmit antennas and receive antenna size when co-scheduled users are not located close to each other.

(b) For ULA setting and transmit antenna array size of 16 and at SNR of 46dB, the sumrate performance of all the BD schemes is overall better than transmit MMSE for the different number of co-scheduled users.

(c) For both ULA and UPA setting with the fixed increase in the number of transmit antennas at the BS, the possibility of co-scheduling maximum number of users in developed BD scheme decreases.

(d) Thus, proposed BD solution cannot be generalized for the usage in a scenario where the total number of receive antennas is very close to the number of transmit antennas. However, with efficient scheduling and user selection methods this challenge can be overcome in a realistic scenario.

8. Summary of analysis of BER:

(a) For QPSK modulation all the precoding schemes have an overall comparable performance for ULA setting with 16 transmit antennas at BS and 8 layers. However, there is a minor difference in between MMSE transmit solution and the other solutions.
(b) In the low SNR regime, it can be noted that BER for channel inversion ZF method is marginally lower than MMSE transmit. BD based schemes perform at par with MMSE transmit in this region.

(c) In the mid-to-high SNR regimes for the same setup, BER curves for all the solutions merge together implying that ZF and BD solutions converge to MMSE transmit performance.

(d) For UPA setting with other setup being same, ZF based solutions (both channel inversion and BD) lack marginally behind than MMSE transmit solution in low SNR regimes. However, the performance of all the precoding schemes is comparable in high SNR region.

(e) Similar results follow for increased dimensions of 64 transmit antenna elements and 16 layers for both UPA and ULA setting.

(f) As a special case of UPA with 16 transmit antenna elements and 14 layers, i.e., when the number of transmit antennas is closely equal to the number of layers, transmit MMSE precoding adapts to the 3D environment well which is created due to UPA setting and outperforms all the ZF based solutions.

6.2 Future work

1. The accuracy of the nullspace for the first subcarrier is limited by the size (Frobenius norm) of its channel matrix. In the implementation of KSQR compute algorithm, when the size of the channel matrix for the first subcarrier is too small (e.g., less than $1e^{-15}$), the nullspace of the matrix is just a Householder reflection of a random vector. The process of Gauss-Newton iteration is not performed owing to this constraint on matrix size which causes incorrect determination of nullspace. However, for the subsequent subcarriers set, due to the addition of the new row one at a time, the nullspace gets reshaped as per the characteristics of the added row. Therefore, for such small channel matrices, use of KSQR compute algorithm is limited and alternative options could be possibly looked.

2. The effect of variation of antenna spacing at the BS with UPA setting can be a part of the future studies.

3. Since the developed BD solution is based on a more general perturbation theory, the same algorithm can be applied to more general scenarios like when the users are in motion. Thus, a more general condition can be derived in which the proposed BD solution could be applied.

4. In this study through simulations we confirm that block-diagonalization (BD) based precoding which is a generalized ZF solution is suboptimal. This effect can especially be seen in low SNR regions where MMSE transmit BF outperforms BD. After a literature review we observe that if an
appropriate regularization term is added to the BD solution, the performance improves in low SNR region as well. Thus if the same term is added to the proposed low complex BD solution, a low complex regularized BD can possibly be implemented.

5. One of the challenges to use the developed solution is when the number of receive antennas is closely equal to the number of transmit antennas. Thus, there is a need to determine an optimal set of users where the sum rate is maximized. In the literature, there has already been some work on this front. Therefore, it might be interesting to check the compatibility of such works and our developed BD based solution.
Chapter 7

Appendix

7.1 Complexity Derivations

We first summarize the complexity for typical Matrix operations used in our algorithm. However, [11] can be referred for their detailed derivations. We assume operation on a complex valued matrix $A$ of dimensions $m \times n$ where $m \geq n$.

- The total flop count for water-filling operation used is $2m^2 + 6m$.
- Complexity of full SVD operation on complex-valued matrices is evaluated by $24mn^2 + 48m^2n + 54m^3$ flops by treating every operation as complex.
- Flops required to compute a matrix-matrix multiplication, given by $8mnp$, where $p$ is number of columns of another matrix $B \in \mathbb{C}^{n \times p}$ which is to be multiplied with $A$.
- Flops required for multiplication of a vector $a \in \mathbb{C}^{m \times 1}$ and a matrix $A$ are $6mn$
- Matrix Inverse operation on $A$ using Cholesky decomposition needs $\frac{7m^3}{3}$ flops.

<table>
<thead>
<tr>
<th>Operation (see Section 4.4.5)</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition, if $|W'Wb|_2 \leq \theta$</td>
<td>$6(n_T \cdot \text{number of col in } W + n_T)$</td>
</tr>
<tr>
<td>Sum (worst case)</td>
<td>$6(n_T \cdot n_T + n_T) = 6(n_T^2 + n_T)$</td>
</tr>
</tbody>
</table>

Table 7.1: Nullspace complexity analysis for constant rank

The value $6(n_T \text{col} + n_T)$ in the complexity analysis provided for the nullspace update using proposed algorithm for constant rank as shown in the Table 4.1.
can be majorly seen as a combination of complexity required for various operations in the KSQR update Alg. 3. In Table 7.1 first row corresponds to the complexity required for checking the condition, where the first term of the product \( n_T \cdot \text{number of col in } W \) denotes the matrix-vector multiplication whereas the second term, i.e., \( n_T \) added to this product denotes complexity required for vector 2 norm calculation. As we perform an additional operation of \( \hat{H}_k \) in Alg. 4, it can be noted that the number of columns in matrix \( W \) is always equal to number of transmit antennas \( n_T \).

<table>
<thead>
<tr>
<th>Operation (see Section 4.4.5)</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition, if (</td>
<td></td>
</tr>
<tr>
<td>Construct ( H ) transform and apply</td>
<td>( 6(2n_T \cdot \text{number of col in } W + 7 \cdot \text{number of col in } W) )</td>
</tr>
<tr>
<td>Sum</td>
<td>( 6(n_T + 3n_T \cdot \text{number of col in } W + 7 \cdot \text{number of col in } W) )</td>
</tr>
</tbody>
</table>

Table 7.2: Nullspace complexity analysis for variable rank

In the case of variable rank, in addition to computations shown for constant rank, two more operations require computations. This includes construction of householder transform and applying it to the previous kernel matrix. Table 7.2 shows the overall complexity analysis. Surprisingly, even though additional operations are performed in this case, the complexity is lower than the complexity for constant rank scenario. The reason for this is, as the rank increases due to insertion of new rows to the unintended channel matrix \( \hat{H} \), its numerical nullity \( \text{nul}_0(\hat{H}_k) \) reduces (please refer to Section 4.4.4 and Alg. 5). Even in flat channel there is minor variation in channel matrix causing the rank to increase and nullity to decrease. Thus, the number of columns in matrix \( W \) reduces from \( n_T - 1 \) (for single row insertion) to \( n_T - (K_{max} - 1)n_{R_k} \) (for insertion of \( (K_{max} - 1)n_{R_k} \) rows).

Table 7.3 represents the derivation for the number of flops required in transmit MMSE beamforming. Please refer Section 3.2.3 for further description.

Considering requirements of flops in typical operations, we derive the complexity required by various operations other than computation of nullspace in our algorithms as follows

- Flops required for waterfilling in all the precoding algorithms is \( 2n_T^2 + 6n_T \)
- Complexity to compute the projection of intended user’s channel estimates on the nullspace (see e.g., Step 5 of Alg. 1) can be simply derived as flops required for the multiplication of matrix \( H_k \in \mathbb{C}^{n_k \times n_T} \) and \( V_k^0 \in \mathbb{C}^{n_T \times \text{col}} \), where \( \text{col} = [(n_T - (K_{max} - 1)n_{R_k}) : n_T - 1] \) and depends on the \( \text{nul}_0(\hat{H}) \). Therefore, this operation requires \( (8n_{R_k}n_T\text{col}) \) flops. For the case of KSQR algorithm and worst case KSQR update algorithm (i.e.,
CHAPTER 7. APPENDIX 7.1. COMPLEXITY DERIVATIONS

<table>
<thead>
<tr>
<th>Operations in transmit MMSE</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>((I_{n_T} + \frac{1}{N_0}HH^H)^{-1}) using Cholesky D</td>
<td>(6(n^2_T K_{max} n_{R_k} + \frac{n^3_T}{3} + 2n^3_T))</td>
</tr>
<tr>
<td>Projecting inverse on (H_k) for (K) users</td>
<td>(6 K_{max}^2 n^2_T n_{R_k})</td>
</tr>
<tr>
<td>Power allocation</td>
<td>(2n^2_T + 6n_T)</td>
</tr>
<tr>
<td>Sum</td>
<td>(6(\frac{7n^3_T}{3} + n^2_T K_{max} n_{R_k} + 2n^2_T + 6n_T + K_{max}^2 n^2_T n_{R_k}))</td>
</tr>
</tbody>
</table>

Table 7.3: Number of FLOPs required per subcarrier to obtain BF weights using transmit MMSE beamforming

constant rank scenario), this operation requires \((8n_{R_k} n^2_T)\) considering the col = \(n_T\).

- In order to compute the range space of the resulting matrix after the above projection operation using SVD we require \(24\text{col}^2 n^2_R + 48\text{col}^2 n_{R_k} + 54\text{col}^3\) flops in accordance to the flops required for a standard full SVD of a matrix with dimensions \(n_R \times \text{col}\) where \(\text{col}\) is as defined above. For constant KSQR update algorithm case and KSQR algorithm the flops required are \(24n_T n^2_{R_k} + 48n^2_T n_{R_k} + 54n^3_T\).

- Finally, we need to compute the flops required to multiply range space obtained from above procedure and nullspace obtained through our algorithm. The flops required for this operation can be derived from matrix-matrix multiplication as mentioned earlier. Thus, \(n_T\text{col} n_{R_k}\) is the total flop count. For constant KSQR update algorithm case and KSQR algorithm the total flops required are \(n^2_T n_{R_k}\).

Similarly, we summarize total flops required for above operations in the case of BD using classical SVD and iterative QRD as \(2n^2_T + 6n_T + 8n_{R_k} n_T (n_T - (K_{max} - 1)n_{R_k}) + 24(n_T - (K_{max} - 1)n_{R_k}) n^2_{R_k} + 48(n_T - (K_{max} - 1)n_{R_k})^2 n_{R_k} + 54(n_T - (K_{max} - 1)n_{R_k})^3 + n_T (n_T - (K_{max} - 1)n_{R_k}) n_{R_k} n_{R_k}n_{R_k}\).

Table 7.4 provides detailed derivation of number of flops required to compute the nullspace for the first subcarrier using KSQR algorithm. Please refer [38] for detailed description of the algorithm.
CHAPTER 7. APPENDIX  7.1. COMPLEXITY DERIVATIONS

Operations in KSQR | Complexity
---|---
$\|H_k\|_\infty$ | $6n_T$
QR($H_k$) | $4n_T^2(3n_R - n_T)$
Obtain $\sigma_n$ from $R$ | $23n_T^2 + 20n_T^2 + 13$
Total Heisenberg QRD | $48n_T^2 + 28n_T^2$
Total Application of Givens | $48n_T^2$
Miscellaneous operations | $12n_T$
Sum | $92n_T^2 + 54n_T^2 + 12n_T^2 n_R + 12n_T$
Considering $n_T = n_R$ | $(104n_T^2 + 54n_T^2 + 12n_T)$
Loop over $K$ users | $F_{KSQR_{compute}} = K_{max}(104n_T^2 + 54n_T^2 + 12n_T)$

Table 7.4: Number of FLOPs required per subcarrier for nullspace determination using KSQR
7.2 Additional Results

Figure 7.1: Sumrate performance for various precoding schemes including BD is mapped against different degrees of spatial correlations at both transmitter (BS) and receivers (users) at SNR=46dB. Channels are generated using exponential correlation model from literature [22]. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 8$. 
Figure 7.2: Tracking of system sumrate for different subcarriers using proposed algorithm at SNR=46dB showcases precision of nullspace approximation. Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$.  

Sumrate Comparison for precoding $(n_T, K, n_{R_k}) = (64, 8, 2)$
CHAPTER 7. APPENDIX

7.2. ADDITIONAL RESULTS

Figure 7.3: Tracking of system sumrate for different subcarriers using proposed algorithm at SNR=46dB showcases precision of nullspace approximation. Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 
Figure 7.4: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 

\[ \text{Sumrate (bits/s/Hz)} \]

\[ \text{SNR in dB} \]
Figure 7.5: Sumrate performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 
Figure 7.6: BER performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Linear Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 

BER Comparison for precoding $(n_T, K, n_{R_k}) = (64, 8, 2)$
Figure 7.7: BER performance for various precoding schemes including BD is mapped as a function of SNR (from -14dB to 46dB). Channels are generated using 3GPP 3D channel model for LTE. Uniform Planar Array alignment is used at the BS. Here $n_T$ is number of transmit antennas, $n_{R_k}$ is number of receive antennas for the user and $K$ is maximum number of users in the system. Maximum number of layers (streams) is $K \times n_{R_k} = 16$. 
Bibliography


