PRECODING DESIGNS UNDER LIMITED CAPACITY FRONTHAUL FOR C-RAN

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AHMET FURKAN ODUNKIRAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

AUGUST 2023

Approval of the thesis:

PRECODING DESIGNS UNDER LIMITED CAPACITY FRONTHAUL FOR C-RAN

submitted by **AHMET FURKAN ODUNKIRAN** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** ______________

Prof. Dr. Ali Özgür Yılmaz
Supervisor, **Electrical and Electronics Engineering, METU** ______________

Examining Committee Members:

Prof. Dr. Tolga Mete Duman
Electrical and Electronics Engineering, Bilkent University ______________

Prof. Dr. Ali Özgür Yılmaz
Electrical and Electronics Engineering, METU ______________

Assist. Prof. Dr. Özlem Tuğfe Demir
Electrical and Electronics Engineering, TOBB ETU ______________

Assist. Prof. Dr. Gökhan Muzaffer Güvensen
Electrical and Electronics Engineering, METU ______________

Assist. Prof. Dr. Elif Tuğçe Ceran Arslan
Electrical and Electronics Engineering, METU ______________

Date: 29.08.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: AHMET FURKAN ODUNKIRAN

Signature :

ABSTRACT

PRECODING DESIGNS UNDER LIMITED CAPACITY FRONTHAUL FOR C-RAN

ODUNKIRAN, AHMET FURKAN
M.S., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Ali Özgür Yılmaz

August 2023, 110 pages

Massive multiple-input multiple-output antenna structures used in 5G systems greatly improve throughput and coverage. However, they also bring challenges in areas such as power consumption, complexity, and hardware.

Centralized or cloud radio access network structures centralize baseband operations in a single unit and allow for the separation of baseband processing and radio frequency processing functions for a MIMO environment and allow for the design of multiple radio units with smaller array sizes compared to the traditional radio access networks which reduce the production cost.

The fronthaul connection between the baseband and radio units in a radio network has strict capacity limits with common processing distribution interfaces between these units. Using different kinds of functional split procedures that are brought by technological improvements helps to reduce the load on fronthaul.

This thesis considers a two-stage precoding approach for the precoding operation in a fronthaul capacity constrained wired fronthaul C-RAN system with multiple access points that have reduced complexities compared to a single access point system. First, a maximization problem that consists of the sum spectral efficiency of the users and its constraints has been stated. Then, the constraints are relaxed and added inside the main maximization problem using mathematical relations and the main maximization problem is expressed in a differentiable form.

In this thesis, an optimization method over the differentiable spectral efficiency expression in a gradient-based form is proposed to maximize the sum spectral efficiency. The performance of the proposed method is discussed under different conditions.

Keywords: Precoding, Fronthaul, Gradient Descent, Cloud Radio Access Network

ÖZ

FRONTHAUL KAPASİTESİ SINIRLI C-RAN YAPILARINDA ÖN KODLAMA TASARIMLARI

ODUNKIRAN, AHMET FURKAN
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Ali Özgür Yılmaz

Ağustos 2023 , 110 sayfa

Kitlesel çok girdili ve çok çıktılı anten yapıları 5G sistemlerinde yüksek verimlilik ve kapsama alanı gibi özellikler sunarak kablosuz iletişim sistemlerinde yüksek performans sağlamaktadır. Fakat, bu sistemler beraberinde güç tüketimi, karmaşıklık ve donanımsal alanlarda zorluk getirir.

Bulut tabanlı radyo ağları, temel bant fonksiyonlarını tek bir merkeze toplar ve radyo frekans bandı fonkisyonlarının temel bant fonksiyonlarından ayrılarak gelekensel çok girdili ve çok çıktılı radyo ağı yapılarına kıyasla daha küçük dizge boyutuna sahip birden fazla radyo unitesi yerleşimine olanak sağlayarak üretim maliyetini düşürür.

Bir radyo ağında temel bant ve radyo ünitesi arasındaki bağlantıyı sağlayan fronthaul bağlantısında bu birimler arasında yaygın iş bölümü arayüzleri kullanılması durumunda ağır kapasite yükü vardır. İlerleyen teknoloji ile birlikte gelen farklı fonksiyonel ayrım standartlarını kullanmak fronthaul üzerindeki yükü azaltır.

Bu tezde, kapasitesi sınırlı kablolu fronthaul bir C-RAN sistemi için iki aşamalı bir ön kodlama yaklaşımı düşünülmüştür. Düşünülen C-RAN sisteminde erişim noktası

sayısı arttırılarak tek erişim noktalı duruma göre daha basit yapılar haline getirilmiştir. Öncelikle, kullacıların toplam spektral verimliliği üzerinden bir maksimizasyon problemi oluşturulmuş, daha sonra bu problemdeki kısıtlamalar gevşetilerek türevlenebilir bir hale getirilmiştir.

Bu tezde, türevlenebilir maksimizasyon problemi üzerinden gradyan bazlı ve kullanıcılar için toplam spektral verimliliği maksimize eden bir metod önerilmiştir. Önerilen metodun perfromansı farklı durumlar altında tartışılmıştır.

Anahtar Kelimeler: Ön kodlama, Fronthaul, Gradyan İniş, Bulut Tabanlı Radyo Ağı

To my mom and dad...

ACKNOWLEDGMENTS

Firstly, I would like to thank and express gratitude to my supervisor Prof. Dr. Ali Özgür Yılmaz, who not only guided me through my research and shaped my views about how research should be done, but also motivated and supported me to study and finish my thesis in my hard times. This thesis would not be finished without his valuable mentorship.

I would also like to offer my thanks to Dr. Fehmi Emre Kadan, who provided valuable insight during my studies. His insight has been an important influence for this thesis study. I would like to also thank my examining committee members for their valuable comments and suggestions.

I would also like to offer my gratitude to Vodafone and BTK for funding my research under 5G and beyond research and development program.

I want to also mention my dear friends Musa Baytok whom we shared stressful course studies together and Safa Çelik whom we also shared stressful course studies together and who has been like a big brother to me during my graduate studies.

I would like to finally thank my mom and dad, Hatice and Mehmet Odunkıran who put a lot of effort into raising me and always being with me in my hardest times. Without their support, I would not be able to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

3GPP	Third Generation Partnership Project
ADC	Analog-to-Digital Converter
AoA	Angle of Arrival
AoD	Angle of Departure
AS	Angular Spread
BBU	Baseband Unit
BS	Base Station
CAPEX	Capital Expenditure
CN	Core Network
CoMP	Coordinated Multi-point
C-RAN	Cloud Radio Access Network
CPRI	Common Public Radio Interface
CSI	Channel State Information
DL	Downlink
D-RAN	Distributed Radio Access Network
DSP	Digital Signal Processing
e-CPRI	Enhanced Common Public Radio Interface
EE	Energy Efficiency
GSM	The Global System for Mobile Communications
i.i.d	Independent Identically Distributed
IQ	In-Phase and Quadrature
LTE	Long-Term Evaluation
MIMO	Multiple Input Multiple Output
MRC	Maximial-ratio combining

MRT	Maximum ratio transmission
MSE	Mean Square Error
MU-MIMO	Multi user MIMO
OPEX	Operating Expense
PDF	Probability Density Function
RF	Radio Frequency
RRU	Remote Radio Unit
RZF	Regularized Zero Forcing
SCFA	Smooth Constraint-Free Approximation
SDN	Software Defined Network
SE	Spectral Efficiency
SINR	Signal-to-Interference-Plus-Noise Ratio
SLNR	Signal-to-Leakage-Plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SQR	Signal-to-Quantization Ratio
SU-MIMO	Single user MIMO
UE	User Equipment
UL	Uplink
ULA	Uniform Linear Array

CHAPTER 1

INTRODUCTION

The advent of 5G systems is certain to bring a significant improvement over contemporary communication networks. According to the research published by Ericsson in [1], the revenue of 5G technology is expected to soar to 31 trillion dollars by 2030. Nonetheless, 5G systems have challenges in areas such as data rate, latency, energy, and cost [2]. Among these challenges, data rate improvement is a paramount concern due to ever-increasing user traffic. In fact, according to the Digital 2023 global overview report released in January 2023 [3], total internet users around the world has reached a staggering amount of 5.16 billion people, corresponding to 64.4% of the global population—an increase around 2% compared to the last year. This means that data rate improvement is necessary to accommodate this increasing demand. In [2], one of the key technologies that are listed to enable higher data rates is increased spectral efficiency (SE) by using multiple-input multiple-output (MIMO) systems.

Multi-user MIMO (MU-MIMO) systems provide an enhanced degree of freedom and data rate by sharing the communication channel between users as elucidated in [4]. In a conventional MU-MIMO network, a base station (BS) equipped with multiple antenna elements establishes communication with multiple users who may possess single or multi-antenna configurations. The transmission from BS to users is called downlink transmission and the transmission from users to BS is called uplink transmission. While MU-MIMO systems bring performance improvements, they may be unable to support excessive data traffic. Therefore, an increase in the number of antennas operating in an area is necessary. One promising approach to achieve such an increase is called massive MIMO where BSs are equipped with very large antenna arrays [5]. Such systems with very large multiple-input and multiple-output antenna arrays have been studied following the work in [6] and continued to be studied in works such as [7] and [8].

In traditional BS configurations with multiple antennas, all the signal processing operations are conducted in a dedicated hardware unit and then the processed signals are transmitted to users by using antenna elements. However, the burgeoning user demands for higher data rates led to the consideration of different network architectures for modern communication systems. In modern telecommunication networks, the network link that connects user equipment (UE) to the core network (CN) is called a radio access network (RAN). In the realm of 4G-based systems, the distributed RAN (D-RAN) is the leading topology where a baseband unit (BBU) is connected to CN through a link called backhaul and to a remote radio unit (RRU) via the link called fronthaul. For the downlink operation, the processed signals in these units are transformed into a radio frequency (RF) signal and then transmitted to UEs in the cell area.

The escalating user needs have led to the requirement of more BSs connected to the CN in traditional RANs [9]. The implications of more base stations are observed most notably by considering two expenditure components which are Capital Expenditure (CAPEX) and Operation Expenditure (OPEX) [10] that operators use to quantify the effectiveness of the network. According to [11], OPEX mainly pertains to expenditures in areas such as network operation cost and it increases as cell cites demand for power consumption increases, while CAPEX mainly encompasses expenditures in areas such as baseband hardware, RF hardware, software licenses and it increases due to increasing number of base stations. As users are expecting to pay less money per bit with time, mobile operators are working to cover these expenditures; hence network architectures other than D-RAN have become a necessity as explained in [11].

Cloud Radio Access Network (C-RAN) architecture is a topology first introduced in [12] tailored to optimize power consumption and expenditures that is proposed within the framework of 5G technology. C-RAN architecture moves functions of BBUs from D-RAN to a centralized unit called the BBU pool which is connected to multiple RRUs via the fronthaul links. This centralization brings improvements in terms of network operation cost due to the smaller number of BBUs needed and decreases CAPEX and OPEX. Also, network operators can share the BBU pool by the introduction of a cloud-based system as explained in [11].

The core structure of C-RAN architecture is outlined in [13]. As it is underscored in [13], for a typical C-RAN configuration, the operation that demands most of the processing resources is the baseband signal processing operation which is moved outside the cell to a centralized or cloud-based BBU pool. Meanwhile, RRU components remain situated inside the cell site and provide essential functions such as filtering, amplification, and analog-to-digital conversion. Moving baseband operations outside the cell site decreases energy consumption, complexity, and cost [13].

1.1 Thesis Motivation

With the emergence of 5G technology, the distribution of processing has become more flexible between BBU and RRU in a RAN. The division of processing work between the BBU and RRU is called a functional split, or simply split. For a 4G-based RAN, the functional split between the BBU and RRU is specified by an interface called common public radio interface (CPRI) as explained in [14] which delineates the connection between BBU and RRU through fiber-optic cables, as depicted in Fig. 1.1.

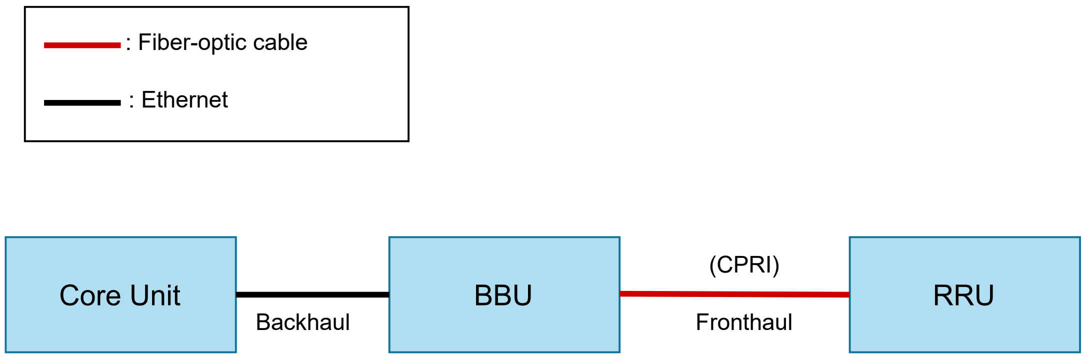


Figure 1.1: Block diagram of the functional split for 4G RAN

CPRI delegates all digital signal processing operations to the BBU while it entrusts digital-to-analog conversion and analog beamforming operations to the RRU.

For 5G, we can employ more flexible functional splits compared to previous generations. Several variations of functional splits for 5G can be observed in Fig. 1.2 which are expounded in [15]. Since our thesis focuses on precoding which is in the physical layer, we zero in on the functional splits that are shown in Fig. 1.2 which consider the splits inside the physical layer and are also called low-level splits [15]. The split definitions in Fig. 1.2 are quite general in that the previously mentioned CPRI connection is studied under Split 8.

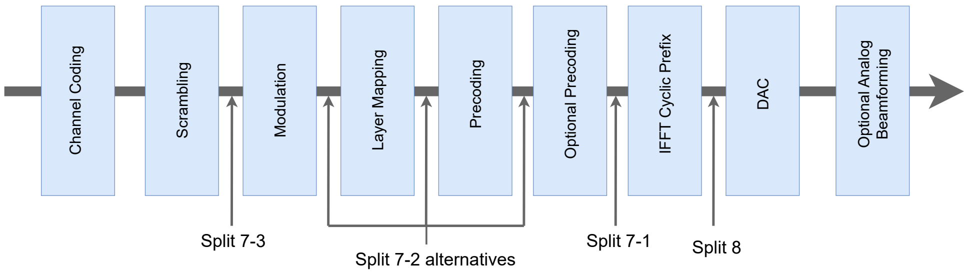


Figure 1.2: Block diagram of functional splits for 5G RAN downlink

Split 8 offloads a significant portion of the processing onto BBU and imposes a constraint on the fronthaul transmission rate up to a value of 9.8 Gbps which could support 3GPP, GSM, LTE [16]. However, it is difficult to accommodate for voracious bit rates required by massive MIMO in 5G solely with CPRI [17]. A way that is proposed to overcome this drawback involves the application of fronthaul compression, detailed in [18]. 5G's enhanced CPRI (e-CPRI) protocol offers a substantial increase in data rate limit up to 25 Gbps [14]. e-CPRI offers to tackle the fronthaul capacity issue by transferring some BBU functions into RRU by considering lower layer splits as in Split 7. As it can be seen in Fig. 1.2, Split 7 can be implemented in different configurations as Split 7-1, Split 7-2, and Split 7-3. With Split 7-2 in C-RAN, the precoding operation can be distributed between the BBU and RRUs as in the rightmost Split 7-2 alternative in Fig. 1.2, which decreases the load on fronthaul capacity.

Different from D-RAN, C-RAN moves BBUs into a centralized unit called BBU pool which can be interconnected with multiple RRUs via fronthaul links. Fronthaul links between RRUs and the BBU pool in C-RAN structures can be established by fiber-optic circuit based transmission, packet based Ethernet protocol, or via a wireless channel as mentioned in [19] and as shown in Fig. 1.3. Multiple RRUs connected to a BBU pool decrease the production cost for antenna units by considering multiple antenna units with smaller array sizes compared to a single antenna unit with a larger array size. It could also allow for sharing resources between RRUs which could form clusters that bring advantages in terms of load and energy [20], while also enabling coordinated multi-point communication (CoMP) [21], [22], [23]. Also, C-RAN allows for distributing RRU placement in a dense way to increase throughput [24]. C-RAN also allows for supporting numerous wireless standards that can be used according to the requirements of users [25]. Furthermore, software-defined radio (SDR) technology can be used to implement hardware functions by using software that makes central processing less dependent on hardware[19], [26].

Our premise in this thesis is that Split 7-2 offers the potential partition of precoding operation between BBU and RRU which reduces the load on the fronthaul capacity. However, for a single RRU, this would entail a very complex RRU structure with high production costs as we load more functions into RRU. To address this challenge, we consider employing multiple RRUs with smaller antenna array sizes that are realizable with C-RAN and reduce RRU complexity while increasing the scalability of the system. We also increase the number of fronthaul links to increase precoding matrix sizes. In our work, we divide the precoding operation between the BBU pool and multiple RRUs according to Split 7-2 for downlink operation and reduce the load on the fronthaul. We consider a sum rate maximization problem for a fronthaul capacity constrained wired fronthaul C-RAN with such a functionality split that serves multiple users with single antennas and propose an iterative and tractable solution to the rate maximization problem. We compare the performances of single and multi-RRU cases and discuss the effect of increasing the number of RRUs while keeping the total number of transmit antennas the same for a C-RAN with unconstrained fronthaul capacity and constrained fronthaul capacity to make comparisons between these cases. We aspire to draw conclusions based on these comparisons over the performance of our proposed method and discuss the effects of an increased number of RRUs and fronthaul links for a fronthaul capacity constrained C-RAN system with functionality split for the precoding operation between the BBU pool and RRUs.

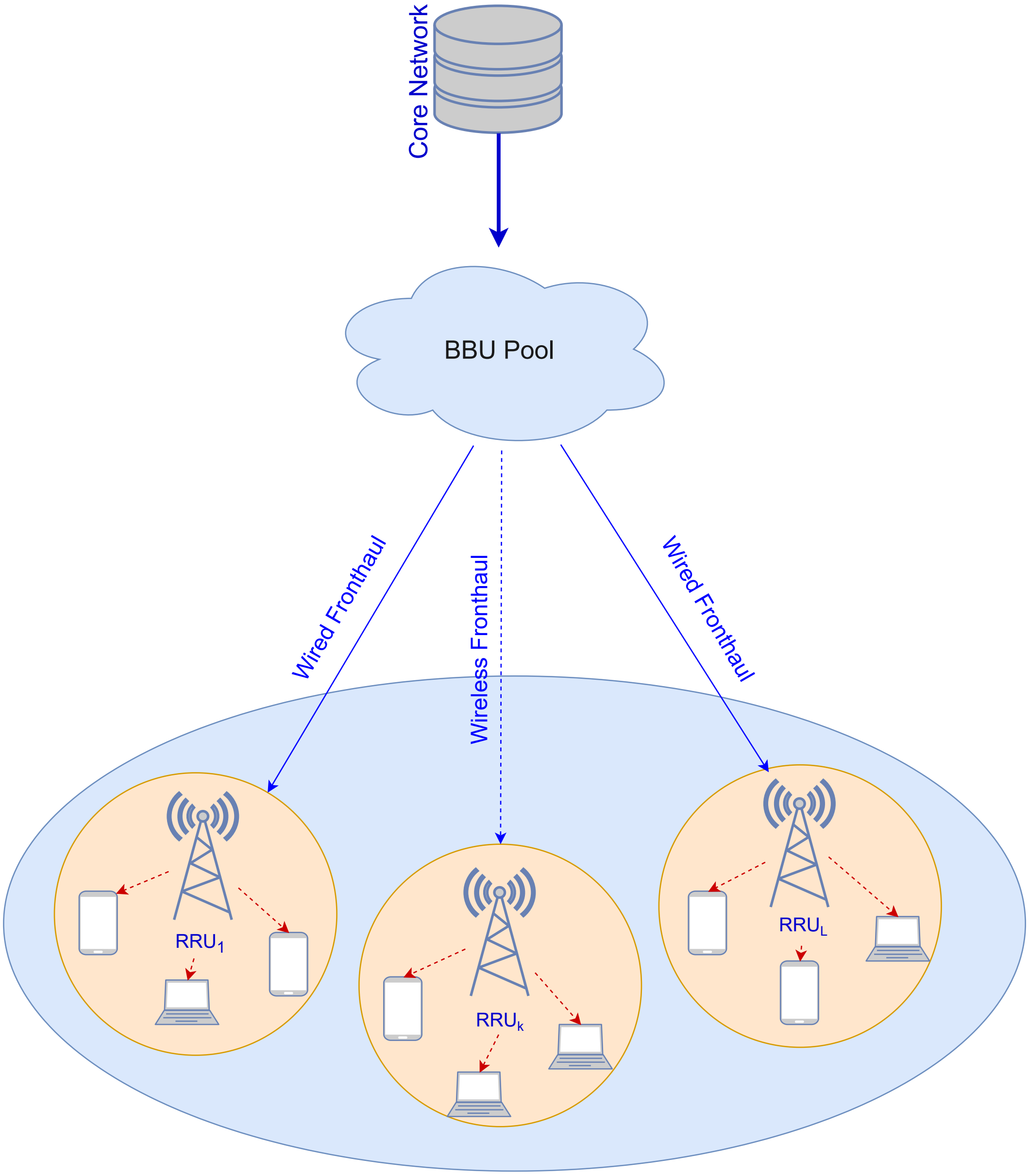


Figure 1.3: Block Diagram of wireless and wired fronthaul C-RAN

1.2 Contributions and Novelties

This thesis considers a C-RAN structure with limited fronthaul capacity and proposes a gradient-based, tractable solution to the sum rate maximization problem. Initial point selection, which is a critical aspect of optimization, is also addressed by drawing insights from several works in literature which are discussed in subsequent chapters. Our contributions to literature in this thesis are listed as follows:

- We tackle the sum rate maximization problem for a limited capacity fronthaul C-RAN system with multiple access points where a two-stage precoding approach is considered.

- To simplify the optimization process, we express the sum rate maximization problem in an easier to differentiate form by relaxing the constraints and transforming the optimization variable into a vectorized expression. This transformation enhances the tractability of the problem and allows a simpler solution.

- We propose an iterative solution to the rate maximization problem under limited fronthaul capacity by employing a gradient-based method.

- By systematically increasing the number of RRUs and fronthaul links, we augment the degree of freedom of our proposed method which increases spectral efficiency.

- We analyze the computational and time complexity of our proposed method which helps to design operations by considering time and transmit data requirements.

1.3 The Outline of the Thesis

In this section, we describe the outline of this thesis.

In Chapter 2, we first describe the development and properties of MIMO systems. Then, we describe several channel models and discuss the concept of the information about channel state in MIMO systems. We also describe the concept of precoding and explore the C-RAN configurations with wired fronthaul links. We conclude the chapter with a discussion on optimization criteria.

In Chapter 3, we discuss the sum rate maximization problem for a C-RAN system with unlimited fronthaul capacity. We first define our problem and then describe our proposed approach for the system. We also explain the initial point selection for the proposed approach in this chapter. We conclude the chapter with complexity calculation and discuss the results.

In Chapter 4, we discuss the sum rate maximization for a C-RAN system with limited fronthaul capacity. We describe the problem and proposed an approach for the system again. We conclude the chapter with complexity calculation and discussion where we also compare the results with the previous chapter.

In Chapter 5, we conclude the thesis and talk about future work.

1.4 Notations

The notations that will be used in this thesis are as follows. Matrices are denoted by bold capital letters so that $\mathbf{A}$ denotes a matrix while vectors are denoted by small bold letters so that $\mathbf{a}$ denotes a vector. $\mathbf{A}^{-1}$ denotes the inverse of matrix $\mathbf{A}$. $\mathbf{A}^T$ and $\mathbf{A}^H$ denotes the transpose and conjugate transpose of $\mathbf{A}$ respectively, while $tr(\mathbf{A})$ denotes the trace operation and a^* denotes the conjugate of complex number a. $\mathbb{E}[.]$ stands for expectation and $vec(\mathbf{A})$ denotes a function where the columns of $\mathbf{A}$ are concatenated to form a vector. $\|\mathbf{a}\|$ denotes the L2 norm of vector $\mathbf{a}$ while $|.|$ denotes the absolute value or L1 norm. $conj(.)$ denotes a function that outputs the conjugate of its complex-valued input. $\det(A)$ denotes the determinant of matrix $\mathbf{A}$. $\lfloor(.)\rfloor$ denotes the floor function where the expression inside the parenthesis is rounded to the closer integer that is less than its value.

CHAPTER 2

DESCRIPTION AND PROPERTIES OF MIMO SYSTEMS, CHANNEL MODELS, PRECODING AND OPTIMIZATION CRITERIA

In this chapter, we provide an overview of MIMO systems, starting by describing their historical development and continuing with mentioning their several properties. Then, we talk about channel models and inspect them under several categories. We also delve into the concept of channel state information in this chapter. The concepts relating the signal power, noise power, interference, and the concept of precoding are also discussed. Finally, we finish the chapter by describing several radio network architectures and discussing several optimization criteria.

2.1 Development of MIMO system

2.1.1 History of MIMO

Wireless communication technology has been an important part of communication technology since James C. Maxwell theoretically proved the electromagnetic theory in 1864, Heinrich R. Hertz transmitted and detected electromagnetic waves in 1888, and Guglielmo Marconi achieved communication with radio waves in 1895 [27]. With the invention of television and the internet, the importance of wireless communication has surged. Therefore, the need for higher capacity, data rates, quality etc. has increased [28]. Due to its numerous advantages such as simpler deployment, lower costs, and better flexibility, the wireless communication technology has gained more popularity [29].

Wireless technology facilitates communication through the atmospheric transmission of electromagnetic signals that fall inside a particular frequency band. Frequency spectrum is becoming more valuable as there surface various competing and complementary wireless technologies [30]. Since the available frequency spectrum is finite, researchers are looking for methods to maximize the usage of the finite frequency spectrum.

The discovery of antenna diversity marked the beginning of MIMO communication which brought great advantage in making use of finite spectrum. In a radio broadcast experiment, engineers Harold H. Beverage and Harold O. Peterson noticed that two receivers that were some distance apart had received a radio signal at different strengths which led to them finding out that this was caused by fading due to multipath propagation and proposed a diversity system in which the received audio signals are combined [31]. This proposed system couldn't work without a separate receiver for each receiver antenna and frequency-switchable radio receivers were built to increase efficiency [31]. In the following years, advancements in hardware and signal processing enabled the development of well-known diversity-combining algorithms such as equal-gain combining, selection combining, and maximal-ratio combining (MRC) [32].

Antenna diversity was first applied in receiver-side applications. According to research by Foschini that was published in the Bell Labs Technical Journal in 1996, increasing the number of transmit and receiver antennas have resulted in significantly higher achieved rates compared to the scenario with single receiver and transmit antenna when the channel's characteristics are known at the receiver but unknown at the transmitter in a fading environment [33]. The impact of using multiple receive and transmit antennas on channel capacity over a Gaussian channel with and without fading is examined by Telatar in [34]. The work in [35] showed that higher achieved rates can be obtained when the information about the state of the channel is also known at the transmitter side. When the channel parameters at the transmitter are unknown, the authors of the study [36] presented a precoding strategy to make use of multi-path channel attributes. MIMO systems have since played a significant role in research and advancements related to wireless communication systems and turned from theory to practice [4]. The MIMO technology is utilized in wireless communication schemes such as 3GPP LTE, 4G, High-speed packet access, and 5G, and is expected to play a crucial role for 6G systems and beyond [37]. There are mainly two types of MIMO systems considered which are single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO).

2.1.2 SU-MIMO

Single-user MIMO systems consider a base station communicating with a single receiver. The downlink structure of a 2×2 SU-MIMO system can be observed in Fig. 2.1. The channel is dedicated to only one user at a time for SU-MIMO systems. Since the channel is reserved for a single user at a time, there can be latency when multiple users try to access the same access point simultaneously. SU-MIMO systems try to achieve diversity gain by using space-time codes and rate gains by transmitting multiple streams over a channel [4].

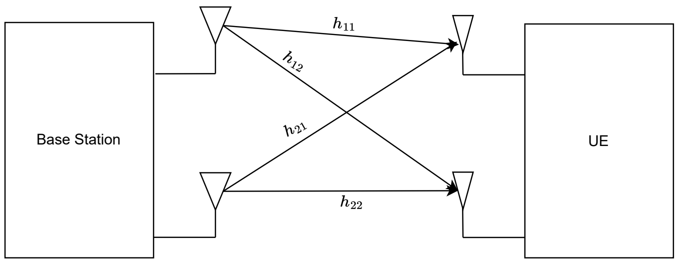


Figure 2.1: SU-MIMO system structure

2.1.3 MU-MIMO

Multiple antenna access points can establish simultaneous communication with different users at the same time in MU-MIMO systems. It spatially multiplexes data and enables base stations to concurrently broadcast or receive data from a number of users who may be equipped with a single antenna or a number of them, giving rise to more degrees of freedom. While the structure of the uplink and downlink is comparable in SU-MIMO systems, the downlink transmission in MU-MIMO is structured differently from the uplink transmission. The uplink structure of an MU-MIMO system with single antenna users can be observed in Fig. 2.2.

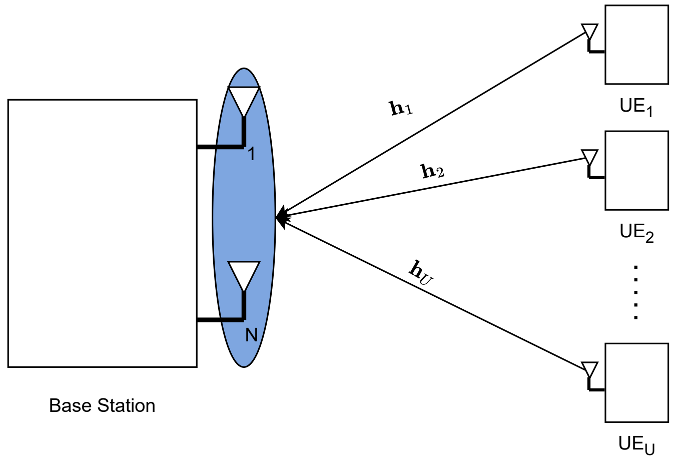


Figure 2.2: Uplink MU-MIMO system structure

The received signal for the system in Fig. 2.2 can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{y}$ is the received vector at the base station equipped with N antennas, $\mathbf{H} = [\mathbf{h}_1\mathbf{h}_2...\mathbf{h}_U]$ is the channel matrix between users and the base station, $\mathbf{s} = [s_1 s_2 ... s_U]^T$ denotes the transmit vector that is sent by users and $\mathbf{n}$ denotes the independent identically distributed (i.i.d) complex Gaussian noise with $\mathcal{CN}(0, \sigma^2\mathbf{I}_N)$ where σ^2 is the noise variance. By recognizing the users as a multiple antenna transmitter transmitting data to a multiple antenna receiver, it is possible to see how this system is similar to SU-MIMO systems in construction. However, transmit antennas in a SU-MIMO system are co-located at a single BS structure in the downlink or at the UE in the uplink. For the uplink in MU-MIMO, the users' antennas don't have to be spatially aligned as they would have if they were to be located on the same device which is different from the SU-MIMO case. The distributed structure of the transmit antennas in uplink MU-MIMO causes signals to arrive at the receive antennas from many directions, allowing the system to function even in the absence of rich scattering.

The downlink MU-MIMO structure for a 2×2 system with single antenna users can be seen in Fig. 2.3.

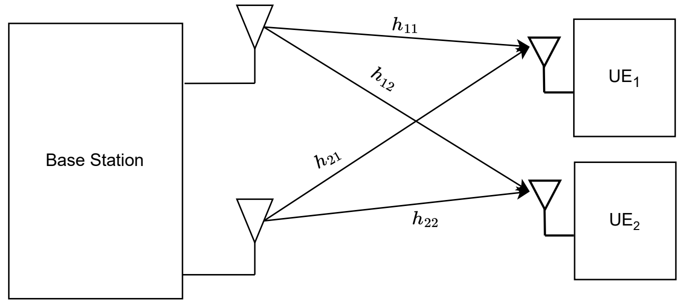


Figure 2.3: Downlink MU-MIMO system structure

The received signal for user u for the above system is expressed as

$$y_u = \mathbf{h}_u^H \mathbf{s} + n_u. \tag{2.2}$$

For the system in Fig. 2.3, we can write a 2×2 channel matrix as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2}. \end{bmatrix} \tag{2.3}$$

In Eqn. 2.2, y_u is the received signal for user u, $\mathbf{h_u}$ is the 2×1 column vector from the downlink channel of MU-MIMO system in Fig. 2.3 and n_u is i.i.d complex Gaussian noise with σ^2 variance.

The fundamental issue with MU-MIMO is that users are often not able to cooperate together owing to their distributed nature. If we assume that the channel $\mathbf{H}$ is known at both the transmitter and receiver, while also assuming that $\mathbf{h}_u$ vectors for all users are orthogonal with each other, then we can express the message vector $\mathbf{s}$ to be transmitted as

$$\mathbf{s} = s_1 \mathbf{h}_1 + s_2 \mathbf{h}_2. \tag{2.4}$$

For this case, the received signal at user u can be expressed as

$$y_u = \|\mathbf{h}_u\|^2 s_1 + n_u. \tag{2.5}$$

This means that we can transmit independent data streams for each user without interference. For the case when $\mathbf{h}_u$ vectors are not orthogonal but linearly independent, we can project them into their orthogonal subspaces and express the message vector as

$$\mathbf{s} = s_1\mathbf{w}_1 + s_2\mathbf{w}_2. \tag{2.6}$$

In Eqn. 2.6, $\mathbf{w}_u$ vectors stand for receiver signature vectors that lie in the orthogonal subspace of all other $\mathbf{h}_i$ vectors except $\mathbf{h}_u$.

Then, the received signal for user u can be expressed as

$$y_u = \mathbf{h}_u^H \mathbf{w}_u s_u + n_u. \tag{2.7}$$

This means that we can also transmit streams without interference for a channel with linearly independent columns and when the channel is known at both the transmitter and the receiver. The multiplication by user signatures above is done at the receiver side and can be called receive beamforming. Denoting transmit signaling vectors also as $\mathbf{w}_u$, we can express the message signal transmitted from the base station as

$$\mathbf{s} = s_1\mathbf{w}_1 + s_2\mathbf{w}_2. \tag{2.8}$$

Then, the received signal at user u for the MU-MIMO system in Fig. 2.3 can be expressed as

$$y_u = \mathbf{h}_u^H \mathbf{w}_u s_u + \sum_{i=1, i\neq u}^{U} \mathbf{h}_u^H \mathbf{w}_i s_i + n_u. \tag{2.9}$$

The second term in Eqn. 2.9 stands for the interference between users. The selection of $\mathbf{w}_u$ vectors for transmit beamforming is important in order to decrease the effect of interference term in Eqn. 2.9. The topic of such a selection is further explained in Section 2.4.2.

2.2 Channel Models in MIMO

MIMO systems employ a large number of antennas at the transmitter and the receiver side. Transmitted signals experience various degradations during the transmission over the channel. The channel between the transmitter and receivers plays a crucial part in the performance of MIMO communication systems. It is an important challenge to properly model the channel which would demonstrate the effects on the signal according to the design needs. Channel models in MIMO systems can be explained in two parts in general which are analytical and physical channel models. Analytical models depend on the statistical characteristics of the channel and they are an easy way to reproduce channel characteristics for specified design needs. Nevertheless, they give limited insight into overall channel characteristics. Analytical models can be investigated by considering the statistical channel properties or propagation framework of the system. In physical models, several parameters such as angle of departure (AoD), angle of arrival (AoA), direction of departure (DoD), and direction of arrival (DoA) play an important part. Although they are useful for simulations, a shortcoming of physical modeling is that the presence of many physical parameters makes it difficult to verify for analytical calculations.

2.2.1 Analytical Models

As mentioned above, analytical channel models try to reproduce channel characteristics by considering the statistical properties or propagation framework. They are useful for analytical calculations during system design. Analytical channel modeling can be done by considering the channel correlations between the channel elements which describe the channel properties in a statistical way. In order to construct such a channel model, we consider channels with zero-mean circularly symmetric complex Gaussian distribution and express the channel covariance matrix for each user equipment as

$$\mathbf{R}_u = \mathbb{E}[\mathbf{h}_u \mathbf{h}_u^H]. \tag{2.10}$$

Adjusting all $\mathbf{h}_u$ vectors into a single vector so that $\mathbf{h} = vec(\mathbf{H})$, we can express the general channel covariance matrix as

$$\mathbf{R}_{\mathbf{H}} = \mathbb{E}[\mathbf{h}\mathbf{h}^H]. \tag{2.11}$$

Using Eqn. 2.11 and the fact the channel is zero-mean circularly symmetric complex Gaussian, we can express the channel distribution as

$$f(\mathbf{h}) = \frac{1}{\pi^{UN} \, |\mathbf{R}_{\mathbf{H}}|} exp(-\mathbf{h}^H \mathbf{R}_{\mathbf{H}}^{-1} \mathbf{h}), \tag{2.12}$$

for a MIMO system with N transmit antennas and U users with a single antenna each. Since we can express the distribution of the channel by only using the channel covariance matrix, it means that the spatial covariance matrix is a sufficient metric to describe the statistical properties of the channel, i.e., second-order statistics are sufficient to describe the channel [38]. We can describe the vectorized channel as

$$\mathbf{h} = \mathbf{R}_{\mathbf{H}}^{1/2} \mathbf{g}, \tag{2.13}$$

where $\mathbf{R}_{\mathbf{H}}^{1/2} \mathbf{R}_{\mathbf{H}}^{1/2} = \mathbf{R}_{\mathbf{H}}$ and $\mathbf{g}$ is a random $NU \times 1$ vector with zero-mean unit variance Gaussian i.i.d elements. This model could be simplified by assuming the user channels as uncorrelated which is useful in rich scattering environments with uniformly distributed components as explained in [38] or by assuming that the channel covariance matrix can be separated by the receiver and the transmitter correlation matrices as in the Kronecker model that was developed under the smart antenna technology in universal broadband wireless networks (SATURN) project as explained in

[39]. We consider a general correlated channel model as in Eqn. 2.13 for the initial point selection as will be explained in Section 3.3. Analytical modeling can also be done by considering the propagation framework as in the finite scatterer model and maximum entropy model. Since we don't use these models in our work, we don't delve into their explanation. One could refer to the works in [39] and [40] for more information about different analytical channel models.

2.2.2 Physical Models

As explained at the beginning of the section, physical channel modeling aims to characterize the channel in a MIMO system in terms of physical parameters such as AoDs, AoAs etc. Physical channel modeling can be done by considering three-dimensional measurements [41], geometrical characteristics of the environment such as the distribution of scatterers as in [42] or by considering the multi-path elements as clusters as in [43] which is called as the Saleh Valenzuela model. Saleh Valenzuela model that is first proposed in [43] considered the propagation only in the elevation plane, hence; it didn't contain information about AoAs and AoDs. This model was extended in the work [44] to also include information about AoDs and AoAs.

In [41], the discrete impulse response for this channel model is expressed as

$$h(t) = \sum_{k} \beta_k e^{j\theta_k} \delta(t - \tau). \tag{2.14}$$

In Eqn. 2.14, k stands for the multi-path number, β_k is the path gain for the multi-path k, τ_k is the propagation delay for the multi-path k and θ_k is the independent uniformly distributed phase shift over the $(0, 2\pi)$ interval.This model was also standardized in [45] as a narrowband model for a clustered environment as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i,l} \alpha_{i,l} \Lambda_r(\phi_{il}^r, \theta_{il}^r) \Lambda_t(\phi_{il}^t, \theta_{il}^t) \mathbf{a}_r(\phi_{il}^r, \theta_{il}^r) \mathbf{a}_t(\phi_{il}^t, \theta_{il}^t). \tag{2.15}$$

In Eqn. 2.15, i is the cluster number, l is the ray number, N_t is the transmit antenna number, N_r is the receiver antenna number, N_{cl} is the number of clusters, N_{ray} is the

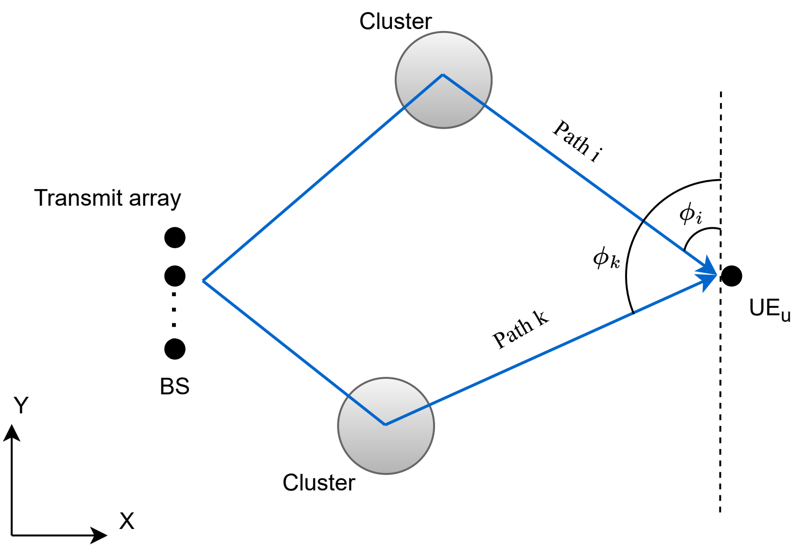


Figure 2.4: Visual representation of the clustered channel model for our simulations

number of rays, $\alpha_{i,l}$ is the complex path gain of the ray l in cluster i, ϕ_{il}^t and ϕ_{il}^r are the elevation AoDs and AoAs and θ_{il}^t and θ_{il}^r are the azimuth AoDs and AoAs respectively. $\Lambda_r(\phi_{il}^r, \theta_{il}^r)$ and $\Lambda_t(\phi_{il}^t, \theta_{il}^t)$ are functions that denote the receive and transmit antenna gains respectively. $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ are the array response vectors for the receiver and transmitter. In [45], the array response vector for a uniform linear array (ULA) with N antenna elements is expressed as

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}}[1, e^{j\frac{2\pi}{\lambda}d\sin\phi}, ..., e^{j(N-1)\frac{2\pi}{\lambda}d\sin\phi}]^T, \tag{2.16}$$

where λ is the wavelength and d is the antenna spacing which can be chosen as $\lambda/2$ for half wave spacing.

In our simulations, we use a simplified version of this clustered channel model as in the work [46]. The visual explanation of our model can be seen in Fig. 2.4.

In our channel model, we don't consider the distance between the users and the transmit antennas or the location and the number of clusters as in Fig. 2.4. We just assume that all users in the cell area experience the same amount of attenuation. The authors in [46] consider only AoDs while constructing the channel, however; we only consider AoAs for users while constructing the channel model between the users and transmit antennas. We express the channel model between a user u and the transmit antennas that we use in our simulations as

$$\mathbf{h}_u = \sum_{i=1}^{Lp} \epsilon_i \mathbf{a}(\phi_i), \tag{2.17}$$

where Lp denotes the total number of multi-paths, ϵ_i is a zero-mean unit variance complex Gaussian channel gain for ith path, and the array response vector for ULA with half-wave spacing at RRUs is given by

$$\mathbf{a}(\phi_\mathbf{i}) = [1 \;\; e^{j\pi \sin(\phi_i)} \;\; e^{j2\pi \sin(\phi_i)} ... \; e^{j(\bar{N}-1)\pi \sin(\phi_i)}]^T, \tag{2.18}$$

where ϕ_i is the Laplacian distributed AoA for the ith path, following a common approach to the model in [46].

2.3 Channel State Information in MIMO

The availability of channel state information (CSI) affects the channel capacity as explained in [47]. Therefore, it is an important concept when designing systems that align with present conditions. The different types of CSI include perfect CSI, statistical CSI and hybrid CSI which are explained below. For the discussions below, consider a narrowband flat fading channel as described in [48] such that

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.19}$$

where $\mathbf{y}$ and $\mathbf{x}$ denote the received and transmitted signal vectors respectively, $\mathbf{H}$ denotes the channel matrix and $\mathbf{n}$ denotes the noise vector with zero mean complex Gaussian $\mathcal{CN}(\mathbf{0}, \mathbf{Q})$ distribution where $\mathbf{Q}$ is the noise covariance matrix.

2.3.1 Perfect CSI

Perfect CSI or instantaneous CSI assumes that instantaneous channel information is known. This corresponds to the knowledge of impulse response such as in Eqn. 2.14 in wireless systems. In ideal conditions $\mathbf{H}$ in Eqn. 2.21 is known perfectly; however, the channel can be expressed due to estimation errors as [49] so that

$$vec(\mathbf{H}_{estimate}) = \mathcal{CN}(vec(\mathbf{H}), \mathbf{R}_{error}), \tag{2.20}$$

where $\mathbf{R}_{error}$ is the covariance matrix of estimation error and $\mathbf{H}_{estimate}$ is the estimated channel. The works such as [50] and [51] assume perfect CSI knowledge. In [50] a low-complexity hybrid precoding scheme is considered for the perfect CSI channel type. In [51], the authors consider a single-user massive MIMO system and propose a precoding scheme to reduce the number of active RF chains needed to approach capacity for a perfect CSI type. Although this type of CSI is unrealistic and complex to implement in real-time applications, it provides a useful approach for theoretical analysis and the results obtained from such works can be applied to an imperfect CSI type with some modifications.

2.3.2 Statistical CSI

In statistical CSI, the statistics or the long-term knowledge of the channel is known instead of the instantaneous information. The systems with this CSI type can employ channel models such as the correlated channel model that we described in Section 2.2.1. In [52], the channel information for this type of CSI is represented as

$$vec(\mathbf{H}) = \mathcal{CN}(\mathbf{0}, \mathbf{R}), \tag{2.21}$$

where $\mathbf{R}$ is the channel covariance matrix. Works in [53, 54] consider statistical CSI.

2.3.3 Hybrid CSI

In real-time applications, the most common way of obtaining CSI is done by sending pilot signals to the transmitters and estimating the channel at receivers. This estimation is later sent to transmitters via feedback. Because of these procedures, there can be imperfections in CSI due to noise, fading or there can be no information at all. Because of this, there can be several kinds of imperfect CSI types. One of these CSI types is the Hybrid CSI where only the distribution of channel and instantaneous observed channel information at the transmitters are known. Works in [55],[46] and [56] consider such a CSI type.

2.4 Concepts of SINR and Precoding in MIMO

2.4.1 SINR in MIMO

In this part, we explain the signal-to-interference-noise power ratio (SINR) notion which is used in rate expressions for MIMO systems. To better understand the concept, observe the downlink MU-MIMO system in Fig. 2.5.

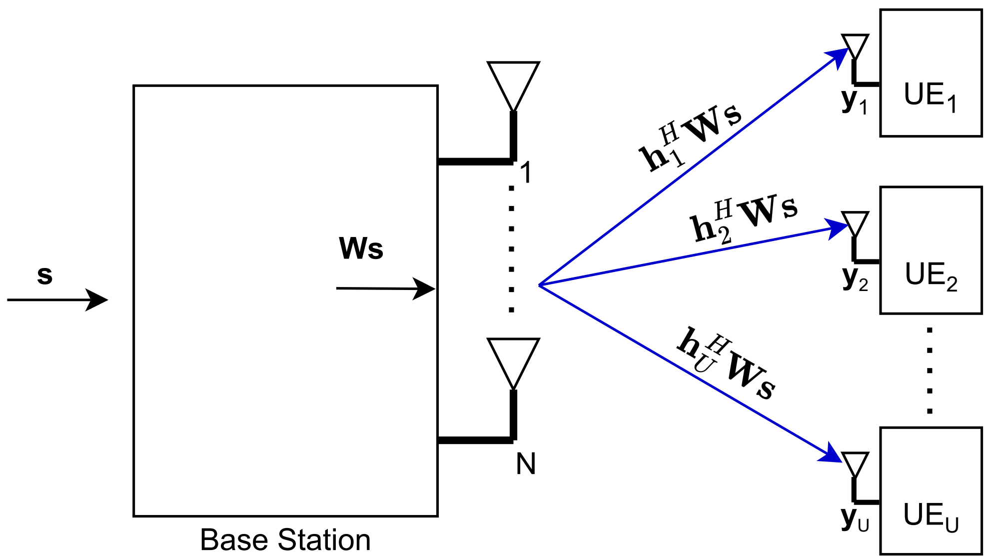


Figure 2.5: Downlink MU-MIMO system with single antenna users with interference

In Fig. 2.5, the zero-mean, unit variance message vector with independent elements $\mathbf{s} = [s_1 s_2 ... s_U]^T$ is weighted at BS with beamforming matrix $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 ... \mathbf{w}_U]$ where each $\mathbf{w}_u$ is the beamforming vector for user u and transmitted over a channel $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 ... \mathbf{h}_U]$ and received as y_u at user u. The received signal at user u can be expressed as

$$y_u = \mathbf{h}_u^H \mathbf{W} \mathbf{s} + n_u, \tag{2.22}$$

where n_u is the zero-mean complex Gaussian noise with σ^2 variance for all user elements. We have $\mathbf{h}_u^H$ as conjugate transpose for the mathematical representation in Eqn. 2.22 since we have defined $\mathbf{h}_u$ vectors as column vectors. The expression in Eqn. 2.22 can be expressed in a more expanded form as

$$y_u = \mathbf{h}_u^H \mathbf{w}_u s_u + \mathbf{h}_u^H \sum_{i=1, i \neq u}^{U} \mathbf{w}_i s_i + n_u. \tag{2.23}$$

The first term at the right-hand side of the equation in Eqn. 2.23 is the desired signal while the second term is the interference due to the presence of multiple users. Without interference, the metric that established the relation between the signal power and the noise power is called signal-to-noise ratio (SNR). The effect of interference is reflected in SINR, which establishes a relationship between the power of the desired signal and the power of the noise combined with the interference power as

$$\mathrm{SINR}_u = \frac{\left|\mathbf{h}_u^H \mathbf{w}_u\right|^2}{\sum_{i=1, i \neq u}^{U} \left|\mathbf{h}_u^H \mathbf{w}_i\right|^2 + \sigma^2}. \tag{2.24}$$

SINR is an important metric for the discussion of optimization problems in MIMO systems which is discussed in Section 2.6.

2.4.2 Precoding in MIMO

In the previous parts, we talked about the multi-user interference in MU-MIMO systems and SINR expression. We have also talked about how user rate is an important metric to improve the quality of a communication system. Assuming all the signal terms are Gaussian distributed, we can express the rate expression for user u in a downlink MU-MIMO system such as in Fig. 2.5 as

$$\text{Rate}_u = \log_2 [1 + \text{SINR}_u]. \tag{2.25}$$

Therefore, decreasing the effect of interference or increasing the effect of signal power in SINR expression (Eqn. 2.24) is beneficial in order to increase user rates. Precoding is a signal processing technique that can be used for maximizing user rate. The selection procedure of beamforming vectors $\mathbf{w}_u$ for user u that maximizes SINR_u in Eqn. 2.24 can be considered as precoding. In the following part, we describe the concept of beamforming and mention several precoding techniques from the literature.

2.4.2.1 Beamforming in MIMO

Beamforming is the process of shaping the radiation pattern of the signal during the transmission to concentrate in a particular direction in order to reduce interference and

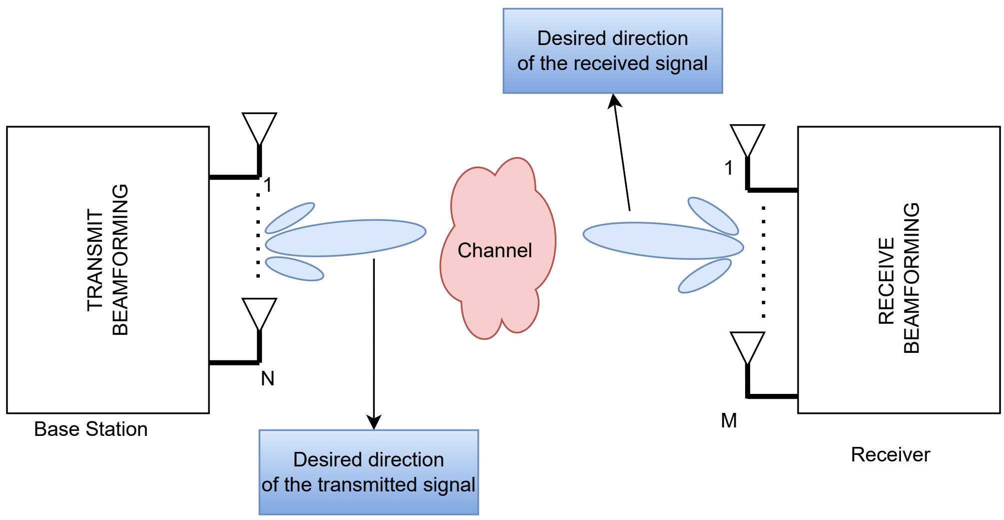


Figure 2.6: Concept of beamforming in MIMO

increase signal power at the receiver. The concept of shaping the signal is illustrated in Fig. 2.6.

The direction finding in Fig. 2.6 could be done by using phase shifters or using digital signal processing techniques in the digital domain. Manipulating the phase and amplitude of the signal via analog components is called analog beamforming and using digital signal processing techniques to manipulate the signal at the baseband is called digital beamforming.

2.4.2.1.1 Analog Beamforming

Analog beamforming refers to the manipulation of the phase and amplitude of the signals by using RF equipment in order to produce the required directivity as in Fig. 2.6. A simple phase-shifting approach to produce directivity is illustrated in Fig. 2.7.

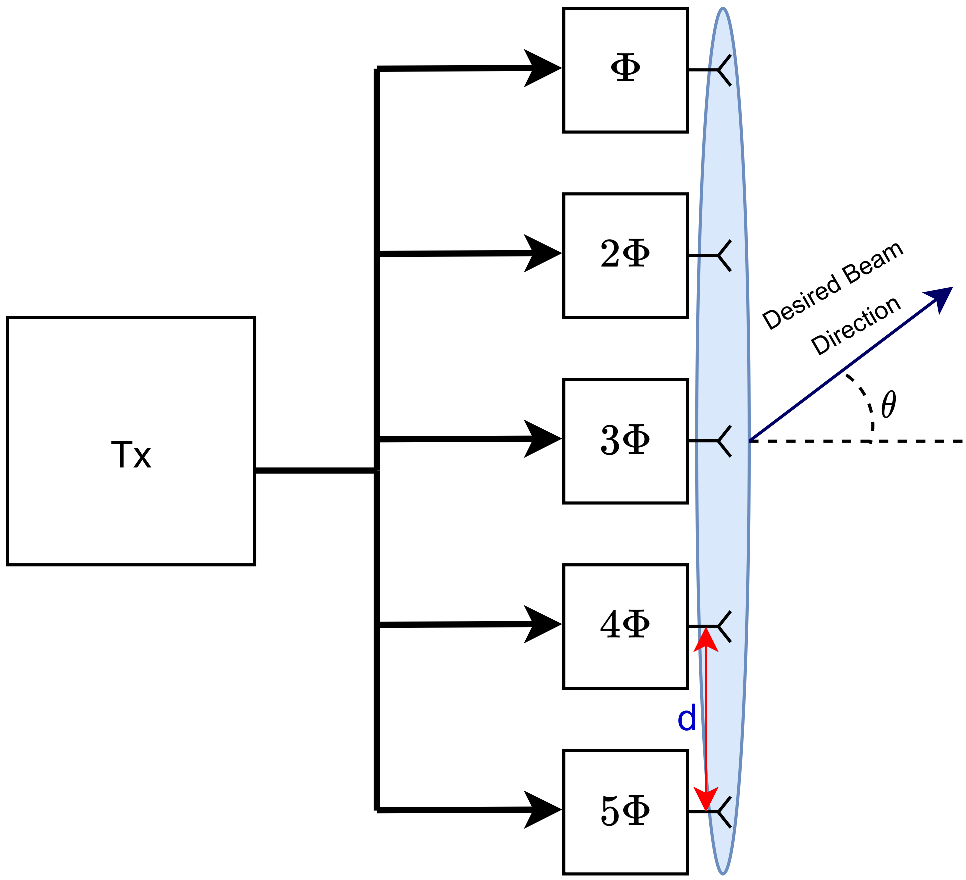


Figure 2.7: Analog beamforming by phase shifting

In Fig. 2.7, Φ is the phase angle and it can be selected as in Eqn. 2.16 so that $\Phi = \frac{2\pi}{\lambda} d \sin\theta$ where d is the antenna spacing and can be chosen as $\frac{\lambda}{2}$ for a wavelength of λ. The beam pattern and direction are adjusted by using the fact that in-phase RF signals are added constructively while out-of-phase RF signals are added destructively [57]. Although analog beamforming introduces a low complexity and low power consuming approach to produce directivity, some important challenges for this type of operation include difficulty to adapt beamforming weights depending on changing channel conditions and user locations [46].

2.4.2.1.2 Digital Beamforming

As explained in the previous part, systems using analog beamforming face challenges trying to adapt to changing channel conditions and scenarios where there are specific user locations. It is more preferable to use Digital beamforming for MU-MIMO systems, especially in environments where dynamic control and precise adjustment of beam patterns are important. The structure of the digital beamforming approach can be seen in Fig. 2.8.

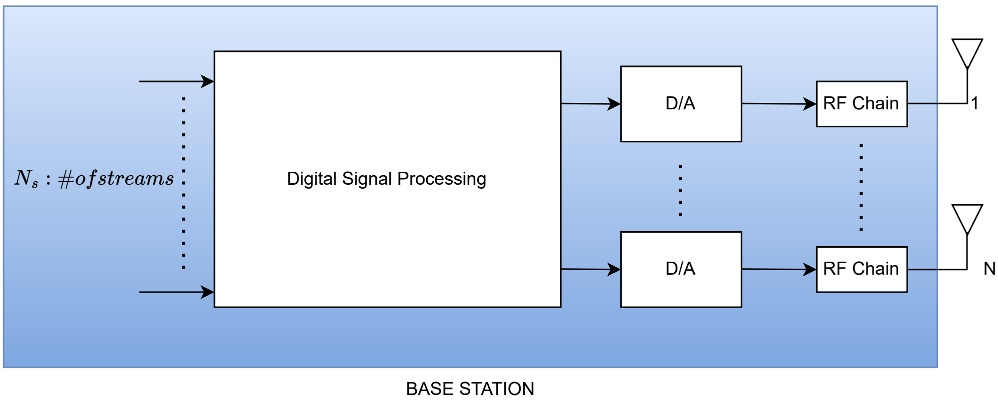


Figure 2.8: Digital beamforming in Downlink MIMO

As shown in the figure, the processing of signals is done digitally at the base station for digital beamforming. The signals are then converted to analog RF waves for transmission in the downlink operation. As it can be seen in Fig. 2.8, each antenna element requires its own RF chain for digital beamforming operation. Although digital beamformers can do the job of analog beamformers, the large number of antenna elements needed for massive MIMO systems could create cost problems for digital beamforming operations. Therefore, some functional splits could be performed at the base station as a measure to mitigate this problem [56]. Hybrid beamforming schemes have been proposed in the literature that utilizes digital beamforming for baseband operations and analog beamforming for RF operations by means of functional splits [56],[50], [51], [55], [46].

2.4.2.2 Precoding Algorithms from Literature

In this part, we describe existing precoding schemes that are widely used in the literature. The schemes that we mention are maximum ratio transmission (MRT), ZF and MMSE algorithms respectively. Since we only use MMSE algorithm in our thesis, we discuss ZF and MRT approaches briefly.

MRT precoding is a low-complexity algorithm that is used commonly. It basically chooses the beamforming weights from the conjugate transpose of user channels under the assumption that the transmitter has the perfect channel knowledge. This scheme could eliminate the interference if all the user channels are orthogonal to each other; however, it doesn't perform well compared to other schemes when this is not the case as it is shown in [58].

In the ZF precoding scheme, we try to project the beamforming vectors onto the orthogonal subspaces of the user channels other than the user that they are associated with. The precoding matrix for this scheme consists of the pseudo-inverse of the channel when the channel information is known at the transmitter. Although ZF precoding is simple to implement, it performs poorly, especially at low SNR as shown in [58].

For MMSE scheme, the orthogonal projection constraint in ZF precoding is relaxed and a "regularization" has been introduced into the channel inversion to improve performance as shown in the work [59]. This scheme is also called the regularized zero-forcing (RZF) method. The beamforming matrix for a downlink MU-MIMO system like in Fig. 2.5 for this scheme from [59] expressed as

$$\mathbf{W}_{MMSE,1} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \mathbf{I}_{\bar{N}})^{-1}, \tag{2.26}$$

for an MU-MIMO system with $\bar{N}$ transmit antennas.

By using the matrix identity [60]

$$\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I} + \mathbf{B}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} = \left(\mathbf{A} + \mathbf{B}\mathbf{B}^{\mathbf{T}}\right)^{-1}\mathbf{B}, \tag{2.27}$$

we can write an equivalent expression to Eqn. 2.26 as

$$\mathbf{W}_{MMSE,2} = (\mathbf{H}^H\mathbf{H} + \mathbf{I}_U)^{-1}\mathbf{H}^H, \tag{2.28}$$

for a MIMO system with U users.

The work in [58] shows that MMSE precoding outperforms MRT and ZF precoding schemes in terms of achievable rate performance. It is a useful scheme when having functional splits for MIMO precoding operations [56, 46]. In our work, we also consider a functional split for the precoding so that we employ two precoding matrices and use the MMSE method to construct one of the precoding matrices when the other precoding matrix is given.

2.5 Radio Access Networks and Development of C-RAN

In Section 2.1, we have talked about how the development of MIMO revolutionized wireless communication systems in order to meet ever-increasing demands. The evolution of communication systems with each generation has brought the need for more flexible communication network types [61]. In this section, we talk about different RAN types and the development of C-RAN through the discussion of C-RAN systems with wireless and wired fronthaul.

2.5.1 Massive MIMO with Traditional Base Station

In this subsection, massive MIMO systems with traditional base stations are described. The development of MIMO systems was described in Section 2.1. With massive MIMO, the number of antennas in a system could exceed 64 antennas [62]. The configuration of a typical base station with a massive MIMO system that serves multiple users with single receives antennas can be seen in Fig. 2.9. In this structure, both the baseband and RF processing operations are done in a co-located fashion and the modified data is sent to the passive antennas via a coaxial cable to be transmitted to a total of U users over a wireless channel.

The list of variables used for equations in this subsection and their explanation can be found in Table 2.1. We have changed the notation for the beamforming matrix from $\mathbf{W}$ which is commonly used in literature to $\mathbf{F}$ from here on in order to avoid confusion for the later parts of the thesis.

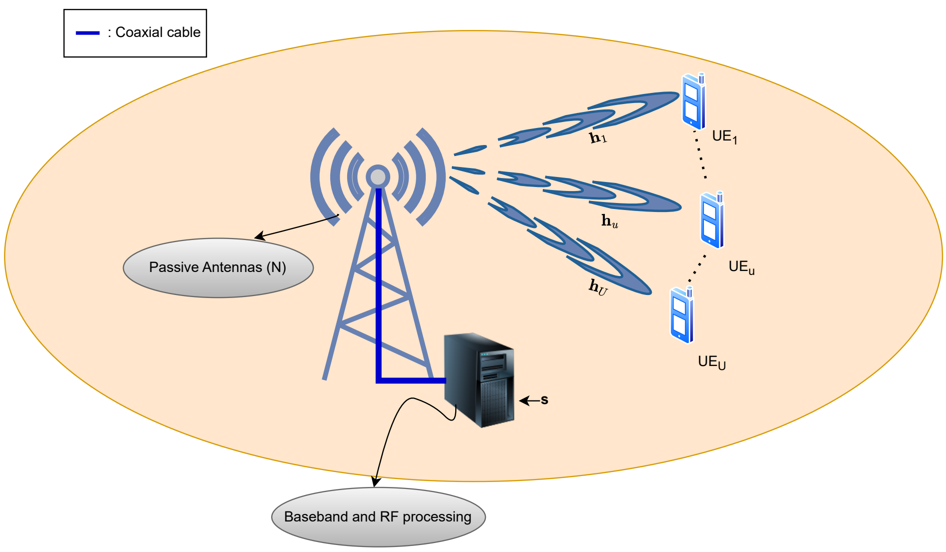


Figure 2.9: Traditional BS structure in a macro-cell area

In Fig. 2.9, the message symbols transmitted from BS to UEs in the downlink operation are denoted by a vector $\mathbf{s} = [s_1 s_2 ... s_U]^T \in \mathbf{C}^U$ which has generally independent elements with unit energy. The channel between BS and u-th user is denoted by a vector $\mathbf{h}_u \in \mathbf{C}^N$ which constructs the columns of total channel matrix as $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 ... \mathbf{h}_U] \in \mathbf{C}^{N \times U}$.

Table 2.1: List of notations for Subsection 2.5.1

Notations	Explanation
U	Number of UEs
N	Number of BS antennas
s_u	Message component for UE u
s	Message vector
n_u	Noise component for UE u
n	Noise vector
$\mathbf{h}_u$	Channel between BS and UE u
H	Channel between BS and UEs
$\mathbf{f}_u$	Beamforming vector of UE u
F	Beamforming matrix

The received signal for the system in Fig. 2.9 can be expressed as

$$\mathbf{y} = \mathbf{H}^H\mathbf{F}\mathbf{s} + \mathbf{n}, \tag{2.29}$$

where $\mathbf{F} = [\mathbf{f}_1\mathbf{f}_2...\mathbf{f}_U]$ is the beamforming matrix in BS and $\mathbf{n}$ is the noise vector which generally follows a circularly symmetric zero-mean Gaussian distribution with identity covariance matrix.

The received signal for the u-th user can be expressed as

$$y_u = \mathbf{h}_u^H\mathbf{f}_u s_u + \sum_{i=1,i\neq u}^{U} \mathbf{h}_u^H\mathbf{f}_i s_i + n_u, \tag{2.30}$$

where $\mathbf{f}_u$ is the beamforming vector between BS and user u and n_u is the corresponding noise component. In multi-user MIMO systems, many users' data streams are transmitted through the same frequency which improves capacity and throughput. Beamforming is used in massive MIMO systems by exploiting a large number of transmit antennas to direct the signal in a certain direction by weighting the signal elements of antenna arrays [63].

2.5.2 Distributed RAN

In this part, we describe a radio network that functions in a distributed manner for baseband and RF operations called distributed radio access network (D-RAN). The general configuration of a D-RAN architecture can be seen in Fig. 2.10.

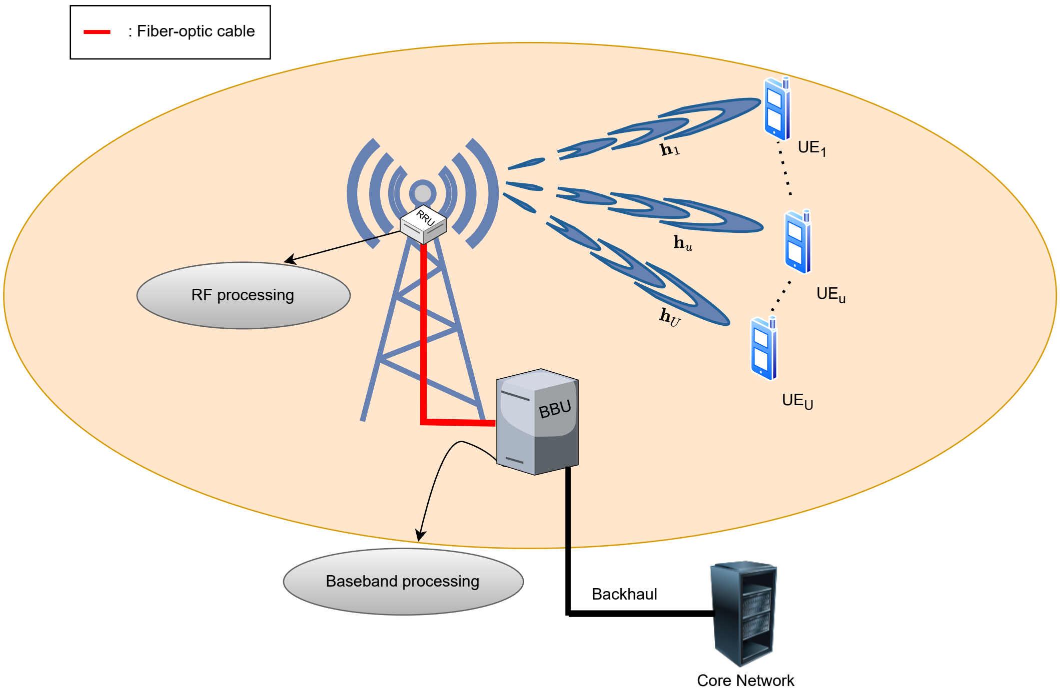


Figure 2.10: D-RAN structure

In Fig. 2.10, BBU where the baseband processing is done is connected to the CN via the link called backhaul which can be wired or wireless depending on the design needs. For this architecture, baseband operations and RF operations are separated between BBU and RRU which can be connected by a fiber-optic cable different from the structure in Fig. 2.9. This configuration allows for the splitting of the beamforming operation between BBU and RRU. With this separation, hybrid beamforming schemes where the digital beamforming is done at the BBU and analog beamforming is done at RRU is possible as shown in [46]. The received signal for a system where a BS is equipped with N transmit antennas that are linked to U users with single antennas as in Fig. 2.10 is expressed in Eqn. 2.31. The notations and their explanation for this part can be found in Table 2.2.

$$\mathbf{y} = \mathbf{H}^H \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} + \mathbf{n}. \tag{2.31}$$

Table 2.2: List of notations for Subsection 2.5.2

Notations	Explanation
U	Number of UEs
N	Number of BS antennas
s_u	Message component for UE u
s	Message vector
n_u	Noise component for UE u
n	Noise vector
$\mathbf{h}_u$	Channel between BS and UE u
H	Channel between BS and UEs
$\mathbf{f}_{BB,u}$	Digital Beamforming vector of UE u
$\mathbf{F}_{BB}$	Digital Beamforming matrix
$\mathbf{F}_{RF}$	Analog Beamforming matrix

Here, $\mathbf{F}_{RF}$ and $\mathbf{F}_{BB}$ are the beamforming matrices for the analog beamforming and digital beamforming operations which are explained in Subsection 2.4.2. Similar to the previous part, message vector $\mathbf{s}$ has independent and unit energy elements while noise vector $\mathbf{n}$ is circularly symmetric Gaussian with zero-mean and identity covariance matrix. The received signal for the user u can then be expressed as

$$y_u = \mathbf{h}_u^H \mathbf{F}_{RF} \mathbf{f}_{BB,u} s_u + \sum_{i=1, i \neq u}^{U} \mathbf{h}_u^H \mathbf{F}_{RF} \mathbf{f}_{BB,i} s_i + n_u. \tag{2.32}$$

As explained in [46], the splitting of functions between analog and digital beamforming reaches close to the fully digital beamforming performance with reduced cost. However, with the developments in hardware, future applications are more prone to be done digitally implemented which would be utilized in new types of network architectures for better results [64].

2.5.3 Centralized RAN

The research for more flexible radio access networks has led to the introduction of C-RAN where BBUs are centralized in a BBU pool or cloud. The BBU pool is connected to RRUs via the link called fronthaul which can be wired or wireless as shown in Fig. 1.3. In this section, we mainly talk about a standard C-RAN system with wired fronthaul links since we also use wired fronthaul C-RAN in our work. With the split between the BBU pool and RRUs in C-RAN, the precoding operation can be distributed between them to share the work and provide flexibility. With the improvements in hardware devices, it is also possible to conduct digital beamforming in RRUs along with BBU pools [64]. The basic structure of the system where multiple RRUs serve multiple users with single antennas in a distributed MIMO fashion can be seen in Fig. 2.11.

In Fig. 2.11, red lines correspond to the fronthaul links between the BBU pool and RRUs. The precoding is divided between the BBU pool and RRUs for this part as $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively. The list of variables used for equations in this subsection and their explanation can be found in Table 2.3.

The arrows in Fig. 2.11 correspond to the channels between RRU m and UE u which is denoted by $\mathbf{h}_{u,m} \in \mathbf{C}^N$. The channel between UE u and RRUs is denoted by $\mathbf{h}_u \in \mathbf{C}^{\bar{N}}$ which concatenates $\mathbf{h}_{u,m}$ vectors as a column vector and constructs the columns of channel matrix as $\mathbf{H} = [\mathbf{h}_1\mathbf{h}_2...\mathbf{h}_U]$. The received signal for the u-th user in such a system can be expressed as

$$y_u = \mathbf{h}_u^H \mathbf{F}_2 \mathbf{F}_1 \mathbf{s} + n_u, \tag{2.33}$$

where n_u is circularly symmetric unit variance zero-mean complex Gaussian noise element for UE u. In Eqn. 2.33, $\mathbf{F}_1$ and $\mathbf{s}$ contains elements for both user u and other users. This creates interference in the received signal. Received signal with this interference can be expressed as

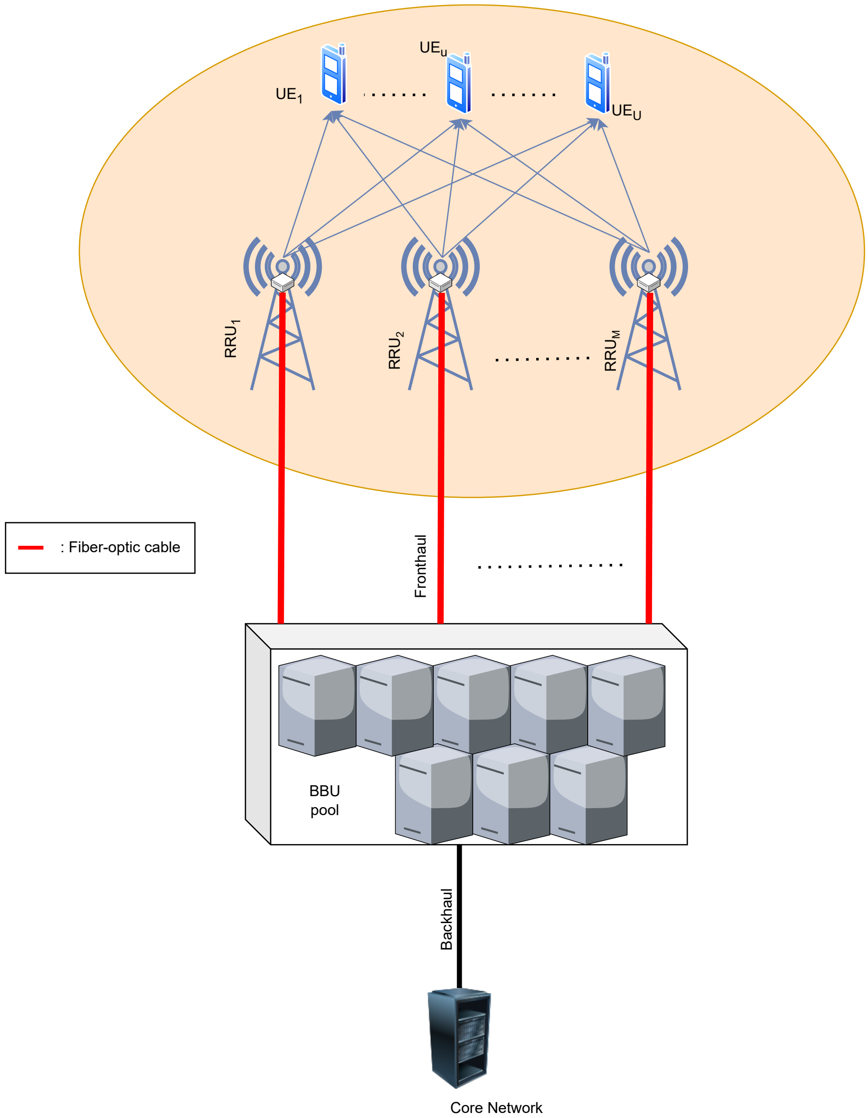


Figure 2.11: Wired fronthaul C-RAN structure

$$y_u = \underbrace{\mathbf{h}_u^H \mathbf{F}_2 \mathbf{f}_{1,u} s_u}_{Desired\, signal} + \underbrace{\mathbf{h}_u^H \mathbf{F}_2 \sum_{i=1, i\neq u}^{U} \mathbf{f}_{1,i} s_i}_{interference} + \underbrace{n_u}_{Thermal\, noise} . \tag{2.34}$$

Here, the desired signal is composed of the user message signal and the beamforming vector for the u^{th} user. The interference component contains signal components and beamforming vectors for other users. In our work, we consider a wired fronthaul C-RAN system with capacity limitations on the fronthaul links. Wireless fronthaul C-RAN is out of scope for this thesis, however; interested readers may refer to the

Table 2.3: List of notations for Subsection 2.5.3

Notations	Explanation
U	Number of UEs
M	Number of RRUs
N	Number of antennas per RRU
$\bar{N}$	Total number of antennas
s_u	Message component for UEu
s	Message vector
n_u	Noise component for UE u
$\mathbf{h}_{u,m}$	Channel vector for RRU m and UE u
$\mathbf{h}_u$	Channel vector for UE u
H	Channel between all transmit antennas and UEs
$\mathbf{f}_{1,u}$	Beamforming vector of UE u at BBU pool
$\mathbf{F}_1$	Beamforming matrix for BBU pool
$\mathbf{F}_2$	Total beamforming matrix for RRUs

work in [65].

2.6 Optimization Criteria

The beamforming structures described in the previous parts can be designed to satisfy several optimization criteria. The aim of the optimization could be maximizing user rates to improve service quality, minimizing transmission power to use energy more efficiently, choosing which users to serve during a certain time or jointly designing an optimization problem that takes multiple criteria into account. In this section, we describe some optimization criteria that are widely used in literature.

2.6.1 Rate maximization

One of the goals of constructing systems such as massive MIMO systems is to increase the data rates for users. For this criterion, the rate can be maximized under constraints such as power constraints or capacity constraints. Some approaches to increase the data rate in the system would be to consider the SINR expression and decrease the interference, or consider maximizing the summation of the rate expressions for all users. Several means to maximize the rate are mentioned below.

2.6.1.1 SINR maximization

In Section 2.4.2, we have mentioned that the rate of a user for massive MIMO systems is proportional to their SINR values. Therefore, increasing SINR for a specific user also increases the rate for that user. However, for the multi-user case, constructing a beamforming scheme to increase the SINR value for a specific user may cause degradation for another user and not necessarily increase the sum rate. A common approach that is used in literature against this kind of situation is to maximize the minimum SINR of the users. The papers [66] and [67] consider this approach while designing beamforming matrices.

2.6.1.2 Sum Rate Maximization

The rate expression for a single user is given in Eqn. 2.25. Summation of these rates for all users gives us the maximum amount of data that we can transfer per second on a unit bandwidth which is also called the spectral efficiency. The sum rate is an important indicator of the performance of a system. The works [56], [68] and [69] consider this maximization criterion under such constraints. In our work, we also study a sum rate maximization problem under limited fronthaul capacity constraints.

2.6.2 Transmit Power Minimization

Optimizing power allocation and designing the precoding matrix accordingly can improve the energy efficiency of a communication system. Therefore, the selection of beamforming weights minimizing transmit power and optimized power allocation under a lower bounded SINR value has been under research in the literature. The works [70] and [71] consider such an optimization criterion under limited SINR for

MIMO systems. We don't use this optimization criterion in our work but it may be an interesting field for future studies.

2.6.3 Admission Control

The performance of a system could be improved by serving only a limited number of users at a time. This is done by selecting the maximum number of users that can be served by considering an SINR threshold and rejecting the users below this threshold. In the literature, an example of an approach considering this criterion can be found in [72]. Although the user admission criterion is not considered for our work, it might be considered in future works to get better insight into our topic.

2.6.4 Joint Optimization

Another type of optimization criterion could be to include several criteria such as power minimization and admission control etc. in a single objective function and jointly optimize them under the specified design needs. The work in [73] considers a transmit power minimization problem that is jointly optimized with the number of users that are served. It might be used for our approach in future works where jointly optimize the criteria that we have mentioned previously.

CHAPTER 3

SUM RATE MAXIMIZATION UNDER UNLIMITED FRONTHAUL CAPACITY

In this chapter, we address the sum rate maximization problem without considering the fronthaul capacity limitations. We first examine a downlink wired fronthaul C-RAN system that is explained in Section 2.5.3 and describe the problem definition. Then, we explain our proposed method for the solution of the sum rate maximization problem without fronthaul capacity constraints. We conclude the chapter by describing an approach from the literature that is important for initial point selection for the proposed approach and discussing the results by presenting our key findings.

3.1 Problem Formulation

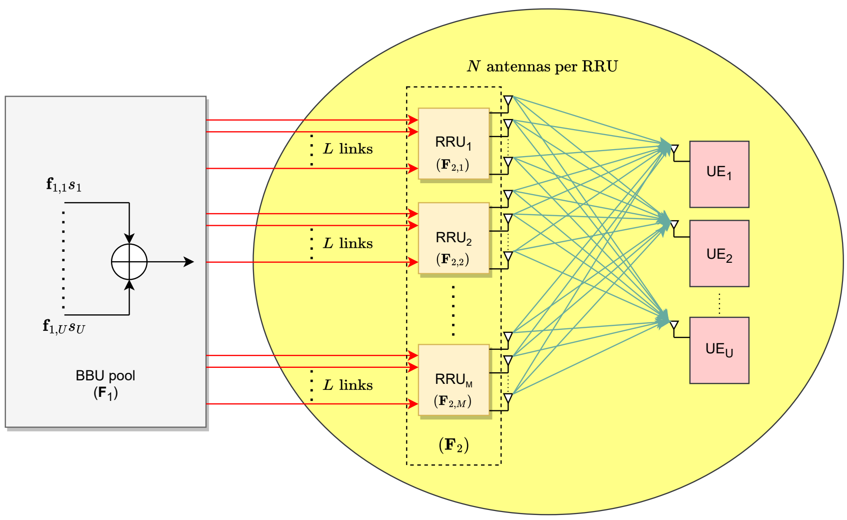


Figure 3.1: Block structure of downlink wired fronthaul C-RAN with unlimited capacity

For the problem definition, we consider a downlink C-RAN system with L active wired fronthaul links between each RRU, M RRUs in total, and a BBU pool. The block structure for the considered system can be seen in Fig. 3.1.

In our work, we split the beamforming operation between the BBU pool and RRUs as $\mathbf{F}_1 \in \mathbb{C}^{(LM)\times U}$ for the BBU pool and $\mathbf{F}_2 \in \mathbb{C}^{\bar{N}\times(LM)}$ for RRUs, where $\mathbf{F}_{2,m} \in \mathbb{C}^{N\times L}$ is the beamforming matrix for RRU m. To incorporate all $\mathbf{F}_{2,m}$ matrices in a single matrix, we express $\mathbf{F}_2$ in a block diagonal form similar to [56] as

$$\mathbf{F}_2 = \begin{bmatrix} \mathbf{F}_{2,1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{2,2} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & . & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & . & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{F}_{2,M} \end{bmatrix}. \tag{3.1}$$

We construct the beamforming matrix $\mathbf{F}_2$ at RRUs only to find the initial point for our proposed method as we will describe in Section 3.3 and we construct it at the BBU pool with our gradient-based method as we will describe in the next section and send the relevant $\mathbf{F}_{2,m}$ matrices to corresponding RRUs from the BBU pool after gradient-descent process is complete. After finding an $\mathbf{F}_2$, we find $\mathbf{F}_1$ by using an MMSE approach as in Eqn. 3.2.

After beamforming in the BBU pool and RRUs, the message symbols are transmitted over a wireless channel $\mathbf{H} = [\mathbf{h}_1\mathbf{h}_2...\mathbf{h}_U] \in \mathbb{C}^{\bar{N}\times U}$ with zero-mean and circularly symmetric Gaussian elements similar to the model in Section 2.2.1. Given the beamforming matrix $\mathbf{F}_2$ in RRUs that is found as will be explained in Section 3.2 and Section 3.3, we express the beamforming matrix $\mathbf{F}_1$ in the BBU pool in an MMSE form similar to Eqn. 2.28 in Section 2.4.2.2 by assuming that, instead of the channel $\mathbf{H}$, the BBU pool utilizes the instantaneous processed channel after precoding in RRUs where $\bar{\mathbf{H}} = \mathbf{H}^H\mathbf{F}_2$ so that the precoding matrix in the BBU pool before normalization is expressed as

$$\bar{\mathbf{F}}_1 = \left[\mathbf{F}_2^H\mathbf{H}\mathbf{H}^H\mathbf{F}_2 + U\frac{\sigma^2}{P_{total}}\mathbf{I}_{LM\times LM}\right]^{-1}\mathbf{F}_2^H\mathbf{H}\ . \tag{3.2}$$

In our work, we assume that the power transmitted to user u from RRUs is equal for all users and divided equally from the total power budget P_{total} Watts. Therefore, we choose the beamforming matrices that satisfy $\|\mathbf{F_2 f_{1,u}}\|^2 = 1$. To have this equality, we scale the columns of $\bar{\mathbf{F}}_1$ with $\mathbf{F}_2$ so that

$$\mathbf{f}_{1,u} = \frac{\bar{\mathbf{f}}_{1,u}}{\sqrt{(\mathbf{F}_2\bar{\mathbf{f}}_{1,u})^H \mathbf{F}_2 \bar{\mathbf{f}}_{1,u}}}, \tag{3.3}$$

for all u values and obtain $\mathbf{F}_1$.

Placement of each $f_{1,(m,l),u}$ in $\mathbf{F}_1$ is shown below as

$$\mathbf{F}_1 = \begin{bmatrix} f_{1,(1,1),1} & f_{1,(1,1),2} & \cdots & \cdots & f_{1,(1,1),U} \\ f_{1,(1,2),1} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ f_{1,(1,L),1} & \cdots & \cdots & \cdots & \vdots \\ f_{1,(2,1),1} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ f_{1,(M,L),1} & f_{1,(M,L),2} & \cdots & \cdots & f_{1,(M,L),U} \end{bmatrix}. \tag{3.4}$$

A list of notations for the considered approach can be found in Table 3.1.

We first consider the received signal for UE u that is jointly served by all RRUs in a cell area in a distributed MIMO fashion similar to y_u in Eqn. 2.34 with a message vector s with zero mean and unit energy message symbols. We apply a total power budget of P_{total} to the system and divide the total power equally among the users as mentioned before. Then, the received signal for the UE u in the system for this part is expressed as

$$y_u = \underbrace{\sqrt{\frac{P_{total}}{U}}\mathbf{h}_u^H \mathbf{F}_2 \mathbf{f}_{1,u} s_u}_{Desired\,signal} + \underbrace{\sqrt{\frac{P_{total}}{U}}\mathbf{h}_u^H \mathbf{F}_2 \sum_{i=1, i\neq u}^{U} \mathbf{f}_{1,i} s_i}_{interference} + \underbrace{n_u}_{Thermal\,noise}. \tag{3.5}$$

Table 3.1: List of notations for Sections 3.1 and 3.2

Notations	Explanation
U	Number of UEs
M	Number of RRUs
L	Number of fronthaul links per-RRU
N	Number of antennas per RRU
$\bar{N}$	Total number of antennas
σ^2	Noise variance
s_u	Message component for UE u
$\mathbf{s}$	Message vector
$\mathbf{h}_{u,m}$	Channel vector for RRU m and UE u
$\mathbf{h}_u$	Channel vector for UE u
$\mathbf{H}$	Channel between all transmit antennas and UEs
$f_{1,(m,l),u}$	Beamforming element of $\mathbf{F}_1$ corresponding to l^{th} link of RRU m and UE u
$\mathbf{F}_{1,m}$	Beamforming matrix in the BBU pool corresponding to RRU m
$\mathbf{F}_1$	Beamforming matrix in the BBU pool
$\mathbf{f}_{2,m,l}$	l^{th} column of $\mathbf{F}_{2,m}$
$\mathbf{f}_{2,m}$	Vectorized form of $\mathbf{F}_{2,m}$
$\mathbf{F}_{2,m}$	Beamforming matrix for RRU m
$\mathbf{F}_2$	Aggregate Beamforming matrix in RRUs

Then SINR expression for the UE u is expressed as

$$\text{SINR}_u = \frac{\frac{P_{total}}{U}\left|\mathbf{h}_u^H \mathbf{F}_2 \mathbf{f}_{1,u}\right|^2}{\frac{P_{total}}{U}\sum_{i=1,i\neq u}^{U}\left|\mathbf{h}_u^H \mathbf{F}_2 \mathbf{f}_{1,i}\right|^2 + \sigma^2}, \tag{3.6}$$

where σ^2 is the variance of the zero-mean circularly symmetric Gaussian noise component in Eqn. 3.5 that is equal for all users and power terms are expressed in Watts inside SINR. Then, we can express the rate for UE u as

$$\text{Rate}_u = \log_2(1 + \text{SINR}_u). \tag{3.7}$$

Our maximization criterion is to find a matrix $\mathbf{F}_2$ that maximizes the sum rate of users. Without fronthaul capacity constraint, the maximization problem can be expressed as

$$(\text{P}_1) \ \max_{\mathbf{F}_2} \textstyle\sum_{u=1}^{U} Rate_u, \tag{3.8}$$

$$\text{s.t.} \ \textstyle\sum_{m=1}^{M} tr(\frac{P_{total}}{U}(\mathbf{F}_{2,m}\mathbf{F}_{1,m})^H \mathbf{F}_{2,m}\mathbf{F}_{1,m}) = P_{total} \quad . \tag{3.9}$$

Here, we utilize all of our available transmit power budget as can be seen in Eqn 3.9, and RRUs distribute the available power budget arbitrarily among themselves.

3.2 Proposed Approach for Unlimited Fronthaul Capacity

In this section, we propose a gradient-based approach to solve (P1). As we have mentioned in Section 1.1, we try to find a differentiable solution to the sum rate maximization problem. In this thesis, we consider a gradient ascent based approach. However, one should keep in mind that this is equivalent to the gradient descent approach that is commonly found in literature if we consider the negative of our maximization problem and decrement the step size for every update. The minimum value for such a gradient descent based approach would be equivalent to the maximum value in the gradient ascent approach. Therefore, we express the equations in a gradient descent way since it is more common in the literature. To make the problem differentiable, we first need to transform the SINR_u expression in Eqn. 3.6 into an infinitely differentiable form with respect to the elements of $\mathbf{F}_2$ and express the constraints with the maximization problem in a single differentiable expression. Since our idea of shifting the constraints into the main problem was based on the smooth constraint-free approximation (SCFA) approach in [73], we denote our approach as SCFA during simulations. Due to the block diagonal nature of $\mathbf{F}_2$, it is hard to differentiate Eqn. 3.6 while $\mathbf{F}_2$ is in a matrix state. Therefore, we express $\mathbf{F}_2$ in a vector form and denote it as f_2 and state Eqn. 3.6 in an equivalent expression consisting of f_2. We describe the columns of $\mathbf{F}_{2,m}$ as

$$\mathbf{F}_{2,m} = [\mathbf{f}_{2,m,1}\mathbf{f}_{2,m,2}...\mathbf{f}_{2,m,L}], \tag{3.10}$$

and we concatenate the columns of $\mathbf{F}_{2,m}$ as

$$\mathbf{f}_{2,m} = [\mathbf{f}_{2,m,1}^T\mathbf{f}_{2,m,2}^T...\mathbf{f}_{2,m,L}^T]^T. \tag{3.11}$$

Vectorized form of $\mathbf{F}_2$ consisting of its nonzero elements is then expressed as

$$\mathbf{f}_2 = [\mathbf{f}_{2,1}^T\mathbf{f}_{2,2}^T...\mathbf{f}_{2,M}^T]^T. \tag{3.12}$$

We then express SINR$_u$ in Eqn. 3.6 with vectorized $\mathbf{f}_2$ as

$$\text{SINR}_u = \frac{\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_u^H\mathbf{b}_u\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2}, \tag{3.13}$$

where

$$\begin{aligned}
\mathbf{A}_u &= \text{diag}(\underbrace{\mathbf{h}_{u,1}^H}_{L-times}, \underbrace{\mathbf{h}_{u,2}^H}_{L-times}, ..., \underbrace{\mathbf{h}_{u,M}^H}_{L-times}), \\
\mathbf{b}_u &= [\underbrace{f_{1,(1,1),u}}_{N-times}, \underbrace{f_{1,(1,2),u}}_{N-times}, ..., \underbrace{f_{1,(M,L),u}}_{N-times}].
\end{aligned} \tag{3.14}$$

Then, we can express the rate in Eqn. 3.7 as

$$\begin{aligned}
\text{Rate}_u &= \log_2(1 + \frac{\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_u^H\mathbf{b}_u\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2}) \\
&= \log_2(\frac{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2 + \mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_u^H\mathbf{b}_u\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2}) \\
&= \log_2(\sum_{i=1}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2) \\
&- \log_2(\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2).
\end{aligned} \tag{3.15}$$

Equivalence of Eqn. 3.7 and Eqn. 3.15 is shown in Appendix A. The expression in Eqn. 3.15 is infinitely differentiable with respect to $\mathbf{f}_2^H$. We assume perfect CSI for gradient operations. Now, we consider a per-RRU power transmit constraint $tr((\mathbf{F}_{2,m}\bar{\mathbf{F}}_{1,m})^H\mathbf{F}_{2,m}\bar{\mathbf{F}}_{1,m}) \leq P_{t,RRU}$, where $P_{t,RRU} = \frac{P_{total}}{M}$, and transform it into a form that is differentiable with respect to $\mathbf{f}_2^H$. We can express this constraint in a differentiable form as

$$\frac{tr((\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T)^H\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T)}{L} \leq P_{t,RRU}, \tag{3.16}$$

where

$$\begin{aligned} \mathbf{C}_m &= diag(\mathbf{0}_{(m-1)L}\mathbf{1}_L\mathbf{0}_{(M-m)L}), \\ \mathbf{D}_m &= \mathbf{C}_m\bar{\mathbf{F}}_1, \\ \mathbf{E}_m &= diag(\mathbf{0}_{(m-1)NL}\mathbf{1}_{NL}\mathbf{0}_{(M-m)NL}). \end{aligned} \tag{3.17}$$

The equivalence of the above expression and Eqn. 3.9 is explained in Appendix B. Now, we use the fact below to shift the per-RRU power transmit constraint into the equation of the main problem.

$$\text{Fact 1}: e^{\frac{x}{\alpha}} \rightarrow \begin{cases} 0\,, \text{if} \;\; x < 0 \\ \infty\,, \text{if} \;\; x > 0 \end{cases} \text{as} \;\; \alpha \rightarrow 0^+. \tag{3.18}$$

To use the fact in Eqn. 3.18, we try to express the constraint in Eqn. 3.16 in an exponential form. To achieve this, we consider the expression below

$$\begin{aligned} c_1 &= \eta_1 \sum_{m=1}^{M} exp(\frac{tr((\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T)^H\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T) - LP_{t,RRU}}{\alpha L}) \\ &= \eta_1 \sum_{m=1}^{M} exp(\frac{tr((vec(\mathbf{D}_m)^T)^H\mathbf{f}_2^H\mathbf{E}_m^H\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T) - LP_{t,RRU}}{\alpha L}) \end{aligned} \tag{3.19}$$

When we consider the expressions in Eqn. 3.18 and Eqn. 3.19, Eqn. 3.19 tends to 0 if the per-RRU power constraint in Eqn. 3.9 is satisfied and tends to (∞) if the constraint is not satisfied for a positive valued η_1. Therefore, when we transfer the constraint into the main maximization problem and the constraint is satisfied, the expression in Eqn. 3.19 is 0 for small positive α. When the constraint is not satisfied, the expression in Eqn. 3.19 is ∞ and since we consider a maximization (minimization in the descent sense) problem, an $\mathbf{f}_2$ that would yield this result would not be considered as a solution. Therefore, this operation would transfer the constraint into the main problem while also maintaining its function.

After transferring the constraint into the main problem, we can express the term to be maximized in Eqn. 3.8 with constraint in 3.9 as below.

$$
\begin{aligned}
\mathrm{T}_1 &= -\sum_{u=1}^{U}[\log_2(\sum_{i=1}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2+\frac{U}{P_{total}}\sigma^2) \\
&- \log_2(\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2+\frac{U}{P_{total}}\sigma^2)] \\
&+ \eta_1\sum_{m=1}^{M}exp(\frac{tr((vec(\mathbf{D}_m)^T)^H\mathbf{f}_2^H\mathbf{E}_m^H\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T)-LP_{t,RRU}}{\alpha L})
\end{aligned}
\tag{3.20}
$$

Now, the expression in Eqn. 3.20 which corresponds to the maximization problem is differentiable with respect to $\mathbf{f}_2^H$. We can express the gradient of T_1 in Eqn. 3.20 with respect to $\mathbf{f}_2^H$ as

$$
\begin{aligned}
&\nabla = \\
&-\sum_{u=1}^{U}(\frac{\sum_{i=1}^{U}\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2+\frac{U}{P_{total}}\sigma^2} \\
&-\frac{\sum_{i=1,i\neq u}^{U}\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2+\frac{U}{P_{total}}\sigma^2}) \\
&+\eta_1\sum_{m=1}^{M}\frac{conj(vec(\mathbf{D}_m)^T)vec(\mathbf{D}_m)\mathbf{f}_2^T\mathbf{E}_m^T conj(\mathbf{E}_m)}{\alpha L}exp(\Gamma_m),
\end{aligned}
\tag{3.21}
$$

where Γ_m is defined as

$$\Gamma_m = \frac{tr((vec(\mathbf{D}_m)^T)^H \mathbf{f}_2^H \mathbf{E}_m^H \mathbf{E}_m \mathbf{f}_2 vec(\mathbf{D}_m)^T) - LP_{t,RRU}}{\alpha L}. \quad (3.22)$$

Our proposed approach is a gradient-based iterative method where the gradient is taken with respect to $\mathbf{f}_2$ and it is updated as

$$\mathbf{f}_2^{(t+1)} = \mathbf{f}_2^{(t)} - \mu^{(t)} \nabla(\mathbf{f}_2^{(t)}). \quad (3.23)$$

In Eqn. 3.23, we update $\mathbf{f}_2$ in a decremental way with a vector $\nabla(.)$ which is the gradient vector in Eqn. 3.21 that takes the value of $\mathbf{f}_2$ at iteration t as input. Here, $\mu^{(t)}$ is the step size which should be adjusted properly for convergence. It can be selected as a constant or it can vary during each iteration for better convergence. One of the methods to choose $\mu^{(t)}$ for each iteration is explained in [74] which results in the below expression for the step size in our approach.

$$\mu^{(t)} = \frac{(\delta \mathbf{f}_2^{(t)})^H \delta \nabla^{(t)}}{(\delta \nabla^{(t)})^H \delta \nabla^{(t)}}, \quad (3.24)$$

where

$$\begin{aligned} \delta \mathbf{f}_2^{(t)} = &\quad \mathbf{f}_2^{(t)} - \mathbf{f}_2^{(t-1)}, \\ \delta \nabla^{(t)} = &\quad \nabla(\mathbf{f}_2^{(t)}) - \nabla(\mathbf{f}_2^{(t-1)}). \end{aligned} \quad (3.25)$$

The description of the proposed approach in algorithm form is given in Algorithm 1. The verbal explanation of Algorithm 1 is that we first initialize the vector $\mathbf{f}_2^{(0)}$, which also determines $\mathbf{F}_2^{(0)}$ by arranging the corresponding elements of $\mathbf{f}_2^{(0)}$, with the approach that is described in Section 3.3. Then, we find the corresponding $\mathbf{F}_1^{(0)}$ by using Eqn. 3.2 and calculate $\mathrm{T}_1^{(0)}$ from Eqn. 3.20. Next, we update $\mathbf{f}_2$ by using Eqn. 3.23 and calculate $\mathbf{F}_1^{(1)}$ with updated $\mathbf{F}_2$. We continue the iterations until $t = t_{max}$ as specified in Algorithm 1 or the difference between $\mathrm{T}_1^{(t+1)}$ at iteration (t+1) and the value of T_1 from previous iterations up to 15 previous iterations is greater than 10^{-5}.

If this difference is less than 10^{-5}, we assume that the convergence is achieved and stop the algorithm. We use the final $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices as the beamforming matrices for our approach.

The selection of the starting point, that is $\mathbf{f}_2^{(0)}$, is an important point to consider for a better convergence. In the next subsection, we describe a method from the literature for the initial point selection.

Algorithm 1 Finding $\mathbf{F}_2$ that maximizes negative of 3.20 with proposed approach

Step 1, Initialization: Set the initial point $\mathbf{f}_2^{(0)}$ as found by the approach in Section 3.3. Define: $\alpha = 10^{-4}, \eta_1 = 1.5, \Delta = 15, t_{max} = 1500$.

Step 2, Gradient Descent:

for $t = 0$ to t_{max} **do**
    Find gradient $\nabla(\mathbf{f}_2^{(t)})$
    Find step size $\mu^{(t)}$ from Eqn. 3.24
    Update beamformer : $\mathbf{f}_2^{(t+1)} = \mathbf{f}_2^{(t)} - \mu^{(t)}\nabla(\mathbf{f}_2^{(t)})$
    Set corresponding elements of $\mathbf{f}_2^{(t+1)}$ as columns of $\mathbf{F}_{2,m}^{(t+1)}$
    Update $\mathbf{F}_1^{(t+1)}$ according to Eqn. 3.2 and Eqn. 3.3
    if ($\max\limits_{t-\Delta \le t_s \le t} \left| \mathrm{T}_1^{(t+1)} - \mathrm{T}_1^{(t)} \right| < 10^{-5}$) **then**
        $t_{max} = t_s$
        break
    end if
end for
$\mathbf{f}_2 = \mathbf{f}_2^{(t_{max})}$
Set corresponding elements of $\mathbf{f}_2$ as columns of $\mathbf{F}_{2,m}$
$\mathbf{F}_1 = \mathbf{F}_1^{(t_{max})}$
return $\mathbf{F}_2, \mathbf{F}_1$

3.3 Exploiting Channel Covariance Matrices for Initial Point Selection

The proposed approach in the previous section updates $\mathbf{f}_2^{(t)}$ at each iteration. We can select the starting point $\mathbf{f}_2^{(0)}$ as random for convenience. However, selecting a specific initial point with proper considerations could lead to faster convergence. Therefore, the selection of an initial value $\mathbf{f}_2^{(0)}$ is an important aspect that affects convergence in the proposed approach. In this section, we describe the selection of such an initial point. The selection procedure is based on the works from the literature [56], [46] for the proposed approach described in Section 3.2. The system that is considered for the work in [46] is a BS with separated BBU and RRU as in Fig. 2.10. The authors of the work propose a hybrid beamforming technique based on dominant eigenvectors of channel covariance matrices for signal-to-leakage ratio (SLNR) maximization. The authors in [56] use Proposition 2 from [46] to derive a beamforming matrix selection approach for the SINR maximization problem for a wired C-RAN system with single RRU and unlimited fronthaul capacity assumptions.

For our work, in light of the approaches that are mentioned above, we start the procedure from the channel covariance matrix expression for user u and RRU m as

$$\mathbf{R}_{u,m} = \mathbb{E}[\mathbf{h}_{u,m}\mathbf{h}_{u,m}^H] \in \mathbb{C}^{N\times N}. \tag{3.26}$$

Then, we sum the covariance matrices for all users which results in the below aggregate covariance matrix.

$$\mathbf{R}_m = \sum_{u=1}^{U} \mathbf{R}_{u,m}. \tag{3.27}$$

Now, we consider L largest eigenvectors of the expression in 3.27 and denote this new matrix as $\mathbf{V}_m$. We also define a new matrix $\mathbf{U} \in \mathbb{C}^{\mathbb{L}\times\mathbb{L}}$ which is an arbitrary semi-unitary matrix so that it satisfies $\mathbf{U}^H\mathbf{U} = \mathbf{I}$. Then, we denote the initial value of the beamforming matrix for RRU m as

$$\mathbf{F}_{2,m}^{(0)} = \mathbf{V}_m\mathbf{U}^H. \tag{3.28}$$

We repeat the above procedure for all RRUs and construct the initial point $\mathbf{f}_2^{(0)}$ as

$$\mathbf{f}_2^{(0)} = \left[vec(\mathbf{F}_{2,1}^{(0)})^T ... vec(\mathbf{F}_{2,M}^{(0)})^T\right]^T. \tag{3.29}$$

The comparison of convergence for the initial point that is found by using the method in this part and a random initial point can be seen in Appendix D.

3.4 Complexity Calculation

In this section, we calculate the computational complexity of our proposed algorithm. We approach the complexity calculation problem by mainly considering two steps. The first step is to calculate the complexity of the initial point selection procedure and the second step is to calculate the complexity of the remaining part of the proposed approach where we consider the rate and per-RRU power transmit constraint expressions in Eqn. 3.15 and Eqn. 3.16.

In the initial point selection procedure, we first consider the number of iterations for the algorithm which is equal to M. Then, we consider the component with the highest order of complexity inside an iteration of the initial point selection algorithm. The two expressions to consider in this algorithm are the vector multiplication inside the channel covariance expression which is $\mathbf{h}_{u,m}\mathbf{h}_{u,m}^H \in \mathbb{C}^{N\times N}$ and the matrix multiplication which is $\mathbf{F}_{2,m}^{(0)} = \mathbf{V}_m\mathbf{U}^H$. With the standard textbook approach, the computational complexity of a $m \times n$ matrix multiplied with a $n \times p$ matrix is $\mathcal{O}(mnp)$ [75]. Since $\mathbf{h}_{u,m}$ is an $N \times 1$ vector, the complexity of vector multiplication is $\mathcal{O}(N^2)$. In the second expression, $\mathbf{V}_m$ is an $N \times L$ matrix while $\mathbf{U}^H$ is an $L \times L$ matrix. Therefore, the computational complexity of the second expression is equal to $\mathcal{O}(L^2N)$. Since N is generally higher than L in our system, we can't outright say $\mathcal{O}(L^2N)$ is greater than $\mathcal{O}(N^2)$. Therefore, the computational complexity of the initial point is $M\left(\mathcal{O}(L^2N) + \mathcal{O}(N^2)\right)$ considering the number of iterations for this part is M.

For the rest of the proposed approach, we first denote the number of iterations as N_{iter}. As mentioned before, rate and per-RRU power transmit constraint are the two expressions that we consider for the complexity calculation of the remaining part. The matrix multiplications inside the rate expression have a computational complexity of $\mathcal{O}(L^2M^2N^2)$. The matrix multiplications inside the per-RRU transmit constraint has the complexity of $\mathcal{O}(L^3M^3N^3)$. Therefore, the total complexity of this part is equal to $N_{iter}(\mathcal{O}(L^2M^2N^2) + \mathcal{O}(L^3M^3N^3))$. Then, we have the total complexity as

$$C_{SCFA,UnlimCap} = M\left(\mathcal{O}(L^2N) + \mathcal{O}(N^2)\right) + N_{iter}(\mathcal{O}(L^2M^2N^2) + \mathcal{O}(L^3M^3N^3)) \quad . \tag{3.30}$$

Omitting the lower order complexity terms, we have the complexity of the proposed approach with unlimited fronthaul capacity as

$$C_{SCFA,UnlimCap} \approx N_{iter}\mathcal{O}(L^3M^3N^3). \tag{3.31}$$

Considering the expression Eqn. 3.31, increasing the number of fronthaul links increases complexity, although it increases spectral efficiency as we will discuss in the next part. Therefore, complexity calculation is important for optimal performance design.

3.5 Results and Discussion

In this chapter, we simulate the performance of the proposed approach under different conditions. The simulations are done in a MATLAB environment with 25 Monte Carlo trials for each simulation and a maximum of 1500 iterations for each gradient descent (or ascent in the negative sense) operation. For the analytical calculations, we have assumed a correlation-based channel model as in Eqn. 2.13.

For the simulations in MATLAB, we use a clustered multi-path channel model that is based on Extended Saleh Valenzuela Model [44] as explained in Section 2.2.2. We assume ULA at the RRUs with half-wave antenna spacing. For simplicity, we assume that all the users experience the same amount of attenuation inversely proportional to the number of total RRU antennas similar to the model in [46] from each multi-path. Then, we can express our channel model for the MATLAB simulations as

$$\mathbf{h}_u = \sum_{i=1}^{Lp} \epsilon_i \mathbf{a}(\phi_i), \tag{3.32}$$

where Lp is the number of multi-paths, ϵ_i is a zero-mean unit variance complex Gaussian channel gain for ith path and the array response vector for ULA with half-wave spacing at RRUs is given by

$$\mathbf{a}(\phi_{\mathbf{i}}) = [1 \;\; e^{j\pi \sin{(\phi_i)}} \;\; e^{j2\pi \sin{(\phi_i)}} ... \, e^{j(\bar{N}-1)\pi \sin{(\phi_i)}}]^T, \tag{3.33}$$

where ϕ_i is the Laplacian distributed AoA for the ith path, following a common approach to the model in [46]. Note that the total transmit power budget, P_{total} is equivalent to SINR expression when $M = 1, L = 1, U = 1, Lp = 1, N = 1$ condition is satisfied. The noise variance is normalized to 0 dBW and the number of multi-paths is set as 10 for the following simulations. MMSE precoder that is expressed as in Eqn. 2.26 for the single RRU structure is also added as a benchmark for performance simulations. The rest of the parameters for simulations can be found in Table 3.2.

Table 3.2: List of simulation parameters and their values for Section 3.5

Parameters	Values
σ^2	0 dBW
η_1	1.5
α	10^{-4}
t$_{max}$	1500 iterations
Δ	15
$P_{total,W}$	$10^{\frac{P_{total,dBW}}{10}}$
Lp	10
$\bar{N}$	64
$\#$ of multi-paths (Lp)	10
Angular Spread for AoA	10

In Fig. 3.2, the user average sum rate performance of the proposed approach and the initial point that is described in Section 3.3 is simulated with respect to the increasing number of fronthaul links per-RRU for a single RRU system. This is actually equivalent to the D-RAN system that is described in Section 2.5.3. In Fig. 3.2, we observe that the performances of the proposed approach and the initial point don't display a considerable improvement until L is increased to a certain point for different numbers of users.

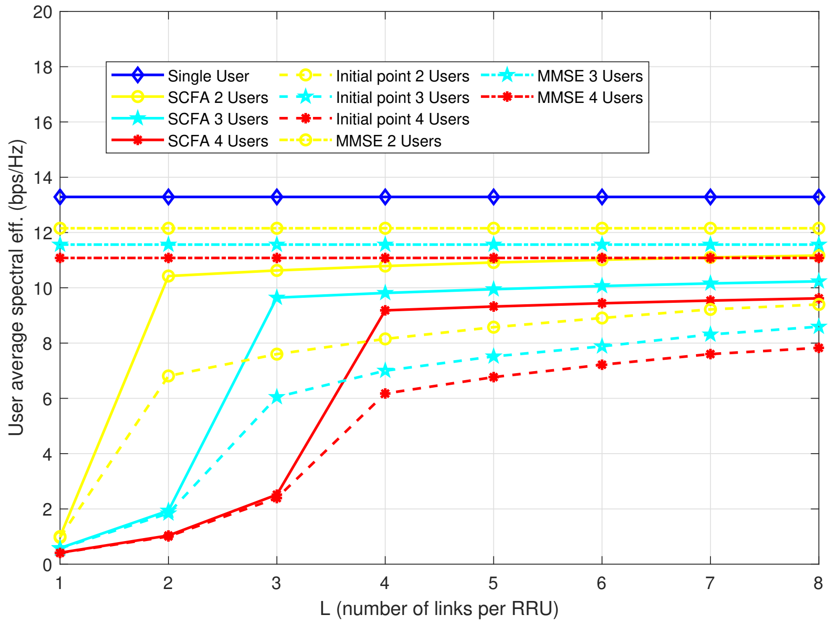


Figure 3.2: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 1$, P_{total} = 12 dBW with unconstrained fronthaul

To better understand the result in Fig. 3.2, consider the message signal after beamforming in the BBU pool and before the transmission to RRUs from fronthaul links in Fig. 3.1 expressed without the power term as $\mathbf{F}_1\mathbf{s}$. We can represent this expression in a more open way by considering the $\mathbf{F}_1$ expression in Eqn. 3.4 by

$$\begin{bmatrix} f_{1,(1,1),1} & f_{1,(1,1),2} & \cdots & \cdots & f_{1,(1,1),U} \\ f_{1,(1,2),1} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ f_{1,(1,L),1} & \cdots & \cdots & \cdots & \vdots \\ f_{1,(2,1),1} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ f_{1,(2,L),1} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ f_{1,(M,L),1} & f_{1,(M,L),2} & \cdots & \cdots & f_{1,(M,L),U} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_U \end{bmatrix}. \tag{3.34}$$

In Eqn. 3.34, each colored block in the left matrix represents the element of $\mathbf{F}_1$ that is forwarded to a separate RRU after multiplication with the message vector s that has unit energy i.i.d zero-mean Gaussian message symbols. Consider the case in Fig. 3.2, where $M = 1$. In this case, $\mathbf{F}_1$ has only one row for the case when $L = 1$. Therefore, the beamforming matrix can't accommodate the case when there is more than a single user. Similarly, when $L = 2$, $\mathbf{F}_1$ has two rows and the beamforming matrix can't accommodate the case when there are more than 2 users. Since $\mathbf{F}_1$ has $M \times L$ rows in total, we can deduce that we need the number of users to satisfy $ML \geq U$ for the beamformer to have a decent performance.

In Section 1.1, we have mentioned that C-RAN reduces the production costs for remote radio units, or access points, by allowing for the deployment of multiple RRUs with smaller array sizes compared to a system with a single RRU as in Section 2.5.3 that has the same number of total antennas. In the subsequent simulations, we investigate whether increasing the number of RRUs brings any performance changes for a C-RAN system with unconstrained capacity fronthaul. We compare these results with the limited capacity fronthaul C-RAN in Section 4.4.2.

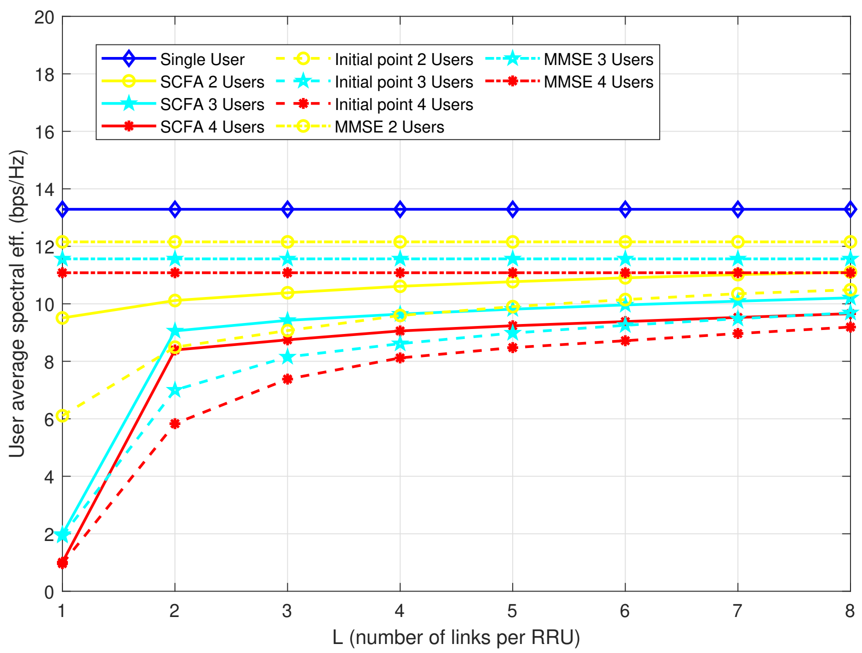


Figure 3.3: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 2$, P_{total} = 12 dBW with unconstrained fronthaul

In Fig. 3.3, the performances of the proposed approach and the initial point are simulated under 12 dBW total transmit power budget for a wired C-RAN system as in Fig. 3.1 with 2 RRUs versus the number of increasing fronthaul links per-RRU. For this case, we have the boundary for users as $2L \geq U$ for the beamformer to have a notable performance. In Fig. 3.3, we can observe that the beamformer can't perform well when $L = 1$ and the number of users is greater than 2. Comparing Fig. 3.2 and Fig. 3.3, another observation that could be made is that the performance of the initial point became closer to the proposed approach as the number of RRUs increased.

In Fig. 3.4, the performances of the proposed approach and the initial point are given for a C-RAN with 4 RRUs versus an increasing number of fronthaul links per-RRU under 12 dbW total transmit power budget. For this case, we have $M = 4$ and $L \geq 1$ so that the condition $ML \geq U$ is satisfied for the number of users up to 4 for all L values. As it can be seen from Fig. 3.4, the proposed approach has a smooth curve for all users among the 4 users. We can also observe that the performance of the initial point has become closer to the proposed approach compared to Fig. 3.3.

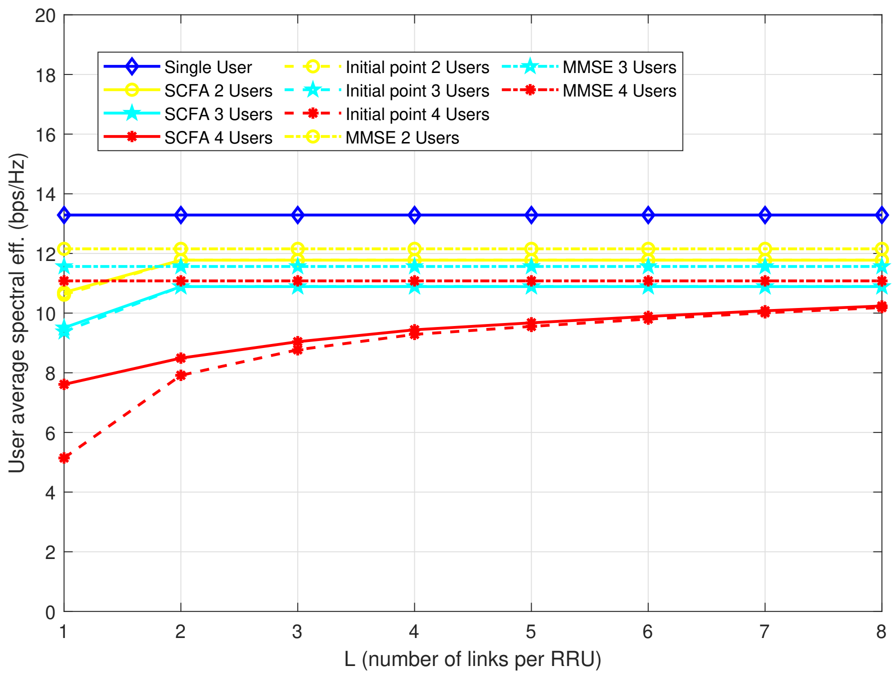


Figure 3.4: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 4$, P_{total} = 12 dBW with unconstrained fronthaul

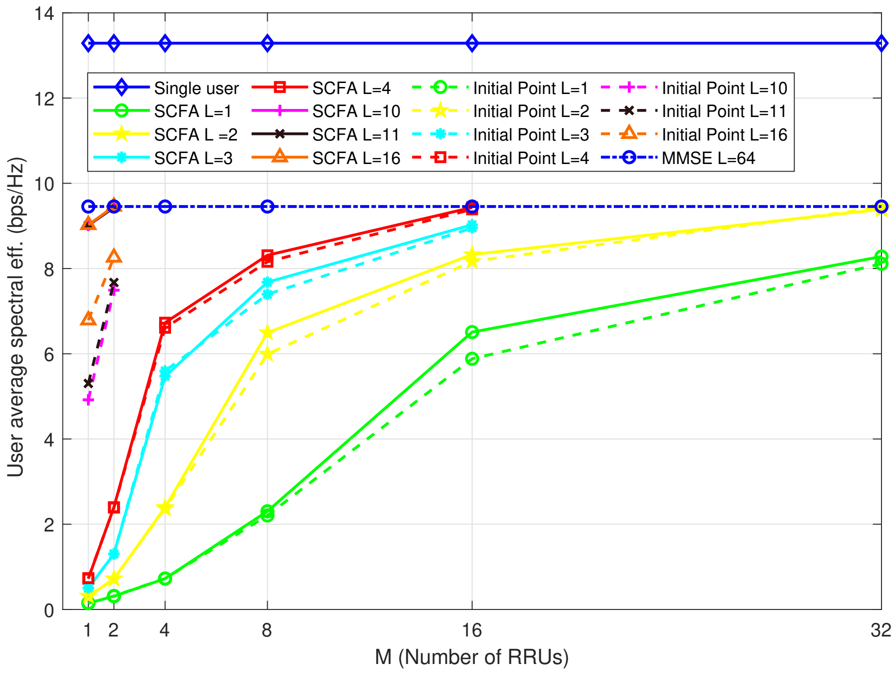


Figure 3.5: Comparison of User Average Spectral efficiency vs. number of RRUs for different number of active fronthaul links per-RRU with $\bar{N} = 64$, $U = 10$, P_{total} = 12 dBW with unconstrained fronthaul

The performance of the proposed approach and the initial point is also simulated for the increasing number of RRUs with active per-RRU fronthaul link numbers are equal to $L = 1$, $L = 2$ and $L = 3$ in Fig. 3.5. The number of users is equal to 10 for the simulation in Fig. 3.5, so the beamformer's condition to perform well is $ML \geq 10$. We can observe that the beamformer started performing well after $M = 16$ for the case when $L = 1$ since it satisfies the condition only for $M = 16$ and $M = 32$ for the axis points in Fig. 3.5. For $L = 2$, we also have decent performance when $M = 8$, and for $L = 3$, the notable performance point starts from $M = 4$ which is in line with the condition described on the number of users. For $L = 3$, we can't have 32 RRUs because each $\mathbf{F}_{2,m}$ has L columns associated with dominant eigenvectors of the correlation matrix for the user channels between user u and RRU m, and while there are 2 antennas per RRU for $M = 32$, we can have at max 2 eigenvectors associated with user channels during initial point selection which is less than 3. The result that we can obtain from these simulations is that since increasing the number of fronthaul links and number of RRUs increase the number of rows for $\mathbf{F}_1$ and the number of columns for $\mathbf{F}_2$, we have an increased degree of freedom during the beamforming operation. Therefore, increasing these values under a system with an unlimited fronthaul capacity results in a sum rate outcome that is closer to the maximum value that the expression in Eqn. 3.20 can take. This leads us to wonder how would the initial point and the proposed approach perform under a fronthaul capacity-limited system which we discuss in the next chapter.

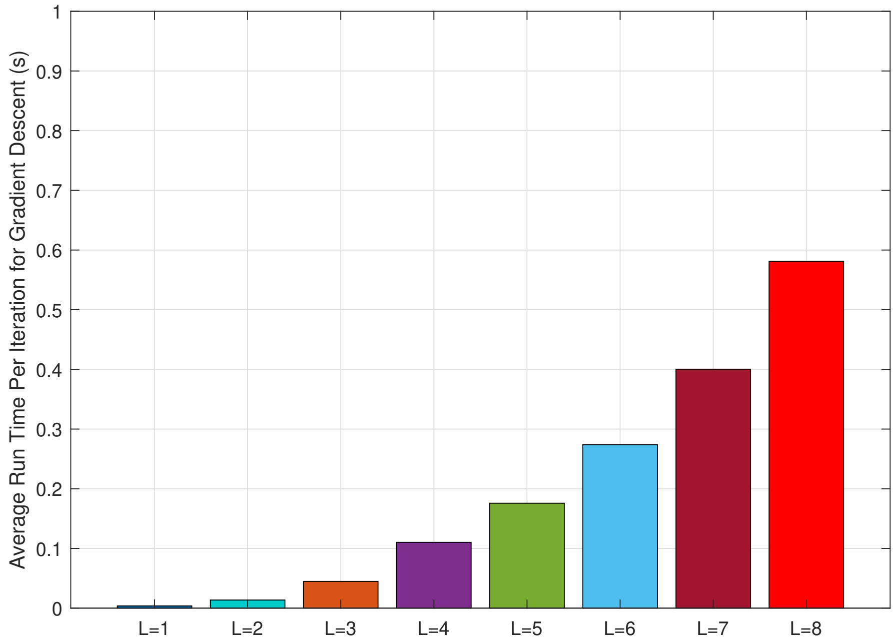


Figure 3.6: Average run time per iteration for the proposed approach vs. the number of active fronthaul links per-RRU, $\bar{N} = 64$, $M = 4$, $U = 4$, P_{total} = 12 dBW

In Section 3.4, we have discussed the complexity calculation for our approach for the unlimited fronthaul capacity case and found the complexity as $N_{iter}\mathcal{O}(L^3M^3N^3)$ approximately. In Fig. 3.6, the run times divided by the number of iterations until convergence and average run times per iteration for the gradient descent are simulated for different numbers of fronthaul links per-RRU. Run times were recorded on a laptop computer with 8 CPUs (Intel Core i7-6700HQ 2.60 GHz) and 8 GB RAM. As can be observed from Fig. 3.6, the run time complexity significantly increases with the increasing number of per-RRU fronthaul links which is in line with the complexity calculations in Section 3.4. Therefore, designing systems with high spectral efficiency that has fewer active fronthaul links would be desirable for the operators in order to have efficient performance.

3.6 Main Outcomes

The main results that we can obtain from this chapter as follows:

- We need $(M \times L) \geq U$ to have beamforming matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ with the appropriate number of rows and columns respectively so that all the users can be accommodated.
- The computational complexity of the proposed approach increases significantly with the increased number of per-RRU fronthaul links, which makes designs with fewer active fronthaul links and higher performance desirable.
- Under no fronthaul capacity constraints, the precoding matrices obtained from both the initial point and the proposed approach exploits the increased degree of freedom from the increased number of rows and columns for $\mathbf{F}_1$ and $\mathbf{F}_2$ i.e, increasing M and L leads to a performance that is closer to the MMSE precoder. This result leads us to further inspect whether this situation changes under a fronthaul capacity limited system where increasing the number of per-RRU active fronthaul links increases the quantization noise due to fronthaul capacity while the capacity is kept fixed. We discuss this case in the following chapter.

CHAPTER 4

SUM RATE MAXIMIZATION UNDER LIMITED FRONTHAUL CAPACITY

4.1 Problem Formulation

In this chapter, we delve into the sum rate maximization problem described in Eqn. 3.8 by also incorporating an additional fronthaul capacity constraint. Building upon the method described in the previous chapter, we use our proposed approach inspired by SCFA in [73] to shift the constraints into the main maximization problem and express the maximization problem in an equivalent and easier to differentiate form to apply gradient descent (or ascent in the negative sense) based maximization. We denote our proposed approach as SCFA in the simulations as the idea to transfer constraints into the main equation came from the work in [73]. A list of notations and their explanations for this chapter can be found in Table 4.1.

We consider a downlink C-RAN system with wired fronthaul similar to the previous chapter which is shown in Fig. 4.1. In this system, the BBU pool is connected to M RRUs each equipped with N antennas and the system has $\bar{N} = N \times M$ antennas in total. RRUs jointly transmit the signal to all users in a cell area in a distributed MIMO fashion with a total of U UEs each with a single antenna in the cell area. Different from the system in the previous chapter, we consider limited capacity in the fronthaul for the system in Fig. 4.1. In Section 1.1, we have mentioned that the CPRI interface that is used for defining the connection between BBU pool and RRUs over optical fiber links has limited data capacity. With the e-CPRI protocol, we can achieve higher fronthaul data rates while also having more flexible functional splits between the BBU and RRUs. In this chapter, we assume that the baseband data is transmitted with I and Q components and the transmitted data rate is expressed as

$2D_mL$ bits/channel use for L fronthaul links between RRU m and the BBU pool. Here, D_m is the number of quantization bits in the fronthaul between BBU pool and RRU m due to compression with uniform scalar quantization (USQ) model [66, 76].

Table 4.1: List of notations for Chapter 4

Notations	Explanation
U	Number of UEs
M	Number of RRUs
L	Number of fronthaul links per-RRU
N	Number of antennas per RRU
$\bar{N}$	Total number of antennas
C_m	Fronthaul capacity between RRU m and the BBU pool
C	Total fronthaul capacity between RRUs and the BBU pool
D_m	Number of quantization bits for the links between the BBU pool and RRU m
σ^2	Noise variance
s_u	Message component for UE u
$\mathbf{s}$	Message vector
$\mathbf{x}$	Message vector after beamforming and power allocation at the BBU pool
$\mathbf{x}_m$	Precoded and power allocated message vector at the BBU pool to be transmitted to RRU m
$\mathbf{q}$	Quantization noise vector
$\mathbf{q}_m$	Quantization noise vector in the fronthaul between RRU m and the BBU pool
$\mathbf{Q}$	Quantization noise variance
$\mathbf{Q}_m$	Quantization noise variance for $\mathbf{q}_m$
$Q_{m,l}$	l^{th} diagonal element of $\mathbf{Q}_m$
$\mathbf{h}_{u,m}$	Channel vector for RRU m and UE u
$\mathbf{h}_u$	Channel vector for UE u
$\mathbf{H}$	Channel between all transmit antennas and UEs
$f_{1,(m,l),u}$	Beamforming element of $\mathbf{F}_1$ corresponding to the l^{th} link of RRU m and UE u
$\mathbf{F}_{1,m}$	Beamforming matrix in the BBU pool corresponding to RRU m
$\mathbf{F}_1$	Beamforming matrix in the BBU pool
$\mathbf{f}_{2,m,l}$	l^{th} column of $\mathbf{F}_{2,m}$
$\mathbf{f}_{2,m}$	Vectorized form of $\mathbf{F}_{2,m}$
$\mathbf{F}_{2,m}$	Beamforming matrix for RRU m
$\mathbf{F}_2$	Aggregate Beamforming matrix in RRUs

We assume that the fronthaul between RRU m and the BBU pool has a capacity up to C_m bits/channel use and the fronthaul has a total capacity of $C = \sum_{m=1}^{M} C_m$ bits/channel use. We express the channel as $\mathbf{H} = [\mathbf{h}_1\mathbf{h}_2...\mathbf{h}_U] \in \mathbb{C}^{\bar{N}\times U}$ that has circularly symmetric zero-mean complex Gaussian elements similar to the correlation-based model in Section 2.2.1 where $\mathbf{h}_u \in \mathbb{C}^{\bar{N}}$ is the channel vector between UE u and RRUs. Also, the channel vector between UE u and RRU m is denoted by $\mathbf{h}_{u,m} \in \mathbb{C}^{N}$ and the spatial covariance matrix of this channel is expressed as $\mathbf{R}_{u,m} = \mathbb{E}[\mathbf{h}_{u,m}\mathbf{h}_{u,m}^H] \in \mathbb{C}^{N \times N}$.

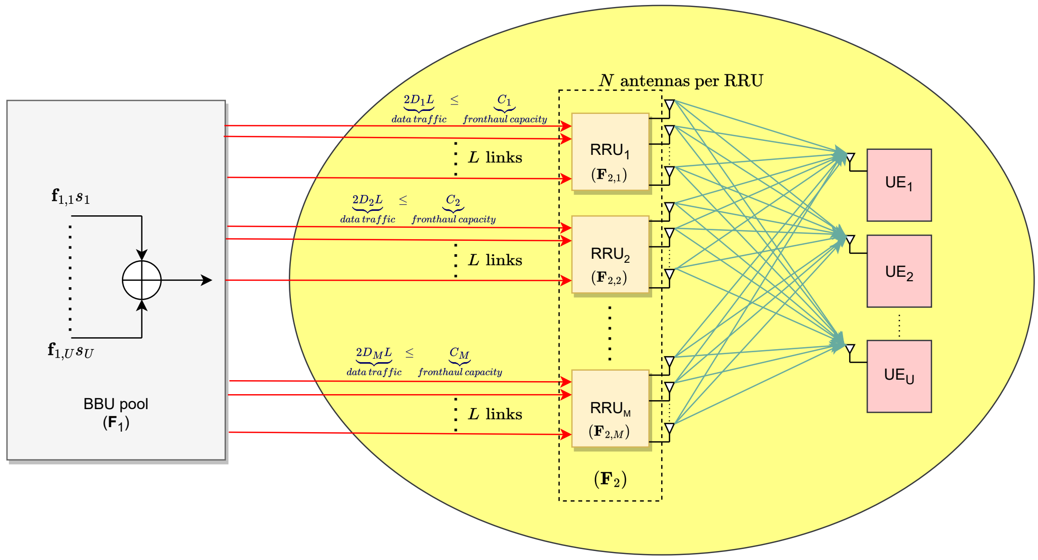


Figure 4.1: Block diagram of downlink C-RAN model with capacity limited fronthaul

Precoding is performed at the BBU pool and RRUs in two stages as in the previous chapter. The beamforming matrix at the BBU pool is denoted by $\mathbf{F}_1 \in \mathbb{C}^{(LM)\times U}$ and the beamforming matrix at RRU m is denoted by $\mathbf{F}_{2,m} \in \mathbb{C}^{N \times L}$. The aggregate $\mathbf{F}_2$ matrix encompasses beamforming matrices for all RRUs in a block diagonal structure similar to Eqn. 3.1 and [56]. For instance, it is in the form of $\mathbf{F}_2 = \begin{bmatrix} \mathbf{F}_{2,1} & \mathbf{0}_{N\times L} \\ \mathbf{0}_{N\times L} & \mathbf{F}_{2,2} \end{bmatrix} \in \mathbb{C}^{\bar{N}\times(2L)}$ when $M = 2$. For a given $\mathbf{F}_2$, we again assume that BBU pool utilizes the processed channel information $\mathbf{H}^H\mathbf{F}_2$ instead of the instantaneous channel information so that $\mathbf{F}_1$ can be expressed in a reduced size with an MMSE form as

$$\bar{\mathbf{F}}_1 = \left[\mathbf{F}_2^H\mathbf{H}\mathbf{H}^H\mathbf{F}_2 + \frac{U}{P_{total}}\sigma^2\mathbf{I}_{LM\times LM}\right]^{-1}\mathbf{F}_2^H\mathbf{H} \ . \tag{4.1}$$

Then, we scale the columns of $\bar{\mathbf{F}}_1$ with $\mathbf{F}_2$ by multiplying them with

$$\mathbf{f}_{1,u} = \frac{\bar{\mathbf{f}}_{1,u}}{\sqrt{(\mathbf{F}_2\bar{\mathbf{f}}_{1,u})^H\mathbf{F}_2\bar{\mathbf{f}}_{1,u}}}, \tag{4.2}$$

for all UEs and obtain $\mathbf{F}_1$.

Let $\mathbf{s} = [s_1 s_2 ... s_U]^T$ be a vector of independent zero mean and unit energy Gaussian message symbols. After precoding and power allocation at the BBU pool, the message signal is expressed as

$$\mathbf{x} = \sqrt{\frac{P_{total}}{U}} \mathbf{F}_1 \mathbf{s} \in \mathbb{C}^{(LM)}, \tag{4.3}$$

where the total power budget P_{total} is divided evenly among all the users. The message signal in Eqn. 4.3 contains $\mathbf{x}_m = \sqrt{\frac{P_{total}}{U}} \mathbf{F}_{1,m} \mathbf{s} \in \mathbb{C}^L$ which are message signals to be transmitted to RRU m. The transmission to RRUs is done in an IQ fashion and USQ is applied to the elements of $\mathbf{x}_m$ vectors. Then, we assume that quantization brings in an additional independent Gaussian zero mean and circularly symmetric noise element where the message signal at RRU m after quantization is expressed as $\bar{\mathbf{x}}_m = \mathbf{x}_m + \mathbf{q}_m$ and aggregate message vector for all RRUs is $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{q}$ where $\mathbf{q} \in \mathbb{C}^{(LM)}$ is the quantization noise vector that contains the vector $\mathbf{q}_m$ for the quantization noise at RRU m. Notice that we uniformly quantize a Gaussian signal, which leads to a greater noise contribution from the quantization compared to optimal quantization of a uniformly distributed signal for uniform quantization [77]. The details of the quantization procedure is explained in Appendix E. The covariance matrix of quantization noise for RRU m is a diagonal matrix expressed as $\mathbf{Q}_m = \mathbb{E}[\mathbf{q}_m \mathbf{q}_m^H]$ and each diagonal element of this matrix is found from [76] as

$$Q_{m,l} = \text{MSE}(D) \sum_{u=1}^{U} \frac{P_{total}}{U} \left| f_{1,m,l,u} \right|^2 . \tag{4.4}$$

The placement of each $f_{1,m,l,u}$ is shown in Eqn. 3.4 and MSE(D) values are shown in Table 4.2 for optimal uniform quantization step sizes for the quantization of a zero-mean unit variance Gaussian signal that are found in [76] for different numbers of quantization bits. After the message signal is transmitted from RRUs with beamforming matrix $\mathbf{F}_2$ also applied, the received signal at the UE u can be expressed as

$$\begin{aligned} y_u &= \mathbf{h}_u^H \mathbf{F}_2 \bar{\mathbf{x}} + n_u \\ &= \underbrace{\sqrt{\frac{P_{total}}{U}} \mathbf{h}_u^H \mathbf{F}_2 \mathbf{f}_{1,u} s_u}_{Desired\, signal} + \underbrace{\sqrt{\frac{P_{total}}{U}} \mathbf{h}_u^H \mathbf{F}_2 \sum_{i=1, i \neq u}^{U} \mathbf{f}_{1,i} s_i}_{interference} \\ &+ \underbrace{\mathbf{h}_u^H \mathbf{F}_2 \mathbf{q}}_{Quantization\, noise} + \underbrace{n_u}_{Thermal\, noise} . \end{aligned} \tag{4.5}$$

In Eqn. 4.5, the covariance matrix of quantization noise is expressed as $\mathbf{Q} = \mathbb{E}[\mathbf{q}\mathbf{q}^H]$ and n_u is the thermal noise with variance σ^2 that is equal among all users. Then, we can express an SINR expression for UE u without considering the correlation between

Table 4.2: Optimal MSE values for different number of quantization bits

D	Optimal Step size	Optimal MSE(D)
1	1.5960	0.3634
2	0.9957	0.1188
3	0.5860	0.0374
4	0.3352	0.0115
5	0.1881	0.00349
6	0.1041	0.00104
7	0.0568	0.00030
8	0.0307	0.00009
9	0.0165	2.50×10^{-5}
10	0.0088	7.00×10^{-6}
11	0.0046	1.94×10^{-6}
12	0.0024	5.35×10^{-7}
13	0.0013	1.46×10^{-7}
14	0.00056	2.6×10^{-8}
15	0.00035	1.1×10^{-8}
16[1]	0.00020	3.3×10^{-9}

[1] The optimal step size for 16 bit quantization was calculated by simulations since it was not provided in [76].

the quantization element and the message symbols as it was also assumed in the work [78] as

$$\text{SINR}_u = \frac{\frac{P_{total}}{U}\left|\mathbf{h}_u^H\mathbf{F}_2\mathbf{f}_{1,u}\right|^2}{\frac{P_{total}}{U}\sum_{i=1,i\neq u}^{U}\left|\mathbf{h}_u^H\mathbf{F}_2\mathbf{f}_{1,i}\right|^2 + \mathbf{h}_u^H\mathbf{F}_2\mathbf{Q}\mathbf{F}_2^H\mathbf{h}_u + \sigma^2}. \tag{4.6}$$

Then, a rate expression for UE u is expressed in the same form as Eqn. 2.25 so that

$$\text{Rate}_u = \log_2(1 + \text{SINR}_u). \tag{4.7}$$

Notice that this is a lower bound for the actual rate since have assumed Gaussian distribution for the quantization elements and haven't considered the correlation between the Gaussian noise and the quantized elements as it was also mentioned in the work [76]. A different rate expression could be obtained by considering Bussgang Decomposition and modeling the additional term as the Gaussian quantization noise and obtaining optimal MSE values for this case as it was done in the work [76]. We can express our maximization problem to find $\mathbf{F}_2$ maximizing the sum rate as

$$(\text{P}_2)\ \max_{\mathbf{F}_2}\sum_{u=1}^{U} Rate_u, \tag{4.8}$$

$$\text{s.t.}\ \sum_{i=1}^{M} tr\left(\frac{P_{total}}{U}(\mathbf{F}_{2,m}\mathbf{F}_{1,m})^H\mathbf{F}_{2,m}\mathbf{F}_{1,m}\right) + tr(\mathbf{F}_{2,m}\mathbf{Q}_m\mathbf{F}_{2,m}^H) = P_{total}, \tag{4.9}$$

$$2D_mL \leq C_m. \tag{4.10}$$

Here, different from (P_1) in Eqn. 3.8, we have additional quantization noise power added transmit power constraint, and a fronthaul capacity constraint in Eqn. 4.10 due to limited capacity fronthaul. We also note that we utilize all of our available transmit power budget and RRUs share available power budget arbitrarily among themselves. In the next part, we describe our proposed approach to solve P_2.

4.2 Proposed Approach Under Capacity Limited Fronthaul

To apply this approach, the maximization problem needs to be infinitely differentiable without constraints. To make the differentiation easier, it is better to transform the maximization parameter $\mathbf{F}_2$ into vector form. Let $\mathbf{F}_{2,m} = \Big[\mathbf{f}_{2,m,1}\mathbf{f}_{2,m,2}...\mathbf{f}_{2,m,L}\Big]$ be the columns of the m^{th} block of $\mathbf{F}_2$. Then, we vectorize $\mathbf{F}_2$ as

$$\mathbf{f}_2 = \Big[\mathbf{f}_{2,1,1}^T...\mathbf{f}_{2,1,L}^T...\mathbf{f}_{2,2,1}^T...\mathbf{f}_{2,2,L}^T...\mathbf{f}_{2,M,L}^T\Big]^T, \tag{4.11}$$

which contains non-zero elements of $\mathbf{F}_2$. Using this form, SINR expression in Eqn. 4.6 can be expressed as

$$\text{SINR}_u = \frac{\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_u^H\mathbf{b}_u\mathbf{A}_u\mathbf{f}_2}{\sum_{i=1,i\neq u}^{U}\mathbf{f}_2^H\mathbf{A}_u^H\mathbf{b}_i^H\mathbf{b}_i\mathbf{A}_u\mathbf{f}_2 + \mathbf{f}_2^H\mathbf{A}_u^H\mathbb{E}[\mathbf{z}^H\mathbf{z}]\mathbf{A}_u\mathbf{f}_2 + \frac{U}{P_{total}}\sigma^2}, \tag{4.12}$$

where

$$\begin{aligned}
\mathbf{A}_u &= \text{diag}(\underbrace{\mathbf{h}_{u,1}^H}_{L-times}, \underbrace{\mathbf{h}_{u,2}^H}_{L-times}, ..., \underbrace{\mathbf{h}_{u,M}^H}_{L-times}),\\
\mathbf{b}_u &= [\underbrace{f_{1,(1,1),u}}_{N-times}, \underbrace{f_{1,(1,2),u}}_{N-times}, ..., \underbrace{f_{1,(M,L),u}}_{N-times}],\\
\mathbf{z} &= [\underbrace{q_1}_{N-times}, \underbrace{q_2}_{N-times}, ..., \underbrace{q_{(LM)}}_{N-times}].
\end{aligned} \tag{4.13}$$

We again assume perfect CSI for gradient operation. Equivalence of Eqn. 4.6 and Eqn. 4.12 is trivial once we consider the equivalence for the unconstrained capacity case in Appendix A and apply a similar transformation to $\mathbf{Q}$ in Eqn. 4.6 as we did with $\mathbf{f}_{1,u}$ in the same equation which is shown in Eqn. 4.13. Then, the rate expression for the UE u can be expressed as

$$
\begin{aligned}
&\text{Rate}_u \\
&= \quad \log_2(\sum_{i=1}^{U} \mathbf{f}_2^H \mathbf{A}_u^H \mathbf{b}_i^H \mathbf{b}_i \mathbf{A}_u \mathbf{f}_2 \\
&+ \quad \mathbf{f}_2^H \mathbf{A}_u^H \mathbb{E}[\mathbf{z}^H \mathbf{z}] \mathbf{A}_u \mathbf{f}_2 + \frac{U}{P_{total}} \sigma^2) \\
&- \quad \log_2(\sum_{i=1, i \neq u}^{U} \mathbf{f}_2^H \mathbf{A}_u^H \mathbf{b}_i^H \mathbf{b}_i \mathbf{A}_u \mathbf{f}_2 \\
&+ \quad \mathbf{f}_2^H \mathbf{A}_u^H \mathbb{E}[\mathbf{z}^H \mathbf{z}] \mathbf{A}_u \mathbf{f}_2 + \frac{U}{P_{total}} \sigma^2). \qquad (4.14)
\end{aligned}
$$

The next step is to transfer the constraints in Eqn. 4.9 and Eqn. 4.10 into the main maximization problem. For this purpose, we first consider a per-RRU power transmit constraint as $tr((\mathbf{F}_{2,m}\bar{\mathbf{F}}_{1,m})^H \mathbf{F}_{2,m}\bar{\mathbf{F}}_{1,m}) + tr(\mathbf{F}_{2,m}\mathbf{Q}_m\mathbf{F}_{2,m}^H) \leq P_{t,RRU}$, where $P_{t,RRU} = P_{total}/M$ and express it in terms of vectorized $\mathbf{f}_2$ as below

$$
\frac{tr(\mathbf{W}_m)}{L} + \frac{tr(\mathbf{G}_m)}{NL} \leq P_{t,RRU}, \qquad (4.15)
$$

where

$$
\begin{aligned}
\mathbf{C}_m &= diag(\mathbf{0}_{(m-1)L}\mathbf{1}_L\mathbf{0}_{(M-m)L}), \\
\mathbf{D}_m &= \mathbf{C}_m\bar{\mathbf{F}}_1, \\
\mathbf{E}_m &= diag(\mathbf{0}_{(m-1)NL}\mathbf{1}_{NL}\mathbf{0}_{(M-m)NL}), \\
\mathbf{G}_m &= \mathbf{E}_m\mathbf{f}_2\mathbb{E}[\mathbf{z}\mathbf{E}_m^T(\mathbf{z}\mathbf{E}_m^T)^H]\mathbf{f}_2^H\mathbf{E}_m^H, \\
\mathbf{W}_m &= (\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T)^H\mathbf{E}_m\mathbf{f}_2 vec(\mathbf{D}_m)^T. \qquad (4.16)
\end{aligned}
$$

Here, the matrix with bold notation $\mathbf{D}_m$ should not be confused with the quantization bits D_m. The equivalence of the above expression and Eqn. 4.9 is explained in Appendix C. Next, to transfer Eqn. 4.15 into the main equation, we use the fact in Eqn. 3.18 and consider the below expression

$$
c_2 = \eta_2 \sum_{m=1}^{M} exp(\frac{Ntr(\mathbf{W}_m) + tr(\mathbf{G}_m) - NLP_{t,RRU}}{\alpha NL}). \qquad (4.17)
$$

The expression in Eqn. 4.17 tends to 0 if the per-RRU power constraint is satisfied and tends to (∞) if the constraint is not satisfied for a positive η_2 value. Therefore, c_2 can be added to the main problem as the transmit power constraint.

To get rid of the fronthaul capacity constraint in Eqn. 4.10, we choose the number of quantization bits such that $D_m = \lfloor \frac{C_m}{2L} \rfloor$ with floor function $\lfloor (.) \rfloor$ for each fronthaul link between the BBU pool and RRU m. After removing the constraints, we can represent Eqn. 4.8 as

$$
\begin{aligned}
\mathrm{T}_2 \;=\; & -\sum_{u=1}^{U}[\log_2(\sum_{i=1}^{U} \mathbf{f}_2^H \mathbf{A}_u^H \mathbf{b}_i^H \mathbf{b}_i \mathbf{A}_u \mathbf{f}_2 \\
& + \; \mathbf{f}_2^H \mathbf{A}_u^H \mathbb{E}[\mathbf{z}^H \mathbf{z}] \mathbf{A}_u \mathbf{f}_2 + \frac{U}{P_{total}} \sigma^2) \\
& - \; \log_2(\sum_{i=1, i\neq u}^{U} \mathbf{f}_2^H \mathbf{A}_u^H \mathbf{b}_i^H \mathbf{b}_i \mathbf{A}_u \mathbf{f}_2 \\
& + \; \mathbf{f}_2^H \mathbf{A}_u^H \mathbb{E}[\mathbf{z}^H \mathbf{z}] \mathbf{A}_u \mathbf{f}_2 + \frac{U}{P_{total}} \sigma^2)] \\
& + \; \eta_2 \sum_{m=1}^{M} exp(\frac{Ntr(\mathbf{W}_m) + tr(\mathbf{G}_m) - NLP_{t,RRU}}{\alpha NL}).
\end{aligned} \tag{4.18}
$$

The expression in Eqn. 4.18 is now differentiable with respect to $\mathbf{f}_2$ and gradient descent is applicable. The selection of the initial point is also important when applying gradient descent. To select the initial point we consider RRUs separately between all users and apply the approach in [46] to construct $\mathbf{f}_2^{(0)}$ whose elements correspond to the blocks of aggregate RRU matrix. Selection of $\mathbf{f}_2^{(0)}$ is described in Algorithm 2.

After determining the initial point, the update equation to calculate $\mathbf{f}_2$ for the next iteration is expressed as

$$
\mathbf{f}_2^{(t+1)} = \mathbf{f}_2^{(t)} - \mu(t) \nabla(\mathbf{f}_2^{(t)}) \tag{4.19}
$$

where $\mu(t)$ is the step size for the iteration and $\nabla(.)$ is the gradient vector which stands for the gradient of Eqn. 4.18 with respect to $\mathbf{f}_2^H$ that takes the value of $\mathbf{f}_2$ at time t as input. Here, $\mu(t)$ can be selected differently for each iteration as in Eqn. 3.24 for better convergence or it can be selected as constant for ease of operation. The comparison between constant and adaptable $\mu(t)$ is discussed in Section 4.4. Also,

Algorithm 2 Algorithm for initial point selection

Input: H,M,N,L,U

Output: $\mathbf{f}_2^{(0)}$

for $m = 1$ to M **do**

  Denote $\mathbf{R}_{u,m} = \mathbb{E}[\mathbf{h}_{u,m}\mathbf{h}_{u,m}^H] \in \mathbb{C}^{NxN}$

  Denote $\mathbf{R}_m = \sum_{u=1}^{U} \mathbf{R}_{u,m}$

  Let $\mathbf{V}_m$ be the matrix consisting of L largest eigenvectors of $\mathbf{R}_m$

  Let $\mathbf{U} \in \mathbb{C}^{LxL}$ be an arbitrary semi-unitary matrix such that $\mathbf{U}^H\mathbf{U} = \mathbf{I}$

  $\mathbf{F}_{2,m}^{(0)} = \mathbf{V}_m\mathbf{U}^H$

end for

$\mathbf{f}_2^{(0)} = \left[vec(\mathbf{F}_{2,1}^{(0)})^T ... vec(\mathbf{F}_{2,M}^{(0)})^T\right]^T$

return $\mathbf{f}_2^{(0)}$

the value of α in Eqn. 4.17 can be adjusted for faster convergence. The algorithm for the proposed approach with proper parameter values chosen is given in Algorithm 3.

We initialize the parameters in step 1 and we apply gradient descent to Eqn. 4.18 with respect to $\mathbf{f}_2^H$ until it converges to the maximum value which returns the optimized beamforming matrices.

4.3 Complexity Calculation

To calculate the computational complexity of the proposed approach, we first need to calculate the complexities of Algorithm 2 and Algorithm 3 separately. For Algorithm 2, the number of iterations is equal to M, and the computational complexity for the Algorithm 2 is expressed as $M\left(\mathcal{O}(L^2N) + \mathcal{O}(N^2)\right)$ as in the previous chapter. To calculate the computational complexity for the Algorithm 3, we consider the number of iterations for the gradient descent algorithm which is N_{iter} and per-RRU power transmit constraint complexity which is $\mathcal{O}(UN^2M^3L^3 + N^3M^3L^3)$. The complexity resulting from this is $\mathcal{O}(N_{iter}(UN^2M^3L^3 + N^3M^3L^3))$. So we can calculate the computational complexity of our approach as

Algorithm 3 Finding $\mathbf{F}_2$ that maximizes the negative of 4.18 with proposed approach

Step 1, Initialization: Set the initial point $\mathbf{f}_2^{(0)}$ as found by the approach in Algorithm 2. Define: $\alpha = 10^{-4}, \eta_2 = 1.5, \Delta = 15, t_{max} = 1500.$

Step 2, Gradient Descent:

for $t = 0$ to t_{max} **do**
    Find gradient $\nabla(\mathbf{f}_2^{(t)})$
    Find step size $\mu^{(t)}$ from Eqn. 3.24 or choose as constant
    Update beamformer : $\mathbf{f}_2^{(t+1)} = \mathbf{f}_2^{(t)} - \mu^{(t)}\nabla(\mathbf{f}_2^{(t)})$
    Set corresponding elements of $\mathbf{f}_2^{(t+1)}$ as columns of $\mathbf{F}_{2,m}^{(t+1)}$ matrices
    Update $\mathbf{F}_1^{(t+1)}$ according to Eqn. 4.1 and Eqn. 4.2
    if ($\max\limits_{t-\Delta \leq t_s \leq t} \left| \mathrm{T}_2^{(t+1)} - \mathrm{T}_2^{(t)} \right| < 10^{-5}$) **then**
        $t_{max} = t_s$
        break
    end if
end for
$\mathbf{f}_2 = \mathbf{f}_2^{(t_{max})}$
Set corresponding elements of $\mathbf{f}_2$ as columns of $\mathbf{F}_{2,m}$ matrices
$\mathbf{F}_1 = \mathbf{F}_1^{(t_{max})}$
return $\mathbf{F}_2$, $\mathbf{F}_1$

$$C_{SCFA,LimCap} = M\left(\mathcal{O}(L^2N) + \mathcal{O}(N^2)\right) + N_{iter}\mathcal{O}((UN^2M^3L^3 + N^3M^3L^3)). \tag{4.20}$$

Ignoring the lower-ordered terms, we have the computational complexity of the proposed approach approximated as

$$C_{SCFA,LimCap} \approx N_{iter}\mathcal{O}((UN^2M^3L^3 + N^3M^3L^3)). \tag{4.21}$$

4.4 Results and Discussion

In this part, we simulate the performance of the proposed approach and the initial point that is found by Algorithm 2 for the fronthaul capacity constrained system in Fig. 4.1 and compare it with the unlimited fronthaul capacity case in the previous chapter. We again employ the same clustered channel model that we have described in Fig. 2.4 and Section 2.2.2 for MATLAB simulations as the previous chapter such that

$$\mathbf{h}_u = \sum_{i=1}^{Lp} \epsilon_i \mathbf{a}(\phi_i), \tag{4.22}$$

where ϵ_i is the zero-mean unit variance complex path gain and array response vector for ULA at RRUs $\mathbf{a}(\phi_i)$ is defined as Eqn. 3.33 with AoAs (ϕ_i) that are Laplacian distributed for each channel path i.

The noise variance σ^2 is normalized to 0 dBW and the number of channel paths is set to 10 unless otherwise specified. Performance of the MMSE precoder as in Eqn. 2.26 with a single RRU and $L = 32$ is also added as a benchmark. However, one should keep in mind that the MMSE precoder doesn't take the finiteness of the fronthaul capacity limitations into account. Hence, it does not perform well under limited fronthaul capacity and is just added as a benchmark precoder that is available

Table 4.3: List of simulation parameters and their values for Section 4.4

Parameters	Values
σ^2	0 dBW
η_2	-1.5
α	10^{-4}
t_{max}	1500 iterations
Δ	15
$P_{total,W}$	$10^{\frac{P_{total,dBW}}{10}}$
C	64 bits/channel use
D	$\lfloor \frac{64}{2LM} \rfloor$
$\bar{N}$	64
# of multi-paths (Lp)	10
Angular Spread for AoAs	10

from the literature for the simulations. The list of parameters for the simulations can be found in Table 4.3. Note that we express the total transmit power budget in dBW but apply it as W inside the SINR expression for the simulations.

4.4.1 Results for limited fronthaul capacity

In this part, we simulate the results for the system with limited fronthaul capacity. For the fronthaul capacity, we set the capacity between the BBU pool and each RRU equal as $C_1 = C_2 = ... = C_M$ so that the quantization bits are also equal as $D_1 = D_2 = ... = D_M$. Then, the total fronthaul capacity is expressed as

$$C = MC_M \text{ bits/channel use.} \tag{4.23}$$

For the simulations, the total fronthaul capacity is kept the same and the number of quantization bits for a total M RRUs is set as

$$D_m = \lfloor \frac{C}{2LM} \rfloor \ , \tag{4.24}$$

for each fronthaul link between RRU m and the BBU pool. We discuss the results where $\mu(t)$ is selected as a constant and the case when it is selected as shown in 3.24 in separate parts.

4.4.1.1 Results with constant step size

In this part, we simulate the performance of the proposed approach with respect to increasing per-RRU fronthaul links while considering a constant step size for the update equation in Eqn. 4.19 and the performance of the initial point in MATLAB with 25 Monte Carlo iterations and a maximum number of gradient iterations for each iteration of Algorithm 3 as $t_{max} = 1500$. We choose the total fronthaul capacity as $C = 64$ bits/channel use so that the fronthaul between the BBU pool and RRU m has a capacity of $C_m = \frac{64}{M}$ bits/channel use and the number of quantization bits are $D_m = \lfloor \frac{64}{2LM} \rfloor$ for a total of M RRUs with each having L fronthaul links between the BBU pool. In the following simulations, we adjust the constant step size for each iteration as $\mu(t) = \mu = 4 \times 10^{-3}$.

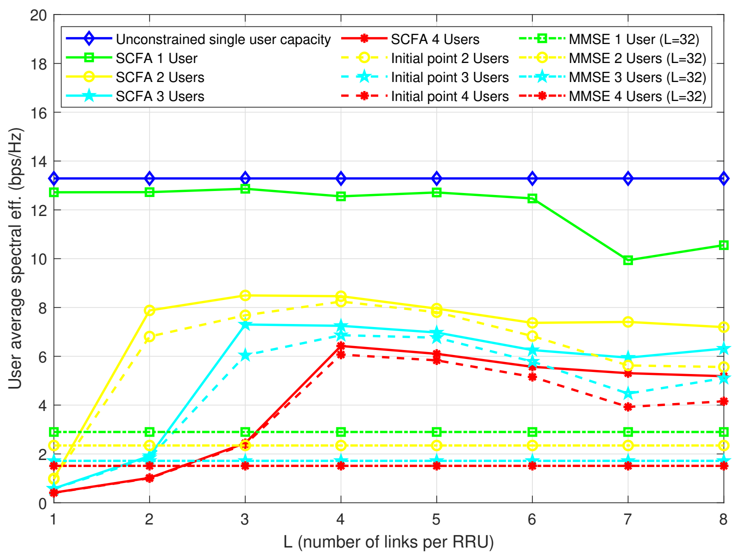


Figure 4.2: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 1$, P_{total} = 12 dBW, $\mu = 0.004$, with $C = 64$ bits/channel use

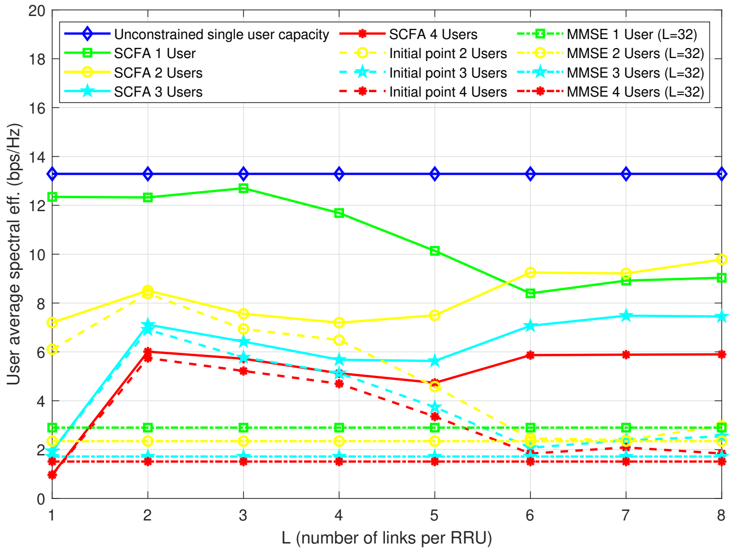


Figure 4.3: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 2$, P_{total} = 12 dBW, $\mu = 0.004$, with $C = 64$ bits/channel use

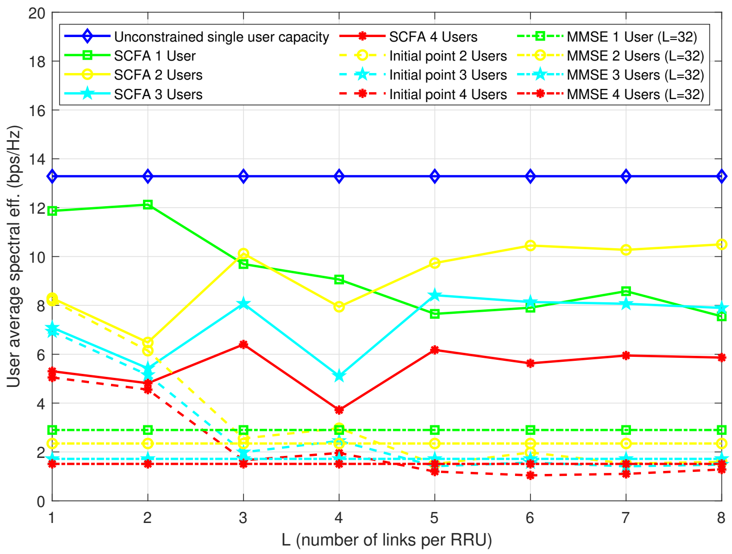


Figure 4.4: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 4$, P_{total} = 12 dBW, $\mu = 0.004$, with $C = 64$ bits/channel use

First of all, we look at the performance of the initial point as the number of total RRUs is increased between the results in Fig. 4.2 with $M = 1$, Fig. 4.3 with $M = 2$ and Fig. 4.4 with $M = 4$ RRUs, while also increasing the number of per-RRU fronthaul links in each figure. Normally, increasing the number of rows for $\mathbf{F}_1$ and increasing the number of columns for $\mathbf{F}_2$ with increasing M and L results in an increased degree of freedom for the beamforming matrix selection which would have a positive effect on the performance. However, we see in Fig. 4.2, Fig. 4.3 and Fig. 4.4 that the performance of the initial point decreases with increased L after the condition $ML \geq U$ mentioned in Section 3.5 is satisfied, despite the increased degree of freedom due to increasing precoding matrix sizes. This can be better understood by considering the number of quantization bits selection $D_m = \lfloor \frac{C_m}{2L} \rfloor$ that we have assumed in order to relax the fronthaul capacity constraint for the maximization problem in Eqn. 4.10. Because of such a selection, increasing the number of RRUs and per-RRU fronthaul links increases the quantization noise. From the results in Fig. 4.2, Fig. 4.3 and Fig. 4.4, we observe that the initial point can't perform well under limited fronthaul capacity, unlike the unlimited fronthaul capacity case.

When we look at the performance of the proposed approach with constant step size, we see that the smoothness of the curves corresponding to the respective number of users decreases with the increasing number of RRUs. This means that the proposed algorithm has trouble converging with constant step size for an increasing number of RRUs. This leads us to the question of how the proposed approach would perform with an adjustable step size for each update operation which we discuss in the next part.

4.4.1.2 Results with adjustable step size

In this part, we simulate the performance of the proposed approach with respect to increasing per-RRU fronthaul links where the step size in Eqn. 4.19 is adjusted at each iteration as in Eqn. 3.24 and the initial point which is the same as in Section 4.4.1.1 is graphed together for different configurations. The simulations are again performed by considering 25 Monte Carlo iterations in MATLAB with a maximum of 1500 iterations for each gradient operation.

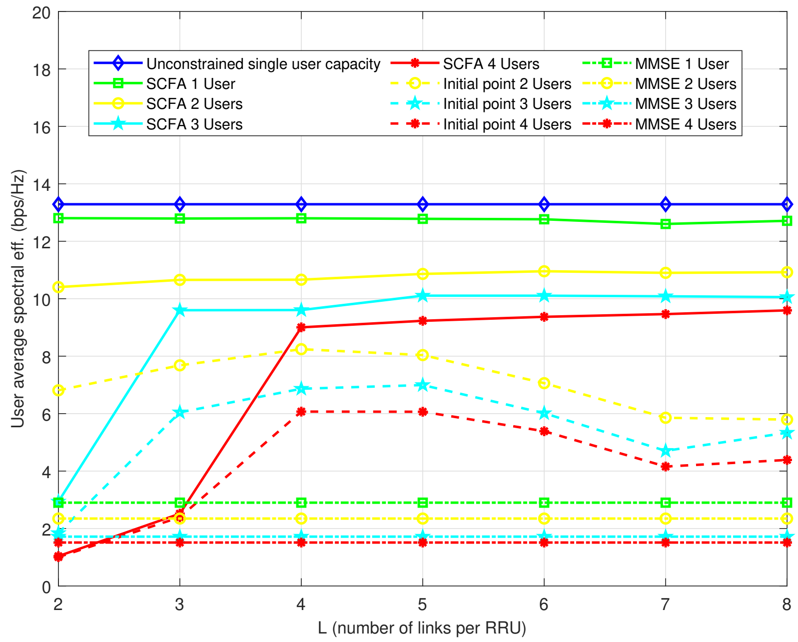


Figure 4.5: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 1$, P_{total} = 12 dBW with $C = 64$ bits/channel use

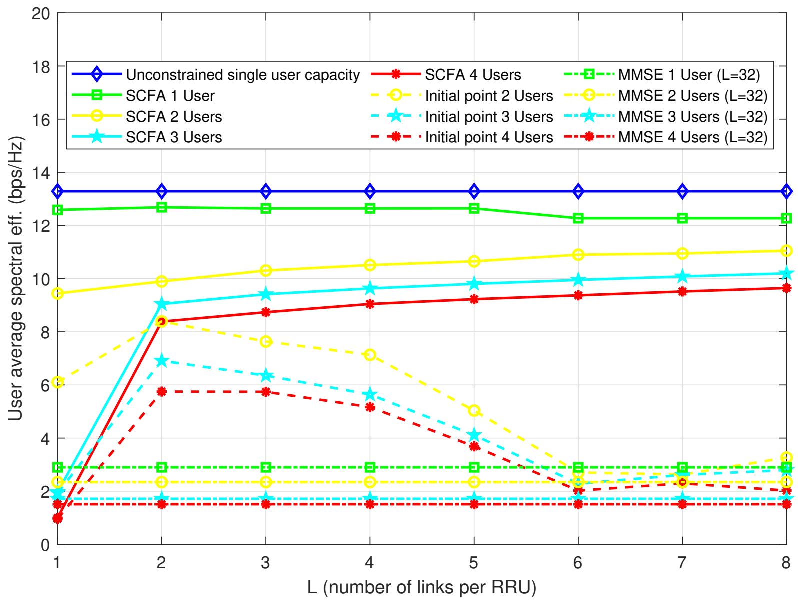


Figure 4.6: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 2$, P_{total} = 12 dBW with $C = 64$ bits/channel use

In Fig. 4.5, we consider a single RRU that is connected to the BBU pool as in D-RAN topology, and the total fronthaul capacity is equal to the fronthaul capacity between this RRU and the BBU pool which is set as $C = 64$ bits/channel use. We can observe that the proposed approach has a smoother curve compared to the constant step size case in Fig. 4.2 and the performance doesn't decrease as the number of per-RRU links, and hence quantization noise, is increased beyond the $ML \geq U$ point.

In Fig. 4.6, we increase the number of RRUs to 2 and the fronthaul capacity between each RRU and the BBU pool becomes $C_m = 32$ bits/channel use for $m = 1, 2$. In this case, we have reduced the RRU complexity compared to single RRU case by decreasing the per-RRU antenna array sizes and reduced the load on fronthaul further between each RRU and the BBU pool by decreasing the number of quantization bits with increased RRUs. For this case, we again observe that the curves for the proposed approach preserve their smoothness and the performance doesn't decrease after the condition $ML \geq U$ is satisfied for the number of users $U = 2, 3, 4$ despite increased quantization noise. For $U = 1$, the performance for the proposed approach decreases after $L = 5$, and $D = \lfloor \frac{64}{20} \rfloor = 3$. Notice that we don't utilize all of the available fronthaul capacity and only utilize 48bits/channel use for $L = 6$.

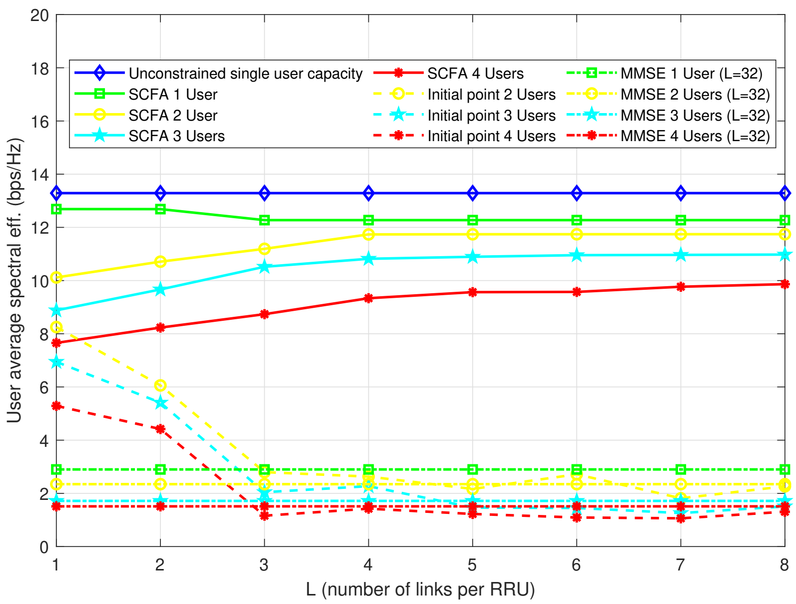
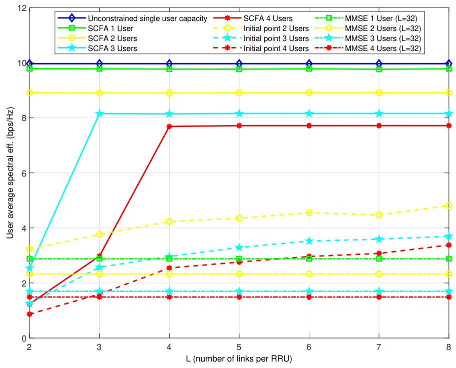
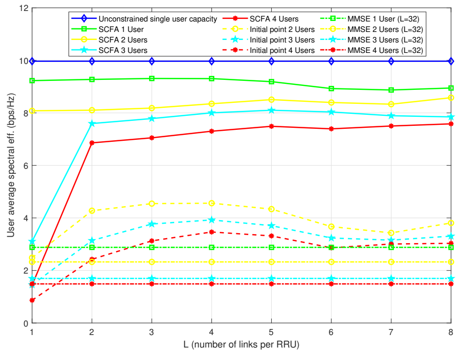


Figure 4.7: Comparison of User Average Spectral efficiency vs. number of fronthaul links per RRU for $\bar{N} = 64$, $M = 4$, P_{total} = 12 dBW with $C = 64$ bits/channel use

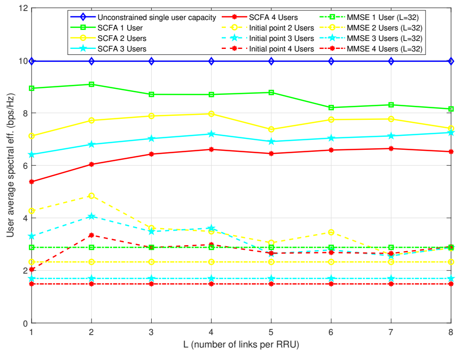
We increase the number of RRUs to 4 in Fig. 4.7 so that the fronthaul capacity between each RRU and the BBU pool becomes $C_m = 16$ bits/channel use for $m = 1, 2, 3, 4$ which further decreases the load on the fronthaul between each RRU and the BBU pool while increasing quantization noise. When we look at the performance of the proposed approach for the user numbers $U = 2, 3, 4$, we again observe that the performance of the proposed approach has a non-decreasing smooth curve that keeps improving up to a certain point since in this case, the condition $4L \geq U$ is satisfied for the number of users up to 4 users for all L values. For $U = 1$, we observe that the performance decreases after $L = 2$ and $D = \lfloor \frac{64}{16} \rfloor = 4$. We note that the point where performance starts decreasing with $L = 3$ corresponds to a number of quantization bits equal to $D = \lfloor \frac{64}{24} \rfloor = 2$ and we need to highlight that in the previous case for $M = 2$, the performance started to decrease after the quantization bits were less than $D = 3$ which lies in between $L = 2$ and $L = 3$ in this case. The first thing we observe from these results is that the proposed algorithm is less robust against increasing quantization noise for the single-user case, therefore; it is preferable to consider a single active per-RRU fronthaul link for single-user cases with our proposed approach. Another thing we observe from these results is that, as we increase the number of RRUs and decrease the RRU complexity, the performance of the proposed approach doesn't differ much from the unconstrained fronthaul capacity case. This leads us to question to what end our proposed algorithm could tolerate increased quantization noise while enjoying reduced RRU complexity.



(a) $\bar{N} = 64$, $M = 1$, P_{total} = 2 dBW with $C = 64$ bits/channel use



(b) $\bar{N} = 64$, $M = 2$, P_{total} = 2 dBW with $C = 64$ bits/channel use



(c) $\bar{N} = 64$, $M = 4$, P_{total} = 2 dBW with $C = 64$ bits/channel use

Figure 4.8: Comparison of the performance of the proposed approach under unlimited and limited fronthaul capacity vs. L for different numbers of RRUs under P_{total} = 2 dBW

When we reduce the total power budget P_{total} as can be seen in Fig. 4.8, we can observe that the performance of the proposed approach has become more susceptible to increasing quantization noise with increasing number of RRUs. Therefore, transmit power budget also affects the performance of our proposed approach.

We investigate the performance of the proposed approach vs. the increasing number of RRUs under different numbers of per-RRU active fronthaul links while the number of users is 10 in Fig. 4.9 while keeping the total number of transmit antennas and the

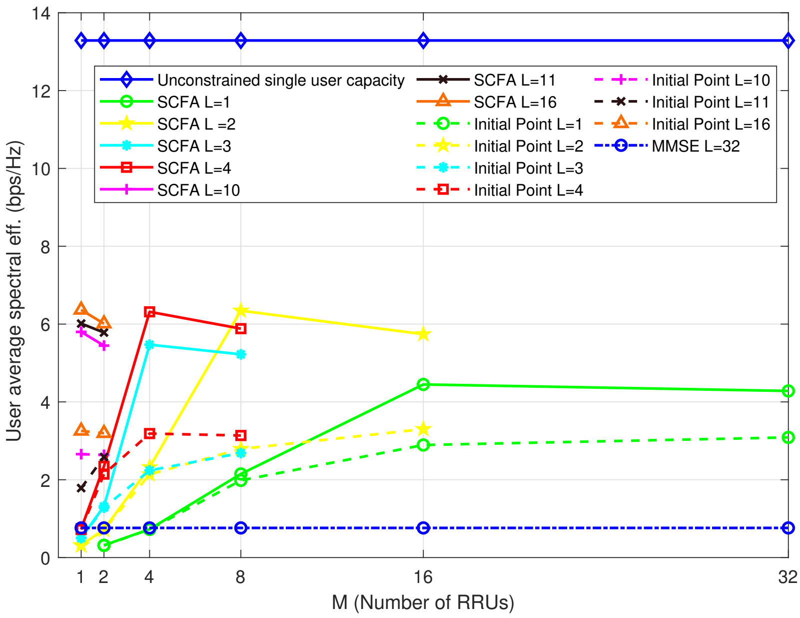


Figure 4.9: Comparison of User Average Spectral efficiency vs. number of RRUs for different number of active per-RRU fronthaul links, $\bar{N} = 64$, $U = 10$ with $C = 64$ bits/channel use and P_{total} = 12 dBW

fronthaul capacity same. For $L = 2$ and $L = 3$, we can draw the performance at most up to a number of RRUs that is equal to 16 and 8 respectively under the given total fronthaul capacity. The reason behind this is that when $M = 16$, and $L = 2$, we have $D = \lfloor \frac{64}{2 \times 16 \times 2} \rfloor = 1$ and after this point, we would have 0-bit quantization which would be meaningless to consider. Similarly, we have 1-bit quantization at $M = 8$ when $L = 3$ and can't have a meaningful result after this point. The same is also true for $L = 4$. When we look at the performance of the proposed approach when $L = 1$, we observe that it has decent performance at $M = 16$ when the condition $ML \geq U$ is satisfied. However, we see that the proposed approach hasn't improved the performance with the parameters that we defined in Table 4.3 at $M = 32$ and $D = 1$ when $L = 1$. When we look at Fig. 4.7, we see that the proposed approach has non-decreasing performance at $L = 8$ and $D = 1$. The beamforming matrix $\mathbf{F}_2$ in both Fig. 4.7 and Fig. 4.9 is a 64×32 matrix for these cases. However, for the case in Fig. 4.7 with $L = 8$, we have $\mathbf{F}_{2,m}$ as a 16×8 matrix for $m = 1, 2, 3, 4$, with 4 $\mathbf{F}_{2,m}$ blocks inside $\mathbf{F}_2$, and for the case in Fig. 4.9, we have $\mathbf{F}_{2,m}$ as a 2×1 matrix for $m = 1, 2, ..., 32$ with 32 $\mathbf{F}_{2,m}$ blocks inside $\mathbf{F}_2$. This means that increasing the number of columns of $\mathbf{F}_{2,m}$ provides a better degree of freedom than increasing the number of $\mathbf{F}_{2,m}$ matrix blocks in $\mathbf{F}_2$, i.e., increasing the number of RRUs while decreasing their complexity consistently doesn't bring consistent performance improvement and is more susceptible to increasing quantization noise. We also see a similar trend when $L = 2$, with the proposed approach having a decent performance when $M = 8$, which satisfies the $ML \geq U$ condition with $D = 2$, and observe a decreased performance when $M = 16$ with $D = 1$ When $L = 3$, we observe that the performance improvement happens around $M = 4$ and with $D = 2$ and it decreases at $M = 8$ with $D = 1$. We observe a similar trend in $L = 4$. However, one should also note that the user averaged performance when $L = 2$ and $M = 8$, and $L = 4$ and $M = 4$ is around 6.3 bps/Hz which is close to the user averaged capacity $\frac{64}{10} = 6.4$ bits/channel use. When $L = 10$ and $L = 11$, we also observe a similar trend and the maximum value occurs at the single RRU case. However, we should note that this would mean a very complex RRU structure and a long operation time. Also, we see that the maximum value when $L = 10$ and $L = 11$ is smaller compared to the maximum value when $L = 2, 3, 4$. This means that we need to increase per-RRU fronthaul links beyond $L = 11$ such as $L = 16$ which takes a lot of time with our proposed approach. Therefore, using multiple RRUs with reduced complexities compared to a single RRU that satisfies the condition $ML \geq U$ in the presence of many users with a limited capacity system could benefit the operators by using our proposed approach.

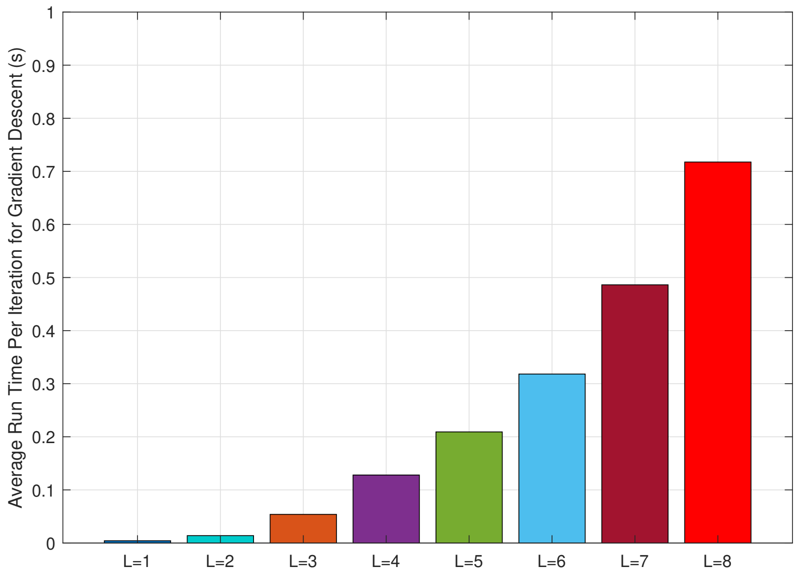
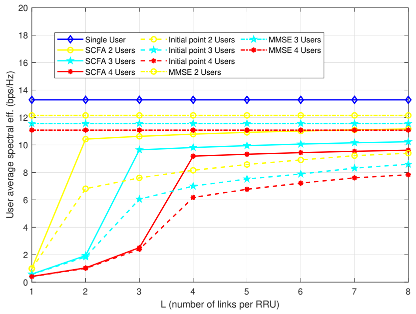


Figure 4.10: Average run time per iteration for the proposed approach vs number of active fronthaul links, $\bar{N} = 64$, $M = 4$, $U = 4$, P_{total} = 12 dBW with $C = 64$ bits/channel use

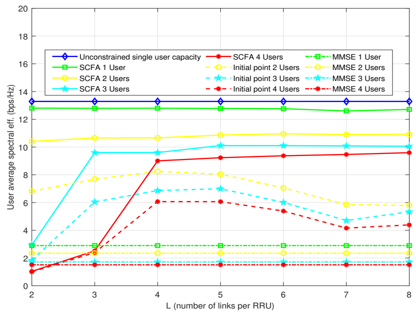
We also observe in Fig. 4.10 that increasing the number of per-RRU fronthaul links brings significant run-time complexity similar to the previous chapter where the results are obtained by the same PC with 8 CPUs and 8 GB RAM. Therefore, selecting designs with higher sum rate performance and fewer active per-RRU fronthaul links that use our proposed approach would be more desirable for the operators. In the next part, we compare the results in this part with the results for the unconstrained fronthaul capacity case in Section 3.5.

4.4.2 Comparison with unlimited fronthaul capacity case

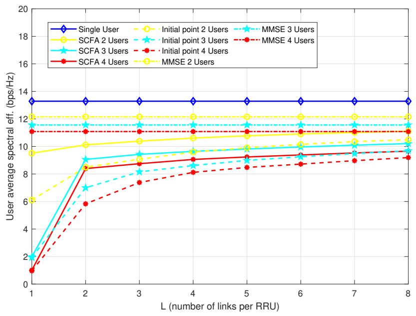
In this part, we compare the performance of the proposed approach and the initial point for the cases with unlimited fronthaul capacity in Section 3.5 and limited fronthaul capacity in this chapter.



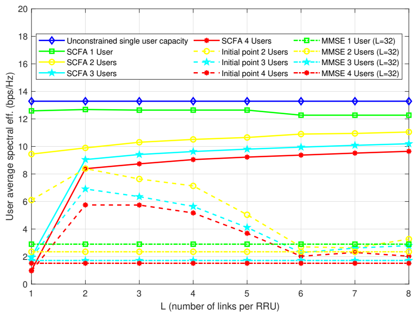
(a) $M = 1$, $\bar{N} = 64$, with unconstrained fronthaul capacity



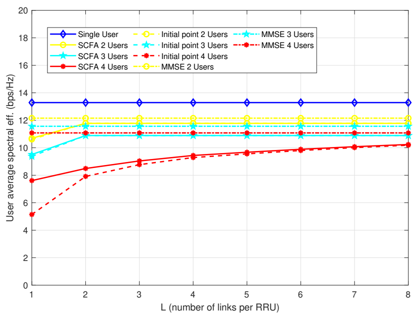
(b) $M = 1$, $\bar{N} = 64$, with $C = 64$ bits/channel use



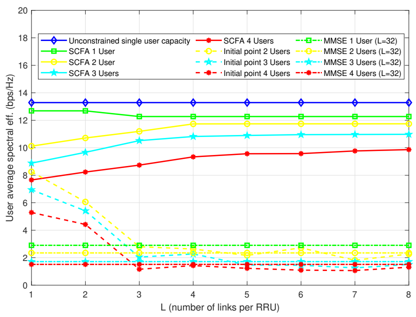
(c) $M = 2$, $\bar{N} = 64$, with unconstrained fronthaul capacity



(d) $M = 2$, $\bar{N} = 64$, with $C = 64$ bits/channel use



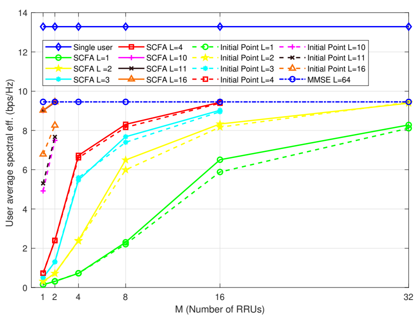
(e) $M = 4$, $\bar{N} = 64$, with unconstrained fronthaul capacity



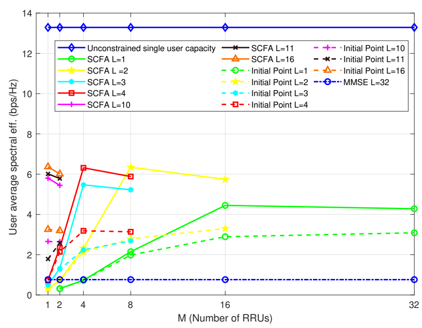
(f) $M = 4$, $\bar{N} = 64$, with $C = 64$ bits/channel use

Figure 4.11: Comparison of the performance of the proposed approach under unlimited and limited fronthaul capacity vs. L for different number of RRUs

First of all, we compare the performances under increasing per-RRU fronthaul links with M =1, 2, and 4. We keep the number of antennas fixed at 64 and the total fronthaul capacity at 64 bits/channel use, and P_{total} = 12 dBW, the same as the previous part. In Fig. 4.11, the left column stands for the unlimited capacity case and the right column stands for the limited capacity case. When we look at the performance of the initial point in the left column, we see that it is able to exploit an increased degree of freedom due to the increasing number of RRUs and get closer to the performance of the MMSE precoder as we increase the number of RRUs. However, when we look at the right column, we see that the performance of the initial point decreases after the condition $ML \geq U$ is satisfied due to increasing quantization noise. When we look at the performance of the proposed approach, we observe that the performance under the limited fronthaul capacity doesn't differ too much from the left column with the unlimited fronthaul capacity and the proposed approach is able to exploit the increasing degree of freedom due to increasing number of per-RRU active fronthaul links for the given configurations. We also note that the data traffic is smaller than the fronthaul capacity rather than equal in some cases and we observe that single-user performance is more susceptible to increasing quantization noise with the proposed approach.



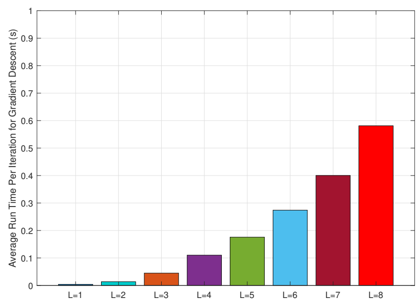
(a) $U = 10$, P_{total} = 12 dBW, $\bar{N} = 64$, with unconstrained fronthaul capacity



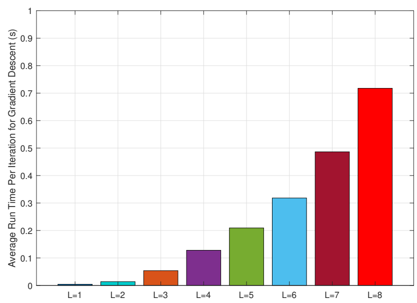
(b) $U = 10$, P_{total} = 12 dBW, $\bar{N} = 64$, with $C = 64$ bits/channel use

Figure 4.12: Comparison of the performance of the proposed approach under unlimited and limited fronthaul capacity vs. M for different number of active per-RRU fronthaul links

However, we observe that this improvement is not constant as can be seen in Fig. 4.12. In Fig. 4.12 (b), we see that the proposed approach has better performance when $L = 2$ at $M = 8$ with $D = 2$ than the case when $L = 4$ and $M = 8$ with $D = 1$ although in both cases we have $\mathbf{F}_2 \in \mathbb{C}^{64\times 32}$ matrix. When we look at the case in Fig. 4.11 (f), we see that the performance of the proposed approach doesn't decrease at $L = 8$ when $M = 4$ with $D = 1$. The difference between these cases is that when $M = 8$, we have $\mathbf{F}_{2,m}$ as a $8 \times L$ matrix, and when $M = 4$, we have $\mathbf{F}_{2,m}$ as a $16 \times L$ matrix for RRU m which is a more complex structure. Since we express $\mathbf{F}_{2,m}$ matrices as a column vector inside $\mathbf{f}_2$ and do the gradient operation over this vector and then distribute the updated $\mathbf{F}_{2,m}$ matrices to RRUs, the proposed approach becomes more susceptible to increased quantization noise when the rows of $\mathbf{F}_{2,m}$ decrease and the number of $\mathbf{F}_{2,m}$ blocks increase after some point with the given parameters. Therefore, it is not possible to constantly increase the number of RRUs while decreasing their complexity and have good performance with our proposed approach.



(a) $M = 4$, $U = 4$, P_{total} = 12 dBW, $\bar{N} = 64$, with unconstrained fronthaul capacity



(b) $M = 4$, $U = 4$, P_{total} = 12 dBW, $\bar{N} = 64$, with $C = 64$ bits/channel use

Figure 4.13: Comparison of average run time per iteration performance of the proposed approach under unlimited and limited fronthaul capacity vs. L

In Fig. 4.13 we compare the average run times per iteration between the unlimited and limited case and observe that the proposed approach under limited fronthaul capacity has a slightly higher time complexity than the unconstrained case as it can be inferred from comparing Eqn. 3.31 and Eqn. 4.21.

4.5 Main Outcomes

A few key points that we can draw from these results are as follows:

- When there is no fronthaul capacity constraint, both the proposed approach and the initial point are able to exploit the increased degree of freedom due to increasing L and M values without a drawback.
- When the system is constrained by the fronthaul capacity, the initial point's performance can't contend with the increasing quantization noise after some point due to increasing M or L values.
- The proposed approach exhibits better convergence with the given parameters by using a variable step size as in Eqn. 3.24, compared to a constant step size.
- While the proposed approach is able to exploit the increased degree of freedom under limited fronthaul capacity due to increased per-RRU fronthaul links with multiple RRUs that have simpler structures compared to a more complex single RRU configuration up to a point, we observe that this improvement is not consistent as we increase the number of RRUs because of the increased quantization noise and reduced RRU complexity. Consequently, achieving continuous performance improvement with our proposed method proves untenable as we indefinitely increase the number of RRUs while reducing their complexities.
- With the proposed approach under limited fronthaul capacity, when $LM \geq U$ condition is satisfied, we observed that the user-averaged spectral efficiency for the increased number of users gets closer to the user-averaged capacity which is the total fronthaul capacity divided by the number of users. This observation suggests potential benefits for operators when the proposed approach is applied with the given parameters under a system featuring a multitude of users.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this thesis, we investigated precoding designs for C-RAN systems with limited fronthaul capacity. The legacy CPRI protocol in a radio network offloads most of the processing onto BBU, resulting in a substantial load on the fronthaul connection. However, the advent of 5G networks and e-CPRI protocol enable more flexible functional splits between BBU and RRU, an example of which is the Split 7-2 which reduces the load on the fronthaul but increases RRU complexity and production cost. In this work, we considered a C-RAN system with a fixed total number of antennas and fronthaul capacity, utilizing multiple RRUs with simpler structures compared to a single complex RRU structure. We initially investigated a multi-user, multi-RRU C-RAN without considering the fronthaul capacity limitations and proposed an iterative, gradient-based solution for a sum-rate maximization problem and discussed the simulation results. After this case, we considered a limited capacity fronthaul C-RAN and modeled the effect of limited capacity as additional quantization noise. We investigated the performance of the proposed approach under limited fronthaul capacity and made comparisons with the previous case.

Our findings demonstrate that utilizing multiple Remote Radio Units (RRUs) of lower complexity is viable with minimal capacity loss in scenarios involving communication with multiple users. This addresses one of the primary concerns within massive MIMO architecture. Furthermore, our proposed algorithm for determining the precoding matrices exhibits quicker convergence when multiple RRUs are employed. This reduction in computational complexity offers an additional advantage of utilizing multiple RRUs. It is worth noting that the selection of the number of RRUs should be made judiciously to ensure quantization with at least 2 bits is achieved in the fronthaul links. This observation introduces a trade-off between the number of RRUs, RRU complexity, and cost along with the resultant spectral efficiency that the overall massive MIMO system can provide.

In the future, this work can be improved further by considering more realistic channel models for the simulations, instead of the simple clustered channel model that we have employed. Also, channel error estimation errors can be taken into account instead of assuming perfect CSI. Additionally, accounting for the user admission criteria and proposing a joint optimization problem where we also consider variable power allocation for each user could lead to a more general precoding design procedure that may better serve needs in real-world practice. Finally, since we haven't considered the correlation between the additional quantization noise term and the quantized signal, the rate expressions that we have obtained in this work are actually lower bounds. Through the use of Bussgang decomposition [79], we can obtain a signal model for quantization noise that is uncorrelated with the quantized signal. This variant of the proposed approach can also be a topic for future work.

REFERENCES

[1] Ericcson, 2020. Available at https://www.ericsson.com/4ac9d8/assets/local/reports-papers/consumerlab/reports/2020/harnessing-the-5g-consumer-potential.pdf.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5g be?,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] S. Kemp, “Digital 2023: Global overview report - datareportal – global digital insights,” Feb 2023. Available at https://datareportal.com/reports/digital-2023-global-overview-report.

[4] D. Gesbert, M. Kountouris, R. W. Heath, C.-b. Chae, and T. Salzer, “Shifting the mimo paradigm,” *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[5] J. Hoydis, S. ten Brink, and M. Debbah, “Massive mimo: How many antennas do we need?,” pp. 545–550, Sept. 2011.

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

[7] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, “Channel estimation and linear precoding in multiuser multiple-antenna tdd systems,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2102–2116, Apr. 2011.

[8] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, “Analysis of the pilot contamination effect in very large multicell multiuser mimo systems for physical channel models,” pp. 3464–3467, May 2011.

[9] A. Hajisami, H. Viswanathan, and D. Pompili, “"cocktail party in the cloud":

Blind source separation for co-operative cellular communication in cloud ran," pp. 37–45, Oct. 2014.

[10] C. M. R. Institute, "C-ran the road towards green ran," Oct. 2011.

[11] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks—a technology overview," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[12] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," *IBM Journal of Research and Development*, vol. 54, no. 1, pp. 4:1–4:12, January-February 2010.

[13] M. Hadzialic, B. Dosenovic, M. Dzaferagic, and J. Musovic, "Cloud-ran: Innovative radio access network architecture," pp. 115–120, Sept. 2013.

[14] J. Mocerino, "5g backhaul/fronthaul opportunities and challenges - ncta technical papers." Available at `https://www.nctatechnicalpapers.com/Paper/2019/2019-5g-backhaul-fronthaul-opportunities-and-challenges`.

[15] M. Grayson, "Comparing lower layer splits for open fronthaul deployments," Apr 2021. Available at `https://blogs.cisco.com/sp/comparing-lower-layer-splits-for-open-fronthaul-deployments`.

[16] C.-L. I, J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li, "Recent progress on c-ran centralization and cloudification," *IEEE Access*, vol. 2, pp. 1030–1039, 2014.

[17] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[18] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Fronthaul compression and precoding design for c-rans over ergodic fading channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5022–5032, 2016.

[19] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Communications*, vol. 22, pp. 152–160, April 2015.

[20] M. Y. Lyazidi, L. Giupponi, J. Mangues-Bafalluy, N. Aitsaadi, and R. Langar, "A novel optimization framework for c-ran bbu selection based on resiliency and price," pp. 1–6, 2017.

[21] H. Ji, C. Sun, and W. Shieh, "Spectral efficiency comparison between analog and digital rof for mobile fronthaul transmission link," *Journal of Lightwave Technology*, vol. 38, no. 20, pp. 5617–5623, 2020.

[22] L. Cheng, M. Zhu, M. Muhammad Usman Gul, X. Ma, and G.-K. Chang, "Adaptive photonics-aided coordinated multipoint transmissions for next-generation mobile fronthaul," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 1907–1914, 2014.

[23] R. Zhang, F. Lu, M. Xu, S. Liu, P.-C. Peng, S. Shen, J. He, H. J. Cho, Q. Zhou, S. Yao, and G.-K. Chang, "An ultra-reliable mmw/fso a-rof system based on coordinated mapping and combining technique for 5g and beyond mobile fronthaul," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4952–4959, 2018.

[24] A. T. Liem, I.-S. Hwang, A. Nikoukar, and A. F. Pakpahan, "Sd-enabled mobile fronthaul dynamic bandwidth and wavelength allocation (dbwa) mechanism in converged twdm-epon architecture," pp. 1–6, 2018.

[25] A. Checko, "Cloud radio access network architecture. towards 5g mobile networks," 2016.

[26] L. Gavrilovska, V. Rakovic, A. Ichkov, D. Todorovski, and S. Marinova, "Flexible c-ran: Radio technology for 5g," pp. 255–264, 2017.

[27] M. Salazar-Palma, T. K. Sarkar, and D. Sengupta, "A brief chronology of the origin and developments of wireless communication and supporting electronics," pp. 1–4, 2009.

[28] R. Dangi, P. Lalwani, G. Choudhary, I. You, and G. Pau, "Study and investigation on 5G technology: A systematic review," *Sensors (Basel)*, vol. 22, Dec. 2021.

[29] U. K. Okpeki, J. O. Egwaile, and F. O. Edeko, “Performance and comparative analysis of wired and wireless communication systems using local area network based on ieee 802.3 and ieee 802.11,” *Journal of Applied Sciences and Environmental Management*, 2019.

[30] M. M. Kassem and M. K. Marina, “Future wireless spectrum below 6 ghz: A uk perspective,” pp. 59–70, 2015.

[31] M. A. Jensen, “A history of mimo wireless communications,” pp. 681–682, 2016.

[32] D. G. Brennan, “Linear diversity combining techniques,” *Proceedings of the IRE*, vol. 47, no. 6, pp. 1075–1102, 1959.

[33] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[34] E. Telatar, “Capacity of multi-antenna gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[35] G. Raleigh and J. Cioffi, “Spatio-temporal coding for wireless communication,” *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, 1998.

[36] S. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[37] I. Selinis, K. Katsaros, M. Allayioti, S. Vahid, and R. Tafazolli, “The race to 5g era; lte and wi-fi,” *IEEE Access*, vol. 6, pp. 56598–56636, 2018.

[38] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, “Survey of channel and radio propagation models for wireless mimo systems,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, p. 019070, Feb 2007.

[39] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach,

"A wideband statistical model for nlos indoor mimo channels," vol. 1, pp. 370–374 vol.1, 2002.

[40] M. Debbah and R. Muller, "Mimo channel modeling and the principle of maximum entropy," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1667–1690, 2005.

[41] B. S. Paul and R. Bhattacharjee, "Mimo channel modeling: A review," *IETE Technical Review*, vol. 25, no. 6, pp. 315–319, 2008.

[42] K. Yu and B. Ottersten, "Models for mimo propagation channels: a review," *Wireless Communications and Mobile Computing*, vol. 2, no. 7, pp. 653–666, 2002.

[43] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, 1987.

[44] J. Wallace and M. Jensen, "Statistical characteristics of measured mimo wireless channel data and comparison to conventional models," vol. 2, pp. 1078–1082 vol.2, 2001.

[45] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[46] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3818–3832, 15 July15, 2017.

[47] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[48] A. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.

[49] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, 2010.

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

[51] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale mimo systems," pp. 2929–2933, 2015.

[52] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic mimo radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, 2002.

[53] K. Zhi, C. Pan, H. Ren, and K. Wang, "Statistical csi-based design for reconfigurable intelligent surface-aided massive mimo systems with direct links," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1128–1132, 2021.

[54] Z. Zheng, C. Zhu, B. Jiang, W. Zhong, and X. Gao, "Statistical channel state information acquisition for massive mimo communications," pp. 1–5, 2015.

[55] A. Liu and V. K. N. Lau, "Impact of csi knowledge on the codebook-based hybrid beamforming in massive mimo," *IEEE Transactions on Signal Processing*, vol. 64, no. 24, pp. 6545–6556, Dec. 2016.

[56] D. M. Kim, J. Park, E. De Carvalho, and C. N. Manchon, "Massive mimo functionality splits based on hybrid analog-digital precoding in a c-ran architecture," pp. 1527–1531, 2017.

[57] M. N. Hamdy, "Beamformers explained," 2020. Available at `https://www.commscope.com/beamformers-explained/`.

[58] W. Tan, W. Huang, X. Yang, Z. Shi, W. Liu, and L. Fan, "Multiuser precoding scheme and achievable rate analysis for massive mimo system," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, p. 210, Aug 2018.

[59] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[60] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012. Available at `http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html/`.

[61] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-ran: Challenges and opportunities," pp. 1–6, 2020.

[62] O. Martínez, E. De Carvalho, and J. Nielsen, "Towards very large aperture massive mimo: A measurement based study," pp. 281–286, 2014.

[63] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.

[64] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive mimo in sub-6 ghz and mmwave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, 2019.

[65] B. Hu, C. Hua, C. Chen, and X. Guan, "Joint beamformer design for wireless fronthaul and access links in c-rans," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2869–2881, 2018.

[66] L. Liu and R. Zhang, "Optimized uplink transmission in multi-antenna c-ran with spatial compression and forward," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5083–5095, Oct. 2015.

[67] Z. Wang, R. Liu, H. Li, M. Li, and Q. Liu, "Hybrid beamforming design for c-ran based mmwave cell-free systems," pp. 1–5, 2020.

[68] X. Li, S. You, L. Chen, A. Liu, and Y. E. Liu, "A new algorithm for the weighted sum rate maximization in mimo interference networks," pp. 147–152, 2015.

[69] S. Hu, Y. Zhang, X. Wang, and G. B. Giannakis, "Weighted sum-rate maximization for mimo downlink systems powered by renewables," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5615–5625, 2016.

[70] R. Zhang, S.-H. Leung, H. Wang, W. Tang, and Z. Luo, "Power minimization precoder design for uplink mimo systems with multi-group noma scheme," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10553–10569, 2021.

[71] X. Zhai, X. Chen, and Y. Cai, "Power minimization for massive mimo over-the-air computation with two-timescale hybrid beamforming," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 873–877, 2021.

[72] H.-T. Wai and W.-K. Ma, "A decentralized method for joint admission control and beamforming in coordinated multicell downlink," pp. 559–563, 2012.

[73] F. E. Kadan and A. Yilmaz, "Beamformer design with smooth constraint-free approximation in downlink cloud radio access networks," *IEEE Access*, vol. 9, pp. 36399–36416, 2021.

[74] J. BARZILAI and J. M. BORWEIN, "Two-Point Step Size Gradient Methods," *IMA Journal of Numerical Analysis*, vol. 8, pp. 141–148, 01 1988.

[75] A. Ambainis, Y. Filmus, and F. Le Gall, "Fast matrix multiplication: Limitations of the coppersmith-winograd method," p. 585–593, 2015.

[76] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink spectral and energy efficiency of cell-free massive mimo with optimal uniform quantization," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 223–245, 2021.

[77] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[78] L. Liu, S. Bi, and R. Zhang, "Joint power control and fronthaul rate allocation for throughput maximization in ofdma-based cloud radio access network," *IEEE Transactions on Communications*, vol. 63, pp. 4097–4110, Nov 2015.

[79] O. T. Demir and E. Bjornson, “The bussgang decomposition of nonlinear systems: Basic theory and mimo extensions [lecture notes],” *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, 2021.

Appendix A

EQUALITY BETWEEN ORIGINAL AND EQUIVALENT RATE EXPRESSIONS

In this part, we show the equality between the original rate expression in Eqn. 3.7 and equivalent expression in Eqn. 3.15. We first show the equality between the nominators of Eqn. 3.6 and Eqn. 3.13 by starting with a simple case such that $U = 1$, $L = 1$, $M = 1$, and $N = 2$ and express the nominator of Eqn. 3.6 as

$$(\mathbf{h}^H \mathbf{F}_2 \mathbf{f}_1)(\mathbf{h}^H \mathbf{F}_2 \mathbf{f}_1)^H, \tag{A.1}$$

without considering the power allocation. We can express this in a more open form as

$$\left(\begin{bmatrix} h_1^* & h_2^* \end{bmatrix} \begin{bmatrix} F_{2,1} \\ F_{2,2} \end{bmatrix} f_1 \right) \left(\begin{bmatrix} h_1^* & h_2^* \end{bmatrix} \begin{bmatrix} F_{2,1} \\ F_{2,2} \end{bmatrix} f_1 \right)^H, \tag{A.2}$$

considering the respective user, fronthaul link, RRU, and RRU antenna numbers that we have defined before. In this expression, the multiplications inside the parenthesis are conjugate transposes of each other. Therefore, we can find an equivalent expression that contains a vectorized $\mathbf{F}_2$ that we can use to differentiate with respect to $\mathbf{f}_2^H$ for only considering a single parenthesis and then multiplying it with its conjugate transpose. For the expression in Eqn. A.2, we have $\mathbf{f}_2 = \mathbf{F}_2$ so we can express the equivalent notation by starting the multiplication with $\mathbf{f}_2^H$ as

$$\left(\begin{bmatrix} F^*_{2,1} & F^*_{2,2} \end{bmatrix} \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix} \begin{bmatrix} f^*_1 \\ f^*_1 \end{bmatrix}\right) \left(\begin{bmatrix} F^*_{2,1} & F^*_{2,2} \end{bmatrix} \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \end{bmatrix} \begin{bmatrix} f^*_1 \\ f^*_1 \end{bmatrix}\right)^H . \tag{A.3}$$

Similarly, for $U = 1$, $L = 2$, $M = 1$, and $N = 2$ we express the nominator of Eqn. 3.6 as

$$\left(\begin{bmatrix} h^*_1 & h^*_2 \end{bmatrix} \begin{bmatrix} F_{2,1,1} & F_{2,1,2} \\ F_{2,2,1} & F_{2,2,2} \end{bmatrix} \begin{bmatrix} f_{1,1} \\ f_{1,2} \end{bmatrix}\right) \left(\begin{bmatrix} h^*_1 & h^*_2 \end{bmatrix} \begin{bmatrix} F_{2,1,1} & F_{2,1,2} \\ F_{2,2,1} & F_{2,2,2} \end{bmatrix} \begin{bmatrix} f_{1,1} \\ f_{1,2} \end{bmatrix}\right)^H . \tag{A.4}$$

We can express Eqn. A.4 with an equivalent that starts with $\mathbf{f}_2^H$ as

$$\left(\begin{bmatrix} F^*_{2,1,1} & F^*_{2,2,1} & F^*_{2,1,2} & F^*_{2,2,2} \end{bmatrix} \begin{bmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_1 & 0 \\ 0 & 0 & 0 & h_2 \end{bmatrix} \begin{bmatrix} f^*_{1,1} \\ f^*_{1,1} \\ f^*_{1,2} \\ f^*_{1,2} \end{bmatrix}\right) \left(\begin{bmatrix} F^*_{2,1,1} & F^*_{2,2,1} & F^*_{2,1,2} & F^*_{2,2,2} \end{bmatrix} \begin{bmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_1 & 0 \\ 0 & 0 & 0 & h_2 \end{bmatrix} \begin{bmatrix} f^*_{1,1} \\ f^*_{1,1} \\ f^*_{1,2} \\ f^*_{1,2} \end{bmatrix}\right)^H . \tag{A.5}$$

Repeating the above steps for increasing U, L, M, and N values, we can obtain the expressions in Eqn. 3.14 and can express Eqn. 3.6 as Eqn. 3.13. Since the rate expression depends on the SINR, we can obtain Eqn. 3.20 by inserting Eqn. 3.13 into Eqn. 3.7.

Appendix B

EQUALITY BETWEEN ORIGINAL AND EQUIVALENT PER RRU TRANSMIT POWER CONSTRAINT EXPRESSIONS

In this part, we show the equivalence between the original per-RRU transmit power expression in Eqn. 3.9 and equivalent expression in Eqn. 3.16.Since Eqn. 3.9 contains trace operation, we first recall that the gradient of a matrix $\mathbf{X}$ for a first-order trace operation can be expressed as

$$\nabla_{\mathbf{X}} \mathrm{Tr}(\mathbf{AXB}) = \mathbf{A}^T \mathbf{B}^T, \tag{B.1}$$

for matrices $\mathbf{A}$ and $\mathbf{B}$. Therefore, our aim is to find an equivalent expression that contains $\mathbf{f}_2$ for Eqn. 3.9 such that

$$\nabla_{\mathbf{f}_2^H} \mathrm{Tr}(\mathbf{A}\mathbf{f}_2^H \mathbf{B}) = \mathbf{A}^T \mathbf{B}^T. \tag{B.2}$$

For a simple case such as $M = 1$, $N = 2$, $L = 1$, and $U = 1$, we can express the matrix multiplication in Eqn. 3.9 as

$$\left(\begin{bmatrix} F_{2,1} \\ F_{2,2} \end{bmatrix} \begin{bmatrix} f_1 \end{bmatrix} \right)^H \left(\begin{bmatrix} F_{2,1} \\ F_{2,2} \end{bmatrix} \begin{bmatrix} f_1 \end{bmatrix} \right). \tag{B.3}$$

In this case, $\mathbf{f}_2 = \mathbf{F}_2$, so we can express Eqn. B.3 as

$$(\mathbf{f}_2 f_1)^H (\mathbf{f}_2 f_1). \tag{B.4}$$

For the case when $M = 1$, $N = 2$, $L = 2$, and $U = 1$, the expression inside the trace in eq 3.9 is

$$\left(\begin{bmatrix} F_{2,1,1} & F_{2,1,2} \\ F_{2,2,1} & F_{2,2,2} \end{bmatrix}\begin{bmatrix} f_{1,1} \\ f_{1,2} \end{bmatrix}\right)^H\left(\begin{bmatrix} F_{2,1,1} & F_{2,1,2} \\ F_{2,2,1} & F_{2,2,2} \end{bmatrix}\begin{bmatrix} f_{1,1} \\ f_{1,2} \end{bmatrix}\right). \tag{B.5}$$

For this expression, trace is equal to

$$\begin{aligned} &|F_{2,1,1}f_{1,1}|^2 + |F_{2,1,2}f_{1,2}|^2 + |F_{2,2,1}f_{1,1}|^2 + |F_{2,2,2}f_{1,2}|^2 + F^*_{2,1,1}f^*_{1,1}F_{2,1,2}f_{1,2} \\ &+ F^*_{2,1,2}f^*_{1,2}F_{2,1,1}f_{1,1} + F^*_{2,2,1}f^*_{1,1}F_{2,2,2}f_{1,2} + F^*_{2,2,2}f^*_{1,2}F_{2,2,1}f_{1,1}. \end{aligned} \tag{B.6}$$

In Eqn. B.6, the fifth term is the conjugate of the sixth term, and the seventh term is the conjugate of the eighth term. So the trace is equal to

$$|F_{2,1,1}f_{1,1}|^2 + |F_{2,1,2}f_{1,2}|^2 + |F_{2,2,1}f_{1,1}|^2 + |F_{2,2,2}f_{1,2}|^2. \tag{B.7}$$

Consider the expression below which contains $\mathbf{f}_2$

$$\left(\begin{bmatrix} F_{2,1,1} \\ F_{2,2,1} \\ F_{2,1,2} \\ F_{2,2,2} \end{bmatrix}\begin{bmatrix} f_{1,1} & f_{1,2} \end{bmatrix}\right)^H\left(\begin{bmatrix} F_{2,1,1} \\ F_{2,2,1} \\ F_{2,1,2} \\ F_{2,2,2} \end{bmatrix}\begin{bmatrix} f_{1,1} & f_{1,2} \end{bmatrix}\right). \tag{B.8}$$

For this expression, the trace is

$$\begin{aligned} &|F_{2,1,1}f_{1,1}|^2 + |F_{2,1,2}f_{1,2}|^2 + |F_{2,2,1}f_{1,1}|^2 + |F_{2,2,2}f_{1,2}|^2 \\ &+ |F_{2,1,1}f_{1,2}|^2 + |F_{2,1,2}f_{1,1}|^2 + |F_{2,2,1}f_{1,2}|^2 + |F_{2,2,2}f_{1,1}|^2. \end{aligned} \tag{B.9}$$

In Eqn. B.9, the summation of the first 4 terms is equal to the summation of the last 4 terms. Therefore, the second trace value is 2, which is equal to L, times the first trace value. Repeating the above steps for different configurations, we observe that the ratio of L is preserved through different L, U, and N values which leads us to the expression in Eqn. 3.16.

Appendix C

EQUALITY BETWEEN ORIGINAL AND EQUIVALENT PER RRU TRANSMIT POWER CONSTRAINT EXPRESSIONS UNDER LIMITED FRONTHAUL CAPACITY

In this part, we show the equivalence of the per-RRU power transmit constraint and the equivalent per-RRU power transmit constraint that we have obtained as Eqn. 4.15. We have already shown the equivalence for this expression when $\mathbf{G}_m = \mathbf{0}$ in Appendix B. We only show the equivalence between $\frac{Tr(\mathbf{G_m})}{NL}$ and $Tr(\mathbf{F}_{2,m}\mathbf{Q}_m\mathbf{F}_{2,m}^H)$ in this part. We first start by considering the matrix multiplication inside the normal expression, which is $\mathbf{F}_{2,m}\mathbf{Q}_m\mathbf{F}_{2,m}^H$, for the simple case; $U = 1$, $L = 1$, $M = 1$, $N = 2$ as

$$\begin{bmatrix} F_{2,1} \\ F_{2,2} \end{bmatrix} \mathbb{E}[qq^*] \begin{bmatrix} F_{2,1}^* & F_{2,2}^* \end{bmatrix}. \tag{C.1}$$

With the values that we have defined, we have $\mathbf{f}_2 = \mathbf{F}_2$ and $\mathbf{Q}_m = \mathbf{Q}$ for this case and the trace of this expression is equal to

$$\mathbb{E}[qq^*](|F_{2,1}|^2 + |F_{2,2}|^2). \tag{C.2}$$

When $M = 1$, $N = 2$, $L = 2$, and $U = 1$, we have the matrix multiplication in Eqn. 4.9 as

$$\begin{bmatrix} F_{2,1,1} & F_{2,1,2} \\ F_{2,2,1} & F_{2,2,2} \end{bmatrix} \mathbb{E}\left[\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \begin{bmatrix} q_1^* & q_2^* \end{bmatrix} \right] \begin{bmatrix} F_{2,1,1}^* & F_{2,2,1}^* \\ F_{2,1,2}^* & F_{2,2,2}^* \end{bmatrix}, \tag{C.3}$$

and since we have i.i.d zero-mean Gaussian distribution for the quantization noise, we obtain the traces of the expression in Eqn. C.3 as

$$\mathbb{E}[q_1 q_1^*] \left|F_{2,1,1}\right|^2 + \mathbb{E}[q_2 q_2^*] \left|F_{2,1,2}\right|^2 + \mathbb{E}[q_1 q_1^*] \left|F_{2,2,1}\right|^2 + \mathbb{E}[q_1 q_1^*] \left|F_{2,2,2}\right|^2 . \tag{C.4}$$

Now, consider the below expression that is containing vectorized $\mathbf{f}_2$

$$\begin{bmatrix} F_{2,1,1} \\ F_{2,2,1} \\ F_{2,1,2} \\ F_{2,2,2} \end{bmatrix} \mathbb{E} \left[\begin{bmatrix} q_1 & q_1 & q_2 & q_2 \end{bmatrix} \begin{bmatrix} q_1^* \\ q_1^* \\ q_2^* \\ q_2^* \end{bmatrix} \right] \begin{bmatrix} F_{2,1,1}^* & F_{2,2,1}^* & F_{2,1,2}^* & F_{2,2,2}^* \end{bmatrix} . \tag{C.5}$$

The trace of this expression is equal to

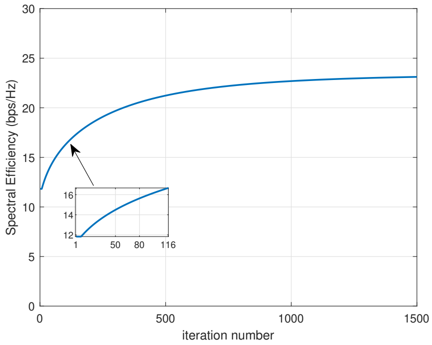
$$2(\mathbb{E}[q_1 q_1^*] + \mathbb{E}[q_2 q_2^*])(\left|F_{2,1,1}\right|^2 + \left|F_{2,2,1}\right|^2 + \left|F_{2,1,2}\right|^2 + \left|F_{2,2,2}\right|^2). \tag{C.6}$$

The result of the expression in Eqn. C.6 is equal to the 4 times the result in Eqn. C.4. When we repeat the above steps for different configurations, we observe that the ratio between $Tr(\mathbf{G_m})$ and $Tr(\mathbf{F}_{2,m} \mathbf{Q}_m \mathbf{F}_{2,m}^H)$ is preserved as $N \times L$, which leads us to the expression in Eqn. 4.15.

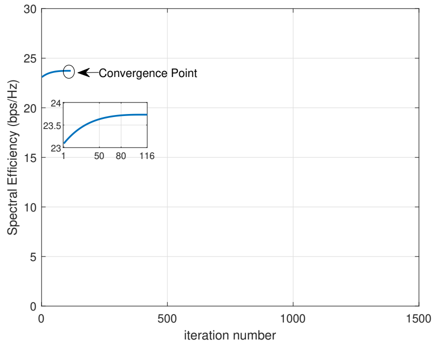
Appendix D

CONVERGENCE OF UTILIZED INITIAL POINT COMPARED TO RANDOM INITIAL POINT

In this part, we show the advantage of the initial point selection procedure that is described in Section 3.3 over a randomly selected initial point from the standard complex normal distribution over the convergence simulations with a single channel realization for the sum rate expression. Also, average iteration numbers when convergence has occurred can be seen in Table D.1, where average iteration numbers are obtained by running 50 Monte Carlo trials and averaging the iteration numbers at the convergence point for different configurations. In Algorithm 1, we have mentioned that convergence is declared if the difference between the last 15 iterations of the proposed approach is smaller than 10^{-5}. Therefore, selecting a sub-optimal initial point that is closer to the local minimum for the gradient-descent (or maximum in the ascent sense) has an advantage over a randomly selected initial point. As observed in the figures below, selecting a random initial point results in a random starting point for the sum rate value which is lower compared to the starting sum rate value for the utilized initial point. The utilized initial point leads to a faster convergence compared to a randomly selected initial point. Therefore, selecting a sub-optimized initial point leads to a faster convergence compared to a randomly selected initial point for the gradient descent algorithm.

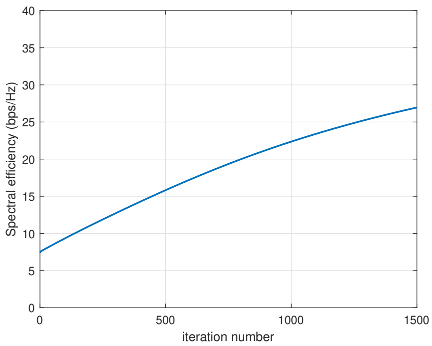


(a) Random initial point, $U = 2$, $M = 8$, $L = 3$, P_{total} = 12 dBW, $\bar{N} = 64$, with unconstrained fronthaul capacity

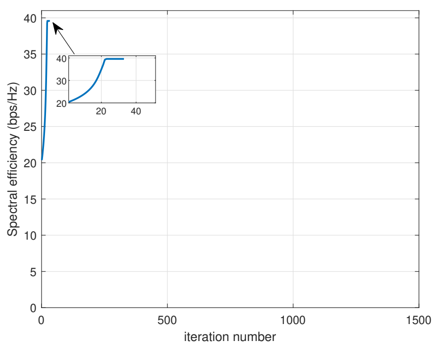


(b) Utilized initial point, $U = 2$, $M = 8$, $L = 3$, P_{total} = 12 dBW, $\bar{N} = 64$, with unconstrained fronthaul capacity

Figure D.1: Comparison of the sum rate convergence with the utilized initial point versus a randomly selected initial point under no fronthaul capacity constraints.



(a) Random initial point, $U = 4$, $M = 16$, $L = 2$, P_{total} = 12 dBW, $\bar{N} = 64$, $C = 64$ bits/channel use



(b) Utilized initial point, $U = 4$, $M = 16$, $L = 2$, P_{total} = 12 dBW, $\bar{N} = 64$, $C = 64$ bits/channel use

Figure D.2: Comparison of the sum rate convergence with the utilized initial point versus a randomly selected initial point when there is limited fronthaul capacity.

Initial point type	U	M	L	C	Avg. iteration
Random	2	8	3	-	1436
Utilized	2	8	3	-	154
Random	4	16	2	-	194
Utilized	4	16	2	-	17
Random	2	8	3	64	1084
Utilized	2	8	3	64	52
Random	4	16	2	64	1484
Utilized	4	16	2	64	212

Table D.1: Comparisons of average iterations until convergence between random and utilized initial points

Appendix E

THE SELECTION OF QUANTIZATION COVARIANCE AND OPTIMALITY DISCUSSION

In our work, we uniformly quantize a given zero-mean Gaussian input source. The quantized message vector that is forwarded to RRU m is expressed as

$$\bar{\mathbf{x}}_m = \frac{P_{total}}{U}\mathbf{F}_{1,m}\mathbf{s} + \mathbf{q}_m. \tag{E.1}$$

Therefore, this implies the uniform quantization of a Gaussian random variable. Note that we don't consider the correlation between additive quantization noise and quantized signal, so the rate we obtain in our work actually serves as a lower bound. For uniform quantization of a Gaussian source, as is the case in our work, we have the Gaussian source PDF as shown in Fig. E.1, which is uniformly quantized by 2^D points.

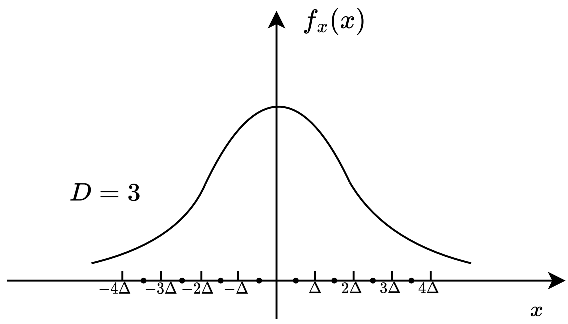


Figure E.1: Gaussian distributed source probability density

In [77], optimal step sizes are found by minimization of MSE expression where MSE is calculated by assuming uniform variance for the Gaussian source from the integral

$$\text{MSE} = 2\sum_{i=1}^{2^D/2}\int_{(i-1)\Delta}^{i\Delta}\left(x-(i-\frac{1}{2}\Delta)\right)^2 f_x(x)dx +$$
$$2\int_{\frac{2^D}{2}\Delta}^{\infty}\left(x-\left(\frac{2^D}{2}-\frac{1}{2}\right)\Delta\right) f_x(x)dx\ . \tag{E.2}$$

The work in [77] finds optimal step size values and hence optimal reconstruction points for $\bar{x}_{m,l}$ for RRU m and link l, numerically by differentiating the expression in Eqn. E.2 with respect to Δ and equating it to zero. We have found optimal MSE values with the optimal step sizes that are given in [76] from the result of Eqn. E.2 and these values can be seen from Table 4.2 for different numbers of quantization bits. For example, for 1 bit quantization, we assign the reconstruction points for $\bar{x}_{m,l}$ as $-\frac{1.596}{2}$ and $\frac{1.596}{2}$ according to the result of Table 4.2.