LAKEHEAD UNIVERSITY

MASTERS THESIS

User-Centric Clustering and Pilot Assignment in Cell-Free Networks

A Stochastic Optimization Approach

Author:
Ahmed ABOULFOTOUH

Supervisor:
Dr. Zubair FADLULLAH

Co-supervisor:
Dr. Mostafa FOUDA

Internal Examiner:
Dr. Muhammad ASADUZZAMAN

External Examiner:
Dr. Al-Sakib Khan PATHAN

A thesis submitted in fulfillment of the requirements for the degree of Masters of Science

in the

Computer Science Department

Declaration of Authorship

I, Ahmed ABOULFOTOUH, declare that this thesis titled, "User-Centric Clustering and Pilot Assignment in Cell-Free Networks" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: *Ahmed Aboulfotouh*

Date: 30th April 2023

LAKEHEAD UNIVERSITY

Abstract

Computer Science Department

Masters of Science

User-Centric Clustering and Pilot Assignment in Cell-Free Networks

A Stochastic Optimization Approach

by Ahmed ABOULFOTOUH

Current *5G* networks, primarily built on the cellular massive MIMO physical layer technology, achieved significant improvement in spectral efficiency as compared to previous generations. Nevertheless, there is always an increasing demand for higher data rates, and more reliable and uniform service. After successful massive MIMO deployments, it has become a natural question, "what will the physical layer in *beyond 5G* and *6G* networks be like?"

Cell-free massive MIMO has emerged as a promising physical layer technology for supporting future deployments in *beyond 5G* and *6G* networks. The main concept is to go beyond the cellular paradigm by employing an ultra dense deployment of small-sized multi-antenna access points (APs) which cooperate to serve users in the coverage area, eliminating the notion of boundaries between cells. The cell-free architecture has shown the capability of providing uniform service within the coverage area, while cellular networks suffer from poor performance at cell edges. It also has better ability to manage interference due to cooperation between APs which is not the case in cellular networks with no cooperation.

The most practical form of this paradigm is *user-centric cell-free massive MIMO*. Instead of allowing all the APs to serve all the users in the network, each user is served by a subset of the APs which ensures that network operation is scalable as the number of users grows. The main objective of this thesis is to provide a structured approach to design the cluster of APs that serve each user which is known as the *user-centric clustering problem*. On the pursuit to solve the clustering problem, there is another problem which is tightly connected to it, the *pilot assignment problem*. Both problems must be solved together to ensure satisfactory network-wide performance.

The thesis provides a mathematical formulation for each of the user-centric clustering and the pilot assignment problems as *stochastic non-linear binary integer programs* which are solved using *sample average approximation* and the *genetic algorithm*. The pilot assignment problem is formulated such that it takes into account the user-centric clusters while choosing the pilot assignments which makes the optimization more accurate and efficient. Numerical experiments show that the resulting solutions outperform heuristic baseline algorithms from the literature, leading to reasonable spectral efficiency gains. Furthermore, we propose an approximate approach to derive closed form expressions of the uplink and downlink SINR which eliminates the need for sampling.

In future work, more effort shall be directed towards finding more practical approaches to realize the optimized solutions of both the user-centric clustering and pilot assignment problems with reasonable time-complexity. For instance, one candidate approach is to use machine learning and AI methods to learn structured approaches to solve both problems, using the optimized solutions as a reference for learning.

Acknowledgements

It is with great pleasure that I acknowledge the individuals who have provided unwavering support and encouragement throughout my academic journey.

Firstly, I express my sincerest gratitude to my supervisor, Dr. Zubair Fadlullah, for his invaluable guidance and assistance throughout this research project. Dr. Fadlullah was always available to offer his expertise and insight, and his dedication and support were instrumental in the successful completion of this work. I feel fortunate to have had such a knowledgeable and committed supervisor.

I would also like to extend my thanks to the professors in the Communication and Information Engineering Department at Zewail City, and those in the Mathematics Department, Dr. Ahmed Abdelsamea and Dr. Hatem Fayed, for their inspiration and intellectual stimulation. I would also like to express my sincere appreciation to Dr. Samy Soliman, who was my mentor in my undergraduate studies. His guidance, encouragement, and support were instrumental in shaping my academic pursuits.

Finally, I would like to express my heartfelt appreciation to my family for their constant love and encouragement. I would like to offer special thanks to my partner, Passant Ibrahim, for her unwavering support and motivation. Her unwavering presence has been a constant source of positivity and inspiration throughout the ups and downs of this academic journey.

Once again, I extend my deepest thanks to all who have supported me along the way, and I am grateful for the opportunity to acknowledge their contributions.

Contents

List of Figures

List of Tables

List of Abbreviations

AP	**A**ccess **P**oint
UE	**U**ser **E**quipment
MIMO	**M**ultiple **I**nput **M**ultiple **O**utput
CDF	**C**umulative **D**istribution **F**unction
SE	**S**pectral **E**fficiency
LTI	**L**inear **T**ime **I**nvariant
TDD	**T**ime **D**ivision **D**uplexing
DCC	**D**ynamic **C**ooperation **C**lustering
FIR	**F**inite **I**mpulse **R**esponse
NLOS	**N**on **L**ine of **S**ight
ULA	**U**niform **L**inear **A**rray
MMSE	**M**inimum **M**ean **S**quare **E**rror
PMMSE	**P**artial **M**inimum **M**ean **S**quare **E**rror
MR	**M**aximum Ratio
PRZF	**P**artial **R**egularized **Z**ero **F**orcing
SINR	**S**ignal to **I**nterference and **N**oise **R**atio
SNR	**S**ignal to **N**oise **R**atio
MOP	**M**ultiobjective **O**ptimization **P**roblem

List of Symbols

L	number of access points in a cell-free network
K	number of users in a cell-free network
N	number of antennas per access point
M	total number of antennas in a cell-free network
$\mathcal{M}_k$	set that includes indices of APs which serve UE k
$\mathbf{D}_{kl}$	clustering matrix that determines whether AP l serves UE k
$\mathbf{I}_N$	N-by-N identity matrix
$\mathbf{0}_{N\times N}$	N-by-N zero matrix
$\mathbf{0}_N$	N-by-1 zero vector
B_c	coherence bandwidth
$\mathcal{N}_{\mathbb{C}}$	complex circularly-symmetric Gaussian distribution
$\mathcal{N}$	Gaussian distribution
tr	trace operator
$\boldsymbol{h}_{kl}$	channel response between AP l and UE k
$\boldsymbol{h}_k$	collective channel between UE k and all APs
$\boldsymbol{y}_l^{\mathrm{ul}}$	received signal at AP l in the uplink
s_i	uplink signal transmitted by UE i
$\boldsymbol{n}_l$	independent additive receiver noise at AP l
p_i	transmit power of UE i in the uplink
$\boldsymbol{v}_{kl}$	receive combining vector assigned by AP l to UE k
$\boldsymbol{u}_l$	signal transmitted by AP l during the downlink
$\boldsymbol{w}_{kl}$	transmit precoding vector assigned by AP l to UE k
$\boldsymbol{y}_k^{\mathrm{dl}}$	received signal at UE k in the downlink
n_k	independent additive receiver noise at UE k
$\mathbf{R}_{kl}$	spatial channel correlation matrix of AP l and UE k
$\mathbf{R}_k$	collective spatial channel correlation matrix of UE k
F_{kl}	shadowing term between AP l and UE k
d_{kl}	distance between UE k and AP l
t_k	index of the pilot assigned to UE k
$\hat{\boldsymbol{h}}_{kl}$	channel estimate of AP l and UE k
$\tilde{\boldsymbol{h}}_{kl}$	channel estimation error of AP l and UE k
$\mathbf{C}_{kl}$	channel estimation error correlation matrix of AP l and UE k
$\hat{\boldsymbol{h}}_k$	collective channel estimate of UE k
$\mathbf{D}_k$	block diagonal clustering matrix of UE k
$\mathrm{SINR}_k^{\mathrm{ul}}$	uplink signal-to-interference-and-noise ratio of UE k
$\mathrm{SINR}_k^{\mathrm{dl}}$	downlink signal-to-interference-and-noise ratio of UE k
$\mathrm{SE}_k^{\mathrm{ul}}$	uplink spectral efficiency of UE k
$\mathrm{SE}_k^{\mathrm{dl}}$	downlink spectral efficiency of UE k
$\boldsymbol{v}_k$	collective receive combining vector of UE k
$\boldsymbol{w}_k$	collective transmit precoding vector of UE k
$\mathcal{S}_k$	set of UEs that are partially served by the same APs as UE k including UE k
$\mathcal{D}_l$	set of UEs served by AP l

$\hat{s}_{kl}$	data estimate of UE k using the received signal at AP l
$\hat{s}_k$	collective data estimate of UE k
$\boldsymbol{x}_k$	binary assignment vector respresenting the cluster of UE k
g_k	performance function of UE k
erf	Gauss error function
T	time interval for which the clusters are kept fixed
n_c	number of coherence frames within a clustering interval T
$\mathbf{T}_k^{\mathrm{ul}}$	uplink signal correlation matrix of UE k
$\mathbf{T}_k^{\mathrm{dl}}$	downlink signal correlation matrix of UE k
$\mathbf{Q}_{ki}^{\mathrm{ul}}$	uplink interference correlation matrix of UE k and UE i
$\mathbf{Q}_{ki}^{\mathrm{dl}}$	downlink interference correlation matrix of UE k and UE i
$\boldsymbol{c}_k$	receive combining correlation vector of UE k
b_k^{MR}	signal part of the uplink SINR for MR combining
$\bar{b}_k^{\mathrm{MR}}$	signal part of the downlink SINR for MR precoding
f_k^{MR}	interference part of the uplink SINR for MR combining
$\bar{f}_k^{\mathrm{MR}}$	interference part of the downlink SINR for MR precoding
$\boldsymbol{a}_k$	binary assignment vector respresenting the pilot allocation of UE k
τ_c	coherence time
τ_{ul}	time dedicated for uplink data transmission within a coherence frame
τ_{dl}	time dedicated for downlink data transmission within a coherence frame
τ_p	time dedicated for pilot transmission within a coherence frame
σ_{ul}^2	uplink receiver noise power
ζ_k	downlink signal intended for UE k
σ_{dl}^2	downlink receiver noise power
β_{kl}	large-scale fading coefficient of AP l and UE k
δ_{ki}	distance between UE k and UE i
θ	nominal elevation angle of incident multipath components
φ	nominal azimuth angle of incident multipath components
$\bar{\theta}$	elevation angle of incident multipath components
$\bar{\varphi}$	azimuth angle of incident multipath components
σ_θ	angular standard deviation of the elevation angle of incident multipath components
σ_φ	angular standard deviation of the azimuth angle of incident multipath components
ϕ_i	ith pilot sequence
$\mathcal{P}_k$	set of UEs that share same pilot as UE k including UE k
η_k	pilot transmit power of UE k
$\mathbf{\Psi}_{t_k l}$	received pilot correlation matrix with index t_k
ρ_k	downlink transmit power assigned to UE k
$\rho_{\max}$	maximum downlink transmit power

Chapter 1

Introduction

This chapter provides a general introduction to the *cell-free networking* paradigm. Section 1.1 introduces the fundamentals of cell-free networks and discusses their benefits over cellular networks. Section 1.2 describes the most practical form of cell-free networks which is *user-centric cell-free massive multiple-input-multiple-output* (MIMO). Our main contributions in this thesis are also highlighted. The purpose of this chapter is to provide a basic background on cell-free networks. Exact system model and mathematical formulations are presented in subsequent chapters.

1.1 Cell-Free Networks

A cell-free network consists of L geographically distributed access points (APs) throughout the coverage area, each equipped with N antennas such that the total number of antennas is $M = LN$. On the contrary to the cellular paradigm, there are no boundaries between the APs and they jointly cooperate to serve K single-antenna users residing in the coverage area. The APs are connected to a central processing unit (CPU) via high-speed fronthaul links, which is responsible for managing AP

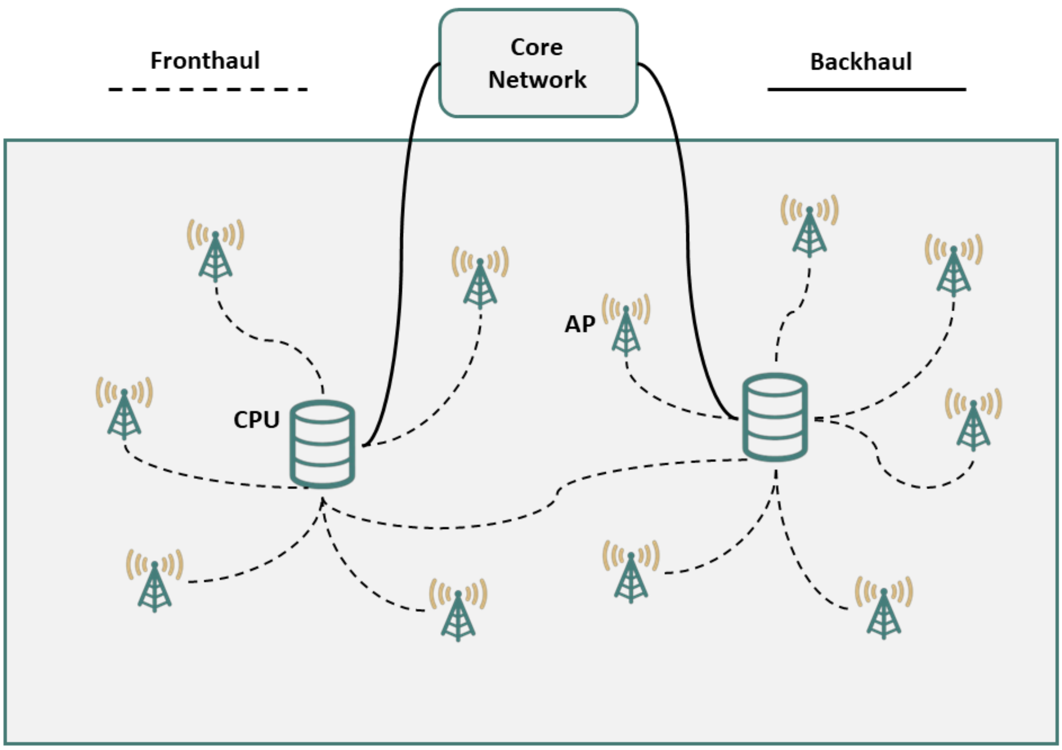

FIGURE 1.1: Illustration of a cell-free network

cooperation. The CPU is connected to the core network using backhaul links which enables communication with the internet or other information sources. The cell-free network is illustrated in figure 1.1. Note that the CPU is a logical entity, there could be more than one physical CPU connected using fronthaul links which act as the logical CPU. As it has been established in [1], there are three main benefits of cell-free networks over cellular networks:

1. Providing higher *signal-to-noise ratio* (SNR) with smaller spatial variations,
2. Better ability to mitigate interference,
3. Transmitting from distributed APs can boost the SNR.

The first benefit comes from the architecture of the network. In a cellular network, a user equipment (UE) located at cell edges suffers from poor performance due to inter-cell interference. While in a cell free network, a UE located anywhere in the network is likely to have multiple APs in close proximity. This allows UEs to receive a reliable and consistent signal, regardless of their location within the network. The second benefit comes from the coordination among APs, to avoid significantly interfering with each other, which is not the case in a traditional cellular network with no cooperation.

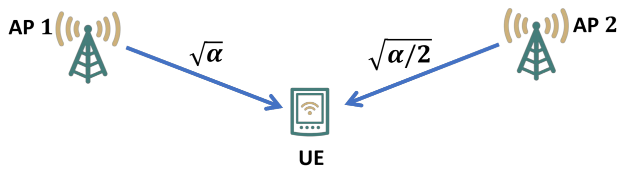

FIGURE 1.2: A UE in the vicinity of two APs

The third benefit is best illustrated with an example. Suppose there is a UE in the vicinity of two APs as illustrated in figure 1.2. The total power budget of each AP is p and the noise power is σ^2. The channel response between AP 1 and the UE is $h_1 = \sqrt{\alpha}$ and the channel response between AP 2 and the UE is $h_2 = \sqrt{\alpha/2}$. Hence, the channel between AP 1 and the UE is better than the channel between AP 2 and the UE. It might seem that the best transmit strategy is to allocate all the power of AP 1 to the UE which results in the following SNR

$$\frac{p\alpha}{\sigma^2} \tag{1.1}$$

However, if we let both APs transmit with half of their power budget, the SNR is

$$\frac{p}{2\sigma^2}(h_1 + h_2)^2 \approx 1.46 \cdot \frac{p\alpha}{\sigma^2} \tag{1.2}$$

The resulting is SNR is higher for the *same total power budget* due to *coherent combination* between signals transmitted from different APs. For a more detailed discussion about the benefits of cell-free networks over cellular networks, the reader is encouraged to refer to [1].

While cell-free networks provide numerous benefits over traditional cellular networks, it is crucial to note that they may not entirely replace cellular networks in all scenarios. In some cases, the signaling overhead associated with the coordination among APs may outweigh the gains in performance, making cellular networks a more viable option.

Therefore, it is expected that both cell-free and cellular networks will coexist, each serving specific use cases based on their unique capabilities and requirements. Cellular networks may be more suitable for applications that require highly reliable and robust network coverage, such as emergency services, while cell-free networks may be better suited for delivering high-bandwidth applications in dense urban areas. Ultimately, the choice between cell-free and cellular networks will depend on the specific needs of each use case, with both networks offering distinct advantages in different scenarios. In section 1.2, we describe user-centric cell-free massive MIMO networks.

1.2 User-Centric Cell-Free Massive MIMO

In cellular networks, the coverage area is divided into cells and each cell has a base station that serves the UEs within the cell. The keyword massive MIMO refers to an operating regime where the total number of base station antennas is *considerably larger* as compared to the number of served UEs. Massive MIMO results in high spectral efficiency gains which led to it becoming the main physical layer technology in the 5G standard. Similarly, the number of APs, in a cell-free massive MIMO network, is considerably larger as compared to the number of UEs $L \gg K$ implying that $M \gg K$.

In the first form of cell-free published in [2], it is assumed that all APs in the network cooperate to serve all the UEs. However, it has been shown that such an approach is both unscalable and impractical in large networks [1]. By limiting the APs which serve each UE, to only the APs that can reach the UE with a non-negligible amount of power, we have the *user-centric* cell-free network [3]–[6]. Each UE is served by a cluster (subset) of APs. The clusters are selected based on the UE needs and they are *allowed to overlap* as illustrated in figure 1.3. The fact that clusters

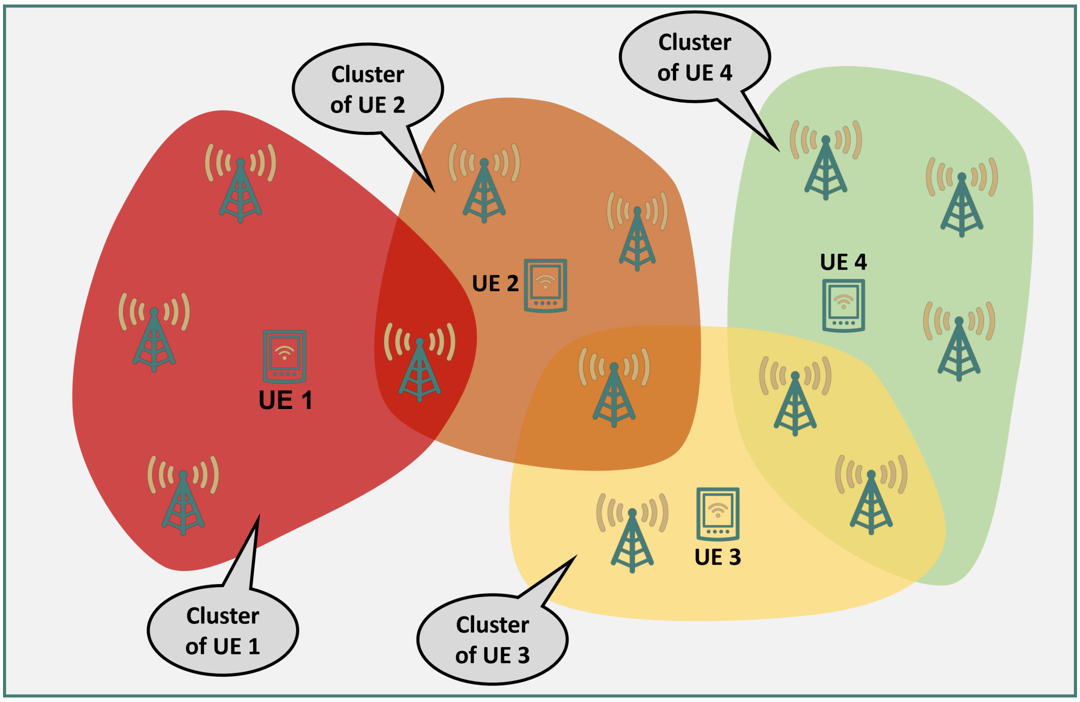

FIGURE 1.3: Illustration of a user-centric cell-free network where each UE is served by a cluster of APs.

can overlap showcases how the network is user-centric. The coverage area cannot be divided into disjoint sets of APs which serve disjoint subsets of UEs.

Choosing the cluster of APs that serves each UE is referred to as the *user-centric clustering problem*. The problem is formalized through the *Dynamic Cooperation Clustering* (DCC) framework [7]. Each UE k is served by a subset of the APs with indices in the set $\mathcal{M}_k \subset \{1, \cdots, L\}$ for $k = 1, \cdots, K$. It is dynamic because the sets are adapted according to time-variant characteristics such as UE locations, channel state, interference situation and service requirements. For notational convenience, a set of diagonal matrices are defined $\mathbf{D}_{kl}$, for $k = 1, \cdots, K$ and $l = 1, \cdots, L$ which determine whether AP l serves UE k.

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if AP } l \text{ serves UE } k \\ \mathbf{0}_{N\times N} & \text{otherwise} \end{cases} \tag{1.3}$$

where $\mathbf{I}_N$ is the N-by-N identity matrix and $\mathbf{0}_{N\times N}$ is the N-by-N zero matrix.

On the pursuit of solving the user-centric clustering problem, there is another problem that is closely related - the *pilot assignment problem*. Channel estimation is a vital aspect of network operation. It involves the transmission of known pilot sequences by each UE to estimate the channel response. APs then use the received pilot sequences, along with a statistical inference model, to estimate the channel. Ideally, the pilot sequence transmitted by each UE is orthogonal to the pilot sequences transmitted by all other UEs to avoid *pilot contamination*. However, this is practically infeasible due to the limited time assigned to pilot transmission. Hence, the number of orthogonal pilots is much smaller than the number of users in any practical network. Therefore, sharing pilot sequences is necessary.

Finding a solution to the pilot assignment problem constitutes choosing the pilot sequence for each UE so that minimal channel estimation error is achieved. The connection between the user-centric clustering and pilot assignment is best illustrated by an example. Suppose that UE 1 and UE 2 are served by *overlapping* subsets of APs, and UE 2 and UE 3 are served by *non-overlapping* subsets of APs. If the same pilot sequence is assigned to UE 1 and UE 2, there will be significant pilot contamination because the two UEs are partially served by the same APs. On the other hand, if the same pilot sequence is assigned to UE 2 and UE 3, pilot contamination will be less significant because there is no shared APs between the two UEs. Therefore, to ensure optimal network performance, both the pilot-assignment and user-centric clustering problems must be addressed together. This is because these two challenges are closely interconnected and solving them in isolation may lead to suboptimal solutions.

The majority of research on the user-centric clustering problem [8]–[13] and the pilot assignment problem [14]–[21] has been centered on designing heuristic algorithms. However, there is a notable lack of optimal benchmarks in these studies. As a result, our focus is on formulating each problem as an optimization problem and developing optimized benchmark solutions that can be used to evaluate heuristic algorithms. By establishing these benchmarks, we can better understand the effectiveness of existing algorithms and develop new approaches that lead to significant performance gains. Moreover, the optimized solutions can be used as a reference for machine learning algorithms.

Our contributions in this thesis are summarized as follows.

- Formulation of the user-centric clustering problem as a stochastic non-linear binary integer program and deriving new signal-to-interference-and-noise ratio (SINR) expressions which are more suited to the problem.

- Formulation of the pilot assignment problem as a stochastic non-linear binary integer program which takes into account the user-centric clusters in evaluating the desirability of pilot assignments.
- Providing nearly optimal benchmarks for both problems through employing sample average approximation and using the genetic algorithm to solve the resulting optimization models.
- Proposing a heuristic approach to derive closed form expressions of the stochastic optimization models which eliminates the need for sampling.

The remainder of the thesis is structured as follows: Chapter 2 introduces the time-division-duplexing (TDD) protocol, block fading model, and develops a system model for the uplink and downlink. It also describes the correlated Rayleigh fading channel model that captures spatial channel correlation characteristics. Chapter 3 discusses the centralized network operation, which includes the design of receive combining vectors and power control in the uplink, as well as the design of transmit precoding vectors and power allocation in the downlink. Chapter 4 defines per UE and network-wide performance metrics and formulates the user-centric clustering problem as a stochastic non-linear binary integer program. Additionally, it proposes a heuristic approach to compute closed forms of the developed stochastic optimization problem model. Chapter 5 provides a formal analysis that highlights the connection between the user-centric clustering problem and the pilot assignment problem. It also formulates the pilot assignment optimization problem as a stochastic non-linear binary integer program. Chapter 6 presents baseline heuristic methods for both problems from the literature and explains the procedure for obtaining optimized solutions using the genetic algorithm. Chapter 7 discusses numerical experiments to evaluate the proposed methodology. Finally, Chapter 8 concludes the thesis and includes proposals for future work.

Chapter 2

Background and System Model

In this chapter, we present the essential background on the system model of user-centric cell-free massive MIMO networks. Section 2.1 presents the block fading model of the wireless channel, including a description of the time-division-duplexing (TDD) protocol. Meanwhile, Section 2.2 describes the correlated Rayleigh fading channel model, which comprehensively captures the large-scale and small-scale fading characteristics, as well as the spatial correlation aspect.

2.1 System Model

2.1.1 Block Fading

The wireless communication channel is commonly viewed as a *linear time-variant system* that receives the transmitted signal, filters it, and outputs the received signal. The linearity of Maxwell's equations makes it a linear system, and its time-variant nature arises from the movements of the transmitter, receiver, and objects within the wireless channel.

Time-invariant systems are inconvenient for analysis. Instead, we focus on a sufficiently short time interval wherein the channel remains relatively constant, thus enabling the usage of a linear time-invariant (LTI) system model. The period in which the channel remains constant is known as the *channel coherence time* τ_c, and its length depends on the velocity of movement. Within this time frame, the channel can be characterized using a finite impulse response filter (FIR), where each tap represents a completely resolvable propagation path with distinct time delay and path loss. The filter exhibits *frequency selectivity*, meaning its response changes with variations in frequency. However, if we concentrate on a narrow frequency range, the frequency response of the channel can be considered constant. This frequency range is referred to as the *channel coherence bandwidth* B_c.

A *channel coherence block* is a block of time and frequency that spans the coherence time and bandwidth, respectively. Within this block, the channel response between a transmit and receive antenna remains constant and frequency-flat, allowing for its description using a single complex scalar. The channel coherence block is depicted in figure 2.1. Assuming that the channel realization within each coherence block is independent of those in other coherence blocks, we obtain the *block fading model*.

2.1.2 Time-Division-Duplexing (TDD) Protocol

To estimate channels efficiently, it is best to utilize a TDD protocol that employs each coherence block for both uplink and downlink transmissions. It is then sufficient to transmit pilots only in the uplink. Based on the received uplink pilots, each AP can then estimate the channels between itself and its served UEs. The channel estimates

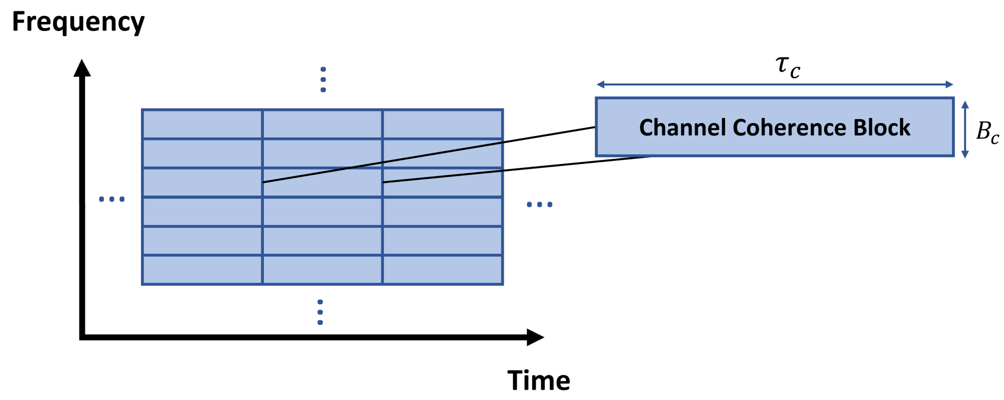

FIGURE 2.1: Illustration of the channel coherence block.

can be utilized for both uplink and downlink transmissions, thanks to the *channel reciprocity*. Each coherence block is split into three parts: τ_p for transmission of pilots, τ_{ul} for transmission of uplink data and τ_{dl} for transmission of downlink data.

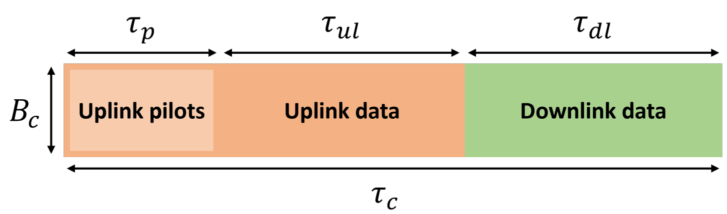

FIGURE 2.2: TDD protocol

The channel response between UE k and AP l in an arbitrary coherence block can be represented by an N-dimensional vector $\boldsymbol{h}_{kl} \in \mathbb{C}^N$ where the nth element represents the coefficient of the channel between the UE and the nth antenna of the AP. The collective channel $\boldsymbol{h}_k \in \mathbb{C}^M$ between UE k and all APs is defined as

$$\boldsymbol{h}_k = \begin{bmatrix} \boldsymbol{h}_{k1} \\ \vdots \\ \boldsymbol{h}_{kL} \end{bmatrix} \tag{2.1}$$

2.1.3 Uplink System Model

During the uplink, each UE transmits its signal to the APs. The received signal $\boldsymbol{y}_l^{\mathrm{ul}} \in \mathbb{C}^N$ at each AP l is a superposition of the signals transmitted by the UEs

$$\boldsymbol{y}_l^{\mathrm{ul}} = \sum_{i=1}^{K} \boldsymbol{h}_{il} s_i + \boldsymbol{n}_l \tag{2.2}$$

where $s_i \in \mathbb{C}$ is the signal transmitted by UE i of power $\mathbb{E}\{|s_i|^2\} = p_i$ and $\boldsymbol{n}_l$ is the independent additive receiver noise $\boldsymbol{n}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{ul}}^2 \mathbf{I}_N)$ with $\mathcal{N}_{\mathbb{C}}$ being the complex circularly-symmetric Gaussian distribution and σ_{ul}^2 is the receiver noise power. The channel is constant during each coherence block, while the signal and the receiver noise take new realizations every transmission symbol.

Utilizing the received signal, AP l applies a receive combining vector $\boldsymbol{v}_{kl} \in \mathbb{C}^N$ to estimate the signal of UE k, resulting in the computed estimate $\hat{s}_{kl}$.

$$
\begin{aligned}
\hat{s}_{kl} &= \boldsymbol{v}_{kl}^H \boldsymbol{y}_l^{\mathrm{ul}} \\
&= \underbrace{\boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl} s_k}_{\text{desired signal}} + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{K} \boldsymbol{v}_{kl}^H \boldsymbol{h}_{il} s_i}_{\text{interference}} + \underbrace{\boldsymbol{v}_{kl}^H \boldsymbol{n}_l}_{\text{noise}}
\end{aligned} \tag{2.3}
$$

However, this computation is done only if AP l serves UE k. Hence, we define the *effective receive combining vector* $\mathbf{D}_{kl}\boldsymbol{v}_{kl}$ which is equal to the receive combining vector $\boldsymbol{v}_{kl}$ if AP l serves UE k ($\mathbf{D}_{kl} = \mathbf{I}_N$) or $\mathbf{0}_N$ otherwise. The estimate $\hat{s}_{kl}$ is generally written as follows

$$
\hat{s}_{kl} = \boldsymbol{v}_{kl}^H \mathbf{D}_{kl} \boldsymbol{y}_l^{\mathrm{ul}} \tag{2.4}
$$

Design of the receive combining vectors should intuitively maximize the power of the desired signal and minimize the interference. Exact algorithms for designing receive combining vectors are discussed in chapter 3.

2.1.4 Downlink System Model

During the downlink, each AP l transmits its signal $\boldsymbol{u}_l$ which is a precoded superposition of signals intended for the UEs served by the particular AP.

$$
\boldsymbol{u}_l = \sum_{i=1}^{K} \mathbf{D}_{il} \boldsymbol{w}_{il} \zeta_i \tag{2.5}
$$

where $\boldsymbol{w}_{il}$ is the transmit precoding vector assigned to UE i by AP l, and ζ_i is the signal intended for UE i which is designed to have unit-power $\mathbb{E}\{|\zeta_i|^2\} = 1$.

The received signal $y_k^{\mathrm{dl}} \in \mathbb{C}$ at each UE k is a superposition of the transmitted signals by the APs

$$
\begin{aligned}
y_k^{\mathrm{dl}} &= \sum_{l=1}^{L} \boldsymbol{h}_{kl}^H \boldsymbol{u}_l + n_k && (2.6) \\
&= \sum_{l=1}^{L} \boldsymbol{h}_{kl}^H \left(\sum_{i=1}^{K} \mathbf{D}_{il} \boldsymbol{w}_{il} \zeta_i \right) + n_k && (2.7) \\
&= \underbrace{\sum_{l=1}^{L} \boldsymbol{h}_{kl}^H \mathbf{D}_{kl} \boldsymbol{w}_{kl} \zeta_k}_{\text{desired signal}} + \underbrace{\sum_{l=1}^{L} \sum_{\substack{i=1 \\ i \neq k}}^{K} \boldsymbol{h}_{kl}^H \mathbf{D}_{il} \boldsymbol{w}_{il} \zeta_i}_{\text{interference}} + \underbrace{n_k}_{\text{noise}} && (2.8)
\end{aligned}
$$

wher $n_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{dl}}^2)$ is the independent additive receiver noise.

The *effective transmit precoding vector* is defined to be $\mathbf{D}_{kl}\boldsymbol{w}_{kl}$ which is equal to the transmit precoding vector $\boldsymbol{w}_{kl}$ if AP l serves UE k and $\mathbf{0}_N$ otherwise. Transmit precoding vectors possess two important features: the direction $\boldsymbol{w}_{kl}/\|\boldsymbol{w}_{kl}\|$ indicating the transmission spatial direction and the average squared norm $\mathbb{E}\{\|\boldsymbol{w}_{kl}\|^2\}$, which determines the average transmit power. Design of transmit precoding vectors should intuitively maximize the power of the desired signal and minimize the interference. Exact algorithms for designing the transmit precoding vectors will be discussed in chapter 3

2.2 Channel Model

2.2.1 Correlated Rayleigh Fading

The most abundant wireless channel model for non-line-of-sight (NLOS) communication is *independent Rayleigh fading* where the channel between AP l and UE k is generated as $\boldsymbol{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \beta_{kl}\mathbf{I}_N)$. In a rich scattering environment, the received signal is the sum of many signal copies arriving from many propagation paths with seemingly random phase shifts which results in *small-scale fading*. The complex Gaussian distribution can be used to model the constructive and destructive interference between the paths motivated by the Central Limit Theorem. The variance β_{kl} of the distribution determines the average channel quality as the UE moves in a small area. In other words, it describes the large scale fading characteristics: path loss and shadowing.

While the uncorrelated Rayleigh fading model is widely used in theoretical studies due to its analytical tractability, empirical measurements indicate that elements of the channel vector $\boldsymbol{h}_{kl}$ are in reality correlated. This is called *spatial correlation* which arises from two phenomena: 1) Some spatial directions are more appropriate for transmitting signals to the UE compared to other directions; and 2) The antenna array geometry, which includes the antenna spacing and shape, enables transmission/reception of signals more effectively in certain directions than in others.

The *correlated Rayleigh fading* model is a stochastic channel model that accounts for spatial correlation. It is particularly applicable in a NLOS rich scattering environment where the channel vector $\boldsymbol{h}_{kl}$ can be modelled using the complex Gaussian distribution. The channel between AP l and UE k is generated as $\boldsymbol{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl})$ where $\mathbf{R}_{kl}$ is the positive semi-definite spatial correlation matrix of AP l and UE k. The correlation matrix $\mathbf{R}_{kl}$, which is defined as $\mathbb{E}\{\boldsymbol{h}_{kl}\boldsymbol{h}_{kl}^H\}$, captures the large-scale fading effects, such as path loss, shadowing, antenna gains, and spatial channel correlation. The Gaussian distribution models the small-scale fading, and the large-scale fading coefficient is determined by the average value of the diagonal elements of $\mathbf{R}_{kl}$.

$$\beta_{kl} = \frac{1}{N}\mathrm{tr}(\mathbf{R}_{kl}) \tag{2.9}$$

where tr is the trace operator. The collective channel of each UE k is generated as $\boldsymbol{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \mathbf{R}_k)$ where $\mathbf{R}_k = \mathrm{diag}(\mathbf{R}_{k1}, \cdots, \mathbf{R}_{kL})$ is the collective spatial channel correlation matrix.

2.2.2 Modeling of Large-Scale Fading

In line with [1], we assume ultra-dense deployment in an *urban* area where the APs are deployed in a plane ten meters above the UEs. This matches the 3GPP Urban Microcell Model defined in [22]. The path loss coefficient is computed as

$$\beta_{kl}\ [\mathrm{dB}] = -30.5 - 36.7\log_{10}\left(\frac{d_{kl}}{1m}\right) + F_{kl} \tag{2.10}$$

where d_{kl} is the three dimensional distance between AP l and UE k. $F_{kl} \sim \mathcal{N}(0, 4^2)$ represents the shadow fading. The shadowing terms are correlated as

$$\mathbb{E}\{F_{kl}F_{ij}\} = \begin{cases} 4^2 2^{-\delta_{ki}/9\ m} & l = j \\ 0 & k \neq i \end{cases} \tag{2.11}$$

where δ_{ki} is the distance between UE k and UE i. Equation (2.11) entails that only the shadowing terms related to the same AP are correlated, while shadowing terms related to different APs are uncorrelated which is sensible because the APs are separated by tens of wavelengths.

2.2.3 Modeling of Spatial Correlation

The spatial correlation matrix is influenced by two primary factors: 1) geometry of the antenna array; 2) angular distribution of the multi-path components. In the case of small-sized access points (APs), which are expected to be used in cell-free networks, a uniform linear antenna array (ULA) is commonly used where the N antennas are equally spaced with half-wavelength spacing. If all the multipaths arrive from the far-field of the array, the (m, l) element of a generic spatial correlation matrix $\mathbf{R}$ can be computed as

$$[\mathbf{R}]_{ml} = \beta \int \int e^{j\pi(m-l)\sin(\bar{\varphi})\cos(\bar{\theta})} f(\bar{\varphi}, \bar{\theta})\, d\bar{\varphi} d\bar{\theta} \tag{2.12}$$

where β is the common large-scale fading coefficient computed from (2.10), $\bar{\varphi}$ denotes the azimuth angle and $\bar{\theta}$ denotes the elevation angle with respect to the broadside of the array. $f(\bar{\varphi}, \bar{\theta})$ is the joint probability density function (PDF) of the azimuth and elevation angles $\bar{\varphi}$, $\bar{\theta}$. Therefore, the integral in (2.12) computes the expected value $\mathbb{E}\{e^{j\pi(m-l)\sin(\bar{\varphi})\cos(\bar{\theta})}\}$ for the multipath components distributed as $f(\bar{\varphi}, \bar{\theta})$. We assume the joint Gaussian PDF defined as

$$f(\bar{\varphi}, \bar{\theta}) = \frac{1}{2\pi\sigma_{\varphi}\sigma_{\theta}} e^{-\frac{(\bar{\theta}-\theta)^2}{2\sigma_{\theta}^2}} e^{-\frac{(\bar{\varphi}-\varphi)^2}{2\sigma_{\varphi}^2}} \tag{2.13}$$

where θ is the nominal elevation angle, φ is the nominal azimuth angle and σ_{φ} and σ_{θ} are the angular standard deviations (ASD) of the azimuth and elevation angles; respectively.

Chapter 3

Network Operation

A cell-free network operation is divided into uplink and downlink operation. Uplink operation involves *power control* and design of receive combining vectors, while downlink operation involves *power allocation* and design of transmit precoding vectors. Receive combining and transmit precoding designs are performed every coherence frame, while power allocation and power control are done when there are significant changes in large-scale fading characteristics. Two operation paradigms exist: centralized and distributed, with centralized processing being the focus of this thesis. In section 3.1, we discuss the minimum mean-square-error (MMSE) channel estimation procedure. Section 3.2 covers centralized uplink operation, including the definition of centralized uplink spectral efficiency, and design procedures for receive combining vectors and power control coefficients. Section 3.3 covers centralized downlink operation, including the definition of centralized downlink spectral efficiency, and design procedures for centralized transmit precoding vectors and power allocation coefficients.

3.1 Channel Estimation

Channel estimation is done by transmitting pilot sequences which are known to the APs. For centralized operation, the APs forward their received pilot signals to the CPU which undergoes channel estimation. Recall from the TDD protocol defined in section 2.1.2, there are τ_p samples dedicated for pilots transmission within each coherence frame. Ideally, each UE is assigned a unique pilot sequence which is orthogonal to the pilot sequences assigned to other UEs. However, because each pilot sequence spans τ_p samples, there are only τ_p mutually orthogonal pilot sequences $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{\tau_p} \in \mathbb{C}^{\tau_p}$ which are designed to have unit-power.

$$\boldsymbol{\phi}_i^H \boldsymbol{\phi}_j = \begin{cases} \tau_p & i = j \\ 0 & i \neq j \end{cases} \tag{3.1}$$

Therefore, if $\tau_p < K$ which is the case in a practical network, pilot reuse is inevitable. Denote the index of the pilot assigned to UE k as $t_k \in \{1, \cdots, \tau_p\}$ and define

$$\mathcal{P}_k = \{i : t_i = t_k, i = 1, \cdots, K\} \tag{3.2}$$

to be the set of UEs that share the same pilot as UE k including UE k. At the start of the uplink, each UE k transmits its assigned pilot sequence. The received pilot signal $\mathbf{Y}_l^{\text{pilot}}$ at AP l is

$$\mathbf{Y}_l^{\text{pilot}} = \sum_{i=1}^{K} \sqrt{\eta_i} \boldsymbol{h}_{il} \boldsymbol{\phi}_{t_i}^H + \mathbf{N}_l \tag{3.3}$$

where η_i is the transmit power of UE i during the pilot transmission phase, and $\mathbf{N}_l \in \mathbb{C}^{N\times\tau_p}$ is the independent receiver noise whose elements are distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma^2_{\mathrm{ul}})$. To estimate the channel between of AP l and UE k, we project the received pilot signal onto $\phi_{t_k}/\sqrt{\tau_p}$ which is the pilot sequence assigned to UE k. This removes interference from UEs which do not share the same pilot sequence as UE k.

$$\begin{aligned}\boldsymbol{y}^{\mathrm{pilot}}_{t_k l} &= \sum_{i=1}^{K} \sqrt{\frac{\eta_i}{\tau_p}} \boldsymbol{h}_{il}\boldsymbol{\phi}^H_{t_i}\boldsymbol{\phi}_{t_i} + \frac{1}{\sqrt{\tau_p}}\mathbf{N}_l\boldsymbol{\phi}_{t_i} \\ &= \underbrace{\sqrt{\eta_k\tau_p}\boldsymbol{h}_{kl}}_{\text{desired channel}} + \underbrace{\sum_{i\in\mathcal{P}_k/\{k\}} \sqrt{\eta_i\tau_p}\boldsymbol{h}_{il}}_{\text{interference}} + \underbrace{\frac{1}{\sqrt{\tau_p}}\mathbf{N}_l\boldsymbol{\phi}_{t_i}}_{\text{receiver noise}}\end{aligned} \tag{3.4}$$

The first term contains the desired channel scaled by $\sqrt{\eta_k\tau_p}$, the second term is the interference generated by the pilot-sharing UEs, and the third term is the receiver noise $\frac{1}{\sqrt{\tau_p}}\mathbf{N}_l\boldsymbol{\phi}_{t_i} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma^2_{\mathrm{ul}}\mathbf{I}_N)$. $\boldsymbol{y}^{\mathrm{pilot}}_{t_k l}$ is sufficient statistics for estimation of $\boldsymbol{h}_{kl}$ because no information is lost about the desired channel or the pilot-sharing UEs. The MMSE estimate of $\boldsymbol{h}_{kl}$ given $\boldsymbol{y}^{\mathrm{pilot}}_{t_k l}$ is

$$\hat{\boldsymbol{h}}_{kl} = \sqrt{\eta_k\tau_p}\mathbf{R}_{kl}\boldsymbol{\Psi}^{-1}_{t_k l}\boldsymbol{y}^{\mathrm{pilot}}_{t_k l} \tag{3.5}$$

where $\boldsymbol{\Psi}_{t_k l}$ is the received pilot correlation matrix defined as

$$\boldsymbol{\Psi}_{t_k l} = \sum_{i\in\mathcal{P}_k} \eta_i\tau_p\mathbf{R}_{il} + \sigma^2_{\mathrm{ul}}\mathbf{I}_N \tag{3.6}$$

The relation between the channel estimate $\hat{\boldsymbol{h}}_{kl}$ and the channel $\boldsymbol{h}_{kl}$ is

$$\boldsymbol{h}_{kl} = \hat{\boldsymbol{h}}_{kl} + \tilde{\boldsymbol{h}}_{kl} \tag{3.7}$$

where $\tilde{h}_{kl}$ is the channel estimation error. The channel estimate and the channel estimation error are independent random variables and their distribution is

$$\hat{\boldsymbol{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \eta_k\tau_p\mathbf{R}_{kl}\boldsymbol{\Psi}^{-1}_{t_k l}\mathbf{R}_{kl}) \tag{3.8}$$

$$\tilde{\boldsymbol{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{C}_{kl}) \tag{3.9}$$

where

$$\mathbf{C}_{kl} = \mathbf{R}_{kl} - \eta_k\tau_p\mathbf{R}_{kl}\boldsymbol{\Psi}^{-1}_{t_k l}\mathbf{R}_{kl} \tag{3.10}$$

The channel estimate $\hat{\boldsymbol{h}}_{kl}$ is computed only if AP l serves UE k.

Since we are interested in centralized operation, having knowledge about the collective channel $\boldsymbol{h}_k$ of each UE k is necessary. Recall that $\mathbf{R}_k \in \mathbb{C}_{M\times M}$ is the block diagonal collective spatial channel correlation matrix defined as $\mathrm{diag}(\mathbf{R}_{k1}, \cdots, \mathbf{R}_{kL})$. The collective channel estimate $\hat{\boldsymbol{h}}_k$ of UE k is defined as

$$\hat{\boldsymbol{h}}_k = \begin{bmatrix} \mathbf{D}_{k1}\hat{\boldsymbol{h}}_{k1} \\ \vdots \\ \mathbf{D}_{kL}\hat{\boldsymbol{h}}_{kL} \end{bmatrix} \tag{3.11}$$

Since we are operating in a user-centric cell-free network, not all the channel estimate $\hat{\boldsymbol{h}}_{kl}$ are known, because each AP serves only a subset of the UEs. Hence, an arbitrary entry $\mathbf{D}_{kl}\hat{\boldsymbol{h}}_{kl}$ in the collective channel estimate is equal to $\hat{\boldsymbol{h}}_{kl}$ if AP l serves UE k and

$\mathbf{0}_N$ otherwise. The collective channel estimate has the following distribution

$$\hat{\boldsymbol{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \eta_k \tau_p \mathbf{D}_k \mathbf{R}_k \mathbf{\Psi}_{t_k}^{-1} \mathbf{R}_k \mathbf{D}_k) \tag{3.12}$$

where $\mathbf{D}_k = \mathrm{diag}(\mathbf{D}_{k1}, \cdots, \mathbf{D}_{kL})$ is the block diagonal clustering matrix of UE k and $\mathbf{\Psi}_{t_k}^{-1} = \mathrm{diag}(\mathbf{\Psi}_{t_k 1}^{-1}, \cdots, \mathbf{\Psi}_{t_k L}^{-1})$ is a block diagonal matrix containing the inverses of the received pilot correlation matrices.

3.2 Centralized Uplink Operation

In centralized uplink operation, the APs transmit their received pilot and data signals to the CPU. The CPU then uses the received pilots for channel estimation and channel estimates are subsequently used for designing the receive combining vectors to perform data detection. The received uplink signals from UE k at its respective serving APs are utilized by the CPU to compute a collective data estimate, denoted as $\hat{s}_k$. This is achieved by summing the per-AP estimates, denoted as $\hat{s}_{kl}$. Recall from section 2.1 that the per AP data estimate is computed as the inner product $\hat{s}_{kl} = \boldsymbol{v}_{kl}^H \mathbf{D}_{kl} \boldsymbol{y}_l^{\mathrm{ul}}$ where $\boldsymbol{v}_{kl}$ is the receive combining vector designed for AP l and UE k. The collective data estimate $\hat{s}_k$ of UE k is

$$\begin{aligned} \hat{s}_k &= \sum_{l=1}^{L} \hat{s}_{kl} \\ &= \sum_{l=1}^{L} \boldsymbol{v}_{kl}^H \mathbf{D}_{kl} \boldsymbol{y}_l^{\mathrm{ul}} = \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{y}_{\mathrm{ul}} \end{aligned} \tag{3.13}$$

with $\boldsymbol{v}_k = \left[\boldsymbol{v}_{k1}^T \cdots \boldsymbol{v}_{k1}^T\right]^T \in \mathbb{C}^{LN}$ is the collective receive combining vector of UE k and $\boldsymbol{y}^{\mathrm{ul}} \in \mathbb{C}^{LN}$ is the collective uplink data signal given by

$$\boldsymbol{y}^{\mathrm{ul}} = \sum_{i=1}^{K} \boldsymbol{h}_i s_i + \boldsymbol{n} \tag{3.14}$$

where $\boldsymbol{n} = \left[\boldsymbol{n}_1^T \cdots \boldsymbol{n}_L^T\right]^T$ is the collective noise vector. We start by defining the spectral efficiency for centralized uplink operation which is the main performance metric of every UE and is used to obtain optimal receive combining design.

3.2.1 Spectral Efficiency

The spectral efficiency SE measured in bits/s/Hz is an evaluation of how efficiently a unit of bandwidth is used. Hence, the data rate of UE k can be defined as

$$\mathrm{Rate}_k = B \cdot \mathrm{SE}_k \tag{3.15}$$

where B is the bandwidth and SE_k is the spectral efficiency achieved by UE k.

In this work, we utilize the spectral efficiency expressions derived using the *use-and-then-forget* bounding technique. This name comes from the fact that channel estimates are utilized to design the receive combining vectors, and are subsequently disregarded before undergoing data detection. The expression is applicable for an arbitrary fading model and user-centric clusters. However, it has the drawback that it underestimates the achievable spectral efficiency, so it can be considered as a lower bound.

An achievable spectral efficiency $\mathrm{SE}_k^{\mathrm{ul}}$ of UE k in centralized uplink operation is

$$\mathrm{SE}_k^{\mathrm{ul}} = \frac{\tau_{\mathrm{ul}}}{\tau_c} \log_2 \left(1 + \mathrm{SINR}_k^{\mathrm{ul}}\right) \tag{3.16}$$

where

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2}{\sum_{i=1}^K p_i \mathbb{E}\left\{\left|\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_i\right|^2\right\} - p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2 + \sigma_{\mathrm{ul}}^2 \mathbb{E}\left\{\|\mathbf{D}_k \boldsymbol{v}_k\|^2\right\}} \tag{3.17}$$

The expectation is computed with respect to the channel realizations.

3.2.2 Centralized Receive Combining

The centralized receive combining vectors $\boldsymbol{v}_k$ for $k = 1, \cdots, K$ are adapted every coherence frame. The design rationale is to optimize the SINR which is defined in equation (3.17). The optimal combining vector which maximizes the SINR is referred to as MMSE combining and can be written as follows.

$$\boldsymbol{v}_k^{\mathrm{MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \mathbf{D}_k (\hat{\boldsymbol{h}}_i \hat{\boldsymbol{h}}_i^H + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{h}_k \tag{3.18}$$

where $\mathbf{C}_i = \mathrm{diag}(\mathbf{C}_{i1}, \cdots, \mathbf{C}_{iL})$ is a block diagonal matrix containing channel estimation error correlation matrices of UE i.

The complexity of computing the MMSE combining vector grows linearly with the number of users K. For the network to be scalable, the complexity associated with computing the combining vectors has to be finite if the number of users grows to infinity. A simplified version of MMSE combining can be obtained through partial MMSE (PMMSE) combining. Define the set

$$\mathcal{S}_k = \{i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN \times LN}\} \tag{3.19}$$

which includes the UEs that are partially served by the same APs as UE k. The PMMSE combining vector can be computed as

$$\boldsymbol{v}_k^{\mathrm{PMMSE}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k (\hat{\boldsymbol{h}}_i \hat{\boldsymbol{h}}_i^H + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{h}_k \tag{3.20}$$

The complexity of computing the PMMSE combining vector grows linearly with $|\mathcal{S}_k|$ which is the cardinality of the set $\mathcal{S}_k$. A further complexity reduction can be achieved by using the partial-regularized zero-forcing combining (PRZF) which is computed as

$$\boldsymbol{v}_k^{\mathrm{PRZF}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \hat{\boldsymbol{h}}_i \hat{\boldsymbol{h}}_i^H \mathbf{D}_k + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{h}_k \tag{3.21}$$

The simplest scalable combining scheme which maximizes the numerator of the uplink SINR defined in (3.17) is the maximum-ratio (MR) combining which is computed as

$$\boldsymbol{v}_k^{\mathrm{MR}} = \mathbf{D}_k \hat{\boldsymbol{h}}_k \tag{3.22}$$

However, this combining scheme is rarely used due its poor interference cancellation performance.

3.2.3 Power Control

Every UE k has a power budget $p_{\max}$ which we assume to be equal for all UEs. During the uplink, every UE transmits its data and has to choose a power p_k in range of 0 to $p_{\max}$. In most of the cell-free massive MIMO literature, the power control coefficients are kept fixed unless there is a significant change in the large-scale fading characteristics. Therefore, they are not adapted to channel realizations. The design rationale is to determine the amount by which each UE should decrease its transmit power to maximize a certain network wide utility function. However, in this thesis, we consider that the UEs transmit with full power in both pilot and data transmission phases which has been shown to be nearly optimal in many scenarios [1].

$$\eta_k = p_k = p_{\max}, \quad k = 1, \cdots, K \tag{3.23}$$

3.3 Centralized Downlink Operation

Recall the signal $\boldsymbol{u}_l$ transmitted by each AP l during the downlink is created as a precoded superposition of the UEs signals

$$\boldsymbol{u}_l = \sum_{i=1}^{K} \mathbf{D}_{il} \boldsymbol{w}_{il} \zeta_i \tag{3.24}$$

where $\boldsymbol{w}_{il}$ is the transmit precoding vector assigned by AP l to UE i, ζ_i is the data signal intended for UE i which is designed to have unit-power $\mathbb{E}\{|\zeta_i|^2\} = 1$ and the effective transmit precoding vector $\mathbf{D}_{il}\boldsymbol{w}_{il}$ is equal to the transmit precoding if AP l serves UE i and $\mathbf{0}_N$ otherwise. Recall from section 3.3 that the received downlink signal y_k^{dl} at UE k is

$$y_k^{\mathrm{dl}} = \sum_{l=1}^{L} \boldsymbol{h}_{kl}^H \boldsymbol{u}_l + n_k \tag{3.25}$$

$$= \sum_{l=1}^{L} \boldsymbol{h}_{kl}^H \left(\sum_{i=1}^{K} \mathbf{D}_{il} \boldsymbol{w}_{il} \zeta_i \right) + n_k \tag{3.26}$$

which can be written in a more compact form as

$$y_k^{\mathrm{dl}} = \sum_{i=1}^{k} \begin{bmatrix} \boldsymbol{h}_{k1} \\ \vdots \\ \boldsymbol{h}_{kL} \end{bmatrix}^H \begin{bmatrix} \mathbf{D}_{i1} \boldsymbol{w}_{i1} \\ \vdots \\ \mathbf{D}_{iL} \boldsymbol{w}_{iL} \end{bmatrix} \zeta_i + n_k$$

$$= \sum_{i=1}^{k} \boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i \zeta_i + n_k \tag{3.27}$$

where $\boldsymbol{h}_k = \left[\boldsymbol{h}_{k1}^T \cdots \boldsymbol{h}_{kL}^T\right]^T \in \mathbb{C}^{LN}$ is the collective channel vector and the collective precoding vector assigned to UE i is $\boldsymbol{w}_i = \left[\boldsymbol{w}_{i1}^T \cdots \boldsymbol{w}_{iL}^T\right]^T \in \mathbb{C}^{LN}$.

3.3.1 Spectral Efficiency

An achievable spectral efficiency $\mathrm{SE}_k^{\mathrm{dl}}$ of UE k in centralized downlink operation derived using the use-and-then-forget bounding technique is

$$\mathrm{SE}_k^{\mathrm{dl}} = \frac{\tau_{\mathrm{dl}}}{\tau_c} \log_2 \left(1 + \mathrm{SINR}_k^{\mathrm{dl}}\right) \tag{3.28}$$

where

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2}{\sum_{i=1}^{K} \mathbb{E}\left\{\left|\boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i\right|^2\right\} - \left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2 + \sigma_{\mathrm{dl}}^2} \tag{3.29}$$

The expectation is computed with respect to the channel realizations.

3.3.2 Centralized Transmit Precoding

The centralized transmit precoding vector is generally defined as

$$\boldsymbol{w}_k = \sqrt{\rho_k} \frac{\bar{\boldsymbol{w}}_k}{\sqrt{\mathbb{E}\{\|\bar{\boldsymbol{w}}_k\|^2\}}} \tag{3.30}$$

where $\bar{\boldsymbol{w}}_k$ determines the direction of the vector and ρ_k determines its magnitude. In other words, choosing ρ_k amounts to determining the power allocated to each UE k.

Design of transmit precoding vectors poses a greater challenge in comparison to the design of receive combining vectors, primarily due to the difference in the mathematical structure of SINR expressions. While receive combining vectors can be optimized on a per UE basis, this is not the case for transmit precoding vectors. This is because the selection of a precoding vector for a specific UE k impacts the performance of other UEs, which is not the case in the uplink. In reality, there is no single set of optimal transmit precoding vectors, as it involves a trade-off between different solutions. Therefore, a commonly utilized heuristic for designing transmit precoding vectors is the *uplink-downlink duality* theorem which is stated as follows.

Let $\{\mathbf{D}_i \boldsymbol{v}_i : i = 1, \cdots, K\}$ and $\{p_i : i = 1, \cdots, K\}$ denote the set of combining vectors and uplink transmit powers; respectively. If the transmit precoding vectors are selected as

$$\boldsymbol{w}_i = \sqrt{\rho_i} \frac{\boldsymbol{v}_i}{\sqrt{\mathbb{E}\{\|\mathbf{D}_i \boldsymbol{v}_i\|^2\}}} \tag{3.31}$$

then there exists a downlink power allocation policy $\{\rho_i : i = 1, \cdots, K\}$ with

$$\sum_{i=1}^{K} \frac{p_i}{\sigma_{\mathrm{ul}}^2} = \sum_{i=1}^{K} \frac{\rho_i}{\sigma_{\mathrm{dl}}^2} \tag{3.32}$$

such that

$$\mathrm{SINR}_k^{\mathrm{ul}} = \mathrm{SINR}_k^{\mathrm{dl}} \tag{3.33}$$

Hence, we use the uplink-downlink duality theorem to choose the direction of the transmit precoding vectors by utilizing the same expressions for the receive combining methods developed in the uplink. However, power allocation is done separately because the power allocation coefficients that come from this theorem are usually impractical to implement.

3.3.3 Centralized Power Allocation

During the downlink, AP l transmits the signal $\boldsymbol{u}_l = \sum_{i=1}^{K} \mathbf{D}_{il}\boldsymbol{w}_{il}\zeta_i$ which is assumed to have a maximum power $\rho_{\max}$. Therefore, the APs have to decide how to split their power among the users. We use the following heuristic to determine the power allocated to each UE which is commonly known as *fractional centralized power allocation* [1].

$$\rho_k = \rho_{\max} \frac{\left(\sqrt{\sum_{l\in\mathcal{M}_k} \beta_{kl}}\right)^{-1} \left(\sqrt{\omega_k}\right)^{-1}}{\max_{l\in\mathcal{M}_k} \sum_{i\in\mathcal{D}_l} \left(\sqrt{\sum_{l\in\mathcal{M}_i} \beta_{il}}\right)^{-1} \left(\sqrt{\omega_i}\right)} \tag{3.34}$$

with

$$\omega_k = \max_{l\in\mathcal{M}_k} \mathbb{E}\left\{\|\bar{\boldsymbol{w}}_{kl}\|^2\right\} \tag{3.35}$$

where $\bar{\boldsymbol{w}}_{kl}$ is the portion of the centralized precoding vector $\bar{\boldsymbol{w}}_k$ assigned to AP l and $\mathcal{D}_l$ is the set of UEs served by AP l.

Chapter 4

User-Centric Clustering Problem Formulation

This chapter provides a comprehensive mathematical formulation for user-centric clustering, presented as an optimization problem. Section 4.1 introduces the notation for the clustering problem and quantifies UE performance. Section 4.2 formulates the user-centric clustering problem as a stochastic multi-objective optimization problem (MOP). A subjective solution is proposed that converts the MOP to a single-objective optimization problem. Section 4.3 introduces new expressions for the signal-to-interference-plus-noise ratio (SINR) that are better suited to the clustering problem. Section 4.4 describes an approximate approach to compute the closed form of the SINR expected value in both the uplink and downlink.

4.1 Preliminaries

Consider a cell-free network with L geographically distributed APs, each equipped with N antennas, which cooperate to serve K users. In the user-centric approach, each UE k is served by a subset of the APs which are called clusters. We define K binary assignment vectors $\boldsymbol{x}_k \in \{0,1\}^L, k = 1, \cdots, K$ to represent the clusters.

$$x_{kl} = \begin{cases} 1 & \text{AP } l \text{ serves UE } k \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

The user-centric clustering problem is concerned with designing the K clusters to maximize a *utility function* or a *performance metric*. There are two viewpoints for assessing the performance of a clustering algorithm:

- Each individual UE performance,
- Network-wide performance which is a collection of simultaneously achievable UE performances.

Each UE k is assumed to have a performance function $g_k : \mathcal{R} \to \mathcal{R}$ of the SINR which measures the degree of satisfaction of the UE by its quality of service. The mathematical structure of the SINR differs between the uplink and downlink, leading to distinct downlink and uplink performances. Accordingly, we define the individual performance of each UE k as the following set:

$$\left\{ g_k(\text{SINR}_k^{\text{ul}}),\ g_k(\text{SINR}_k^{\text{dl}}) \right\} \tag{4.2}$$

The performance function generally depends on the current service being used. For example, voice traffic requires short delays and a minimum information rate that is constantly available (while higher rates are unnecessary). On the contrary, internet traffic can accept long delays and variations in the information rate, while the satisfaction is strictly increasing with the information rate. Candidates to the performance function are:

- *Information rate* $g_k(\text{SINR}_k) = \alpha_1 \log(1 + \alpha_2 \text{SINR}_k)$ where α_1 and α_2 are tunable parameters that account for different channel coding schemes or modulation types.
- *Bit-error rate* of a 16-QAM constellation $g_k(\text{SINR}_k) = {}^3\!/_8 \cdot \text{erf}\left(\sqrt{{}^2\!/_5 \cdot \text{SINR}_k}\right)$ where erf is the Gauss error function.

Design of user-centric clusters is done based on large-scale fading characteristics. Therefore, it is only adapted if there are significant changes in the positioning of the UEs leading to variation in the large-scale fading characteristics. Denote the time interval for which the clusters are kept fixed as $T = n_c \tau_c$ where n_c is the number of coherence frames within T which we refer to as the clustering interval. With each coherence frame, there is a new channel realization for each UE, resulting in changes in the SINR value.

4.2 Multiobjective Optimization Problem

Without loss of generality, the user-centric clustering problem can be formulated as the following multi-objective optimization problem (MOP), inspired by the approach presented in [23].

$$\max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K} \left\{ g_k\left(\text{SINR}_k^{\text{ul}}(n)\right), g_k\left(\text{SINR}_k^{\text{dl}}(n)\right) : k = 1, \cdots, K \wedge n = 1, \cdots, n_c \right\} \tag{4.3}$$

where $\text{SINR}_k^{\text{ul}}(n)$ and $\text{SINR}_k^{\text{dl}}(n)$ are the uplink and downlink SINR at time n; respectively. The MOP can be interpreted as searching for the clusters $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$ that maximize the performance of all UEs during the clustering interval T.

Since the performances of different UEs are coupled, there is generally not a single transmit strategy that simultaneously maximizes the performance of all UEs. Furthermore, the clusters are designed before the start of the clustering interval T. Hence, the performance metrics of the UEs defined by the set in (4.3) are unknown. In such case, we need to deal with the uncertainty in the objective functions. Optimization under uncertainty refers to this branch of optimization where there are uncertainties involved in the data or the model and is popularly known as stochastic programming or stochastic optimization.

Most of the approaches to solve stochastic optimization problems convert the problem into a deterministic one and then common optimization techniques can be used. One way to address the uncertainty is to optimize the expected value of the performance functions rather than the instantaneous value. If we implement that approach to the objective functions defined in (4.3), we get

$$\max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K} \left\{ \mathbb{E}\left\{ g_k\left(\text{SINR}_k^{\text{ul}}\right)\right\}, \mathbb{E}\left\{ g_k\left(\text{SINR}_k^{\text{dl}}\right)\right\} : k = 1, \cdots, K \right\} \tag{4.4}$$

where $\mathbb{E}$ is the expectation operator. The expectation is computed with respect to the time index n which we dropped for convenience in the notation.

As with any MOP, there are many operating points[1] that we can choose from. Therefore, we need a way to assess the desirability of each operating point. The common approach is choose an *aggregate system utility* function $f(\boldsymbol{g})$ which takes an operating point $\boldsymbol{g} = (g_1, \cdots, g_K)$ as an input and outputs a scalar value where g_i is the value of objective i. Candidates of the aggregate system utility function are

- *Weighted arithmetic mean* $f(\boldsymbol{g}) = \sum_k w_k g_k$
- *Weighted geometric mean* $f(\boldsymbol{g}) = \prod_k g_k^{w_k}$
- *Weighted max-min fairness* $f(\boldsymbol{g}) = \min_k {}^{g_k}/_{w_k}$

The weights w_k can be used to prioritize certain objectives over others. We adopt the ergodic spectral efficiency $g_k(\mathrm{SINR}_k) = \log_2(1 + \mathbb{E}\{\mathrm{SINR}_k\})$ as a performance function and the sum spectral efficiency as the aggregate system utility function which is equivalent to the weighted arithmetic mean with all the weights set to one. This results in the following single objective optimization problem

$$\max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K} \sum_{k=1}^{K} \log_2 \left(\left(1 + \mathbb{E}\left\{\mathrm{SINR}_k^{\mathrm{ul}}\right\}\right) \left(1 + \mathbb{E}\left\{\mathrm{SINR}_k^{\mathrm{dl}}\right\}\right) \right) \tag{4.5}$$

4.3 Mathematical Expressions of the SINR

The SINR expressions defined in chapter 3 are an implicit function of the clustering assignment vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$. To make the analysis and classification of the optimization problem simpler, we modify the SINR expressions such that they are an explicit function of the clustering assignment vectors. We define the following sets of matrices

$$\left\{\mathbf{T}_k^{\mathrm{ul}} \in \mathbb{C}^{L\times L} | k = 1, \cdots, K\right\} \tag{4.6}$$

$$\left\{\mathbf{T}_k^{\mathrm{dl}} \in \mathbb{C}^{L\times L} | k = 1, \cdots, K\right\} \tag{4.7}$$

$$\left\{\mathbf{Q}_{ki}^{\mathrm{ul}} \in \mathbb{C}^{L\times L} | k, i = 1, \cdots, K\right\} \tag{4.8}$$

$$\left\{\mathbf{Q}_{ki}^{\mathrm{dl}} \in \mathbb{C}^{L\times L} | k, i = 1, \cdots, K\right\} \tag{4.9}$$

$$\left\{\mathbf{Q}_{ki}^{\mathrm{ul}} \in \mathbb{C}^{L\times L} | k, i = 1, \cdots, K\right\} \tag{4.10}$$

$$\left\{\boldsymbol{c}_k \in \mathbb{C}^{L} | k = 1, \cdots, K\right\} \tag{4.11}$$

We refer to $\mathbf{T}_k^{\mathrm{ul}}$ as the *uplink signal correlation matrix* of UE k, $\mathbf{T}_k^{\mathrm{dl}}$ as the *downlink signal correlation matrix* of UE k, $\mathbf{Q}_{ki}^{\mathrm{ul}}$ as the *uplink interference correlation matrix* between UE k and UE i, $\mathbf{Q}_{ki}^{\mathrm{dl}}$ as the *downlink interference correlation matrix* between UE k and UE i and $\boldsymbol{c}_k$ as the *receive combining correlation vector* of UE k. The entries of the matrices

[1] An operating point is any solution of the MOP which assigns a value for each of the design variables, hence, assigning a value to each of the objectives.

are defined as follows.

$$\left[\mathbf{T}_k^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{v}_{kn}^H\boldsymbol{h}_{kn}\}^*\mathbb{E}\{\boldsymbol{v}_{km}^H\boldsymbol{h}_{km}\} + \mathbb{E}\{\boldsymbol{v}_{kn}^H\boldsymbol{h}_{kn}\}\mathbb{E}\{\boldsymbol{v}_{km}^H\boldsymbol{h}_{km}\}^*\right) \tag{4.12}$$

$$\left[\mathbf{T}_k^{\mathrm{dl}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{h}_{kn}^H\boldsymbol{w}_{kn}\}^*\mathbb{E}\{\boldsymbol{h}_{km}^H\boldsymbol{w}_{km}\} + \mathbb{E}\{\boldsymbol{h}_{kn}^H\boldsymbol{w}_{kn}\}\mathbb{E}\{\boldsymbol{h}_{km}^H\boldsymbol{w}_{km}\}^*\right) \tag{4.13}$$

$$\left[\mathbf{Q}_{ki}^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{v}_{kn}^H\boldsymbol{h}_{in})^*(\boldsymbol{v}_{km}^H\boldsymbol{h}_{im}) + (\boldsymbol{v}_{kn}^H\boldsymbol{h}_{in})(\boldsymbol{v}_{km}^H\boldsymbol{h}_{im})^*\}\right) \tag{4.14}$$

$$\left[\mathbf{Q}_{ki}^{\mathrm{dl}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{h}_{kn}^H\boldsymbol{w}_{in})^*(\boldsymbol{h}_{km}^H\boldsymbol{w}_{im}) + (\boldsymbol{h}_{kn}^H\boldsymbol{w}_{in})(\boldsymbol{h}_{km}^H\boldsymbol{w}_{im})^*\}\right) \tag{4.15}$$

$$[\boldsymbol{c}_k]_n = \mathbb{E}\{\boldsymbol{v}_{kn}^H\boldsymbol{v}_{kn}\} \tag{4.16}$$

Using the definitions above, the single objective user-centric clustering optimization problem can be written as follows.

$$\max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K} \sum_{k=1}^{K} \log_2\left(\left(1+\mathbb{E}\left\{\mathrm{SINR}_k^{\mathrm{ul}}\right\}\right)\left(1+\mathbb{E}\left\{\mathrm{SINR}_k^{\mathrm{dl}}\right\}\right)\right) \tag{4.17}$$

with

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k\boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{ul}}\boldsymbol{x}_k}{\boldsymbol{x}_k^T(\sum_{i=1}^K p_i\mathbf{Q}_{ki}^{\mathrm{ul}} - p_k\mathbf{T}_k^{\mathrm{ul}})\boldsymbol{x}_k + \sigma_{\mathrm{ul}}^2\boldsymbol{x}_k^T\boldsymbol{c}_k} \tag{4.18}$$

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k}{\sum_{i=1}^K \boldsymbol{x}_i^T\mathbf{Q}_{ki}^{\mathrm{dl}}\boldsymbol{x}_i - \boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k + \sigma_{\mathrm{dl}}^2} \tag{4.19}$$

The proof of the modified SINR expressions in equations (4.18) and (4.19) is available in appendix A. The problem is classified as a binary integer non-linear program which belongs to the class of *NP*-complete problems. It is usually hard to compute closed forms of the SINR expected values. However, we can generate random realizations of the SINR and use them to estimate their expected values.

$$\mathbb{E}\left\{\mathrm{SINR}_k^{ul}\right\} \approx \frac{1}{N_r}\sum_{n=1}^{N_r}\mathrm{SINR}_k^{ul} \tag{4.20}$$

$$\mathbb{E}\left\{\mathrm{SINR}_k^{dl}\right\} \approx \frac{1}{N_r}\sum_{n=1}^{N_r}\mathrm{SINR}_k^{dl} \tag{4.21}$$

where N_r is the number of generated realizations. The technique is known as *sample average approximation*. The procedure by which random realizations of the SINR are generated will be discussed in chapter 7,

4.4 Expected Value of SINR

While expected values of the SINR can be estimated through simulation, closed-form expressions for these expectations are preferable. Closed forms facilitate the analysis of the optimization problem and reduce optimization costs by eliminating the need for sampling. We recall the centralized uplink and downlink SINR expressions defined in (3.17) and (3.29), which we reproduce here for convenience. The downlink SINR is written as

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\left|\mathbb{E}\left\{\boldsymbol{h}_k^H\mathbf{D}_k\boldsymbol{w}_k\right\}\right|^2}{\sum_{i=1}^K \mathbb{E}\left\{\left|\boldsymbol{h}_k^H\mathbf{D}_i\boldsymbol{w}_i\right|^2\right\} - \left|\mathbb{E}\left\{\boldsymbol{h}_k^H\mathbf{D}_k\boldsymbol{w}_k\right\}\right|^2 + \sigma_{\mathrm{dl}}^2} \tag{4.22}$$

which is a function of the clustering matrices $\{\mathbf{D}_k : k = 1, \cdots, K\}$, channel realizations $\{\boldsymbol{h}_k : k = 1, \cdots, K\}$, power allocation coefficients $\{\rho_k : k = 1, \cdots, K\}$ and transmit precoding vectors $\{\boldsymbol{w}_k : k = 1, \cdots, K\}$. While the uplink SINR is written as

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k \left| \mathbb{E}\left\{ \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k \right\} \right|^2}{\sum_{i=1}^K p_i \mathbb{E}\left\{ \left| \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_i \right|^2 \right\} - p_k \left| \mathbb{E}\left\{ \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k \right\} \right|^2 + \sigma_{\mathrm{ul}}^2 \mathbb{E}\left\{ \|\mathbf{D}_k \boldsymbol{v}_k\|^2 \right\}} \tag{4.23}$$

which is a function of the clustering matrices $\{\mathbf{D}_k : k = 1, \cdots, K\}$, channel realizations $\{\boldsymbol{h}_k : k = 1, \cdots, K\}$, power control coefficients $\{p_k : k = 1, \cdots, K\}$ and receive combining vectors $\{\boldsymbol{v}_k : k = 1, \cdots, K\}$. The clustering matrices, power control, and power allocation coefficients remain fixed throughout the transmission and can be considered as deterministic constants. The randomness arises from the randomness in channel realizations; hence, the randomness in channel estimates. These estimates are used to design both the receive combining vectors and the transmit precoding vectors. Thus, to compute a closed-form expression of the uplink and downlink SINR, we must specify the receive combining and transmit precoding design methods, respectively.

The simplest scalable form of centralized combining and precoding is the MR. We derive a closed form of the uplink SINR in case of MR combining in section 4.4.1, and the downlink SINR in case of MR precoding in section 4.4.2

4.4.1 Closed Form of the Uplink SINR for MR Combining

Recall that the uplink SINR defined in (4.18) can be written as

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k \boldsymbol{x}_k^T \mathbf{T}_k^{\mathrm{ul}} \boldsymbol{x}_k}{\boldsymbol{x}_k^T (\sum_{i=1}^K p_i \mathbf{Q}_{ki}^{\mathrm{ul}} - p_k \mathbf{T}_k^{\mathrm{ul}}) \boldsymbol{x}_k + \sigma_{\mathrm{ul}}^2 \boldsymbol{x}_k^T \boldsymbol{c}_k}$$

The randomness in this expression comes from randomness in $\mathbf{T}_k^{\mathrm{ul}}$, $\mathbf{Q}_{ki}^{\mathrm{ul}}$ and $\boldsymbol{c}_k$. The rest of the variables are deterministic. Our approach to derive closed forms for the SINR is to compute closed forms of the random matrices, which leads to a closed form of the SINR. Throughout the analysis, we assume perfect channel estimation $\hat{\boldsymbol{h}}_k = \boldsymbol{h}_k$ for $k = 1, \cdots, K$. Hence. the centralized MR combining vector can be written as

$$\boldsymbol{v}_k^{\mathrm{MR}} = \boldsymbol{h}_k \tag{4.24}$$

The portion of the centralized combining vector $\boldsymbol{v}_{kl}^{\mathrm{MRC}}$ assigned to AP l is

$$\boldsymbol{v}_{kl}^{\mathrm{MR}} = \boldsymbol{h}_{kl} \tag{4.25}$$

Closed form of the uplink signal correlation matrix $\mathbf{T}_k^{\mathrm{ul}}$

As defined in (4.12), an entry in the uplink signal correlation matrix $\mathbf{T}_k^{\mathrm{ul}}$ is

$$\left[\mathbf{T}_k^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{v}_{kn}^H \boldsymbol{h}_{kn}\}^* \mathbb{E}\{\boldsymbol{v}_{km}^H \boldsymbol{h}_{km}\} + \mathbb{E}\{\boldsymbol{v}_{kn}^H \boldsymbol{h}_{kn}\} \mathbb{E}\{\boldsymbol{v}_{km}^H \boldsymbol{h}_{km}\}^* \right) \tag{4.26}$$

By substituting equation (4.25) into (4.26), we get

$$\left[\mathbf{T}_k^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{h}_{kn}^H\boldsymbol{h}_{kn}\}^*\mathbb{E}\{\boldsymbol{h}_{km}^H\boldsymbol{h}_{km}\} + \mathbb{E}\{\boldsymbol{h}_{kn}^H\boldsymbol{h}_{kn}\}\mathbb{E}\{\boldsymbol{h}_{km}^H\boldsymbol{h}_{km}\}^*\right) \tag{4.27}$$

$$= \begin{Bmatrix} \mathrm{tr}\left(\mathbf{R}_{kn}\right)^2 & n = m \\ \mathrm{tr}\left(\mathbf{R}_{kn}\right)\mathrm{tr}\left(\mathbf{R}_{km}\right) & n \neq m \end{Bmatrix} \tag{4.28}$$

where $\mathbf{R}_{kn}$ is the spatial channel correlation matrix of AP n and UE k which is assumed to be known and tr is the trace operator.

Closed form of the uplink interference correlation matrix $\mathbf{Q}_{ki}^{\mathrm{ul}}$

As defined in (4.14), an entry in the uplink interference correlation matrix $\mathbf{Q}_{ki}^{\mathrm{ul}}$ is

$$\left[\mathbf{Q}_{ki}^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{v}_{kn}^H\boldsymbol{h}_{in})^*(\boldsymbol{v}_{km}^H\boldsymbol{h}_{im}) + (\boldsymbol{v}_{kn}^H\boldsymbol{h}_{in})(\boldsymbol{v}_{km}^H\boldsymbol{h}_{im})^*\}\right) \tag{4.29}$$

By substituting equation (4.25) into (4.29), we get

$$\left[\mathbf{Q}_{ki}^{\mathrm{ul}}\right]_{nm} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{h}_{kn}^H\boldsymbol{h}_{in})^*(\boldsymbol{h}_{km}^H\boldsymbol{h}_{im}) + (\boldsymbol{h}_{kn}^H\boldsymbol{h}_{in})(\boldsymbol{h}_{km}^H\boldsymbol{h}_{im})^*\}\right) \tag{4.30}$$

$$= \begin{Bmatrix} \mathrm{tr}\left(\left(\mathbf{I}_{N^2} + \mathbf{K}_{(N,N)}\right)\mathbf{R}_{kn}\otimes\mathbf{R}_{kn}\right) & k = i,\ n = m \\ \mathrm{tr}\left(\mathbf{R}_{kn}\right)\mathrm{tr}\left(\mathbf{R}_{km}\right) & k = i,\ n \neq m \\ \mathrm{tr}\left(\mathbf{R}_{kn}\mathbf{R}_{in}\right) & k \neq i,\ n = m \\ 0 & k \neq i,\ n \neq m \end{Bmatrix} \tag{4.31}$$

where $\otimes$ the Kronecker product and $\mathbf{K}_{(N,N)}$ is the N^2-by-N^2 commutation matrix.

Closed form of the receive combining correlation vector $\boldsymbol{c}_k$

As defined in (4.16), an entry in the receive combining correlation vector $\boldsymbol{c}_k$ is

$$[\boldsymbol{c}_k]_n = \mathbb{E}\{\boldsymbol{v}_{kn}^H\boldsymbol{v}_{kn}\} \tag{4.32}$$

By substituting equation (4.25) into (4.32), we get

$$[\boldsymbol{c}_k]_n = \mathbb{E}\{\boldsymbol{h}_{kn}^H\boldsymbol{h}_{kn}\} = \mathrm{tr}\left(\mathbf{R}_{kn}\right) \tag{4.33}$$

The closed form expressions for $\mathbf{T}_k^{\mathrm{ul}}$ in (4.28), $\mathbf{Q}_{ki}^{\mathrm{ul}}$ in (4.31) and $\boldsymbol{c}_k$ in (4.33) can be used to directly compute the uplink SINR for the case of MR combining.

4.4.2 Closed Form of the Downlink SINR for MR Precoding

Recall that the downlink SINR defined in (4.19) can be written as

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k}{\sum_{i=1}^K \boldsymbol{x}_i^T\mathbf{Q}_{ki}^{\mathrm{dl}}\boldsymbol{x}_i - \boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k + \sigma_{\mathrm{dl}}^2} \tag{4.34}$$

We utilize the same uplink approach to derive closed forms for the downlink SINR by computing closed forms of the random matrices $\mathbf{T}_k^{\mathrm{dl}}$ and $\mathbf{Q}_{ki}^{\mathrm{dl}}$. Throughout the analysis, we assume perfect channel estimation $\hat{\boldsymbol{h}}_k = \boldsymbol{h}_k$ for $k = 1, \cdots, K$. Hence.

the centralized MR precoding vector can be written as

$$\boldsymbol{w}_k^{\mathrm{MR}} = \sqrt{\rho_k}\frac{\boldsymbol{h_k}}{\sqrt{\mathrm{tr}\left(\mathbf{R}_k\right)}} \tag{4.35}$$

The portion of the centralized combining vector $\boldsymbol{w}_{kl}^{\mathrm{MR}}$ assigned to AP l is

$$\boldsymbol{w}_{kl}^{\mathrm{MR}} = \sqrt{\rho_k}\frac{\boldsymbol{h}_{kl}}{\sqrt{\mathrm{tr}\left(\mathbf{R}_k\right)}} \tag{4.36}$$

Using the same approach for the uplink, the closed form of $\mathbf{T}_k^{\mathrm{dl}}$ and $\mathbf{Q}_{ki}^{\mathrm{dl}}$ can be written as follows.

$$\left[\mathbf{T}_k^{\mathrm{dl}}\right]_{nm} = \begin{Bmatrix} \frac{\rho_k}{\mathrm{tr}(\mathbf{R}_k)}\mathrm{tr}\left(\mathbf{R}_{kn}\right)^2 & n = m \\ \frac{\rho_k}{\mathrm{tr}(\mathbf{R}_k)}\mathrm{tr}\left(\mathbf{R}_{kn}\right)\mathrm{tr}\left(\mathbf{R}_{km}\right) & n \neq m \end{Bmatrix} \tag{4.37}$$

$$\left[\mathbf{Q}_{ki}^{\mathrm{dl}}\right]_{nm} = \begin{Bmatrix} \frac{\rho_k}{\mathrm{tr}(\mathbf{R}_k)}\mathrm{tr}\left(\left(\mathbf{I}_{N^2} + \mathbf{K}_{(N,N)}\right)\mathbf{R}_{kn} \otimes \mathbf{R}_{kn}\right) & k = i,\ n = m \\ \frac{\rho_k}{\mathrm{tr}(\mathbf{R}_k)}\mathrm{tr}\left(\mathbf{R}_{kn}\right)\mathrm{tr}\left(\mathbf{R}_{km}\right) & k = i,\ n \neq m \\ \frac{\sqrt{\rho_k\rho_i}}{\mathrm{tr}(\mathbf{R}_k)}\mathrm{tr}\left(\mathbf{R}_{kn}\mathbf{R}_{in}\right) & k \neq i,\ n = m \\ 0 & k \neq i,\ n \neq m \end{Bmatrix} \tag{4.38}$$

The closed form expressions for $\mathbf{T}_k^{\mathrm{dl}}$ in (4.37) and $\mathbf{Q}_{ki}^{\mathrm{dl}}$ in (4.38) can be used to directly compute the downlink SINR for the case of MR precoding.

4.4.3 $\alpha - \mu$ SINR expressions

Receive combining and transmit precoding methods, beyond maximum ratio (MR), often have complex mathematical expressions. Consequently, deriving closed-form expressions of the SINR for such methods is analytically intractable. To address this issue, we propose an approximate analytical approach that leverages the closed-form SINR expressions derived for MR combining and precoding.

The the design rationale of MR is to maximize the energy of the signal which lies in the numerator of the SINR expression. In the uplink, we use the SINR expression defined in equation (3.17) which is written as

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k\left|\mathbb{E}\left\{\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_k\right\}\right|^2}{\sum_{i=1}^K p_i\mathbb{E}\left\{\left|\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_i\right|^2\right\} - p_k\left|\mathbb{E}\left\{\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_k\right\}\right|^2 + \sigma_{\mathrm{ul}}^2\mathbb{E}\left\{\|\mathbf{D}_k\boldsymbol{v}_k\|^2\right\}} \tag{4.39}$$

Formally, the MR combining vector $\boldsymbol{h}_k$ (if we assume perfect channel estimation) maximizes the signal term $\left|\mathbb{E}\{\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_k\}\right|^2$. However, the drawback of this approach is its poor interference cancellation ability. To achieve higher SINR, better combining methods balance signal amplification and interference cancellation. Therefore, we can approximate the SINR of any other *better performing* combining method as

$$\mathrm{SINR}_k^{\mathrm{ul}} \approx \frac{\alpha_k^{\mathrm{ul}} b_k^{\mathrm{MR}}}{\mu_k^{\mathrm{ul}} f_k^{\mathrm{MR}}} \tag{4.40}$$

with

$$\boldsymbol{v}_k = \boldsymbol{v}_k^{\mathrm{MR}} \tag{4.41}$$

$$b_k^{\mathrm{MR}} = p_k \left| \mathbb{E}\left\{ \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k \right\} \right|^2 \tag{4.42}$$

$$f_k^{\mathrm{MR}} = \sum_{i=1}^{K} p_i \mathbb{E}\left\{ \left| \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_i \right|^2 \right\} - p_k \left| \mathbb{E}\left\{ \boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k \right\} \right|^2 + \sigma_{\mathrm{ul}}^2 \mathbb{E}\left\{ \| \mathbf{D}_k \boldsymbol{v}_k \|^2 \right\} \tag{4.43}$$

where $\alpha_k^{\mathrm{ul}} \in [0,1]$ controls how much of the maximum signal power is kept, and $\mu_k^{\mathrm{ul}} \in [0,1]$ controls how much of the maximum interference power is kept. The values of α_k and μ_k vary from combining method to another and from one UE to another. The case of $\alpha_k = 1$ and $\mu_k = 1$ corresponds to MR combining.

Using the same rationale, the downlink SINR can be approximated as

$$\mathrm{SINR}_k^{\mathrm{dl}} \approx \frac{\alpha_k^{\mathrm{dl}} \bar{b}_k^{\mathrm{MR}}}{\mu_k^{\mathrm{dl}} \bar{f}_k^{\mathrm{MR}} + \sigma_{dl}^2} \tag{4.44}$$

with

$$\boldsymbol{w}_k = \boldsymbol{w}_k^{\mathrm{MR}} \tag{4.45}$$

$$\bar{b}_k^{\mathrm{MR}} = \left| \mathbb{E}\left\{ \boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k \right\} \right|^2 \tag{4.46}$$

$$\bar{f}_k^{\mathrm{MR}} = \sum_{i=1}^{K} \mathbb{E}\left\{ \left| \boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i \right|^2 \right\} - \left| \mathbb{E}\left\{ \boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k \right\} \right|^2 \tag{4.47}$$

where $\alpha_k^{\mathrm{dl}}, \mu_k^{\mathrm{dl}} \in [0,1]$.

We refer to the expressions in equations (4.40) and (4.44) as the $\alpha - \mu$ SINR expressions. It remains an open problem how to choose the coefficients $\left(\alpha_k^{\mathrm{ul}}, \mu_k^{\mathrm{ul}}, \alpha_k^{\mathrm{dl}}, \mu_k^{\mathrm{dl}}\right)$ such that the uplink and dowlink SINR are approximated with a high degree of accuracy. Intuitively, they depend on the combining/precoding method and the large-scale fading characteristics of the UE. Nevertheless, we run simulations to estimate these coefficients in an effort to demonstrate the tightness of the approximation.

Chapter 5

Clustering Aware Pilot Assignment Problem Formulation

This chapter presents a thorough mathematical formulation for the pilot assignment problem that considers the user-centric clusters as a factor in assessing the desirability of pilot assignments. Section 5.1 explains the relationship between the pilot assignment and user-clustering problems. Section 5.2 specifies the optimization criterion, which is the channel estimation error, and provides a detailed explanation of the formulation of the optimization problem for *clustering aware pilot assignment*.

5.1 Connection between Pilot Assignment and User-Centric Clustering

Recall from section 3.1 that channel estimation is done by transmitting pilot sequences during the uplink. Ideally, the UEs are assigned mutually orthogonal pilot sequences which eliminates pilot contamination. However, there are only τ_p mutually orthogonal pilot sequences and in any practical network the number of users is larger than the number of pilots $\tau_p < K$. Hence, pilot reuse is necessary.

To estimate the channel $\boldsymbol{h}_{kl}$ between of AP l and UE k, the received pilot signal $\boldsymbol{y}_{t_k l}$ is used which can be written as

$$\boldsymbol{y}_{t_k l}^{\text{pilot}} = \underbrace{\sqrt{\eta_k \tau_p}\boldsymbol{h}_{kl}}_{\text{desired channel}} + \underbrace{\sum_{i \in \mathcal{P}_k / \{k\}} \sqrt{\eta_i \tau_p}\boldsymbol{h}_{il}}_{\text{interference}} + \underbrace{\frac{1}{\sqrt{\tau_p}}\mathbf{N}_l \boldsymbol{\phi}_{t_k}^H}_{\text{receiver noise}} \tag{5.1}$$

where $\mathcal{P}_k$ was defined to be the set of pilot-sharing UEs including UE k. From the expression in equation (5.1), there are four ways to improve channel estimation performance:

1. Increase the energy of the desired channel $\eta_k \tau_p \|\boldsymbol{h}_{kl}\|^2$ by boosting the transmit power η_k,
2. Decrease the interference energy by limiting the transmit powers of pilot sharing UEs η_i for $i \in \mathcal{P}_k / \{k\}$,
3. Choose AP l to serve UE k if $\|\boldsymbol{h}_{kl}\|^2$ is large enough,
4. Reduce the number of pilot-sharing UEs.

The first two strategies appear to be conflicting as prioritizing one user by increasing its transmit power while limiting the power of other pilot-sharing users may lead to a considerable decrease in network-wide performance. The third strategy involves the selection of user-centric clusters, which is the problem we aim to solve in this thesis. The fourth strategy pertains to pilot assignment, which is a separate problem. Nevertheless, both pilot and cluster assignment are closely intertwined, as evident from the arguments presented here.

5.2 Problem Formulation

The main purpose of transmitting pilots during the uplink is to utilize the received pilots in channel estimation. Hence, it is sensible to adopt the channel estimation error as the optimization criterion. Recall from section 3.1 that the channel between AP l and UE k can be written as follows

$$\boldsymbol{h}_{kl} = \hat{\boldsymbol{h}}_{kl} + \tilde{\boldsymbol{h}}_{kl} \tag{5.2}$$

where $\hat{\boldsymbol{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl})$ is the channel estimate and $\tilde{\boldsymbol{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{C}_{kl})$ is the channel estimation error with

$$\boldsymbol{\Psi}_{t_k l} = \sum_{i \in \mathcal{P}_k} \eta_i \tau_p \mathbf{R}_{il} + \sigma_{\text{ul}}^2 \mathbf{I}_N \tag{5.3}$$

$$\mathbf{C}_{kl} = \mathbb{E}\{\tilde{\boldsymbol{h}}_{kl} \tilde{\boldsymbol{h}}_{kl}^H\} = \mathbf{R}_{kl} - \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl} \tag{5.4}$$

We define a set of binary assignment vectors $\boldsymbol{a}_k \in \{0,1\}^{\tau_p},\ k = 1, \cdots, K$ to represent the pilot allocation such that

$$a_{kj} = \begin{cases} 1 & \text{pilot } j \text{ is assigned to UE } k \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Using the assignment vectors, we can rewrite equation (5.3) and (5.4) as follows

$$\boldsymbol{\Psi}_{jl} = \sum_{i=1}^{K} \eta_i \tau_p \mathbf{R}_{il} a_{ij} + \sigma_{\text{ul}}^2 \mathbf{I}_N \tag{5.6}$$

$$\mathbf{C}_{kl} = \mathbf{R}_{kl} - \eta_k \tau_p \sum_{j=1}^{\tau_p} \mathbf{R}_{kl} \boldsymbol{\Psi}_{jl}^{-1} \mathbf{R}_{kl} a_{kj} \tag{5.7}$$

$$j \in \{1, \cdots, \tau_p\}$$

We adopt the expected value of the L_2 norm of the channel estimation error as the optimization criterion which is written as follows

$$\begin{aligned} \mathbb{E}\{\tilde{\boldsymbol{h}}_{kl}^H \tilde{\boldsymbol{h}}_{kl}\} &= \text{tr}(\mathbf{C}_{kl}) \\ &= \text{tr}(\mathbf{R}_{kl}) - \text{tr}\left(\eta_k \tau_p \sum_{j=1}^{\tau_p} \mathbf{R}_{kl} \boldsymbol{\Psi}_{jl}^{-1} \mathbf{R}_{kl} a_{kj}\right) \end{aligned} \tag{5.8}$$

where tr is the trace operator. By substituting (5.6) into (5.8)

$$\mathbb{E}\{\tilde{\boldsymbol{h}}_{kl}^H \tilde{\boldsymbol{h}}_{kl}\} = \text{tr}(\mathbf{R}_{kl}) - \eta_k \tau_p \sum_{j=1}^{\tau_p} \text{tr}\left(\mathbf{R}_{kl} \left(\sum_{i=1}^{K} \eta_i \tau_p \mathbf{R}_{il} a_{ij} + \sigma_{\text{ul}}^2 \mathbf{I}_N\right)^{-1} \mathbf{R}_{kl}\right) a_{kj} \tag{5.9}$$

As discussed in Section 5.1, there is a strong connection between the clustering and pilot assignment problems. Consequently, our formulation of the pilot assignment problem considers the user-centric clusters. Hence, we refer to this as the *clustering aware pilot assignment* problem formulation. Assuming that the clusters $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$ are known, we can formulate the clustering aware pilot assignment optimization problem as follows.

$$\min_{\boldsymbol{a_1},\cdots,\boldsymbol{a_k}} \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbb{E}\{\tilde{\boldsymbol{h}}_{kl}^H \tilde{\boldsymbol{h}}_{kl}\} x_{kl} \tag{5.10}$$

$$\mathbb{E}\{\tilde{\boldsymbol{h}}_{kl}^H \tilde{\boldsymbol{h}}_{kl}\} = \text{tr}(\mathbf{R}_{kl}) - \eta_k \tau_p \sum_{j=1}^{\tau_p} \text{tr}\left(\mathbf{R}_{kl} \left(\sum_{i=1}^{K} \eta_i \tau_p \mathbf{R}_{il} a_{ij} + \sigma_{\text{ul}}^2 \mathbf{I}_N \right)^{-1} \mathbf{R}_{kl} \right) a_{kj} \tag{5.11}$$

$$\sum_{i=1}^{\tau_p} a_{ki} = 1 \tag{5.12}$$

$$\forall k \in \{1, \cdots, K\}$$

The objective function in (5.10) is the sum of channel estimation errors between each AP l and UE k, only if AP l serves UE k. If we ignore the user-centric clusters (assuming $x_{kl} = 1$ for all k and l), the objective function becomes the sum of all channel estimation errors between every AP l and UE k, regardless of whether or not AP l serves UE k. However, by taking into account the user-centric clusters, we gain more degrees of freedom and potentially improve overall performance. The constraints in equation (5.12) ensure that each UE is assigned only one pilot. Like the user-centric clustering problem, the pilot assignment problem is a binary integer non-linear program, which is known to be *NP*-complete. However, a closed-form expression of the expected value $\mathbb{E}\{\tilde{\boldsymbol{h}}_{kl}^H \tilde{\boldsymbol{h}}_{kl}\}$, given in equation (5.11), is available, eliminating the need for sampling.

Chapter 6

Solution Methodology

This chapter discusses the methodology for solving the problems of user-centric clustering and pilot assignment. In section 6.1, we present a baseline algorithm for jointly solving the pilot assignment and user-centric clustering problems. In section 6.2, we describe the genetic algorithm that we employ to obtain the optimized solution.

6.1 Baselines

This section outlines the baseline algorithm that will be used for comparison with the optimized solution. Specifically, we utilize the joint pilot assignment and clustering algorithm proposed in [1].

Algorithm 1: Basic pilot assignment and clustering algorithm

Initalization:
  $\boldsymbol{a}_1 = \cdots = \boldsymbol{a}_K = \mathbf{0}_{\tau_p}$ and $\boldsymbol{x}_1 = \cdots = \boldsymbol{x}_K = \mathbf{0}_N$
for $k = 1$ **to** τ_p **do**
  $a_{kk} \leftarrow 1$ ▷ Assign orthogonal pilots to first τ_p UEs
end
for $k = \tau_p + 1$ **to** K **do**
  $l \leftarrow \underset{l \in \{1,\cdots,L\}}{\texttt{argmax}} \ \beta_{kl}$ ▷ Find best AP for UE k
  $\tau \leftarrow \underset{t \in \{1,\cdots,\tau_p\}}{\texttt{argmin}} \ \sum_{\substack{i=1 \\ t_i = t}}^{k-1} \beta_{il}$ ▷ Find the pilot with least interference at AP l
  $a_{k\tau} \leftarrow 1$ ▷ Assign pilot τ to UE k
end
for $l = 1$ **to** L **do**
  for $t = 1$ **to** τ_p **do**
    $i \leftarrow \underset{k \in \{1,\cdots,K\}:\ t_k = t}{\texttt{argmax}} \ \beta_{kl}$ ▷ Find the UE that AP l serves best on pilot t
    $x_{il} \leftarrow 1$ ▷ Assign AP l to serve UE i
  end
end
Output: Pilot Assignments $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_K$ and Clusters $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$

Algorithm 1 employs a greedy strategy for pilot assignment and user-centric clusters design. Initially, it assigns orthogonal pilot sequences to the first τ_p UEs. For the remaining UEs, each UE k is assigned a pilot that causes the least interference at the best AP. The best AP for UE k is the one with the highest average channel gain β_{kl} with UE k. Subsequently, each AP l serves only τ_p UEs to avoid pilot-sharing on the same AP. Each UE k is served by AP l on pilot t, if it has the highest channel gain β_{kl} among all the UEs sharing the same pilot.

6.2 Optimized Solution

Both the user-centric clustering and pilot assignment problems are classified as *non-linear binary integer programs*. There is generally no optimization algorithm that can find the global optimum in a reasonable timeframe. Two potential optimization algorithms are readily available for solving these problems: *surrogate optimization* and *genetic algorithm*. In our experiments, surrogate optimization was incapable of efficiently improving the value of the objective function, therefore, we only focus on the genetic algorithm.

6.2.1 Genetic Algorithm

The genetic algorithm is a *structured random search* optimization method inspired by natural selection which drives biological evolution. The procedure by which the algorithm generally works is summarized in the following steps.

1. Create a random initial population of solutions, which satisfy the problem constraints.
2. Evaluate the *fitness* value of all the individuals in the population. In genetic algorithm terminology, the fitness value is the objective function value.
3. Every time-step, the algorithm uses the current population to create a new population. Any individual in the new population is referred to as a child of the current population. There are three different types of children:
   (a) *Elite Children* are the individuals in the current population with the best *fitness* value.
   (b) *Crossover Children* are created by combining two individuals in the current population.
   (c) *Mutation Children* are created by making a random modification in one of the individuals in the current population.
4. The algorithm stops when a stopping condition is met. There are several options for the stopping condition:
   (a) maximum number of function evaluations,
   (b) fitness value threshold,
   (c) maximum time,
   (d) minimum change in the objective function value for a specific number of generations.

Chapter 7

Results and Discussion

This chapter presents numerical experiments designed to assess the effectiveness of optimized solutions for the user-centric clustering and pilot assignment problems, relative to the baseline algorithm outlined in section 6.1. Section 7.1 introduces the simulation setup utilized in this thesis and delineates the process of simulating the network. Section 7.2 presents the numerical experiments conducted to evaluate the performance of the algorithms.

7.1 Simulation Setup

To evaluate the performance of the optimized solutions, we define a simulation setup which will be fixed throughout the experiments. The total coverage area is 0.5 km $\times$ 0.5 km, the number of APs is $L = 30$, each equipped with $N = 4$ antennas, and the number of users is $K = 12$. Each coherence block extends for $\tau_c = 200$ samples and the length of pilot sequences is $\tau_p = 5$. Both APs and UEs are uniformly distributed throughout the coverage area, and wrap-around[1] topology is used to avoid cell-edge effects. The full parameters of the simulation setup are summarized in table 7.1.

Parameter	Value
Network Area	0.5 km $\times$ 0.5 km
AP distribution	Uniformly distributed
Users distribution	Uniformly distributed
Number of APs	30
Number of users	12
Number of antennas per AP	4
Samples per coherence block	200
Samples per pilot	5
Bandwidth	20 MHz
Receiver noise power	-94 dBm
Maximum uplink transmit power	100 mW
Maximum downllink transmit power	200 mW

TABLE 7.1: Parameters of the simulation setup

[1] Wrap-around topology means the north and south edges are connected, and also the west and east edges are connected such that there is more than one way to draw a straight line between two coordinates in the coverage area. Whenever a distance is computed, the shortest distance is chosen. This is done to ensure that all the AP and UEs are at the center of a large network.

The procedure by which the network is simulated is described in the following steps.

1. Deploy the APs and UEs randomly within the coverage area,
2. Compute the wrap-around distance between the APs and UEs,
3. Calculate the large-scale fading coefficients using the 3GPP urban microcell model defined in section 2.2.2,
4. Compute the spatial correlation matrices $\mathbf{R}_{kl}$ for every AP l and UE k,
5. Compute the power allocation coefficients using the fractional centralized power allocation heuristic defined in 3.3.3,
6. Generate channel realizations and compute the corresponding receive combining and transmit precoding vectors,
7. Use the channel realizations, receive combining and transmit precoding vectors to compute samples of the SINR in both the uplink and downlink,
8. Compute the optimized clusters and pilot assignment by solving the optimization problems using the genetic algorithm,
9. Calculate the desired performance metrics,
10. Repeat the procedure for the desired number of network layouts.

7.2 Numerical Results

This section describes two sets of experiments, the first of which, discussed in Section 7.2.1, aims to evaluate the performance of the optimized clustering compared to baseline clustering. Consequently, the pilot assignments remain fixed for this set of experiments. The second set, elaborated in Section 7.2.2, aims to evaluate the performance of optimized pilot assignments relative to the baseline pilot assignments. Therefore, the clustering remains fixed for this set of experiments.

7.2.1 Evaluation of Clustering Solutions

This experiment assesses the performance of optimized clustering solutions relative to the baseline, which is defined in Section 6.1. Figures 7.1 and 7.2 display the cumulative distribution function (CDF) of uplink and downlink spectral efficiency achieved by UE k for different combining/precoding schemes, respectively. The optimized clusters exhibit superior performance to the baseline clusters in the uplink for all combining schemes and in the downlink for all precoding schemes. To further illustrate this observation, we present the 90% likely spectral efficiency for both the uplink and downlink in Figures 7.3 and 7.4, respectively. The 90% likely SE is computed by drawing a horizontal line CDF $= 0.1$ on Figures 7.1 and 7.2 and computing the SE at the intersection point with each curve. The optimized clusters achieve higher 90% likely SE values. We observe that the uplink experiences a more significant improvement in performance than the downlink. This can be attributed to the different mathematical structures of the SINR, which makes solving the downlink problem more challenging. For instance, choosing all APs to serve a particular UE in the uplink does not impact the performance of other UEs, whereas doing the same in the downlink can significantly degrade the performance of other UEs.

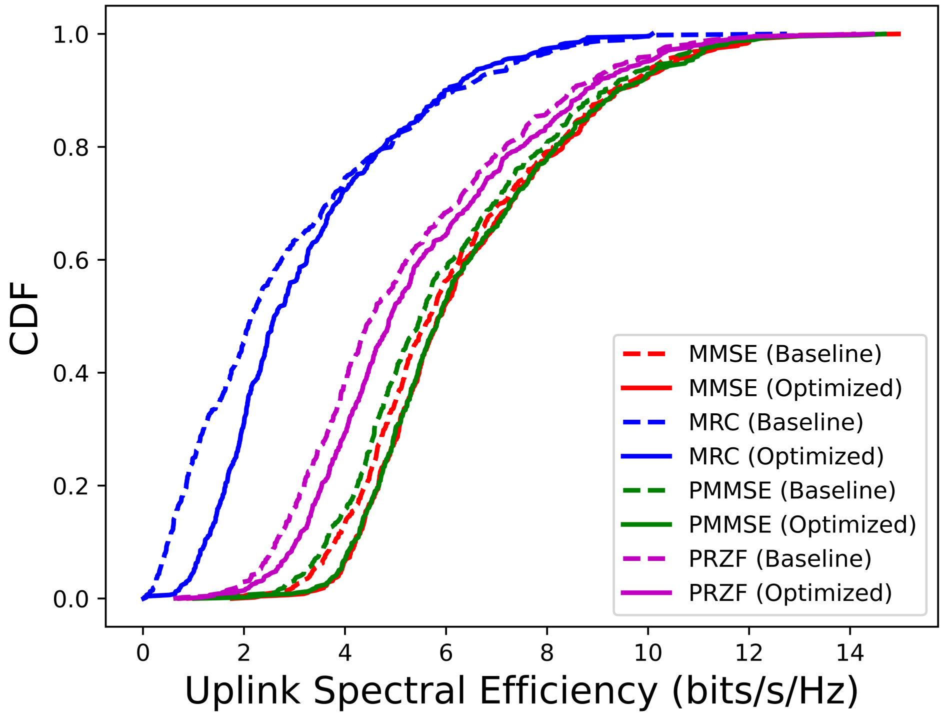

FIGURE 7.1: CDF of the uplink spectral efficiency of UE k

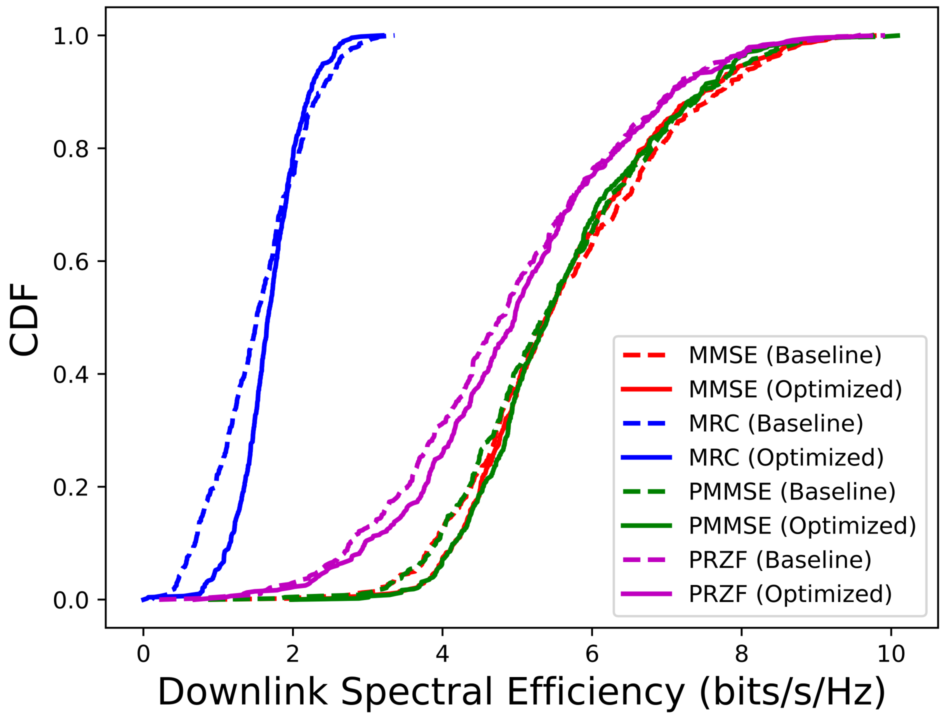

FIGURE 7.2: CDF of the downlink spectral efficiency of UE k

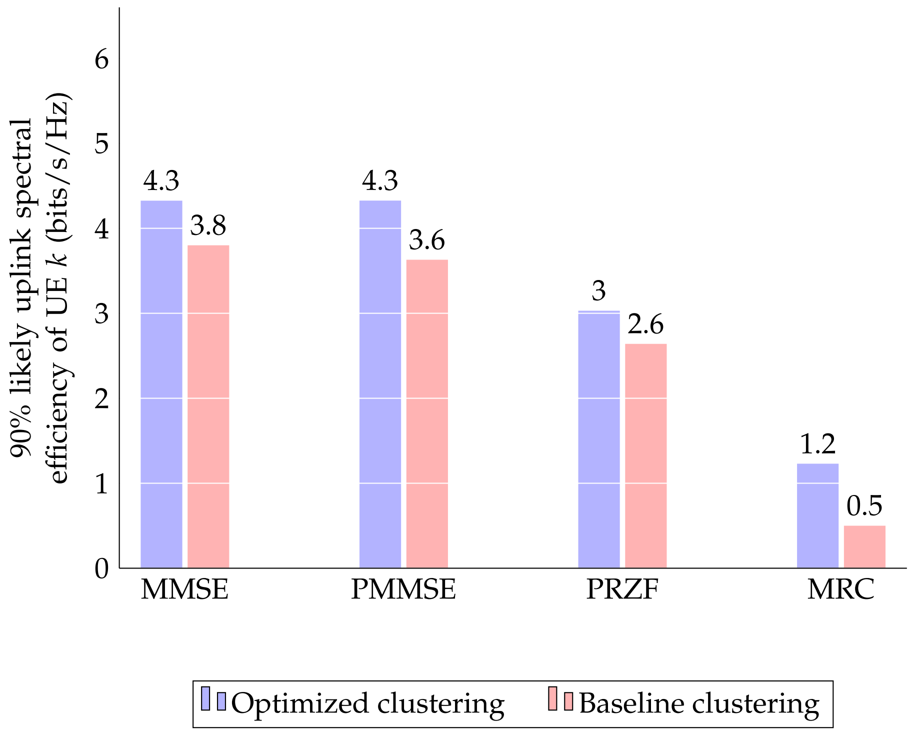

FIGURE 7.3: Comparison of the 90% likely uplink spectral efficiency of UE k achieved by the optimized and baseline clustering solutions for fixed pilot assignment.

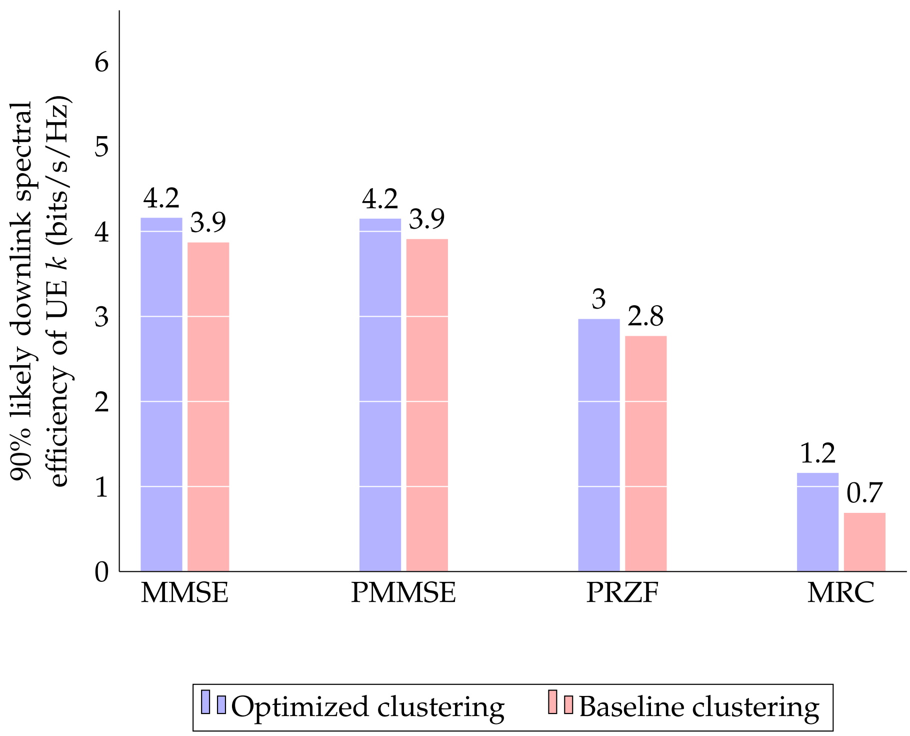

FIGURE 7.4: Comparison of the 90% likely downlink spectral efficiency of UE k achieved by the optimized and baseline clustering solutions for fixed pilot assignment.

7.2.2 Evaluation of Pilot Assignment Solutions

We conduct a performance evaluation of the optimized pilot assignment solution and compare it with the baseline pilot assignment defined in section 6.1. The evaluation is based on the CDF of the uplink and downlink spectral efficiency achieved by a random UE k using different combining/precoding schemes, as depicted in Figures 7.5 and 7.6, respectively. The results reveal that the optimized pilot assignment outperforms the baseline pilot assignment in the uplink for all combining schemes and in the downlink for all precoding schemes except for MR. Furthermore, Figures 7.7 and 7.8 demonstrate the 90% likely uplink and downlink SE, respectively. The optimized pilot assignments show superior performance in both the uplink and downlink. It is worth noting that even small enhancements in spectral efficiency can translate to significant gains in information rate. For instance, a 0.1 increase in spectral efficiency corresponds to a 0.1×20 Mbps increase in the information rate.

7.3 Tightness of $\alpha - \mu$ SINR Expressions

Recall the $\alpha - \mu$ SINR expressions defined in section 4.4.3 as a closed form approximation of the uplink and downlink SINR for any combining or precoding method. We run a simulation to estimate the values of the coefficients $(\alpha_k^{\mathrm{ul}}, \mu_k^{\mathrm{ul}}, \alpha_k^{\mathrm{dl}}, \mu_k^{\mathrm{udl}})$ which takes the following steps

1. Generate a set of channel realizations,
2. Compute centralized combining and precoding vectors,
3. Compute the uplink SINR for each of the combining methods,
4. Compute the downlink SINR for each of the precoding methods,
5. Compare the SINR achieved by each combining/precoding method to the SINR achieved by MR combining/precoding to estimate $(\alpha_k^{\mathrm{ul}}, \mu_k^{\mathrm{ul}}, \alpha_k^{\mathrm{dl}}, \mu_k^{\mathrm{udl}})$.

After the estimation is done, another phase of the simulation starts.

1. Generate a new set of channel realizations,
2. Compute centralized combining and precoding vectors,
3. Compute the uplink SINR for each of the combining methods,
4. Compute the downlink SINR for each of the precoding methods,
5. Compute the approximate uplink SINR for each of the combining methods using the $\alpha - \mu$ expression,
6. Compute the approximate downlink SINR for each of the precoding methods using the $\alpha - \mu$ expression.

The result is illustrated in figures 7.9 and 7.10 which demonstrates the approximation is tight. The spectral efficiency of UE k and its approximation are nearly an exact match. However, there is a need for an efficient way to set the coefficients $(\alpha_k^{\mathrm{ul}}, \mu_k^{\mathrm{ul}}, \alpha_k^{\mathrm{dl}}, \mu_k^{\mathrm{udl}})$.

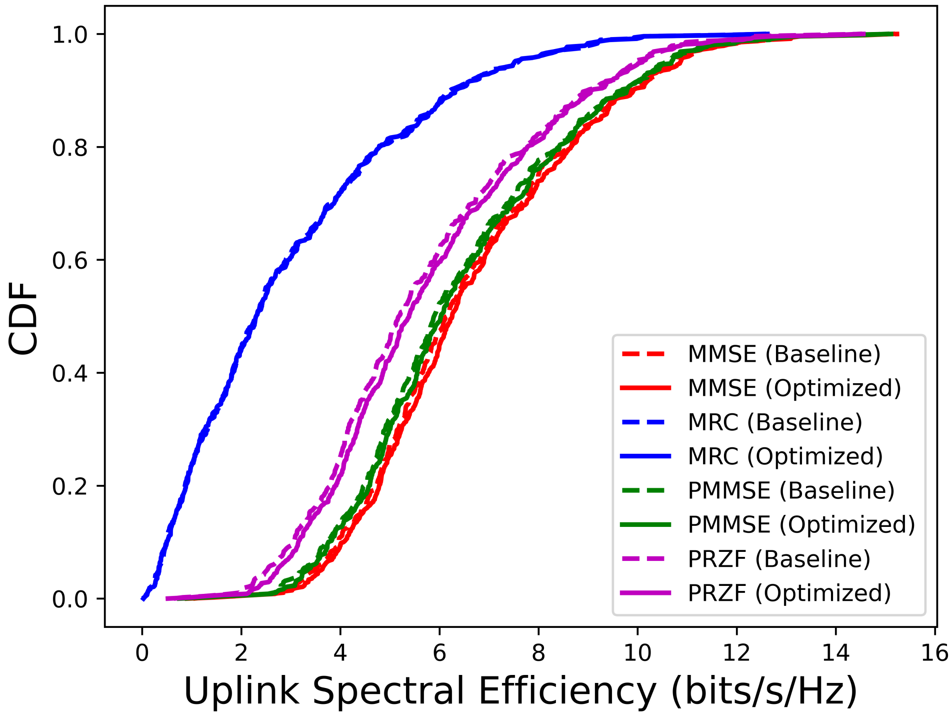

FIGURE 7.5: CDF of the uplink spectral efficiency of UE k

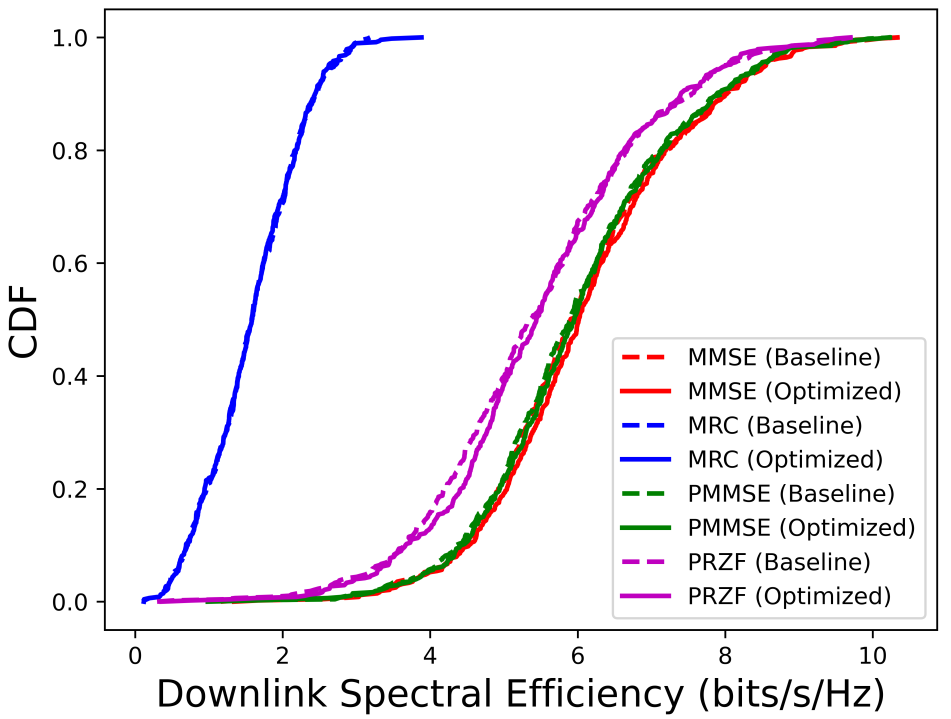

FIGURE 7.6: CDF of the downlink spectral efficiency of UE k

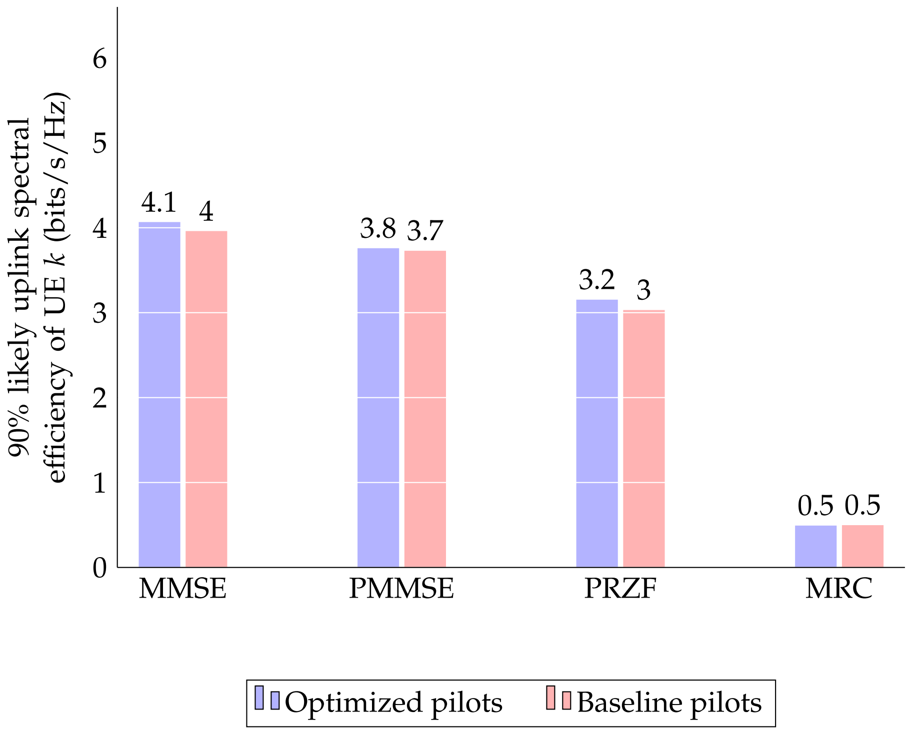

FIGURE 7.7: Comparison of the 90% likely uplink spectral efficiency of UE k achieved by the optimized and baseline pilot assignment solutions for fixed clustering.

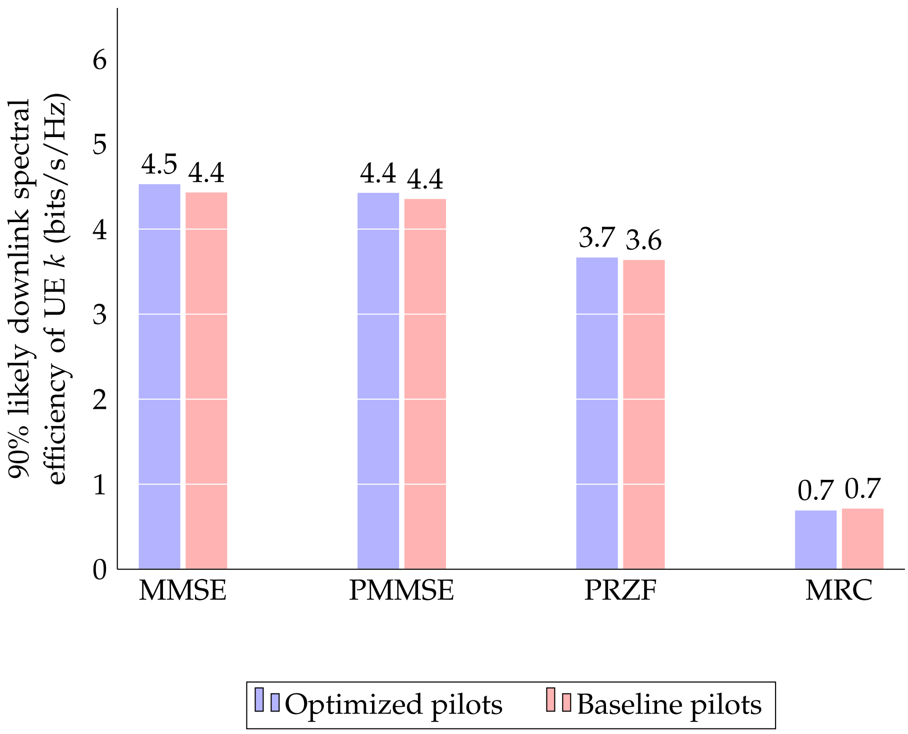

FIGURE 7.8: Comparison of the 90% likely downlink spectral efficiency of UE k achieved by the optimized and baseline pilot assignment solutions for fixed clustering.

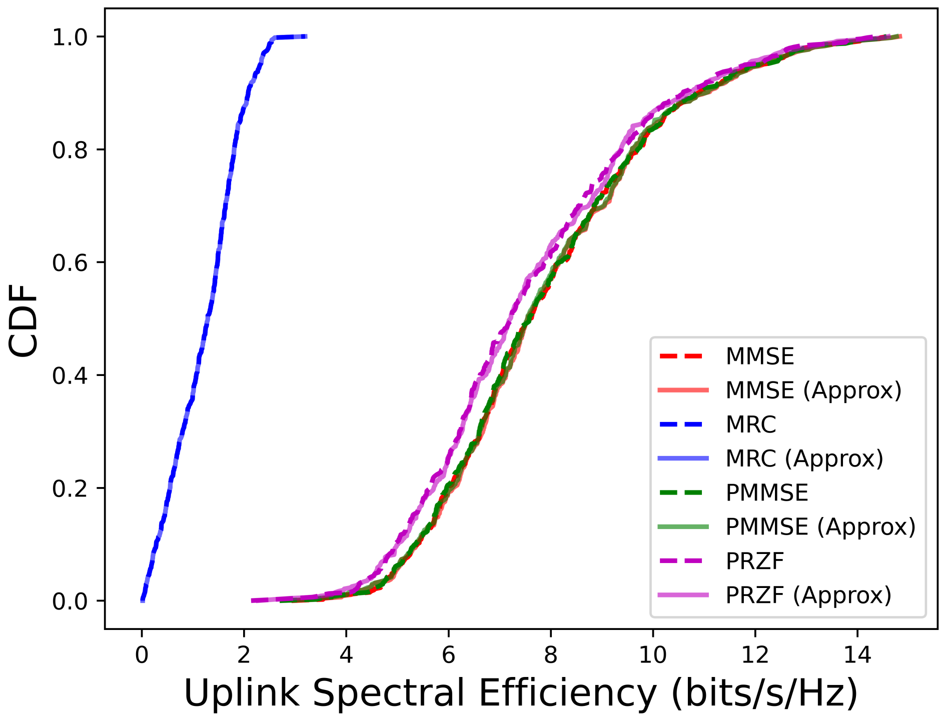

FIGURE 7.9: Tightness of the $\alpha - \mu$ uplink spectral efficiency expressions

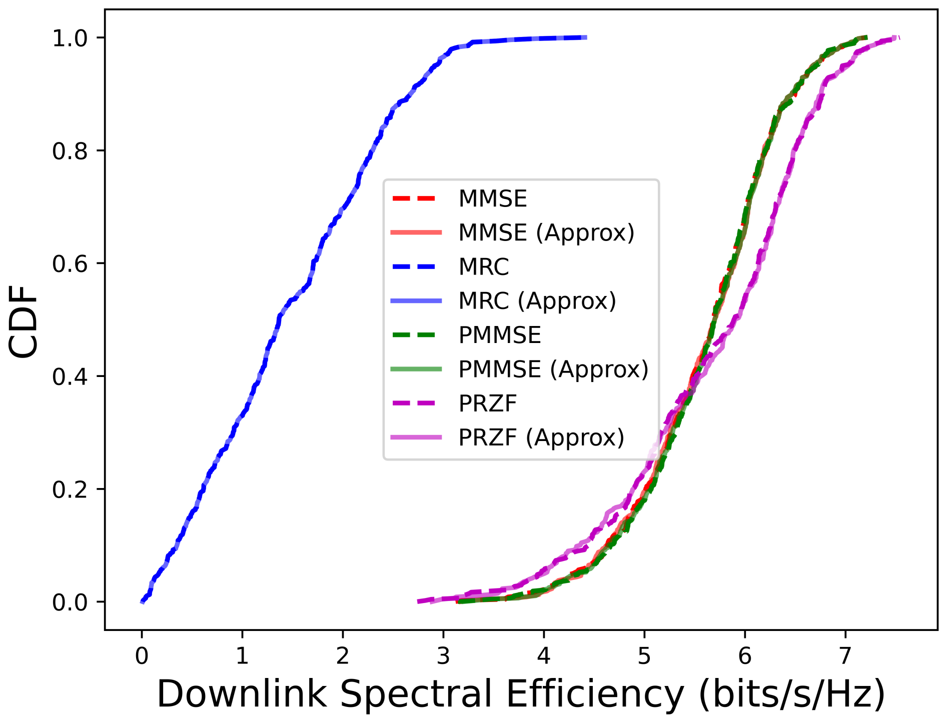

FIGURE 7.10: Tightness of the $\alpha - \mu$ downlink spectral efficiency expressions

Chapter 8

Conclusion and Future Work

In this thesis, the user-centric clustering and pilot assignment problems in cell-free networks were studied, where solving both problems together was deemed necessary as described in chapter 5. The main motivation for approaching these problems is the lack of benchmarks and general formulations, as well as the subjectively designed objective functions and heuristic algorithms used in most literature. Stochastic non-linear binary integer programs were formulated for both the user-centric clustering in Chapter 4 and the pilot assignment problem in Chapter 5. The pilot assignment formulation was developed to consider user-centric clusters when evaluating the desirability of the pilot assignment, making it more efficient. The user-centric clustering problem was solved by applying sample average approximation and using the genetic algorithm, while the pilot assignment problem was solved directly using the genetic algorithm. However, no algorithm is known to guarantee optimal solutions. Numerical experiments in chapter 7 showed that the optimized solutions outperformed baseline solutions, resulting in reasonable spectral efficiency gains. Additionally, approximate analytical expressions for the uplink and downlink SINR were developed in Section 4.4.3 to eliminate the need for sampling. It was demonstrated that these expressions were sufficiently tight, but there was no clear way of setting their parameters.

There are several research directions which we recommend for future work:

1. Development of more efficient algorithms to solve the optimization problems, which can approach the optimized solutions with a reasonable time complexity. One promising approach could be to employ machine learning models that use the optimized solutions as reference data for training.

2. Exploration of the proposed optimization approach in larger networks, non-uniform distribution of UEs within the coverage area, and using different system utility and performance functions.

3. Development of a heuristic approach to determine the parameters of $\alpha - \mu$ SINR expressions, thereby eliminating the need for sampling in the user-centric clustering problem.

4. The present study assumes an ideal scenario where all UEs synchronize and join the network simultaneously. However, in practical situations, the network can operate while new UEs continue to join. Therefore, addressing the practical aspects of the problem, such as scalability of the clustering algorithm, accommodating new joining UEs, establishing the first connection, and frequency of cluster adaptation, is critical.

Appendix A

Uplink and Downlink SINR Expressions

This appendix aims to simplify the analysis and classification of the user-centric clustering optimization problem by manipulating the uplink and downlink SINR expressions. Specifically, the expressions are modified to become explicit functions of binary assignment vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$, which represent the user-centric clusters.

A.1 Uplink SINR

Recall from chapter 3 that the uplink SINR of UE k can be written as

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2}{\sum_{i=1}^{K} p_i \mathbb{E}\left\{\left|\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_i\right|^2\right\} - p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2 + \sigma_{\mathrm{ul}}^2 \mathbb{E}\left\{\|\mathbf{D}_k \boldsymbol{v}_k\|^2\right\}} \tag{A.1}$$

The signal part $p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2$ can be written as

$$\begin{aligned} p_k \left|\mathbb{E}\left\{\boldsymbol{v}_k^H \mathbf{D}_k \boldsymbol{h}_k\right\}\right|^2 &= \left|\mathbb{E}\left\{\sum_{l=1}^{L} \boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl} x_{kl}\right\}\right|^2 \\ &= \left|\sum_{l=1}^{L} \mathbb{E}\left\{\boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl}\right\} x_{kl}\right|^2 \end{aligned} \tag{A.2}$$

Using the following identity

$$\left|\sum_{i=1}^{N} a_i\right|^2 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i^* a_j + a_i a_j^* \tag{A.3}$$

where $a_i \in \mathbb{C}$, we rewrite the signal part in (A.2) as follows

$$\begin{aligned} &\left|\sum_{l=1}^{L} \mathbb{E}\left\{\boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl}\right\} x_{kl}\right|^2 = \\ &\frac{1}{2} \sum_{l=1}^{L} \sum_{j=1}^{L} \left(\mathbb{E}\left\{\boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl}\right\}^* \mathbb{E}\left\{\boldsymbol{v}_{kj}^H \boldsymbol{h}_{kj}\right\} + \mathbb{E}\left\{\boldsymbol{v}_{kl}^H \boldsymbol{h}_{kl}\right\} \mathbb{E}\left\{\boldsymbol{v}_{kj}^H \boldsymbol{h}_{kj}\right\}^*\right) x_{kl} x_{kj} \end{aligned} \tag{A.4}$$

By using the uplink signal correlation matrix $\mathbf{T}_k^{\text{ul}}$ of UE k defined in (4.12) which we rewrite here for convenience.

$$\left[\mathbf{T}_k^{\text{ul}}\right]_{lj} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{v}_{kl}^H\boldsymbol{h}_{kl}\}^*\mathbb{E}\{\boldsymbol{v}_{kj}^H\boldsymbol{h}_{kj}\} + \mathbb{E}\{\boldsymbol{v}_{kl}^H\boldsymbol{h}_{kl}\}\mathbb{E}\{\boldsymbol{v}_{kj}^H\boldsymbol{h}_{kj}\}^*\right) \tag{A.5}$$

The signal part can be written as the following quadratic form

$$p_k\left|\mathbb{E}\left\{\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_k\right\}\right|^2 = p_k\boldsymbol{x}_k^T\mathbf{T}_k^{\text{ul}}\boldsymbol{x}_k \tag{A.6}$$

The interference term $p_i\mathbb{E}\left\{\left|\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_i\right|^2\right\}$ of UE i can be written as

$$p_i\mathbb{E}\left\{\left|\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_i\right|^2\right\} = p_i\mathbb{E}\left\{\left|\sum_{i=1}^{L}\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il}x_{kl}\right|^2\right\} \tag{A.7}$$

Using the identity defined in (A.3), we can rewrite (A.7) as

$$p_i\mathbb{E}\left\{\left|\sum_{i=1}^{L}\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il}x_{kl}\right|^2\right\} = $$

$$\frac{p_i}{2}\sum_{l=1}^{L}\sum_{j=1}^{L}\mathbb{E}\left\{\left(\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il}\right)^*\left(\boldsymbol{v}_{kj}^H\boldsymbol{h}_{ij}\right) + \left(\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il}\right)\left(\boldsymbol{v}_{kj}^H\boldsymbol{h}_{ij}\right)^*\right\}x_{kl}x_{kj} \tag{A.8}$$

By using the uplink interference correlation matrix $\mathbf{Q}_{ki}^{\text{ul}}$ between UE k and UE i defined in (4.14) which we rewrite here for convenience.

$$\left[\mathbf{Q}_{ki}^{\text{ul}}\right]_{lj} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il})^*(\boldsymbol{v}_{kj}^H\boldsymbol{h}_{ij}) + (\boldsymbol{v}_{kl}^H\boldsymbol{h}_{il})(\boldsymbol{v}_{kj}^H\boldsymbol{h}_{ij})^*\}\right) \tag{A.9}$$

The interference term of UE i can be written as

$$p_i\mathbb{E}\left\{\left|\boldsymbol{v}_k^H\mathbf{D}_k\boldsymbol{h}_i\right|^2\right\} = p_i\boldsymbol{x}_k^T\boldsymbol{Q}_{ki}^{\text{ul}}\boldsymbol{x}_k \tag{A.10}$$

The noise term $\sigma_{\text{ul}}^2\mathbb{E}\left\{\|\mathbf{D}_k\boldsymbol{v}_k\|^2\right\}$ can be written as

$$\sigma_{\text{ul}}^2\mathbb{E}\left\{\|\mathbf{D}_k\boldsymbol{v}_k\|^2\right\} = \sigma_{\text{ul}}^2\mathbb{E}\left\{\left\|\begin{bmatrix}\boldsymbol{v}_{k1}x_{k1}\\ \vdots\\ \boldsymbol{v}_{kL}x_{kL}\end{bmatrix}\right\|^2\right\} \tag{A.11}$$

$$= \sum_{l=1}^{L}\mathbb{E}\left\{\boldsymbol{v}_{kl}^H\boldsymbol{v}_{kl}\right\}x_{kl} \tag{A.12}$$

Using the receive combining correlation vector $\boldsymbol{c}_k$ of UE k defined in (4.16), which we rewrite here for convenience.

$$[\boldsymbol{c}_k]_l = \mathbb{E}\{\boldsymbol{v}_{kl}^H\boldsymbol{v}_{kl}\} \tag{A.13}$$

The noise term can be written as

$$\sigma_{\text{ul}}^2\mathbb{E}\left\{\|\mathbf{D}_k\boldsymbol{v}_k\|^2\right\} = \sigma_{\text{ul}}^2\boldsymbol{x}_k^T\boldsymbol{c}_k \tag{A.14}$$

Finally, using equations (A.6), (A.10) and (A.14), the uplink SINR of UE k can be written as

$$\mathrm{SINR}_k^{\mathrm{ul}} = \frac{p_k \boldsymbol{x}_k^T \mathbf{T}_k^{\mathrm{ul}} \boldsymbol{x}_k}{\boldsymbol{x}_k^T (\sum_{i=1}^K p_i \mathbf{Q}_{ki}^{\mathrm{ul}} - p_k \mathbf{T}_k^{\mathrm{ul}}) \boldsymbol{x}_k + \sigma_{\mathrm{ul}}^2 \boldsymbol{x}_k^T \boldsymbol{c}_k} \tag{A.15}$$

which is an explicit function of the clustering assignment vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$.

A.2 Downlink SINR

Recall from chapter 3 that the downlink SINR of UE k can be written as

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2}{\sum_{i=1}^K \mathbb{E}\left\{\left|\boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i\right|^2\right\} - \left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2 + \sigma_{\mathrm{dl}}^2} \tag{A.16}$$

The signal part $\left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2$ can be written as

$$\begin{aligned} \left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2 &= \left|\mathbb{E}\left\{\sum_{l=1}^L \boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl} x_{kl}\right\}\right|^2 \\ &= \left|\sum_{l=1}^L \mathbb{E}\left\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\right\} x_{kl}\right|^2 \end{aligned} \tag{A.17}$$

We rewrite the signal part in (A.17) as follows

$$\begin{aligned} &\left|\sum_{l=1}^L \mathbb{E}\left\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\right\} x_{kl}\right|^2 = \\ &\quad \frac{1}{2}\sum_{l=1}^L \sum_{j=1}^L \left(\mathbb{E}\left\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\right\}^* \mathbb{E}\left\{\boldsymbol{h}_{kj}^H \boldsymbol{w}_{kj}\right\} + \mathbb{E}\left\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\right\} \mathbb{E}\left\{\boldsymbol{h}_{kj}^H \boldsymbol{w}_{kj}\right\}^*\right) x_{kl} x_{kj} \end{aligned} \tag{A.18}$$

By using the downlink signal correlation matrix $\mathbf{T}_k^{\mathrm{dl}}$ of UE k defined in (4.13) which we rewrite here for convenience.

$$\left[\mathbf{T}_k^{\mathrm{dl}}\right]_{lj} = \frac{1}{2}\left(\mathbb{E}\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\}^* \mathbb{E}\{\boldsymbol{h}_{kj}^H \boldsymbol{w}_{kj}\} + \mathbb{E}\{\boldsymbol{h}_{kl}^H \boldsymbol{w}_{kl}\} \mathbb{E}\{\boldsymbol{h}_{kj}^H \boldsymbol{w}_{kj}\}^*\right) \tag{A.19}$$

The signal part can be written as the quadratic form

$$\left|\mathbb{E}\left\{\boldsymbol{h}_k^H \mathbf{D}_k \boldsymbol{w}_k\right\}\right|^2 = \boldsymbol{x}_k^T \mathbf{T}_k^{\mathrm{dl}} \boldsymbol{x}_k \tag{A.20}$$

The interference term $\mathbb{E}\left\{\left|\boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i\right|^2\right\}$ of UE i can be written as

$$\mathbb{E}\left\{\left|\boldsymbol{h}_k^H \mathbf{D}_i \boldsymbol{w}_i\right|^2\right\} = \mathbb{E}\left\{\left|\sum_{i=1}^L \boldsymbol{h}_{kl}^H \boldsymbol{w}_{il} x_{il}\right|^2\right\} \tag{A.21}$$

Using the identity defined in (A.3), we can rewrite equation (A.21) as

$$\mathbb{E}\left\{\left|\sum_{i=1}^{L} \boldsymbol{h}_{kl}^{H}\boldsymbol{w}_{il}x_{il}\right|^2\right\} = \\ \frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{L}\mathbb{E}\left\{\left(\boldsymbol{h}_{kl}^{H}\boldsymbol{w}_{il}\right)^*\left(\boldsymbol{h}_{kj}^{H}\boldsymbol{w}_{ij}\right)+\left(\boldsymbol{h}_{kl}^{H}\boldsymbol{w}_{il}\right)\left(\boldsymbol{h}_{kj}^{H}\boldsymbol{w}_{ij}\right)^*\right\}x_{il}x_{ij} \quad \text{(A.22)}$$

By using the downlink interference correlation matrix $\mathbf{Q}_{ki}^{\mathrm{dl}}$ between UE k and UE i defined in (4.15) which we rewrite here for convenience.

$$\left[\mathbf{Q}_{ki}^{\mathrm{dl}}\right]_{lj} = \frac{1}{2}\left(\mathbb{E}\{(\boldsymbol{h}_{kl}^{H}\boldsymbol{w}_{il})^*(\boldsymbol{h}_{kj}^{H}\boldsymbol{w}_{ij})+(\boldsymbol{h}_{kl}^{H}\boldsymbol{w}_{il})(\boldsymbol{h}_{kj}^{H}\boldsymbol{w}_{ij})^*\}\right) \quad \text{(A.23)}$$

The interference term of UE i can be written as

$$\mathbb{E}\left\{\left|\boldsymbol{h}_{k}^{H}\mathbf{D}_i\boldsymbol{w}_i\right|^2\right\} = \boldsymbol{x}_i^T\mathbf{Q}_{ki}^{\mathrm{dl}}\boldsymbol{x}_i \quad \text{(A.24)}$$

Finally, using equations (A.20) and (A.24), the downlink SINR of UE k can be written as

$$\mathrm{SINR}_k^{\mathrm{dl}} = \frac{\boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k}{\sum_{i=1}^{K}\boldsymbol{x}_i^T\mathbf{Q}_{ki}^{\mathrm{dl}}\boldsymbol{x}_i-\boldsymbol{x}_k^T\mathbf{T}_k^{\mathrm{dl}}\boldsymbol{x}_k+\sigma_{\mathrm{dl}}^2} \quad \text{(A.25)}$$

which is an explicit function of the clustering assignment vectors $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K$.

Bibliography

[1] O. T. Demir, E. Bjornson, and L. Sanguinetti, *Foundations of user-centric cell-free massive Mimo*. Now Publishers, 2021.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017. DOI: 10.1109/TWC.2017.2655515.

[3] E. Björnson and L. Sanguinetti, "Scalable cell-free massive mimo systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020. DOI: 10.1109/TCOMM.2020.2987311.

[4] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, 2019. DOI: 10.1186/s13638-019-1507-0. [Online]. Available: https://doi.org/10.1186\%2Fs13638-019-1507-0.

[5] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo: Uniformly great service for everyone," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 201–205. DOI: 10.1109/SPAWC.2015.7227028.

[6] E. Björnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020. DOI: 10.1109/TWC.2019.2941478.

[7] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell ofdma transmission," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6086–6101, 2011. DOI: 10.1109/TSP.2011.2165706.

[8] C. F. Mendoza, S. Schwarz, and M. Rupp, "Cluster formation in scalable cell-free massive mimo networks," in *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2020, pp. 62–67. DOI: 10.1109/WiMob50308.2020.9253391.

[9] S. Biswas and P. Vijayakumar, "Ap selection in cell-free massive mimo system using machine learning algorithm," in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 158–161. DOI: 10.1109/WiSPNET51692.2021.9419450.

[10] R. Wang, M. Shen, Y. He, and X. Liu, "Performance of cell-free massive mimo with joint user clustering and access point selection," *IEEE Access*, vol. 9, pp. 40 860–40 870, 2021. DOI: 10.1109/ACCESS.2021.3056051.

[11] Y. Al-Eryani, M. Akrout, and E. Hossain, "Multiple access in cell-free networks: Outage performance, dynamic clustering, and deep reinforcement learning-based design," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1028–1042, 2021. DOI: 10.1109/JSAC.2020.3018825.

[12] C. D'Andrea and E. G. Larsson, *User association in scalable cell-free massive mimo systems*, 2021. arXiv: 2103.05321 [cs.IT].

[13] C. Wei, K. Xu, X. Xia, *et al.*, "User-centric access point selection in cell-free massive mimo systems: A game-theoretic approach," *IEEE Communications Letters*, vol. 26, no. 9, pp. 2225–2229, 2022. DOI: 10.1109/LCOMM.2022.3186350.

[14] X.-T. Dang, T. Lai-Thuc, A.-T. Nguyen, T. Vu-Huy, N. H. Anh Tran, and H.-D. Han, "A genetic algorithm based pilot assignment strategy for cell-free massive mimo system," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, 2021, pp. 93–98. DOI: 10.1109/ICCE48956.2021.9352116.

[15] Y. Hao, J. Xin, W. Tao, S. Tao, L. Yu-xiang, and W. Hao, "Pilot allocation algorithm based on k-means clustering in cell-free massive mimo systems," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020, pp. 608–611. DOI: 10.1109/ICCC51575.2020.9345177.

[16] R. Sabbagh, C. Pan, and J. Wang, "Pilot allocation and sum-rate analysis in cell-free massive mimo systems," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422575.

[17] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive mimo wireless networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1–6. DOI: 10.1109/ICCW.2018.8403508.

[18] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive mimo systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 392–396. DOI: 10.1109/CompComm.2018.8780756.

[19] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive mimo based on the hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, 2021. DOI: 10.1109/LWC.2020.3020003.

[20] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive mimo," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1644–1647, 2015. DOI: 10.1109/LCOMM.2015.2409176.

[21] H. Liu, J. Zhang, X. Zhang, A. Kurniawan, T. Juhana, and B. Ai, "Tabu-search-based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2286–2290, 2020. DOI: 10.1109/TVT.2019.2956217.

[22] 3GPP, *Further advancements for e-utra physical layer aspects (release 9)*, 3GPP TS 36.814., 2017.

[23] E. Björnson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. 2013.