MASSIVE MULTIPLE ACCESS IN FUTURE WIRELESS NETWORKS:

DYNAMIC ARCHITECTURES AND MACHINE LEARNING-BASED DESIGN

By

YASSER AL-ERYANI

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF MANITOBA
Electrical and Computer Engineering

JUNE 2021

Supervisor: Ekram Hossain

Abstract

During the last few decades, wireless communication technologies and services have radically changed the way we live and interact at the personal, social, local and global levels. Such changes were mainly driven by the continuous emergence of innovative wireless communication services and products. These services and products represents a direct upshot of enduring research outcomes within the area. Nevertheless, the blessing of such innovation was accompanied by extremely high demands in forms of data traffic, per-user transmission rate, minimum transmission delay and in the number of wireless devices per unit area. Tackling these issues through cellular network densification was faced by many technical issues related to high interference levels, tedious user scheduling processes, and complicated network resource allocation algorithms. Trying to address these imperative technical issues in future wireless networks, this thesis develops several innovative enabling techniques for massive wireless multiple access. Specifically, we commence this work by introducing a new concept of partial spectrum overlapping among active users equipment (UEs). The proposed scheme represents a trade-off between fully orthogonal multiple access schemes (e.g. time division multiple access [TDMA], frequency division multiple access (FDMA) and orthogonal frequency division multiple access (OFDMA)) and that of non-orthogonal multiple access (NOMA). Second, we develop several innovative dynamic cell-free network architectures that support massive wireless connectivity through adaptive access points (APs)/base stations (BSs) coordination and/or cooperation. The proposed network models are then evaluated under different state-of-the-art enabling wireless techniques such as millimeter wave (mmWave) channel links and massive multiple-input multiple-output (mMIMO) sys-

tems. Furthermore, the performance of the proposed architectures is investigated through the derivation of several closed-form expressions of exact and/or asymptotic performance metrics (example, probability of outage, asymptotic outage, instantaneous rate and outage-capacity). Finally, for practical control and monitoring of the proposed access techniques and network models, we develop several low-complexity deep reinforcement learning (DRL)-based modeling frameworks that can efficiently learn the solution of several combinatorial optimization problems related to network partitioning (clustering) and uplink/downlink beamforming. This is achieved through innovative nested DRL designs that utilizes continuous and discrete deep neural networks (DNN) agents based on the nature of the problem. Several operating scenarios of the proposed techniques are evaluated through extensive Monte-Carlo simulations (Matlab and Python) with practical parameters and assumptions.

Dedication

To the beautiful soul of my mom Huda, who passed away during final stages of writhing this work, to my dad, Fadhl, my all times role model, to Maryam, my love, my life and my light and to my little champion, Radin.

ACKNOWLEDGMENT

Acknowledgement is due to the University of Manitoba for giving me this precious opportunity to resume my PhD degree. I would like to express deep gratefulness and appreciation to my Thesis advisor, Prof. Ekram Hossain for his continuous help, guidance, and encouragement throughout the course of this work. He spent a lot of his precious time helping me and advising me at each step. Additionally, I would like also to thank my Thesis committee members: Prof. Gabriel Thomas, Dr. Noman Mohammed and Prof. Abraham Fapojuwo for their great help and cooperation, which contributed significantly to the improvement of this work. Finally, my heartfelt gratitude goes to my parents, my wife, my sisters, and brothers for their encouragement, prayers, and moral support.

Contents

List of Tables

List of Figures

List of Abbreviations

A/D:	Analog to Digital	**MA:**	Multiple Access
AP:	Access Point	**MAP:**	maximum a posteriori
AI:	Artificial Intelligence	**MIMO:**	Multiple-Input Multiple-Output
AWGN:	Additive Whihte Gaussian Noise	**ML:**	Machine Learning
BBU:	Baseband Unit	**mmWaves:**	millimeter Waves
BS:	Base Station	**MMSE:**	Minimum Mean Square Error
CDF:	Cumulative Distribution Function	**MTC:**	Machine Type Communications
CoMP:	Coordinated Multi-Point	**NCC:**	Network Cloud Controller
CPU:	Central Processing Unit	**NOMA:**	Non-Orthogonal Multiple Access
C-RAN:	Cloud Radio Access Network	**OFDMA:**	Orthogonal Frequency MA
D/A:	Digital to Analog	**OMA:**	Orthogonal Multiple Access
DAS	Distributed Antenna System	**PDF:**	Probability Density Function
DDPG:	Deep Deterministic Policy Gradient	**PG:**	Policy Gradient
DDQN:	Deep Double Q-Learning	**P2P:**	Point-to-Point
DRL:	Deep Reinforcement Learning	**RAN:**	Radio Access Network
eMBB:	Enhanced Mobile Broad Band	**RRH**	Remote Radio Hub
FDMA:	Frequency Division Multiple Access	**rv:**	Random Variable
HetNet:	Heterogeneous Network	**SAC:**	Soft Actor-Critic
eAP:	Enhanced Access Point	**SARSA:**	State-Action-Reward-State-Action
eCP:	Edge Cloud Processor	**SIC:**	Successive Interference Cancellation
ICI:	Inter-Cell Interference	**SINR:**	Signal to Interference Plus Noise Ratio
IUI:	Inter-User Interference	**TDMA:**	Time Division Multiple Access
i.i.d:	Independent and Identically Dist.	**TRD:**	Transmit-Receive Diversity
i.n.d	Independent & Not Identically Dist.	**UE:**	User Equipment
INI:	Inter-NOMA Interference		
INCI:	Inter-NOMA-Cluster Interference		
ISNI:	Inter-Subnetwork Interference	**WSR:**	Weighted Sum Rate

Chapter One

Introduction

The last few decades have witnessed a revolutionary ameliorate on the services and standards of wireless communication networks. This was preceded by continuous improvements on network architectures and various network optimization protocols and algorithms. Evidently, moving from each generation of wireless communication to the next one, an important milestones have been characterized by supporting an exponentially increasing number of users equipment (UEs) and providing significantly higher transmission rates with better reliability. As such evolution continues, the next 6G/beyond 6G (B6G) generation of wireless networks will have to provide an enormous transmission rates with extreme efficiency and reliability. These imperative demands and many others will require radical solutions in the design of network architectures and operating protocols and algorithms. Furthermore, with the recent occurrence of critical wireless services such as autonomous driving and virtual reality, the demands for extremely ultra-reliable low-latency (eURLL) services with high data rates has become a prominent need [1]. This has posed a serious technical issues regarding intolerable high interference levels, high data traffics and computational and/or hardware complexity [2]. **Table 1.1** shows the main promises that the ongoing developed 5G network need to fulfils and how these promises suffices an important aspects. Deficiencies mentioned in **Table 1.1** are expected to be tackled at the beyond 5G (B5G)/6G generations of wireless communications. To cope with these challenges, network designers have been continuously developing

Table 1.1 5G 3GPP New Radio

Promises	**Deficiencies**
Support up to 20 Gpbs at the 20-90 GHz band (2Gbps below 6GHz)	New bands works only on certain hotspots: (stadiums,concerts,trains, etc.)
Support mMTC for industrial applications	Complicated multiuser transmission techniques (mobility, interference mitigation, etc.)
Provide URLL applications ($<$ 1ms latency).	Insufficient for recently emerging technologies (autonomous driving, remote health applications)
Support cloud implementation and edge computing.	Introduces huge amount of control traffic data, especially for massive network.

several enabling technologies that are based on more robust/dynamic allocation of network resources, diversity enhancement and complexity reduction.

For instant, future multiple access (MA) schemes are expected to enable several UEs to access the same frequency bands at the same time and within the same network geographical area. Such a technology has been recently referred to as non-orthogonal MA scheme (NOMA). NOMA scheme has emerged as a promising technology to enhance spectrum efficiency of both uplink and downlink cellular wireless networks [3–5]. With power-domain NOMA, at a certain frequency sub-band, signals of multiple UEs are superimposed in the power domain such that the received signal for each UE has a distinct power level. At the NOMA receiver end, successive interference cancellation (SIC) is used to cancel signal components with higher weights than the desired signal (starting from the signal with the highest weight) [6]. In theory, NOMA can significantly enhances the spectral efficiency compared to that of orthogonal multiple access schemes (OMA), at the expense of receiver complexity and processing delay [7, 8].

Furthermore, significant recent improvements have been achieved within areas related to diversity-enhancing technologies such as massive multiple input multiple output (mMIMO) [9], cell-free network architectures [10], eURLL [11], massive machine type communication (mMTC) [12], meta-surfaces [13] and the recent exploration of new radio bands within the $24 - 300$ GHz ranges [14]. In fully centralized cell-free network, all access points (APs) of

the network are simultaneously serving all active (recognized) UEs using the same frequency bands. This include signal transmission/reception, power control, traffic management, etc. Precoding/decoding of downlink/uplink signals is then implemented to mitigate the inter-user interference (IUI) caused by simultaneous utilization of the same time/frequency resources. Cell-free architecture is found to enhance the network coverage and transmission performance. However, such a performance enhancement comes with the expense of high computational complexity, high network traffic and increased interference levels. The complexity of signal processing however, can be reduced by using scalable, dynamic, and low-complexity designs [15], and also by using suitable network- and UE-centric architectures and algorithms [16], [17,18]. One efficient technique to mitigate high interference levels is by adding more robustness in beamsteering and spatial multiplexing techniques. This can be best implemented through the adoption of millimeter wave (mmWave) band over distributed APs of cell-free network [19]. Interestingly, cell-free network architecture is found to provide an efficient solution for the poor scattering and high path loss nature of mmWave bands. Furthermore, mMIMO systems represents a dire need for mmWave network to provide more directional beams with improved signal strength [20]. This mmWave-MIMO cell-free network architecture requires significant central computational capabilities and significantly higher control signaling among network entities. Such a computational complexity requirement becomes more deteriorating as the number of wireless devices within the network increases.

In addition to the aforementioned enabling technologies, a recent emergence of machine-learning (ML)-based solutions and models have been proposed [21, 22]. Within the context of cell-free network, several learning-based algorithms are being developed for multiuser beamforming and channel estimation [23–26]. For instance, the authors in [23] proposed a practical channel estimation framework based on a denoising convolutional neural network for mmWave cell-free network. In [26], an unsupervised-learning was proposed for power control in cell-free network. Furthermore, a great potentials of deep reinforcement learning (DRL)

in solving combinatorial problems that include both discrete and continuous optimization variables has been found on [15, 27].

1.1 Literature Review

The concept of simultaneous transmission of data at the same frequency/time resources date back to the development of code division multiple access schemes (CDMA) [28]. In CDMA however, discrepancy among different UEs was achieved by assigning an orthogonal sequences of data for each UE. When no orthogonal sequences are assigned to UEs that access the same time/frequency resources, UEs are referred to as power domain NOMA UEs [29]. Furthermore, The 'marriage' between CoMP and NOMA has introduced an attractive solution to compensate for the excessive spectrum usage in CoMP [30] and increases the number of UEs per NOMA clusters [31]. It also makes it possible to generalize the concept of CoMP of cooperation among APs to serve all UEs within the network instead of only cell-edge UEs. Several researches on the literature focused on utilizing NOMA scheme under a variety of APs cooperation layouts [32–36]. In [32], authors proposed a network NOMA for uplink CoMP that enhances the system spectral efficiency by forming NOMA clusters of two UEs (one cell-center UE and other cell-edge UE). A detailed performance analysis of NOMA-enabled CoMP scheme was presented in [33] where authors showed that enabling NOMA scheme over CoMP network with the appropriate UE clustering gives better performance to that with random clustering. In [34], downlink UEs were assumed to be divided into a number of NOMA clusters with SIC operation applied at every cluster members and the achievable per-UE transmission rate was calculated. The authors in [35] proposed an adaptive NOMA/OMA selection scheme such that the downlink per-UE transmission rate is maximized. It was also found that the performance of NOMA scheme outperforms that of cell-free OMA when the number of UEs is relatively high. Additionally, in [36], a spectral efficiency maximization algorithm for uplink NOMA-enabled cell-free network was proposed

where the authors showed that a better performance can be achieved by controlling the per-UE transmission power.

Additionally, recent advancements in network design have proposed the utilization of cell-free network architectures as an enabling technology for network densification with increased coverage and transmission rate. In cell-free (or cell-less) wireless network, a large number of wireless devices in a geographical area will be served simultaneously in non-orthogonal MA scenarios by a large number of distributed APs [37]. Conventionally, distributive APs coordinate/cooperate with each other through a centralized processing pool [10]. Such a APs cooperation includes channel state information (CSI) estimation, uplink(downlink) decoding(beamforming) and several transmission controlling protocols and algorithms. The state-of-the-art of cell-free network focus on estimation of CSI [23, 38–40], uplink/downlink beamforming [41–44], fronthaul imperfections [45–47], and scalable cell-free network design [10, 16, 48, 49]. All these works and many others in the literature have mainly aimed to solve part of a three substantial technical issues in cell-free networking namely: i) estimation error caused by non-orthogonal pilot contamination, ii) high computational and hardware complexity resulted from centralized processing, and iii) increased amount of control signaling among cooperating APs (i.e, traffic overhead).

For instance in [40] authors have designed a joint uplink/downlink pilot training scheme that uses orthogonal subsets of pilots in the downlink in instead of using channel reciprocity concept while in [50], authors developed a semi-blind channel estimation of uplink cell-free massive MIMO network utilizing an enhanced K-means clustering (E-KMC) algorithm. In [41], inspired by multiuser MIMO beamforming techniques, the authors have proposed a downlink conjugate beamforming and zero-forcing precoding scheme for a fully centralized downlink cell-free network. It was shown in this work that the zero-forcing (ZF) technique outperforms the conjugate beamforming method at the expense of increased computational complexity. However, when the number of UEs and/or APs increases, the complexity behind

using ZF beamforming increases significantly due to the requirements of matrix inversion. Accordingly, in [43] a modified conjugate beamforming technique is proposed. In this work, the requirement for matrix inversion was tackled at the expense of CSI coordination among distributed APs. Another research track of CSI estimation [23]/beamforming [10, 51, 52] technique parallel to conventional model-based analysis has studied the utilization of several deep learning (DL) model-free techniques in optimizing several cell-free network parameters. For example, authors in [23] have developed a channel estimation technique for mmWave-enabled massive cell-free network using the supervised learning-based denoising convolutional neural network. To further utilize the flexibility offered by model-free DL optimization, authors in [10] has solved a joint problem of AP clustering and uplink beamforming in massive cell-free network using DRL techniques. In addition to avoiding the requirements for long training sequences for beamforming matrices optimization, authors of this work proposed the utilization of deep deterministic policy gradient algorithm (DDPG) with continuous action space.

To further reduce the complexity of centralized data processing, [16] proposes a partitioned cell-free wireless network architecture. AP clustering/partitioning[1] is achieved in a UE-centric manner where each active UE select a set of best serving APs. The proposed scheme enables an efficient design of practical mMTC systems by compensating the effect of inter-cluster interference. The compensation follows by network partitioning and enabling multi-level SIC at each receiver [31]. Another low-complexity design of cell-free network architecture appears in [10]. The core idea is to reduce the dimensionality of beamforming matrices by a dynamic clustering of APs. Each cluster then represents a single multi-antenna AP (transmit/receive diversity). However, with the beamforming problem scaled by AP cluster, a central single CPU has to tackle signals received by all active UEs, simultaneously.

Network-Centric vs UE-Centric Network Clustering: Cell-free network represented a

[1]We use the terms clustering(cluster) and partitioning(partition) interchangeably.

very promising solution for problems caused by network densification throughout hierarchical heterogeneous design [53]. However, fully centralized cell-free network, on the other hand, were stumbled by significantly high hardware complexity and processing requirements [10]. Recently, several works in the literature have discussed different layouts for scalable low complexity cell-free network design that are based on either UE-centric or Network-centric clustering. In UE-centric clustering, each active UE within the network coverage area selects the best N APs to be simultaneously served by [16, 54, 55]. On the other hand, a non-overlapping sets of APs may choose a distinct sets of UEs to simultaneously transmit/receive data to/from throughout an Network-centric clustering scheme [10,18]. UE-centric approach may be the most intuitive candidate, since it allows different UEs to connect to their best serving APs. However, this mechanism has two major drawbacks that makes it undesirable for practical design. The first drawback is that when UEs are given the freedom to pick their best serving APs, they may belong to an overlapping clusters which will make fully centralized detection/processing and coordination as a must. Second, the utilization of SIC-based detection may be highly inappropriate in UE-centric approach since different UEs will have different order in each NOMA cluster which add significant amount of complexity to NOMA transmitters and receivers [16]. Accordingly, in this work, we adopt the most general network-centric clustering when partitioning UE and eAPs to different subnetworks.

All the aforementioned low-complexity design, however, sacrifice the performance gain of centralized processing. The complexity of solving the beamforming problem in a centralized manner (e.g. to obtain the beamforming vectors at a centralized processing unit) can however be reduced by using a distributed learning or processing approach while the detection of the transmitted data is still performed at the central unit. Such a solution has been recently investigated in the literature in [25, 56]. Where in [25], the authors utilize supervised learning to solve the beamforming problem in cell-free network. They locate a complete neural network optimizer in each AP. Every AP then obtains the local CSI knowledge by estimating only the large-scale fading while considering the small-scale fading as a constant.

As a neoteric rate-enhancement milestone, researchers have initiated the utilization of a new radio (NR) high transmission bands for mobile wireless services (within the 24-300 GHz rang) [14, 57, 58]. Furthermore, the adoption of mmwave with cell-free network has received a recent attention in the literature [19, 59, 60]. Interestingly, cell-free network were found to provide an efficient solution for the poor scattering nature and high path loss of mmWave bands [59]. Due to the propagation issues related to mmWave channels such as high pathloss an poor scattering, massive multiple-input multiple-output (mMIMO) systems are usually used in mmWave-supported wireless devices [20, 61]. This mmWave-MIMO cell-free network architecture requires significant central computational capabilities and significantly higher control signaling among network entities. Such a computational complexity requirement becomes more deteriorating as the number of wireless devices within the network increases.

1.2 Theoretical Background

1.2.1 Performance Metrics for Wireless Communication Link

For any wireless communication link, there are several metrics that have been used in evaluating the performance, reliability and efficiency of the communication link namely: i) spectral efficiency, ii) probability of outage, iii) bit error rate, iv) communication latency and v) instantaneous transmission rate. In this section, we briefly discuss three important performance metrics that have been used throughout the chapters of this thesis work.

Probability of Outage

In Information theory, the term "*outage probability*" of a communication channel is defined as the probability that a given information rate is not supported, because of variable channel capacity. Or alternatively, outage probability is defined as the probability that the signal-

to-interference plus noise ratio (SINR) of a received signal falls below a certain threshold value. This threshold value is the minimum SINR required by the receiver to be able to decode the received message signal, at a certain transmission rate. Mathematically, the outage probability at the $S \to D$ link is defined as

$$P_{\text{out}} = \mathrm{P}\left(\gamma_D \leq \gamma_{\text{th}}\right), \tag{1.1}$$

where γ_D is the end-to-end SINR between S and D and γ_{th} is a predefined threshold value related to the minimum required transmission rate. The existence of closed-form expression of the probability of outage of any communication system depends on the ability to derive a closed-form expression for the cumulative distributive function of γ_D, denoted by $F_{\gamma_D}(\gamma)$.

Spectral Efficiency

In communication theory, the term "*spectral efficiency*" refer to the information rate that can be transmitted over a given bandwidth in a specific communication system. Alternatively, spectral efficiency is the ratio of the throughput over the utilized bandwidth i.e,

$$r = \frac{\text{Channel Throughput (b/s)}}{\text{Channel Bandwidth (Hz)} \times \text{Channel Utilization (\%)}}. \tag{1.2}$$

This can be written as a function of the link SINR value as follows

$$r = \mathbb{E}_{\boldsymbol{H}}\left[\log_2\left(1+\gamma_D\right)\right] \qquad \text{(bps/Hz)}, \tag{1.3}$$

where, $\boldsymbol{H}$ is the channel matrix at the $S \to D$ link. Usually, averaging over $\boldsymbol{H}$ in (1.3) is very difficult due to the high complexity of the SINR expression γ_D. Accordingly, researchers tend to either simplify the analysis through a series of assumptions or uses the so called "*instantaneous transmission rate*".

Instantaneous Transmission Rate

Instantaneous transmission rate is a performance metric that directly measures the system transmission rate at a time slot-bases, assuming a constant CSI. Considering the assumption

that CSI remains constant during one time slot, the metric in (1.3) can be modified as

$$r = \log_2 (1 + \gamma_D) \qquad \text{(bps/Hz)}, \tag{1.4}$$

1.2.2 Resource Allocation in Massive Wireless Network

The newly emerging wireless network architectures have opened the door for new technical challenges related to the allocation of different network resources (transmission power, frequency bands, the right to connect, etc). One major issue is represented by applying these optimization problems to massive network while guaranteeing good performance, short processing delay and affordable computational and hardware complexities[2]. This urge wireless network researchers to investigate the development of advanced optimization techniques (mathematical and algorithmic) that comply with B5G/6G requirements and standards.

Traditional Optimization Techniques

In traditional resource allocation techniques, optimization problems usually solved by first formulating it in a way such that the objective function represents a linear function of the optimized variable(s) (could be a single element, a vector or a matrix of elements) and such that all accompanying constraints represented by an affine functions (or inequalities) of the optimization variables. If the objective function is convex and relative constraints are affine, then the resource optimization problem can be easily solved through the theory of convex optimization that guarantees the existence of a global solution of the problem (ex., transmission power, allocated frequency , etc.). In some scenarios where the objective function is not linear, researcher may tend to relax it into a linear function using some conditions such as second order Taylor series expansion and duality property. Such an objective function relaxation or reformation comes at the expense of a certain performance

[2]It is expected that the group of smart meters, road security devices and consumer electronic devices will reach several millions UEs per km^2 [62].

degradation.

Furthermore, with the scenario of massive wireless connectivity, it is quite rarely that resource allocation problems can be formulated by a tractable forms that are easily solvable using convex optimization-based methods. This is due to the fact that simultaneous utilization of the network resources presents a serious problems related to interference alignment, security-related constraints, transmission power efficiency, etc. These types of optimization problems are usually formulated as a non-convex optimization problems. Such a non-convex problems can be solved sub-optimally by variety of algorithmic-based optimization techniques. For example, UE clustering algorithms (such as greedy algorithms, metaheuristic algorithms, genetic algorithms, etc.) [63], power allocation algorithms (such as equal power allocation algorithms, iterative power control algorithms, exhaustive search algorithms, etc.) [64].

Machine Learning-Based Optimization Methods

When the optimization problem contains a mix of discrete and continuous optimization variables (i.e, combinatorial optimization problem), conventional optimization strategies fail to provide an efficient solution without the need of exhaustive search (ES) or look up strategies. As an example, performance optimization problems (transmission rate, power efficiency, frequency allocation, etc.) that are related to multiuser-multicell 6G network consist of several discrete target variables such as UEs and/or AP clustering configurations and frequency bands allocation indices and contains several continuous optimization variables such as transmission power values, beamforming matrices, etc. Utilizing ES-based methods in solving these problems with massive number of variables will results in significant amount of computational requirements longer processing delay. This call for the need of unconventional optimization methods that are able to jointly optimize discrete and continuous variables without the need of going through all possible discrete values. Accordingly,

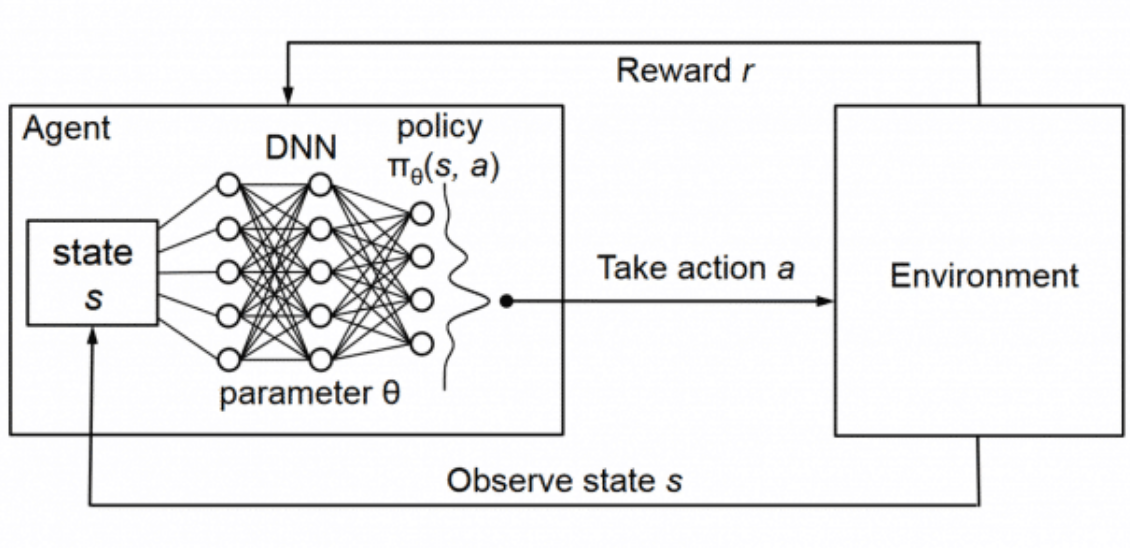

Figure 1.1 Basic block diagram of DRL system.

different ML-based methods are being developed to solve several optimization problems related to B5G/6G network with massive connectivity [15]. There are two main ML-based optimization approaches namely; i) resource allocation through supervised learning and ii) optimization techniques through DRL algorithms. For wireless communication network however, the availability of efficient long training and testing sequences is very unlikely. This is due to the highly dynamic wireless communication mediums that is susceptible to several propagation parameters. For this reason, we will focus in the utilization of DRL algorithms for efficiently optimizing several important wireless resource allocation problems.

DRL: Theoretical Preliminaries

The concept of reinforcement learning (RL) refers to the learning process of an agent interacting with its environment after receiving certain observations. The environment provides a reward to the agent for every interaction and the RL agent aims to select the right action for the next interaction in order to maximize the discounted reward over a time horizon (**Fig. 1.1**). This problem can be formulated as a *Markov decision process* (MDP). An MDP is a tuple ($\mathcal{S}$, $\mathcal{A}$, $\mathcal{P}$, $\mathcal{R}$, ζ), where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is a finite set of actions from

which the agent can choose, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0,\, 1]$ is a transition probability in which $\mathcal{P}(\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{s}')$ defines the probability of observing state $\boldsymbol{s}'$ after executing action $\boldsymbol{a}$ in the state $\boldsymbol{s}$, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the expected reward after being in state $\boldsymbol{s}$ and taking action $\boldsymbol{a}$, and $\zeta \in [0, 1)$ is the discount factor. To solve the MDP, RL algorithms have been developed to learn and find a discrete value function or a "policy". Such a discretization can lead to lack of generalization and significantly increase the problem dimensionality. Therefore, DRL algorithms based on function approximation by deep neural networks (DNNs) have been proposed. DRL algorithms can be classified into three types:

- ***Value-Based*** methods: Such as deep Q-learning (DQL) and state-action-reward-state-action (SARSA) which only learn the so-called value function to find a policy. In value-based algorithm, each state is assigned a certain value through a state-value function $V^{\pi}(\boldsymbol{s})$ (also known as the expected return function when starting at a certain state $\boldsymbol{s}$). The state-value function is defined as

$$V^{\pi}\left(\boldsymbol{s}\right) = \mathbb{E}\left[r|\boldsymbol{s}, \pi\right], \tag{1.5}$$

  where π is the followed policy and can be found such that

$$V^{*}\left(\boldsymbol{s}\right) = \max_{\pi} V^{\pi}\left(\boldsymbol{s}\right), \quad \forall\ \boldsymbol{s} \in \boldsymbol{\mathcal{S}}. \tag{1.6}$$

  Given $V^{*}\left(\boldsymbol{s}\right)$, the optimal policy π^{*} is found by selecting the best action that maximizes $\mathbb{E}_{\boldsymbol{s}' \sim \mathcal{T}(\boldsymbol{s}'|\boldsymbol{s},\boldsymbol{a})}\left[V^{*}\left(\boldsymbol{s}'\right)\right]$, where $\mathcal{T}\left(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}\right)$ is the transition dynamics that is usually unavailable in RL. Hence, the value function is replaced by a quality state-action-value function $Q^{\pi}\left(\boldsymbol{s}, \boldsymbol{a}\right)$ which is different from V^{π} by the fact that a random initial action $\boldsymbol{a}_0$ is provided and the policy π is only counted from the succeeding state, i.e

$$Q^{\pi}\left(\boldsymbol{s}, \boldsymbol{a}\right) = \mathbb{E}\left[r|\boldsymbol{s}, \boldsymbol{a}, \pi\right]. \tag{1.7}$$

  The learning of the Q^{π} network is deployed using the Bellman equation with the following recursive form [65]

$$Q^{\pi}\left(\boldsymbol{s}, \boldsymbol{a}\right) = \mathbb{E}_{\boldsymbol{s}'}\left[r' + \gamma Q^{\pi}\left(\boldsymbol{s}', \pi\left(\boldsymbol{s}'\right)\right)\right]. \tag{1.8}$$

This means that the quality function can be improved by bootstrapping (i.e using current values of Q^{π} to improve our estimate). This modeling has triggered the foundation of Q-learning [66] and SARSA [67] algorithms:

$$Q^{\pi}\left(\boldsymbol{s},\boldsymbol{a}\right)\leftarrow Q^{\pi}\left(\boldsymbol{s},\boldsymbol{a}\right)+\alpha\delta, \tag{1.9}$$

where α is the learning rate and $\delta=Y-Q^{\pi}\left(\boldsymbol{s},\boldsymbol{a}\right)$ is the temporal difference error with Y represents a target (as in standard regression problems). Using Q-learning agent, the target Y directly approximate Q^{*} by setting $Y=r+\gamma\underset{j}{\max}Q^{\pi}\left(\boldsymbol{s}',\boldsymbol{a}\right)$ (off-policy agent). On the other hand, the SARSA algorithm improves the estimate of Q^{π} by deriving a behavioural policy from Q^{π}. This is achieved by setting $Y=r+\gamma Q^{\pi}\left(\boldsymbol{s}',\boldsymbol{a}'\right)$ (on-policy agent).

- ***Policy-Based*** methods such as policy gradient (PG) algorithm which learn the policy directly by following the gradient with respect to the policy. In Policy gradient algorithms, the modeling and optimization of a certain policy is conducted directly through a parameterized function, $\mu_{\theta}\left(\boldsymbol{a},\boldsymbol{s}\right)$. The value of the objective function (the reward) directly depends on the policy. The value of the reward function for PG-based clustering agents is given by

$$\begin{aligned}J(\theta)&=\sum_{\boldsymbol{s}\in\boldsymbol{\mathcal{S}}}\left(\boldsymbol{s}\right)V^{\mu}\left(\boldsymbol{s}\right)\\&=\sum_{\boldsymbol{s}\in\boldsymbol{\mathcal{S}}}d^{\mu}\left(\boldsymbol{s}\right)\sum_{j\in\mathbb{Z}}\mu_{\theta}\left(\boldsymbol{a}|\boldsymbol{s}\right)Q^{\mu}\left(\boldsymbol{s},\boldsymbol{a}\right),\end{aligned} \tag{1.10}$$

where $d^{\mu}\left(\boldsymbol{s}\right)$ is the stationary state distribution of Markov chain. Note that the gradient of $J\left(\theta\right)$ (denoted by $\nabla_{\theta}J\left(\theta\right)$) depends both on the selected actions $\boldsymbol{a}_{\mathrm{c}}$ and the stationary distribution $d^{\mu}\left(\boldsymbol{s}\right)$ which makes the environment unknown since it will be very difficult to estimate the effect of steady states from instantaneous actions perspectives. In this work, we also use policy gradient (PG) algorithm that significantly

simplifies the computation of objective gradient by reforming $J(\theta)$ and removing its dependence on $d^{\mu}(\boldsymbol{s})$. This can be expressed as [68, Sec. 13.2]

$$\begin{aligned} \nabla_{\theta} J(\theta) &= \nabla_{\theta} \sum_{\boldsymbol{s} \in \boldsymbol{\mathcal{S}}} d^{\mu}(\boldsymbol{s}) \sum_{j \in \mathbb{Z}} \mu_{\theta}(\boldsymbol{a}|\boldsymbol{s}) Q^{\mu}(\boldsymbol{s}, \boldsymbol{a}) \\ &\propto \sum_{\boldsymbol{s} \in \boldsymbol{\mathcal{S}}} d^{\mu}(\boldsymbol{s}) \sum_{j \in \mathbb{Z}} \mu_{\theta}(\boldsymbol{a}|\boldsymbol{s}) \nabla_{\theta} Q^{\mu}(\boldsymbol{s}, \boldsymbol{a}). \end{aligned} \tag{1.11}$$

The general policy gradient method has a significantly high variance. Accordingly, many following algorithms were proposed that keep the bias unchanged while reducing the variance of (5.20). As a result, (5.20) can be written in the following general form

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\mu}\left[Q^{\mu}(\boldsymbol{s}, \boldsymbol{a}) \nabla_{\theta} \ln \mu_{\theta}(\boldsymbol{a}|\boldsymbol{s})\right]. \tag{1.12}$$

This general form is used as a foundation of several different PG algorithms with the following form

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\mu_{\theta}}\left[\sum_{t=0}^{T-1} G_t \nabla_{\theta} \log \mu_{\theta}(\boldsymbol{a}|\boldsymbol{s})\right]. \tag{1.13}$$

The PG-based agent for network partitioning process can be then trained through the following steps:

i- Initialize the actor $\mu(\boldsymbol{s})$ with random weights θ_{μ}.

ii- For each training Episode (i.e, every T training steps), generate the experience by following $\mu(\boldsymbol{s})$: the actor generate probability values for each possible cell-free partition, then the agent randomly select an action based on a certain probability distribution. This process continues for T steps.

iii- At each step of a certain episode, calculate the return value G_t using the discount future reward as follows

$$G_t = \sum_{l=t}^{T} \zeta^{l-1} r_l. \tag{1.14}$$

iv- Find the accumulative sum of the actor network gradients during one whole learning episode as

$$d\theta_{\mu} = \sum_{t=1}^{T} G_t \nabla_{\theta_{\mu}} \ln \mu \left(\boldsymbol{s} | \theta_{\mu} \right) \tag{1.15}$$

v- Update actor using

$$\theta_{\mu} \leftarrow \theta_{\mu} + \alpha d\theta_{\mu} \tag{1.16}$$

- ***Actor-Critic***: In PG-based DRL algorithm, the value function G_t is estimated based on a preassigned policy. However, the estimation of G_t through predefined objective is found to introduce a relatively high gradient variance which in term reduces the quality of of action selection. In order to tackle the high variance problem, researchers has proposed the utilization of a second DNN that can accurately learn the value of G_t [69]. In this work, we use a modified actor-critic (AC) algorithm that enable both actor and critic networks to have discrete action spaces [70]. The standard DQL method represents the most popular actor-critic update algorithm in the literature due to the availability of a mature theory. Basically, the DQL update equation at time t for a network agent with parameters θ^{Q} after taking action $\boldsymbol{a}$ in state $\boldsymbol{s}$ and observing the immediate reward r' and resulting state $\boldsymbol{s}'$ is:

$$\begin{aligned} Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q\prime}) &= Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q}) + \nu \left[r' + \zeta \max_{\boldsymbol{a}'} Q(\boldsymbol{s}', \boldsymbol{a}' \,|\, \theta^{Q}) - Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q}) \right] \\ &= Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q}) + \nu \left[r' + \zeta \max_{\boldsymbol{a}'} Q(\boldsymbol{s}', \operatorname*{argmax}_{\boldsymbol{a}'} Q(\boldsymbol{s}', \boldsymbol{a}' \,|\, \theta^{Q}) \,|\, \theta^{Q}) - Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q}) \right], \end{aligned} \tag{1.17}$$

where ν is the learning rate. Computing the term $\max_{\boldsymbol{a}'} Q(\boldsymbol{s}', \operatorname{argmax}_{\boldsymbol{a}'} Q(\boldsymbol{s}', \boldsymbol{a}' \,|\, \theta^{Q}) \,|\, \theta^{Q})$ introduces a systematic overestimation of the Q-values during the learning that is accentuated by the use of bootstrapping, i.e. learning estimates from estimates. The Q-learning update in (4.17) uses the same Q-network $Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q})$ both to select and to evaluate an action. After highlighting the overestimation bias in experiments across different Atari game environments, Hasselt et al. [71] decoupled the action selection and evaluation by introducing two deep Q-networks, a Q network and a target network

Q' with different parameters θ^Q and θ'^Q, respectively, to avoid the maximization bias. The Q' network is used for action selection while the Q network is used for action evaluation. This is known as *deep double Q-learning algorithm (DDQL)*. The DDQL update equation of the network can be expressed as:

$$\begin{aligned} Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^{Q\prime}) &= Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^Q) \\ &+ \nu \Bigg[r' + \zeta \max_{\boldsymbol{a}'} Q(\boldsymbol{s}', \operatorname*{argmax}_{\boldsymbol{a}'} Q'(\boldsymbol{s}', \boldsymbol{a}' \,|\, \theta'^Q) \,|\, \theta^Q) - Q(\boldsymbol{s}, \boldsymbol{a} \,|\, \theta^Q) \Bigg]. \end{aligned} \tag{1.18}$$

The parameters θ'^Q of the Q' network periodically hard-copy the parameters θ^Q of Q network after t_0 time steps using the Polyak averaging method with parameter $\tau \in [0, 1]$:

$$\theta'^Q_{t+t_0} = (1 - \tau)\, \theta'^Q + \tau\, \theta^Q. \tag{1.19}$$

The DDQL network show a better performance that standard DQL [71]; however, due to the discretization requirements of the DNN outputs (the action space $\mathcal{A}$), it results in a huge expansion of the action space dimensionality when used in the optimization of an objective function of continuous dependent variables. *This dimensionality issue makes it an unattractive solution for solving the beamforming problem under massive number of UEs and APs.* However, it is a relevant candidate for the clustering problem of the APs since it avoids the need for an extremely inefficient exhaustive search method. This motivates us to utilize the "DDPG" policy for the beamforming design problem.

DDPG belongs to the class of actor-critic algorithms. It concurrently learns a Q-function network approximation $Q(\boldsymbol{s}, \boldsymbol{a}|\theta^Q)$ called the critic, and a policy network approximation $\mu(\boldsymbol{s}|\theta^\mu)$ called the actor. The Q-function network is trained using the Bellman equation, while the policy network is learnt using the Q-function. Unlike the DQL policies which output the probability distribution $\pi(\boldsymbol{a}|\boldsymbol{s})$ across a discrete action space $\mathcal{A}$, the policy network of DDPG directly maps states to actions. Specifically, at

every time step t, it maximizes its loss function defined as:

$$J(\theta) = \mathbb{E}\Big[Q(\boldsymbol{s}, \boldsymbol{a}) \mid \boldsymbol{s}, \boldsymbol{a} = \pi(\boldsymbol{a}|\boldsymbol{s})\Big] \tag{1.20}$$

and updates its weights θ by following the gradient of (4.20):

$$\nabla J_{\theta^\mu}(\theta) \approx \nabla_{\boldsymbol{a}} Q(\boldsymbol{s}, \boldsymbol{a})\, \nabla \mu(\boldsymbol{s}|\theta^\mu). \tag{1.21}$$

This update rule represents the deterministic policy gradient (DPG) theorem, rigorously proved by Silver et al. in the supplementary material of [72]. The term $\nabla_{\boldsymbol{a}} Q(\boldsymbol{s}, \boldsymbol{a})$ is obtained from a Q-network $Q(\boldsymbol{s}, \boldsymbol{a}|\theta^Q)$ called the critic by backpropagating its output w.r.t. the action input $\mu(\boldsymbol{s}|\theta^\mu)$. When the number of actions is very large, this actor-critic training procedure solves the intractability problem of DQN [73] by using the following approximation:

$$\max_{\boldsymbol{a}} Q(\boldsymbol{s}, \boldsymbol{a}) \approx Q(\boldsymbol{s}, \boldsymbol{a}|\theta^Q)|_{\boldsymbol{a}=\mu(\boldsymbol{s}|\theta^\mu)}. \tag{1.22}$$

Similar to DQN, two tricks are employed to stabilize the training of the DDPG actor-critic architecture, namely, 1) the experience replay buffer R to train the critic, and 2) target network for both the actor and the critic which are updated using the polyak averaging in the same way it was done in (4.19).

1.3 Enabling Techniques for Future Wireless Networks

In this section, we briefly discuss some enabling techniques for future wireless networks that have been receiving a noticeable interests in recent literature.

1.3.1 Massive Wireless Connectivity

In B5G/6G networks, the amount of transmitted data will continue in its massive growth. One new feature of B5G/6G networks is that massive amount of data will be shifted from

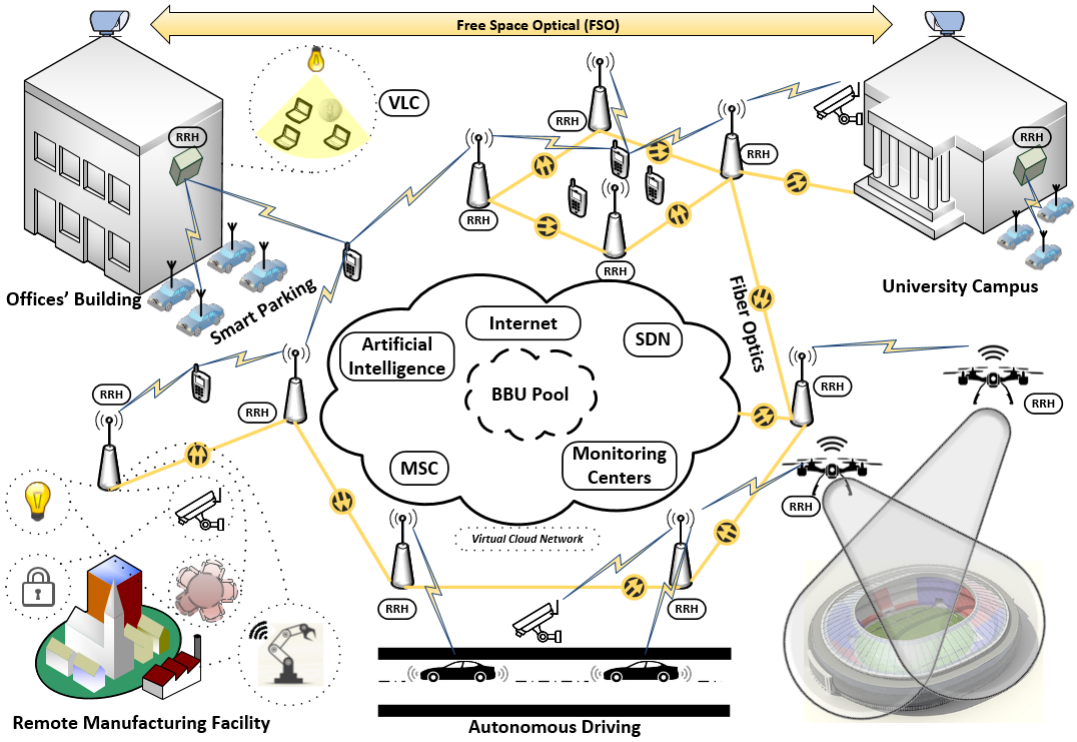

Figure 1.2 Massive connectivity in future wireless networks.

a centralized, big data into distributed small data. This feature is a direct consequence of the emergence of distributed wireless technologies with massive transmission data in a distributed form (see Fig. 1.2). Furthermore, the next generation wireless network will have to serve a massive number of terminal UEs within small geographical areas and this will give rise to dense/ultra-dense deployment of APs with overlapping coverage areas. In such a scenario, devices will be served simultaneously by multiple APs (e.g. through multipoint transmissions and multipoint UE associations), which will be required for efficient hand-off/hand-over, frequency allocation and interference management.

1.3.2 Non-Orthogonal Multiple Access

The idea of power-domain NOMA is that at a certain frequency sub-band, signals of multiple UEs are superimposed in the power domain such that the received signal for each UE has a distinct power level. At the NOMA receiver end, SIC is used to cancel signal components with

higher weights than the desired signal (starting from the signal with the highest weight) [6]. In theory, NOMA enhances the spectral efficiency significantly compared to that of OMA at the expense of receiver complexity and processing delay [7, 8]. Generally, NOMA can be applied either in the uplink or downlink of a wireless system (Fig. 1.3). For example, in

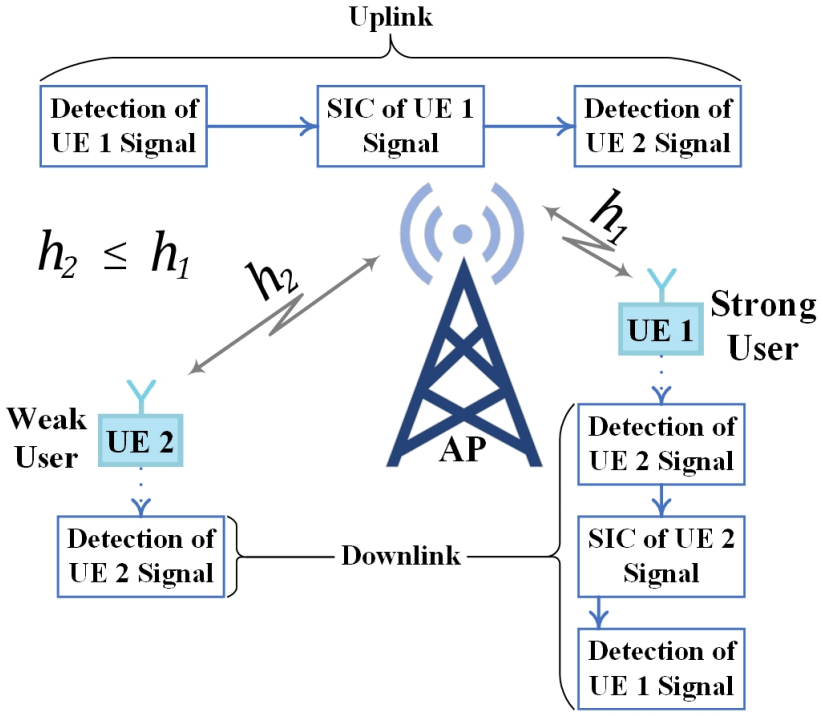

Figure 1.3 Two UEs Power domain NOMA: uplink and downlink.

downlink NOMA with K UEs per NOMA cluster, the received signal at the k-th UE is given by

$$\begin{aligned} y_k &= P\sum_{i=1}^{K}\sqrt{w_i P}h_k x_i + n_k \\ &= \underbrace{\sqrt{w_k P}h_k x_k}_{\text{Desired Signal}} + \underbrace{\sum_{i=1, i\neq k}^{K}\sqrt{w_i P}h_k x_i}_{\text{IUI}} + n_k, \end{aligned} \tag{1.23}$$

where h_k is the the channel fading coefficient between the k-th UE and the AP, x_k is the signal transmitted to the k-th UE such that $\mathbb{E}\left[|x_k|^2\right] = 1$, P is the power budget at the AP, w_k is the fraction of power allocated to the k-th UE and n_k is the additive white Gaussian noise

(AWGN) at the input of the k-th UE. If the CSI of UEs are ordered such that $h_1 \leq \ldots \leq h_K$ and power fractions for UEs are assigned such that $w_1 \geq \ldots \geq w_K$, the k-th UE will be able to remove the interfering signal from $k-1$ UEs with lower channel gain. Accordingl,y the signal to interference plus noise ratio (SINR) at the input of the k-th UE detector is given by

$$\gamma_k = \frac{w_k P |h_k|^2}{\sum_{i=k+1}^{K} w_i P |h_k|^2 + N_k}, \tag{1.24}$$

where N_k is the power of the AWGN at the input of the k-th UE. On the other hand, for uplink NOMA, the received signal related to the k-th UE at the AP is given by

$$\begin{aligned} y_k &= P \sum_{i=1}^{K} \sqrt{P} h_i x_i + n_k \\ &= \underbrace{\sqrt{P} h_k x_k}_{\text{Desired Signal}} + \underbrace{\sum_{i=1, i\neq k}^{K} \sqrt{P} h_i x_i}_{\text{IUI}} + n, \end{aligned} \tag{1.25}$$

where n is the AWGN at the input of the AP. If UEs are ordered such that $h_1 \geq \ldots \geq h_K$ and under the assumption that all UEs are transmitting with a constant power P, the AP will use SIC to first decode signals of UEs with higher gain, and then remove it from the overall signal. Accordingly, the SINR value related to the k-th UE decoder is given by

$$\gamma_k = \frac{P |h_k|^2}{\sum_{i=k+1}^{K} P |h_i|^2 + N_k}. \tag{1.26}$$

Note that the SINR in (1.24) has the same channel gain at both numerator and denominator, unlike that of (1.26) which has different channel gains. This feature of the SINR values will play a major role in the design of uplink/downlink beamforming schemes for NOMA systems as will be shown in subsequent chapters.

1.3.3 Cell-free Wireless Network Architectures

Unlike wireless cellular network, cell-free network architecture enables all APs of the network to simultaneously serve all UEs within the network coverage area using the same time/frequency resources (Fig. 1.4). Utilizing the very fast backhauling links among differ-

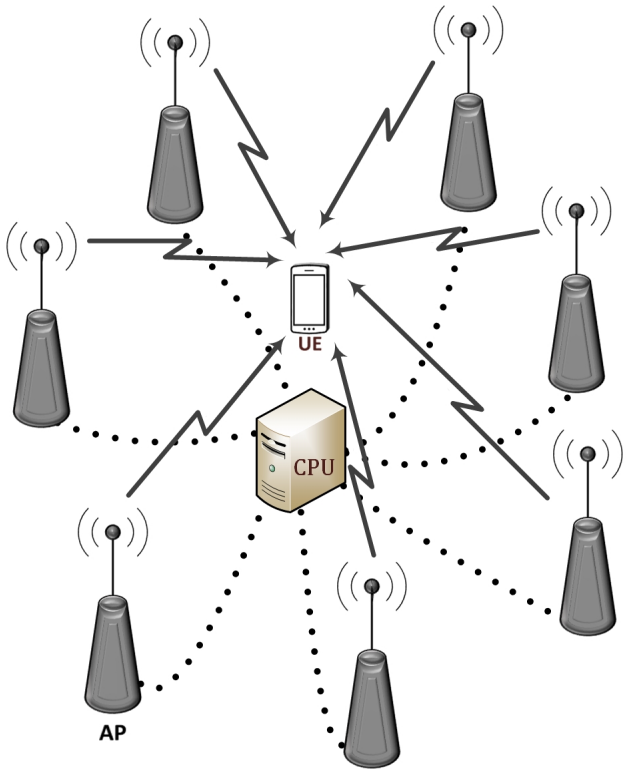

Figure 1.4 Cell-free network model.

ent APs, the overall network will appear as a cell-free distributed massive MIMO system from end-device point of view. Specifically, all APs will be aware of all active devices within their vicinity. APs may be considered as remote radio heads (RRHs) as in the case of C-RANs [74]. Every device may be served by more than one RRH either by transmission coordination or transmission multiplexing. It may be useful to view this cell-free architecture as a generalized version of the well known CoMP network at which cooperating APs jointly serve all devices within their coverage area (cell-edge and cell-centre devices). This can be enabled by the utilization of very fast CPUs that assign resources to different terminal devices while

data processing may be conducted at the BBU as in the case of CRANs. With complete coordination among different RRHs, interference management can be performed optimally or near optimally throughout some centralized or distributed optimization methodologies. Let us consider a fully centralized downlink cell-free network with M APs and K UEs, the received downlink signal at the k-th UE at any arbitrary time slot is given by

$$\begin{aligned} y_k &= \sum_{m=1}^{M}\sum_{i=1}^{K}\sqrt{w_{im}P}h_{km}x_i + n_k, \\ &= \underbrace{\sum_{m=1}^{M}\sqrt{w_{km}P}h_{km}x_k}_{\text{Desired Signal}} + \underbrace{\sum_{m=1}^{M}\sum_{i=1,i\neq k}^{K}\sqrt{w_{im}P}h_{km}x_i}_{\text{IUI Signal}} + n_k, \end{aligned} \tag{1.27}$$

where w_{km} is the beamforming fraction assigned to the $m \to k$ link such that $\sum_{i=1}^{K} w_{km} \leq 1$, P is the power budget at each AP, h_{km} is the channel gain coefficient at the $m \to k$ link and n_k is the AWGN at the input of the k-th UE. Accordingly, the SINR at the k-th UE is given by

$$\gamma_{k,\text{Downlink}} = \frac{||\sum_{m=1}^{M}\sqrt{w_{km}P}h_{km}||^2}{||\sum_{m=1}^{M}\sum_{i=1,i\neq k}\sqrt{w_{im}P}h_{km}||^2 + N_k}. \tag{1.28}$$

In the uplink, the signal collected from distributed APs is sent into a module of K detectors at a certain central processing unit (CPU). To detect the uplink signal related to the k-th UE, the detector multiply the overall signal received by each AP by a certain precoding fraction $w_{mk} \in [0 \;\; 1]$, accordingly, the received uplink signal at the input of the k-th UE detector is given by

$$\begin{aligned} y_k &= \sum_{m=1}^{M}\sqrt{w_{mk}}\left(\sqrt{P}\sum_{i=1}^{K}h_{mi}x_i + n_k\right), \\ &= \underbrace{\sum_{m=1}^{M}\sqrt{w_{mk}P}h_{mk}x_k}_{\text{Desired Signal}} + \underbrace{\sum_{m=1}^{M}\sqrt{w_{mi}}\sum_{i=1,i\neq k}^{K}\sqrt{P}h_{mi}x_i}_{\text{IUI Signal}} + \underbrace{\sum_{m=1}^{M}\sqrt{w_{mk}}n_m}_{\text{AWGN}}, \end{aligned} \tag{1.29}$$

where h_{mk} is the channel gain coefficient at the $k \to m$ link. Accordingly, the SINR at the input of the k-th UE detector in the CPU is given by

$$\gamma_{k,\text{Uplink}} = \frac{||\sum_{m=1}^{M} \sqrt{w_{mk}P} h_{mk}||^2}{||\sum_{m=1}^{M} \sqrt{w_{mk}} \sum_{i=1, i\neq k} \sqrt{P} h_{mi}||^2 + \sum_{m=1}^{M} w_{mk} N_k}. \tag{1.30}$$

1.3.4 Boosting Link Performance through Time-Diversity

Multiple APs/antenna schemes has been a main icon of wireless networks since the very beginning of multiuser transmission. In this section, we give a brief discussion that shows the great potentials of dynamically changing communication links based on the stochastic properties of the propagation media. In order to achieve this, lets consider the following arbitrary CSI matrix of a 5×5 MIMO system (drawn from a random Rayleigh distribution):

$$\boldsymbol{H} = \begin{bmatrix} -0.2725 & -0.3538 & 0.0335 & 0.0229 & -0.9792 \\ 1.0984 & -0.8236 & -1.3337 & -0.2620 & -1.1564 \\ -0.2779 & -1.5771 & 1.1275 & -1.7502 & -0.5336 \\ 0.7015 & 0.5080 & 0.3502 & -0.2857 & -2.0026 \\ -2.0518 & 0.2820 & -0.2991 & -0.8314 & 0.9642 \end{bmatrix}. \tag{1.31}$$

Under the assumption of maximum transmission power of $P = 38$ dBm and optimal MIMO precoding (water-filling power allocation), this MIMO system can achieve a Shannon normalized rate of 12.5502 bps/Hz. Let us now assume that we want to split the 5×5 MIMO system into two independent 4×4 and 1×1 MIMO subsystems. As an example, consider the two possible MIMO antenna clustering sets

$$\mathcal{C}_1 = \{\overbrace{\{\{m_1, m_2, m_4, m_5\}, \{n_1, n_2, n_3, n_4\}\}}^{4\times 4 \text{ MIMO Subsystem}}, \overbrace{\{\{m_3\}, \{n_5\}\}}^{1\times 1 \text{ MIMO Subsystem}} \}.$$

$$\mathcal{C}_2 = \{\overbrace{\{\{m_1, m_2, m_3, m_4\}, \{n_2, n_3, n_4, n_5\}\}}^{4\times 4 \text{ MIMO Subsystem}}, \overbrace{\{\{m_5\}, \{n_1\}\}}^{1\times 1 \text{ MIMO Subsystem}} \}.$$

By splitting P into $\frac{4}{5}P$ and $\frac{1}{5}P$ for the 4×4 and 1×1 MIMO subsystems, respectively, the achieved normalized transmission rates for configurations $\mathcal{C}_1$ and $\mathcal{C}_2$ are shown in **Table 1.2**

Table 1.2 Illustrative example

Configuration	4×4	1×1	Overall
$\mathcal{C}_1$	7.2492 bps/Hz	0.4428 bps/Hz	7.6920 bps/Hz
$\mathcal{C}_2$	9.6700 bps/Hz	2.6582 bps/Hz	12.3282 bps/Hz

below. As can be noticed from **Table 1.2**, the selection of antenna elements significantly affects the performance of the new MIMO subsystems. If such a selection is optimized on a time slot-basis to exploit the time-varying nature of the channel, a significant performance gain, compared to that of static MIMO splitting, technique can be achieved. Such a large performance gain can be utilized in optimizing a variety of network architectures/technologies in a time slot-bases. For example, dynamically changing the set of cooperating APs in cell-free network, one can significantly scale-down the system complexity while achieving an excellent performance. Applying this concept of time diversity over massive cell-free network is one major contribution of this thesis work.

1.3.5 mmWave Communications

one peculiar property of every generation of wireless communication is the addition of new frequency bands. As a neoteric development to the next wireless networks, researchers are investigating the utilization of the millimeter wave (mmWave) band of frequencies for very high speed hot-spots. The mmWave band is referred to the band of frequencies at which the signal wavelengths is in the millimeter range (tens or hundreds of millimeters). This includes all frequencies within the $24 - 300$ GHz band. Utilizing these frequencies for wireless transmission can offer a great link capacities, as high as multi-gigabit-per-second (Gbps). Despite the enormous available bandwidth potential, mmWave signal transmissions suffer from fundamental technical challenges like severe path loss, sensitivity to blockage, directivity, and narrow beamwidth, due to its short wavelengths [75]. To overcome these challenges, an inno-

vative techniques should be designed to comply with the new physical characteristics of the newly utilized spectrum. For instance, multi-antennas systems may provide robust beamsteering capabilities for the low scattering nature of mmWaves. Another way to boost up mmWave communication is the utilization of hybrid beamforming techniques under mMIMO techniques. In hybrid beamforming, the antenna radiations are directed toward main components of the mmWaves through analog beamsteering while IUI components are mitigated through digital beamforming techniques.

1.4 Motivation

In previous sections, it was shown how the emergence of new services and technologies have resulted in a dramatic increase in the number of wireless devices per unit area. The dense distribution of these wireless devices has changed the notion of large data transmission at network centres into a massive small data distributed among network coverage area. This has created a serious technical issues related to spectrum scarcity, wireless interference, resource allocation and power efficiency. It also imposed a dire need of solving these issues in a coordinated way, either by centralized network architectures or through distributed optimization problems that solve a centralized main objective. Furthermore, spectrum scarcity can be significantly resolved by utilizing more bands within the realizable ranges. However, with the recent introduction of the $24 - 300$ GHz range, it has became true that the whole usable frequency spectrum is almost completely utilized. As a result, recent works in the literature proposed a non-orthogonal utilization of the available spectrum through the well-known NOMA scheme. Power domain NOMA, however, has opened the door for increased hardware and computational complexities, especially for large-scale NOMA clusters. To solve this issue, we have designed a partially orthogonal MA scheme namely delta-orthogonal multiple access (D-OMA). The idea of the proposed D-OMA is that different NOMA clusters are allowed to partially overlap at the frequency domain under some quality of service (QoS)

constraints. The spectrum gained by the partial overlapping allows us to decrease the number of wireless devices per NOMA cluster while maintaining the same QoS, which will results in a significant hardware and computational complexities.

To further enhance the quality of signal transmitted/received to/from massively distributed UEs, we have taken a step forward by allowing distributed APs to cooperate in serving all UEs within the network coverage area. Such a network structure is know in the literature as cell-free network architectures and it add excellent robustness to transmitted data in term of interference mitigation and control. However, as we previously illustrated in the literature review, the centralized simultaneous processing of date in large-scale cell-free network requires a huge processing capabilities with significantly high control traffic. Such a complexity issues increases even more with the emergence of mmWave-supported mMIMO at each transmitting/receiving node. The complexity of centralized signal processing however, can be reduced by using scalable, dynamic, and low-complexity design [10], and also by using suitable network- and UE-centric architectures and algorithms, [17, 18]. Nevertheless, the aforementioned low-complexity models sacrifice the performance gain of centralized processing. Accordingly, major revision and improvement of the previously mentioned design is required to reduce the amount of performance degradation of such models. In this context, we have developed several wireless network architectures that enable cell-free networking at the large-scale and with minimized performance degradation, lower complexity and with shorter processing delay. As the main idea of the developed network architectures rely behind the ability of the network to self-adapt/self-optimize its architecture at a time slot bases, the development of several innovative methodologies that efficiently conduct such tasks has became a dire need, especially at the large-scale paradigm. Motivated by this need, we have designed and investigated several novel DRL models that jointly solve the problem of network clustering/partitioning and uplink/downlink beamforming. Such a models enable simultaneous online training and operation with minimize processing delay and hardware

and computational complexities.

1.5 Thesis Objectives

This thesis consists of one primary objective and a set of secondary objectives as follows.

Primary Objective: The main objective of this thesis is to design and analyze several algorithms and techniques that enables massive multiple wireless access for future wireless networks.

Secondary Objectives: While working on this thesis toward our primary objective, several secondary objectives have been raising out as a group of enablers for the primary goal. Here, we enumerate some of the secondary objectives of this thesis work.

- *Tractable Expressions for Performance Analysis:* In this regard, we aim to derive a set of closed-form expressions of several performance metrics for the developed network architectures.

- *Provisioning of Self-Organizing Networks through Innovative DRL Models:* In this objective, we aim to set up the main framework for several DRL models that can be efficiently used in the optimization and organization of future large-scale wireless networks.

- *Enabling Distributed Processing over Massive Cell-Free Network:* In this objective, we aim to design and discuss several network architectures and DRL algorithms that efficiently enable distributed uplink/downlink beamforming without severely degrading the performance of centralized processing.

1.6 Thesis Contributions and Organization

The contribution of this thesis can be divided into four complementary research tracks as follows:

- **Track I:** In the first research track, we design a novel MA scheme that represents a trade-off between fully OMA scheme and that of NOMA (**Chapter 2**). The designed access scheme (denoted by the D-OMA scheme) allows adjacent NOMA cluster to partially overlap in the frequency domain to increase the spectral efficiency of the wireless network. The gained spectral efficiency increases the degree of freedom for reducing the size of each NOMA cluster, which results in significant reduction in hardware and computational complexities. Furthermore, in order to further enhance the per-UE performance under the D-OMA scheme, we allow adjacent APs to simultaneously serve all D-OMA UEs by coordinating their processing capabilities through CPU units.

- **Track II:** In this thesis track, we generalized the D-OMA concept for full-spectrum overlapping. The inter-NOMA-cluster interference (INCI) is then compensated by proposing different forms of APs cooperation. For this cooperation scheme, we derive several exact and asymptotic closed-form expressions of different performance metrics for wireless cell-free network. Specifically, we have derived exact expressions of the probability of outage, ergodic capacity and outage-capacity for the fully centralized cell-free network as well as the scalable UE-centric clustered cell-free architecture.

- **Track III:** As the concept of cell-free network has been fully studied and evaluated in **Track II**, in the third track of this thesis, we have developed several innovative techniques for scaling down the size of cell-free network while maintaining a satisfactorily per-UE performance. This is achieved through three innovative network designs related to UE-centric AP clustering, network-centric AP clustering, and self-organizing network architectures (**Chapter 3**, **Chapter 4** and **Chapter 5**, respectively).

- **Track IV:** In the fourth track of this thesis, we develop several innovative frameworks for optimizing different resources of large-scale wireless cell-free network (designed in **Track II** and **Track III**). Specifically, we first investigate the utilization of continuous space actor-critic algorithms in solving the problem of uplink/downlink beamforming over cell-free network. Upon introducing the concept of APs/UEs clustering and network partitioning, we first design several discrete space DRL clustering algorithms that can efficiently cluster APs and/or UEs into a group of subsystems without the need of going through all possible clustering configurations at each time the problem is solved. Second, we develop an innovative hybrid discrete-continuous space DRL models that can jointly scale/partition the cell-free network and conduct uplink/downlink beamforming through an interactive DRL subsystems (**Chapter 3 and Chapter 4**). Furthermore, this thesis track introduces the concept of hierarchical DRL subsystems with mixed DRL-convex optimization problems (**Chapter 5**).

The rest of this thesis is organized as follows. In **Chapter 2**, the proposed D-OMA scheme is presented. **Chapter 3** Introduces the concept of G-CoMP that allows different NOMA clusters to be served by multiple APs through a UE-centric clustering approach. In **Chapter 4**, we introduce a novel network-centric AP clustering approach that scale the overall cell-free network into a group of cooperating distributed antenna systems (DASs) and then solve the problem of AP clustering and uplink beamforming through hybrid DRL design. **Chapter 5** introduces the proposed self-organizing mmWave cell-free network along with the developed hybrid beamforming scheme and hierarchical DRL design. Finally, in **Chapter 6**, we conclude this thesis by some insights about the developed network architectures and algorithms and also pave the road for some important research trends in the future.

Chapter Two

Multiple Access Using Partially Overlapping Spectrum: The D-OMA Scheme

2.1 Introduction

For future massive multiple access, the utilization of the available spectrum through pure OMA schemes will not be sufficient. On the other hand, pure NOMA methods will not have the flexibility to support massive wireless connectivity for devices with different service requirements [3]. Therefore, new multiple access/resource allocation and interference management methods will be required for cell-free networks given the limited spectrum resources. In this chapter, we design a novel multiple access scheme that represents a trade-off between fully orthogonal OFDMA and that of NOMA scheme with fully overlapping spectrum. The reset of this chapter is organized as follows. In **Section 2.2** we present a brief technical comparisons between NOMA and OMA schemes. The technical issues behind large-scale NOMA clusters is discussed in **Section 2.3**. In **Section 2.4**, a detailed technical description of the proposed D-OMA scheme is presented. The concept of (D-OMA)-enabled cell-free network

is presented in **Section 2.5** while **Section 2.6** concludes this work.

2.2 NOMA Vs OMA: Performance and Complexity Perspectives

Orthogonal utilization of spectrum has been used in cellular generations 1G through 5G. Due to the orthogonality among different subcarriers and the relatively high bandwidth separation requirements among them, OFDMA, which is used on 4G networks, may not provide an efficient solution for future generation networks. Therefore, the NOMA technique has been adopted lately by the 3GPP release-16 standards (5G) [5]. The main idea of NOMA is to superimpose many signals at the power domain within the same sub-band and uses SIC at the receiver end to filter out the undesired interfering signals. Using NOMA, every single OMA sub-band can serve multiple devices simultaneously and in this process a higher portion of transmission power is given to those with lower link quality (Fig. 2.1). With

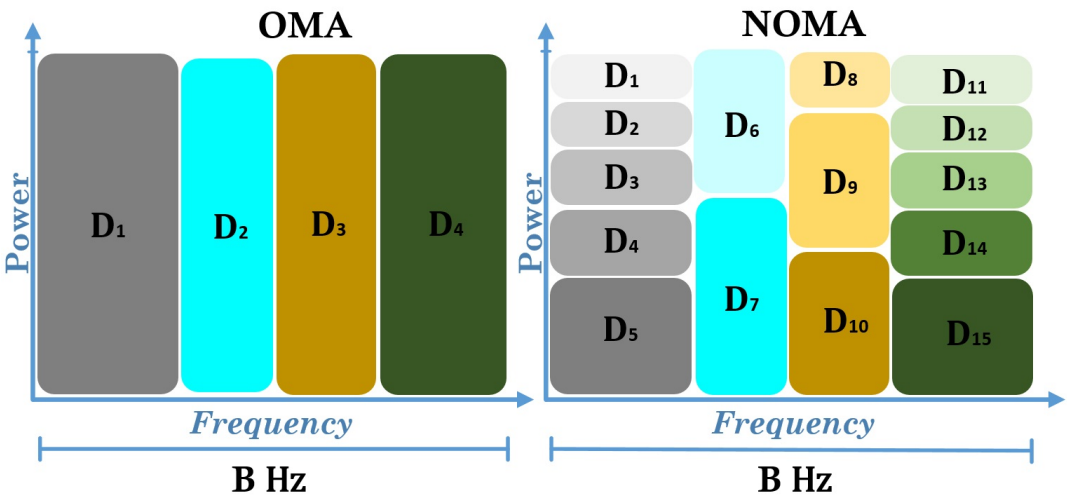

Figure 2.1 An example of OMA Versus NOMA schemes.

NOMA deployment, every sub-band will serve a single NOMA cluster. The devices within a certain cluster will suffer from two types of interference, namely; i) the INI caused by

residual unfiltered NOMA interference signal that is caused by other NOMA devices within the same cluster and the ii) INCI that is caused by using the same sub-band by other nearby clusters. Despite these two interference components, NOMA scheme is found to outperform its OMA counterpart, especially at large NOMA cluster sizes (see **Fig.** 2.2), especially from overall network performance prospective. From a UE prospective however, it's found

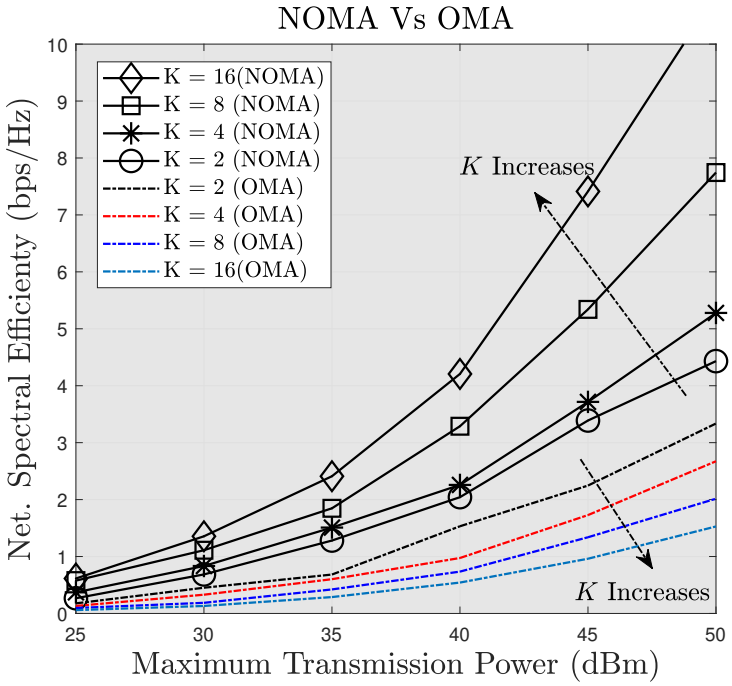

Figure 2.2 Network spectral efficiency: NOMA Vs OMA

that the per-UE spectral efficiency of NOMA scheme decreases as the number of UEs per NOMA cluster increases (see **Fig.** 2.3). The size of a NOMA cluster can be considered as a design parameter to achieve trade-off among several factors, namely; i) the data rate requirements of the UEs, ii) the complexity level at NOMA receivers, iii) the overall power budget per NOMA cluster and iv) the NOMA UE immunity to INI, INCI and SIC-based error propagation.

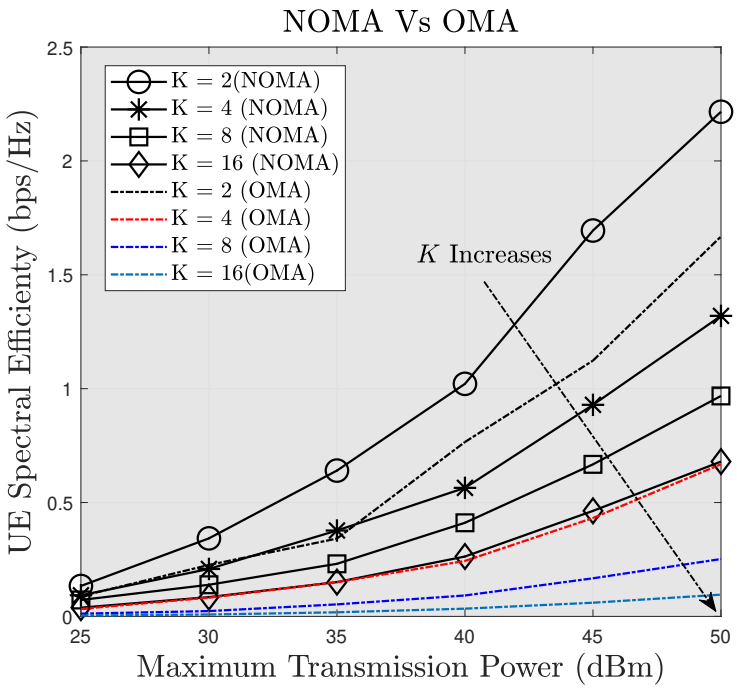

Figure 2.3 UE spectral efficiency: NOMA Vs OMA

2.3 Technical Issues

Small NOMA clusters are preferable when we target low hardware and computational complexity, short processing delay and low SINR requirements. However, smaller NOMA clusters suffers from relatively low spectral efficiency which contradict with the main benefits behind non-orthogonal utilization of spectrum. Large NOMA clusters, on the other hand, provide significantly higher spectrum efficiency than its OMA counterpart as will as NOMA scheme with 2 UEs per NOMA cluster. This comes with the cost of increased hardware and computational complexity due to multilevel SIC operations as well as well as increased processing delay and higher SINR requirements (see **Fig.** 2.4). Motivated by the aforementioned discussion, we developed a new MA scheme that represents a trade-off between OMA and NOMA schemes.

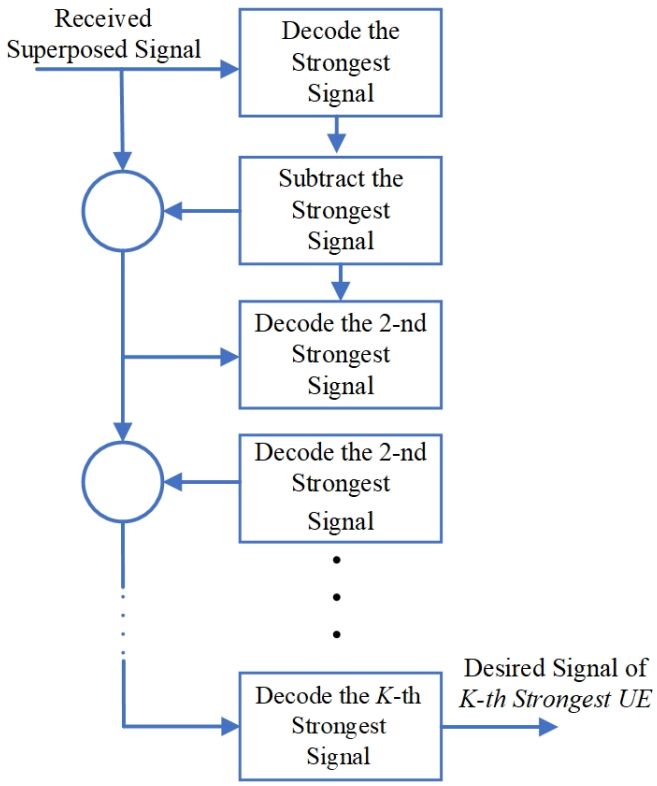

Figure 2.4 K-UEs NOMA cluster.

2.4 Partially Overlapping Spectrum: The D-OMA

The idea of the proposed MA scheme is that adjacent NOMA clusters are allowed to overlap in frequency for a certain fraction, denoted by δ (and hence the name D-OMA). The spectral efficiency acquired by the partial overlapping of NOMA clusters is then utilized to increase the degree of freedom in the size of NOMA cluster. Accordingly, the number of NOMA UEs per cluster can be reduced without affecting the overall network spectral efficiency.

2.4.1 Two-UEs D-OMA

Here, we start by introducing the simplest form of the D-OMA scheme where two UEs are allowed to transmit/receive through a partially overlapping subcarriers (**Fig.** 2.5). For the 2-UEs downlink D-OMA, the signal received at U_1 is first used to decode $\delta\%$ of the interfering

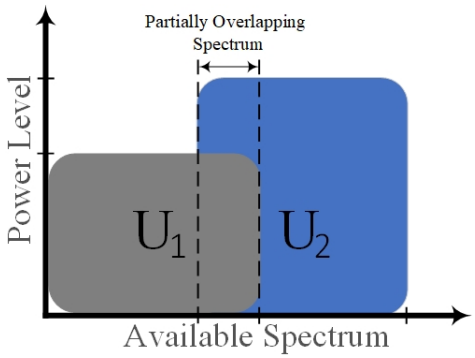

Figure 2.5 Two-UEs D-OMA scheme.

signal of U_2, subtract it from the overall received signal and then decode the signal for U_1. At UE U_2, the $\delta\%$ of U_1 interfering signal is neglected and considered as an untreated interference signal beside the AWGN component. The spectral efficiency gain achieved by 2U-DOMA scheme will be arbitrary small compared to the hardware and complexity requirements. In the next section, we show how powerful the proposed scheme can be when NOMA clusters has arbitrarily large number of UEs per cluster (> 2).

2.4.2 Multiuser D-OMA

The best utilization of the proposed D-OMA scheme can be achieved when different overlapping NOMA clusters are equipped with relatively large number of NOMA UEs (i.e, > 2). Accordingly, for multiuser D-OMA scheme, the overall spectrum will be divided into an group of overlapping sub-bands with each sub-band assigned a certain number of NOMA UEs to form a single NOMA cluster (**Fig. 2.6**). At each NOMA sub-band, SIC operations will be conducted at the receivers side to remove the INI caused by UEs belonging to the same cluster. The INCI component caused by UEs from other clusters will be considered as unfiltered interference component added to UEs along with the AWGN component. However, the amount of INCI component can be easily controlled by changing the overlapping percentage among different subcarriers. Lets consider a single AP downlink D-OMA system

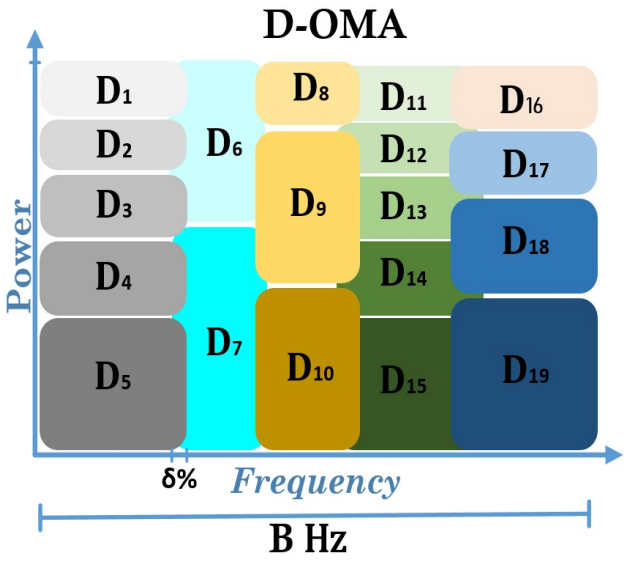

Figure 2.6 An example of the proposed D-OMA scheme.

with K UEs. Furthermore, let the available spectrum is splitted to N subcarriers each act as a NOMA cluster with K_n UEs such that $\sum_{n=1}^{N} K_n = K$. Accordingly, the achieved SINR at the k_n-th UE in the n-th NOMA cluster after applying SIC operations can be given as[1]

$$\gamma_{k_n} = \frac{w_{k_n} P |h_{k_n}|^2}{\underbrace{\sum_{i_n=k_n+1}^{K_n} w_{i_n} P |h_{k_n}|^2}_{\text{INI}} + \underbrace{\sum_{l \in \{n-1, n+1\}} \delta_{nl} \sum_{k_l=1}^{K_l} w_{k_l} P |h_{k_l}|^2}_{\text{INCI}} + \underbrace{N_{k_n}}_{\text{AWGN}}}, \tag{2.1}$$

where δ_{nl} is the amount of spectrum overlapping between the n-th and l-th NOMA clusters. Lets consider a D-OMA system with $K = 2$, $N = 2$ and $K_1 = K_2 = 6$ (**Fig. 2.7**).

It's important to emphasize that the proposed D-OMA scheme offers to performance gain factors as follows:

- Increased in overall network spectrum efficiency due to fitting more NOMA clusters within the same frequency band (**Fig. 2.2**).

[1]For more detailed illustration about UEs ordering and SIC operation, please refer to **Sec. 1.3.2** and enclosed discussion.

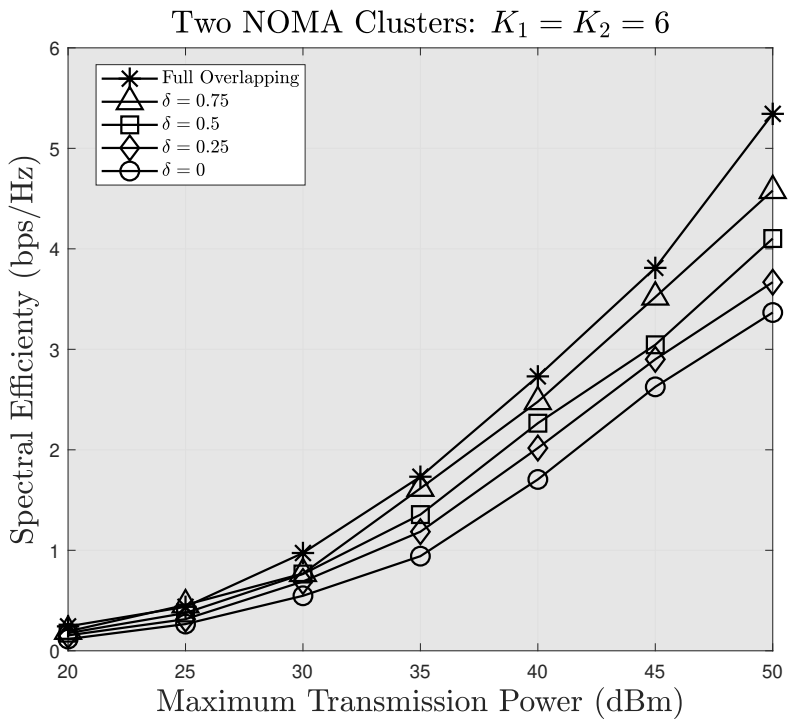

Figure 2.7 Performance Gain of the D-OMA scheme (Eq. (1.4)).

- Reduced hardware and computational complexity due to the ability of reducing the number of UEs per NOMA cluster without affecting the network spectral efficiency.

We, however, showed previously how the per-UE spectral efficiency can significantly degrade as the number of UEs per NOMA cluster increases. This has motivated us to consider different levels of APs cooperation/coordination throughout the rest of work of this thesis.

2.5 NOMA-Enabled Cell-Free Networks

To compensate for the per-UE loss in spectral efficiency caused by large NOMA cluster and the INCI component of the D-OMA scheme, different APs of the network may cooperate to serve all UEs within the network coverage area. Under this assumption, the SINR value at

(2.1) can be rewritten as

$$\gamma_{k_n} = \frac{\sum_{m=1}^{M} w_{k_n m} P |h_{k_n m}|^2}{\underbrace{\sum_{m=1}^{M} \sum_{i_n=k_n+1}^{K_n m} w_{i_n m} P |h_{k_n m}|^2}_{\text{INI}} + \underbrace{\sum_{m=1}^{M} \sum_{l \in \{n-1, n+1\}} \delta_{nl} \sum_{k_l=1}^{K_l} w_{k_l m} P |h_{k_l m}|^2}_{\text{INCI}} + \underbrace{N_{k_n}}_{\text{AWGN}}}, \quad (2.2)$$

where $w_{k_n m}$ is the power fraction assigned to the k_n-th UE from the m-th AP. As an example, we showed how the per-UE rate decreases as the number of UEs per NOMA cluster increases (**Fig.** 2.3). If we allow nearby APs to cooperate to serve all UEs of the same NOMA cluster, the per-UE transmission rate can be improved significantly as shown in **Fig.** 2.8. 2.3). The performance enhancement resulted from APs cooperation can then be used to reduces

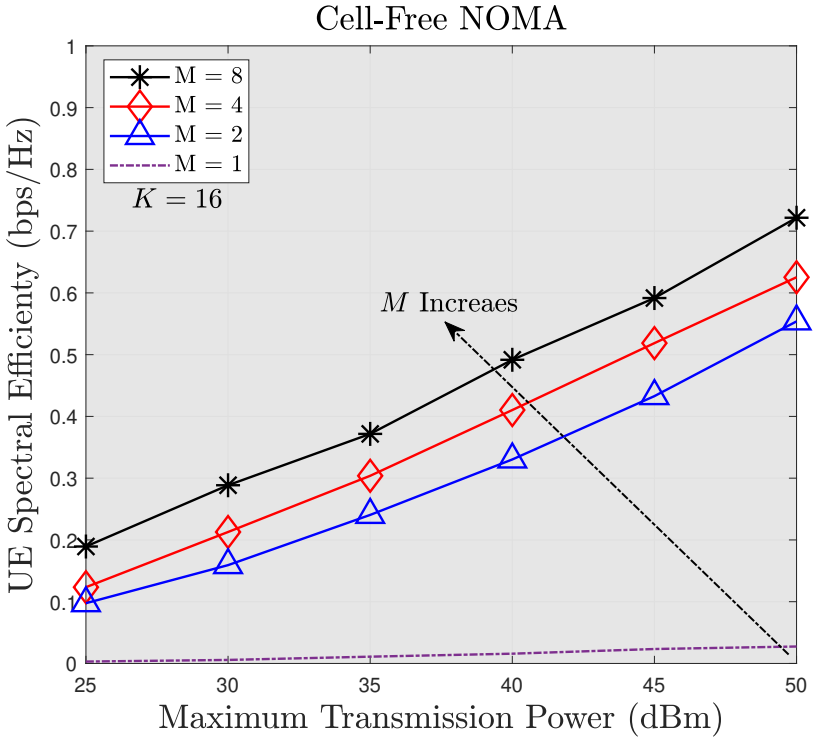

Figure 2.8 Example of performance gain achieved by APs

the number of UEs per NOMA cluster which will decreases hardware and computational complexity of terminal devices significantly.

2.6 Conclusion

In this chapter, we have proposed a novel MA scheme for future B5G/6G communication networks. The proposed scheme represents a trade-off between the OFDMA with its strict orthogonality demand of its bands and that of the NOMA with completely overlapping frequency bands. Furthermore, we showed that by allowing nearby APs to simultaneously serve all active UEs within their coverage area, the per-UE performance can be significantly enhanced. Such a performance enhancement can help in decreasing the number of UEs per NOMA cluster, which will decreases the hardware and computational complexities of terminal devices significantly while maintaining an acceptable performance. In the next three chapters, we propose and design several innovative network architectures that aim to decrease the computational complexity behind centralized processing of signals collected from the set of cooperating APs.

Chapter Three

UE-Centric AP Clustering: The GCoMP Architecture

3.1 Introduction

In **Chapter 2**, we have proposed a new multiple access scheme, the D-OMA, that represents a trade-off between fully orthogonal OFDMA and that of power domain NOMA with full spectrum overlapping. We also showed how the proposed D-OMA scheme add new degree of freedom that enable reducing the number of NOMA UEs at each cluster and hence, significantly decreases the hardware and computational complexities at NOMA terminal UEs. The introduction of additional interference signal caused by NOMA clusters overlapping in the D-OMA (the INCI component) however, may have a direct impact on the per-UE performance. In the following three chapters, we take a step forward by allowing several adjacent APs to simultaneously serve all UEs within their coverage area. Starting by UE-centric clustering approaches, in this chapter we design a novel downlink UE-centric AP clustering scheme at which each UE is served by variable number of APs under NOMA power allocation. The rest of this chapter is organized as follows. **Section 3.2** presents the general system model. The n-th order clustering protocol for the GCoMP-enabled NOMA

is presented in **Section 3.3**. **Section 3.5** presents the outage and the outage capacity of the proposed system. The clustering scheme and the corresponding power allocation method for the proposed system are presented in **Section 3.6**. Numerical results are presented in **Section 3.7** while **Section 3.8** conclude this chapter.

3.2 System Model and Assumptions

Consider a downlink wireless network with M single-antenna APs and K single-antenna UEs that are located at fixed locations within a certain coverage area as shown in Fig. 3.1. All APs are connected to each other through a fast backhaul link and have the ability of collaborate with each other at the baseband and radio levels. This is similar to a C-RAN architecture in which distributed RRHs are connected to a single mega BBU that has the ability of jointly processing signals from different RRHs by dealing with distributed RRHs as a virtual $K \times M$ MIMO system [74]. We also assume that perfect control signaling is possible among the cooperative APs and the distributed UEs within their coverage area due to the existence of fast backhauling links that connect all of the cooperative APs together. We mean by perfect control signaling is that all APs can be communicated from any point within the network coverage area.

Every UE can be served by n APs ($1 \leq n \leq M$), which is referred to as the order of cooperation (or clustering order) and each AP can serve k UEs ($0 \leq k \leq K$). The UEs associated with a AP form a NOMA cluster using the same frequency band and also all the NOMA clusters served by all the APs use the same band. Therefore, every UE is a member of n clusters and not all the clusters need to have the same number of UEs. Note that, with this system model, $n = 1$ corresponds to traditional NOMA (i.e, without CoMP). For NOMA, we assume that every AP applies superposition coding in the power domain in its transmission band and it has its own power budget. Additionally, full CSI is assumed to be available

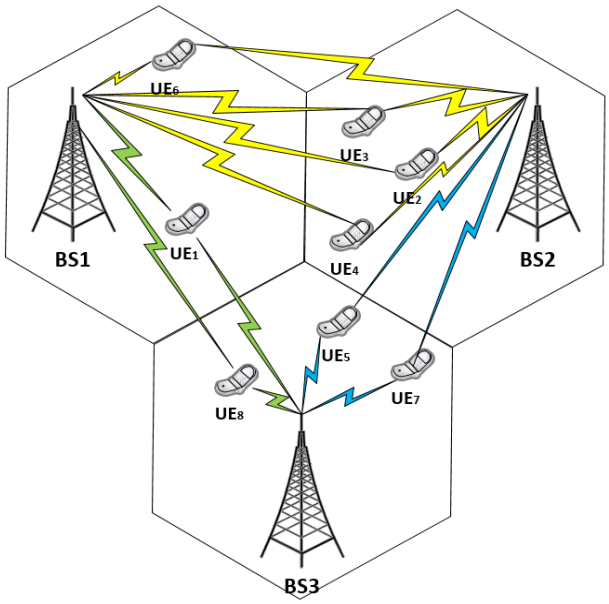

Figure 3.1 Example of the proposed network model (same colors refer to a single NOMA cluster).

at all APs while CSI of every NOMA cluster is assumed to be available at all UEs of that cluster. The downlink channel gain between the m-th AP and the k-th UE is denoted by h_{km}. We assume that channel gains h_{km} are independent but non-identically distributed (i.n.d) and follow complex Gaussian distribution, i.e, $h_{km} \sim \mathcal{CG}(0, \sigma_{km})$, where σ_{km} is the standard deviation of the h_{km}. Thus, the power gain $|h_{km}|^2$ follows an exponential distribution, i.e, $|h_{km}|^2 \sim$ Exp$(1/2\sigma_{km}^2)$. At every AP, perfect SIC operation is assumed such that interference to every UE that is caused by UEs within the same NOMA cluster with higher channel gains are perfectly filtered out. Finally, when compared with a conventional CoMP system, we assume that the frequency reuse factor is always greater than one (e.g. 3, 7). This assumption is justified by the fact that such networks would be practical for crowded urban environments in which distributed APs would be close to each other which will cause severe interference to CoMP cell-centre UEs. Fig. 3.2 illustrates a simple example of different networking models (with $M = 2$ and $K = 3$) when compared to the proposed GCoMP-NOMA scheme. In this figure, different colors are used to show the different frequency bands (channels) used by the different APs and UEs.

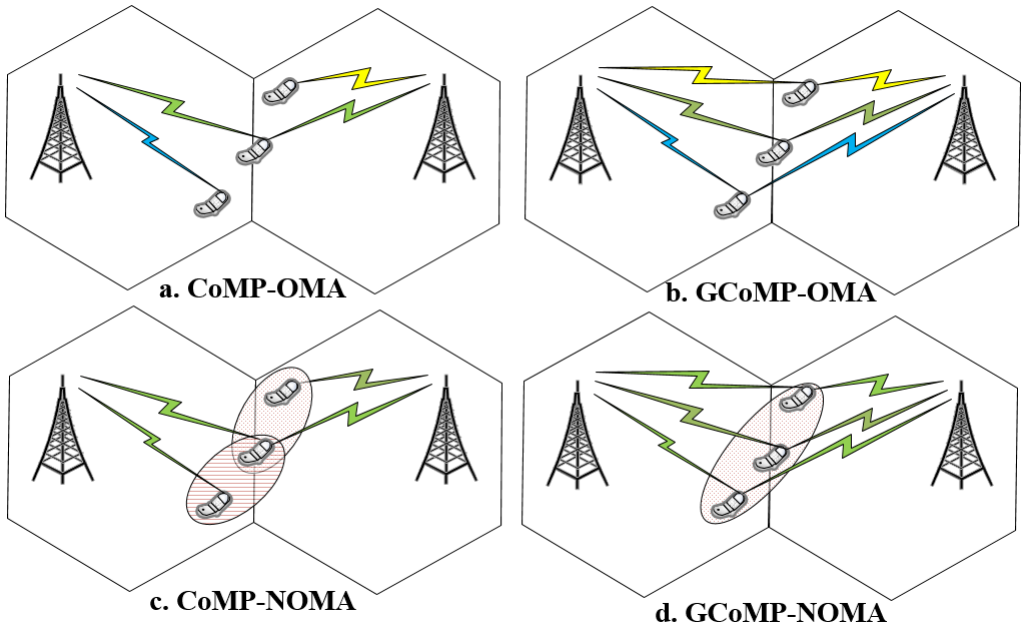

Figure 3.2 Examples of different network models.

3.3 An n-th Order AP Clustering for GCoMP

This is a variable order clustering scheme that enables multiple UEs to utilize a single frequency band with the help of joint cooperation among several connected APs within their vicinity. First, let $\Psi_{n,m}$ denote the set of indices of UEs connected to the m-th AP using the same frequency band, where $n = 1, \ldots, M$ and $m = 1, \ldots, M$. The clustering order n denotes the number of APs connected to every UE at a certain time instant. For example, if a (2-nd)-order clustering is used, every UE within the network will be connected with two serving APs. For the moment, given a certain channel gain matrix of the network $\mathbf{H} \in \mathbb{C}^{K \times M}$, the task is to find the set of UEs connected to the m-th AP under the n-th order clustering, $\Psi_{n,m}$. Here, $\mathbf{H}$ is a matrix containing channel gains from all APs to all UEs within the coverage area of the network.

Algorithm 1 illustrates the proposed method that is used to find the clusters of UEs for every AP under n-th order clustering.

Algorithm 1 : The n-th Order UE-centric clustering protocol.

Require: $\mathbf{H} \in \mathbb{C}^{K \times M}, n, M, K$

$\Psi_{i,m} = \Phi, \forall i = 1, \ldots, n$ and $m = 1, \ldots, M$

for $i = 1 : n$ **do**

  for $k = 1 : K$ **do**

    $m^* = \max_m(\mathbf{H}(k, :))$

    $\Psi_{i,m^*} = \Psi_{i,m^*} \bigcup \{k\}$ and $\mathbf{H}(k, m^*) = -1$

  end for

end for

$\Psi_{n,m} = \bigcup_{i=1}^{n} \Psi_{i,m}, m = 1, \ldots, M$

return $\Psi_{n,\{1,\ldots,M\}}$

The main idea of **Algorithm 1** is than for a network with M APs and K UEs randomly distributed among the network coverage area, every UE will be served by n APs related to the best channel gain while the signals received from other $M - n$ APs is considered as INCI signals. Note that this algorithm has a linear complexity that depends on the dimensionality of the $\boldsymbol{H}$ matrix i.e, M and K. Using the proposed clustering scheme, cooperation among distributed APs is exploited for all UEs (both cell-edge and cell-centre UEs) within a certain network coverage area. Additionally, for a certain frequency band, since every UE is a member of n clusters, NOMA transmission power coefficients for every cluster must be allocated properly such that the desired received signal at a certain UE can be extracted by multilevel SIC operations.

3.4 GCoMP-NOMA Transmission Model

At every NOMA cluster, the m-th AP will transmit $x_m = \sum_{k_m=1}^{|\Psi_{n,m}|} \sqrt{w_{k_m M} P_m} s_{k_m m}$, where $s_{k_m m}$ is the message for the k-th UE (k_m is the ordered index of the k-th UE in $\Psi_{n,m}$ and

k is an arbitrary index for UEs using the same sub-band within a certain network coverage area) with $\mathbb{E}\left[|s_{k_m m}|^2\right] = 1$, P_m is the overall transmission power budget at the m-th AP (assigned for a certain sub-band) and $w_{k_m m}$ is the NOMA power allocation coefficient from the m-th AP to the k-th UE with index k_m in $\Psi_{n,m}$ such that $\sum_{k_m=1}^{|\Psi_{n,m}|} w_{k_m m} \leq 1$. Generally, the values of the coefficients $w_{k_m m},\ \forall\ k_m = 1, \ldots, |\Psi_{n,m}|$ and $m = 1, \ldots, M$ can be derived optimally by considering joint optimization problem for all clusters within the network that uses the same frequency band. A practical technique for optimizing the transmission power coefficients for the proposed GCoMP-enabled NOMA system will be presented later in this chapter.

First, let us define the received signal at the k-th UE as

$$y_k = \sum_{\forall\ i \in \mathbb{S}_{\mathrm{c},k}} h_{ki} x_i + \sum_{\forall\ j \in \mathbb{S}_{\mathrm{nc},k}} h_{kj} x_j + n_k, \tag{3.1}$$

where n_k is the AWGN component at the input of the k-th UE with power spectral density (PSD) N_k and $\mathbb{S}_{\mathrm{c},k}(\mathbb{S}_{\mathrm{nc},k})$ is the set contains the indices of serving (not-serving) APs for the k-th UE. As can be noticed from (3.1), within a certain cluster set $\Psi_{n,m}$, every UE is an element of a random set of clusters (beside $\Psi_{n,m}$). Therefore, every UE in $\Psi_{n,m}$ will have a different inter-cell interference component that must be taken into consideration when establishing power fractions for NOMA transmission. Therefore, let us assume that for the m-th AP with a cluster set $\Psi_{n,m}$, the set of power gains between the m-th AP and all UEs in $\Psi_{n,m}$ are normalized by inter-cell interference caused to every UE within the cluster and then arranged in an ascending order, i.e, $\frac{|h_{1_m m}|^2}{I^1_{\mathrm{INCI}}} \leq \ldots \leq \frac{|h_{k_m m}|^2}{I^m_{\mathrm{INCI}}} \leq \ldots \frac{|h_{|\Psi_{n,m}|m}|^2}{I^{|\Psi_{n,m}|}_{\mathrm{INCI}}}$, where I^k_{INCI} is the inter-NOMA-cluster interference component resulting from the utilization of the same frequency band by other APs and is defined as $I^k_{\mathrm{ICI}} = \sum_{\forall\ i \in \mathbb{S}_{\mathrm{nc},k}} \Phi_i |h_{ki}|^2$ with $\Phi_i = P_i \sum_{j_i=1}^{|\Psi_{n,i}|} w_{j_i i}$[1]. Accordingly, for successful SIC operation at every UE, the transmission power allocation coefficients must be allocated such that the overall received power of the

[1]Remember that $\sum_{j_i=1}^{|\Psi_{n,i}|} w_{j_i i} = 1$, however, we keep it in the analysis to illustrate the operations in the proposed model.

desired signal has a distinct power level from other combinations of undesired inter-NOMA interference signals (INI) with a certain power gap that depends on the sensitivity of the SIC hardware unit. Accordingly, the design of SIC unit at every UE, which will require a multi-level SIC operation, will be challenging[2].

We assume that the power gains between the k-th UE and different serving APs are ordered such that $|h_{k1}|^2 \geq \ldots \geq |h_{kM}|^2$. Hence, the SINR at the input of the k-th UE is given by

$$\gamma_k = \frac{\sum_{i=1}^{n} \Lambda_i^k |h_{ki}|^2}{\underbrace{\sum_{j=1}^{n} \Delta_j^k |h_{kj}|^2}_{I_{\text{INI}}} + \underbrace{\sum_{l=n+1}^{M} \Phi_l |h_{kl}|^2}_{I_{\text{INCI}}} + \underbrace{N_k}_{\text{AWGN}}}, \tag{3.2}$$

where $\Lambda_i^k = w_{k_i i} P_i$, $\Delta_j^k = P_j \left(\sum_{t_j=k^*+1}^{|\Psi_{n,j}|} w_{t_j j} + \Theta \right)$, $\Phi_l = P_l \sum_{t_l=1}^{|\Psi_{n,l}|} w_{t_l l}$, $k^* = \max_k (k_m)$, I_{INI} denotes the unfiltered INI component within the same NOMA cluster and Θ is the residual INI interference caused by the multi-level SIC. The level of SIC is the number of required demodulation-subtraction operations within the SIC unit of a certain UE. Generally, Θ lies between $\Theta = 0$ (when a certain UE has an identical ordering rank over all associated clusters) and $\Theta = \sum_{j=1}^{n} \left(\sum_{l_j=k_j+1}^{k_j+n} w_{l_j j} \right) P_j |h_{kj}|^2$ (when a certain UE has different ordering over all associated clusters). However, the value of Θ can be set to zero with a proper sub-optimal design of the proposed system as will be discussed in the subsequent section.

Under this communication setup, the k-th UE will detect the first $k^* - 1$ signals using a first round of SIC, and then store the signal components that contain its desired signal for a next level of SIC procedure (multi-level SIC is required only when $\Theta > 0$). The available SINR at the input of the k-th UE when detecting the signal component of the δ-th UE

[2]While the purpose of this part of the work is to give an insight on the performance of multiuser NOMA under the proposed GCoMP-enabled NOMA model, a practical low-complexity GCoMP-enabled NOMA model will be presented along with the power allocation method later in this chapter.

$(1 \leq \delta \leq k^*)$ is given by

$$\gamma_{\delta \to k} = \frac{\sum_{l=1}^{n} \Lambda_l^{\delta} \frac{|h_{kl}|^2}{\sum_{t=n+1} \Phi_t |h_{kt}|^2 + N_k}}{\sum_{j=1}^{n} \Delta_j^{\delta} \frac{|h_{kj}|^2}{\sum_{t=n+1} \Phi_t |h_{kt}|^2 + N_k} + 1}. \tag{3.3}$$

The difference between (3.2) and (3.3) appears at $\Lambda_j^{\delta} = w_{\delta,j} P_j$ and $\Delta_j^{\delta} = P_j \sum_{i_j=\delta+1}^{|\Psi_{n,j}|} w_{i_j j}$ and depends on the number of signals to be decoded and subtracted from the overall received signal. However, it is clear that $\gamma_{\delta \to k}$ is an increasing function of δ.

Illustrative Example

To illustrate the main idea behind the GCoMP-enabled NOMA, we present a simple example with three APs ($M = 3$) and eight UEs ($K = 8$) that operate under 2-nd order clustering (i.e, $n = 2$). Let us first assume a single realization of the channel gain matrix $\mathbf{H}$ as

$$|\mathbf{H}|' = \begin{bmatrix} & \text{UE}_1 & \text{UE}_2 & \text{UE}_3 & \text{UE}_4 & \text{UE}_5 & \text{UE}_6 & \text{UE}_7 & \text{UE}_8 \\ \text{AP}_1 & 0.9 & 1 & 0.45 & 0.7 & 0.39 & 1.2 & 0.38 & 0.89 \\ \text{AP}_2 & 0.1 & 0.98 & 0.35 & 0.65 & 0.93 & 0.72 & 0.91 & 0.3 \\ \text{AP}_3 & 0.43 & 0.78 & 0.21 & 0.19 & 0.95 & 0.31 & 0.99 & 0.56 \end{bmatrix}. \tag{3.4}$$

Applying **Algorithm 1** to $\mathbf{H}$, the decreasingly ordered cluster elements of $\Psi_{2,1}$, $\Psi_{2,2}$ and $\Psi_{2,3}$ are given in Table 3.1. For example, the received signal at UE_1 is given by

Table 3.1 Cluster elements under 2-nd-order clustering

$\Psi_{2,1}$	UE_1	UE_2	UE_3	UE_4	UE_6	UE_8
$\Psi_{2,2}$	UE_2	UE_3	UE_4	UE_5	UE_6	UE_7
$\Psi_{2,3}$	UE_5	UE_7	UE_1	UE_8	NA	NA

$$\begin{aligned} y_1 = &\left(\sqrt{w_{11}P_1}h_{11} + \sqrt{w_{13}P_3}h_{13}\right) s_1 + \left(\sqrt{w_{81}P_1}h_{11} + \sqrt{w_{83}P_3}h_{13}\right) s_8 + \sqrt{w_{21}P_1}h_{11}s_2 \\ &+ \sqrt{w_{31}P_1}h_{11}s_3 + \sqrt{w_{41}P_1}h_{11}s_4 + \sqrt{w_{51}P_1}h_{11}s_5 + \sqrt{w_{61}P_1}h_{11}s_6 + \sqrt{w_{71}P_1}h_{11}s_7 + n_1. \end{aligned} \tag{3.5}$$

To extract s_1 from y_1, a set of SIC operations have to be conducted such that the overall power of the desired signal s_1 is separated by a certain SIC receiver sensitivity power gap (P_s) from other undesired signal components. The number of signals to be decoded before extracting s_1 will depend on the set of constraints required to extract s_2 through s_8 at UE_2 through UE_8. To further illustrate, let us consider the case of full-order clustering for which every UE will be connected with all serving APs. Accordingly, for the given $\mathbf{H}$, the cluster sets are given in Table 4.2. Therefore, the received signal at UE_1 is given by

Table 3.2 Cluster sets under full order clustering

$\Psi_{3,1}$	UE_1	UE_2	UE_3	UE_4	UE_6	UE_8	UE_5	UE_7
$\Psi_{3,2}$	UE_2	UE_3	UE_4	UE_5	UE_6	UE_7	UE_1	UE_8
$\Psi_{3,3}$	UE_5	UE_7	UE_1	UE_8	UE_2	UE_3	UE_4	UE_6

$$
\begin{aligned}
y_1 = &\left(\sqrt{w_{11}P_1}h_{11} + \sqrt{w_{12}P_2}h_{12} + \sqrt{w_{13}P_3}h_{13}\right) s_1 \\
&+ \left(\sqrt{w_{21}P_1}h_{11} + \sqrt{w_{22}P_2}h_{12} + \sqrt{w_{23}P_3}h_{13}\right) s_2 \\
&+ \left(\sqrt{w_{31}P_1}h_{11} + \sqrt{w_{32}P_2}h_{12} + \sqrt{w_{33}P_3}h_{13}\right) s_3 \\
&+ \left(\sqrt{w_{41}P_1}h_{11} + \sqrt{w_{42}P_2}h_{12} + \sqrt{w_{43}P_3}h_{13}\right) s_4 \\
&+ \left(\sqrt{w_{5,1}P_1}h_{11} + \sqrt{w_{52}P_2}h_{12} + \sqrt{w_{53}P_3}h_{13}\right) s_5 \\
&+ \left(\sqrt{w_{61}P_1}h_{11} + \sqrt{w_{62}P_2}h_{12} + \sqrt{w_{63}P_3}h_{13}\right) s_6 \\
&+ \left(\sqrt{w_{71}P_1}h_{11} + \sqrt{w_{72}P_2}h_{12} + \sqrt{w_{73}P_3}h_{13}\right) s_7 \\
&+ \left(\sqrt{w_{81}P_1}h_{11} + \sqrt{w_{82}P_2}h_{12} + \sqrt{w_{83}P_3}h_{13}\right) s_8 + n_1.
\end{aligned}
\tag{3.6}
$$

In order to be able to extract s_1 from y_1, a set of SIC operations will be conducted. Specifically, for successful SIC operations, a subset of the following constraints must be satisfied:

$$
|\mathcal{N}_i - \sum_{\forall j \in \mathcal{A}, i \notin \mathcal{A}} \mathcal{N}_j| \ \geq P_s, i = 1, \ldots, 8 \text{ and } j = 1, \ldots, 8, \tag{3.7}
$$

where $\mathcal{A}$ is the set of indices $\{1, \ldots, K\}$ and $\mathcal{N}_k = w_{k1}P_1|h_{11}|^2 + w_{k2}P_2|h_{12}|^2 + w_{k3}P_3|h_{13}|^2$. Note that a set of different constraints has to be satisfied at every UE and transmission power for NOMA must be allocated such that all constraints at all UEs are satisfied. Furthermore, the number of SIC operations required at every UE will depend on its overall power level (weight) compared to other distinct signals (the more the power the fewer will be the number of required SIC operations). Multilevel SIC operation can be outfitted to work in a single-level SIC receiver. However, such a design would result in high complexity since the number of SIC operations to be conducted will be large. On the other hand, to minimize the processing delay, a bank of n SIC receivers can be used at every UE to extract the desired signal, this will roughly decrease the processing delay by a factor of n (n is the number of cooperative APs [clustering order]) at the expense of increasing complexity.

For simplicity of analysis, we first adopt a constant power allocation scheme for NOMA under the assumption of the availability of perfect SIC. However, the effect of imperfect estimation of CSI is evaluated at subsequent section. In particular, the m-th AP applies a constant power allocation for its own cluster members served through NOMA based on the following relation:

$$w_{i_m m} = \begin{cases} \frac{1}{2^i} & i = 1, \ldots, |\Psi_{n,m}| - 1, \\ w_{|\Psi_{n,m}|-1,m} & i = |\Psi_{n,m}|. \end{cases} \tag{3.8}$$

Note that for this constant power allocation scheme, the condition $\sum_{i_m=1}^{|\Psi_{n,m}|} w_{i_m m} = 1$ holds. It also follows the NOMA concept by allocating higher power portion to UEs with lower channel gains. Besides, the notion of allocating higher power fractions to NOMA UEs with weaker channel gains is also satisfied. Note that, the optimal power allocation scheme for a modified practical GCoMP-enabled NOMA system will be presented in Section 3.6.

3.5 Outage and Capacity Performance

In this section, a closed-form expression for the outage probability for a UE under the GCoMP-enabled NOMA scheme is derived. Generally, the k-th UE will be in outage if it fails to successfully decode at least one of the k-th higher weight signal components. This can be mathematically expressed as[3]

$$\mathrm{P}_{\mathrm{out}}^{k} = 1 - \mathrm{P}\left(\bigcap_{\delta=1}^{k} \mathrm{E}_{\delta\to k}^{\mathrm{c}}\right), \tag{3.9}$$

where $\mathrm{E}_{\delta\to k}$ is the event that the k-th UE has failed to decode the δ-th signal component under a certain performance requirements and $\mathrm{E}_{\delta\to k}^{\mathrm{c}}$ is the complement of $\mathrm{E}_{\delta\to k}$. This can be mathematically expressed as

$$\mathrm{E}_{\delta\to k} \triangleq \begin{cases} \mathrm{P}\left(\gamma_{\delta\to k} \leq \gamma_{\mathrm{th}}^{\delta}\right) & \delta = 1, \\ \mathrm{P}\left(\gamma_{\delta\to k} \leq \gamma_{\mathrm{th}}^{\delta} \mid \gamma_{\delta-1\to k} \leq \gamma_{\mathrm{th}}^{\delta-1}\right) & \delta > 1, \end{cases} \tag{3.10}$$

where $\gamma_{\mathrm{th}}^{\delta} = 2^{\bar{R}_{\delta}} - 1$. The value $\bar{R}_{\delta}$ is the minimum transmission rate required by the δ-th UE (assuming that every UE has identical ordering over all connected clusters). For simplicity of analysis, we will assume that all UEs have the same rate requirements, i.e, $\gamma_{\mathrm{th}}^{\delta} = \gamma_{\mathrm{th}}$. Deriving a closed-form expression for (3.9) is possible, however, the final expression is found to be complicated and difficult to be programmed. To simplify the analysis, we first calculate the average INCI at the m-th UE and substitute it into (3.3). **Theorem** 1 defines the average value of I_{INCI}^{k} for the proposed GCoMP-enabled NOMA scheme.

Theorem 1. *The average inter-NOMA-cluster interference at the k-th UE under the GCoMP-enabled NOMA scheme is given by*

$$\bar{I}_{INCI}^{k} = \sum_{i_{n+1},\ldots,i_K}^{\{1,2,\ldots,M\}} \prod_{l=n+1}^{M} \frac{\lambda_{ki_l}}{\sum_{d=n+1}^{l} \lambda_{ki_d} + \sum_{q=1}^{n} \lambda_{ki_q}} \left(\sum_{t=n+1}^{K} \frac{t}{\sum_{q=n+1}^{t} \lambda_{ki_q} + \sum_{q=1}^{n} \lambda_{ki_q}} \right), \tag{3.11}$$

[3]Here, we have assumed that the higher levels of SIC operations are conducted ideally such that Θ is set to zero. This is an assumption that complies with the practical design of the proposed system as will be shown later in this chapter.

where $\lambda_{kt} = 1/2\Phi_t\sigma_{kt}$ *and* $\{i_{n+1}, \ldots, i_M\}$ *are distinct indices that take values from* $\{1, \ldots, M\}$.

Proof. See **Appendix A**. ■

Remark: Theorem 1 considers only the ordering of APs with respect to any arbitrary UE at $\Psi_{n,\{1,\ldots,M\}}$. This is true since $\bar{I}^k_{\text{INCI}}$ is constant for every UE even when the k-th UE is decoding the δ-th UE signal ($\delta < k^*$).

Substituting $\bar{I}^k_{\text{INCI}}$ into (3.3), the outage probability of (3.9) can be expressed as

$$P^k_{\text{out}} = \mathbb{P}\left(\sum_{i=1}^{n} z_{ki} \leq \frac{\gamma_{\text{th}} P}{\max_\delta\left(\Lambda^\delta - \gamma_{\text{th}}\Delta^\delta\right)}\right), \tag{3.12}$$

where $z_{ki} \sim \text{Exp}(\alpha_{ki})$, $\alpha_{ki} = \left(\bar{I}^k_{\text{INCI}} + N_k\right)/2P\sigma^2_{ki}$ and we have assumed that all APs are transmitting using the same maximum transmission power budget, i.e, $P_m = P$, $\forall\ m = 1, \ldots, M$. Additionally, the maximum value of $\left(\Lambda^\delta - \gamma_{\text{th}}\Delta^\delta\right)$ changes at every time slot based on the instantaneous CSI and the NOMA transmission power allocation method. Accordingly, the probability of outage of the k-th UE under the GCoMP-enabled NOMA scheme is defined in **Theorem** 2.

Theorem 2. *The probability of outage of the* k*-th UE under the* (n*-th*)*-order clustering of the GCoMP-enabled NOMA scheme is given by*

$$P^{(n)}_{out} = \sum_{k=1}^{\Omega}\sum_{S_k}\prod_{l=1}^{k} F^{(n)}_{\gamma_{\kappa_l}}\left(\gamma'_{th}\right)\prod_{q=k+1}^{\Omega}\left[1 - F^{(n)}_{\gamma_{\kappa_q}}\left(\gamma'_{th}\right)\right], \tag{3.13}$$

where the summation extends over all permutations $(\kappa_1, \ldots, \kappa_\Omega)$ *of* $1, \ldots, \Omega$ *for which* $\kappa_1 < \cdots < \kappa_k$, $\kappa_{k+1} < \cdots < \kappa_\Omega$ *and* $\Omega = \max_m|\Psi_{n,m}|$ *such that* $\{\Psi_{n,m}|k \in \Psi_{n,m}\}$. *The CDF* $F^{(n)}_{\gamma_\kappa}(\gamma)$ *is given by*

$$F^{(n)}_{\gamma_\kappa}(\gamma) = \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} J_1(\kappa, i)\left[\sum_{t_1=1}^{n}\frac{\eta^\kappa_{t_1}}{\rho^\kappa_{t_1}}\left(1 - e^{-\rho^\kappa_{t_1}\gamma}\right) - \sum_{h_1=1}^{M-n}(-1)^{h_1}\sum_{j_1\leq\ldots\leq j_{h_1}}^{\{n,\ldots,M\}}\sum_{t_2=1}^{n}\frac{\eta^\kappa_{t_2}}{\rho^\kappa_{t_2}}\left(1 - e^{-\rho^\kappa_{t_2}\gamma}\right)\right], \tag{3.14}$$

where $J_1(\kappa, i) = \left(\prod_{q=1}^{n} \frac{\alpha_{\kappa,iq}}{q}\right)$, $\gamma_{th}^{'} = \frac{\gamma_{th}P}{\max_{\delta}\left(\Lambda^{\delta} - \gamma_{th}\Delta^{\delta}\right)}$ *and* $\eta_{t_1}^{\kappa}$, $\rho_{t_1}^{\kappa}$, $\eta_{t_2}^{\kappa}$ *and* $\rho_{t_2}^{\kappa}$ *are defined in Appendix B.*

Proof. See **Appendix B**. ■

Remark: The effect of cooperation among APs within GCoMP-enabled NOMA scheme appears in terms of the increased SINR and decreased I_{INCI}^{k} per UE. However, a better performance enhancement can be achieved by optimizing NOMA coefficients for all APs simultaneously as one matrix, as will be discussed in a subsequent section. One particular case of significant importance is that when all APs within a certain geographical area cooperate to serve a set of UEs in their vicinity using the same sub-band. This can be considered as a full-order clustering of the proposed scheme, i.e, $n = M$. **Corollary 1** presents a simpler expression of this particular scenario under the assumption of i.i.d. channel gains.

Corollary 1. *Under full-order clustering ($n = M$), with i.i.d. channel gains, the probability of outage for the GCoMP-NOMA scheme given in* ***Theorem 2*** *reduces to*

$$P_{out}^{(M)} = \sum_{k=1}^{K} \binom{K}{k} \gamma(M, \alpha\gamma_{th'})^{k} \left(1 - \gamma(M, \alpha\gamma_{th}^{'})\right)^{Q}, \tag{3.15}$$

where $\alpha = N/2P\sigma^2$ $(\bar{I}_{INCI} = 0)$, $Q = K - k$, $\gamma(x, y)$ *is the normalized lower incomplete gamma function [76, Eq. 6.5.2] and* $\Gamma(x)$ *is the gamma function.*

Proof. This Corollary can easily be proven by repeating the same procedure of **Theorem 2** under the given assumptions. ■

Furthermore, to obtain more insights about the performance of the proposed system in terms of the diversity order and coding gain, **Corollary 2** presents an approximate expression of the probability of outage of the k-th UE under full-order clustering at the high SINR regime.

Corollary 2. *Under full-order clustering with i.i.d. channel gains and high SINR regime* ($\bar{\gamma} \to \infty$)*, the probability of outage of the GCoMP-enabled NOMA scheme can be expressed as*

$$P_{out}^{(M)} \approx \left(\left(\frac{M\Gamma(M)}{K(\gamma_{th}^{'})^{M}} \right)^{1/K} \bar{\gamma} \right)^{-M}, \tag{3.16}$$

where $\bar{\gamma} = 2P\sigma^2/N$.

Proof. This can be proven by utilizing the series representation of $\gamma\left(M, \alpha\gamma_{\text{th}}^{'}\right)$ in [76, Eq. 6.5.29], substituting in (3.15), and then taking the first term (the dominant term). ■

Remarks:

- At high SINR, $P_{\text{out}}^{(M)} \approx (G_c\bar{\gamma})^{-G_d}$, where G_c and G_d are the coding gain and diversity order, respectively [77]. Accordingly, under full-order clustering we have $G_c = \left(M\Gamma(K)/K(\gamma_{\text{th}}^{'})^{M}\right)^{1/M}$ and $G_d = M$.
- It can be noticed from **Corollary 2** that, besides $\gamma_{\text{th}}^{'}$, the outage performance of the system is significantly affected by the number of NOMA UEs used per cluster (K) in terms of the decreased coding gain. This negative impact caused by increasing the NOMA cluster size can be significantly annihilated by increasing M (when $M \to \infty$, G_c approaches 1).

Outage Capacity: We evaluate the achievable transmission rate per UE of the proposed GCoMP-enabled NOMA. However, a closed-form expression of the ergodic capacity of the proposed system is found to be very complicated and does not carry any significant insights. Nevertheless, we evaluate the so called ϵ-*outage capacity* under the high SINR regime, where ϵ is the maximum allowable outage to achieve a capacity of C_ϵ [78].

Proposition 1. *Under full-order clustering with i.i.d. channel gains and high SINR regime* ($\bar{\gamma} \to \infty$)*, the* ϵ*-outage capacity for a UE under GCoMP-enabled NOMA is approximated by*

$$C_\epsilon \approx \log_2\left(1 + \sqrt[M]{\frac{\epsilon M\Gamma(M)}{K}}\bar{\gamma}\right). \tag{3.17}$$

Proof. This can be directly proven by using the definition of ϵ-*outage capacity* in [78] and utilizing the approximate outage expression from **Corollary 2**. ■

Imperfect CSI: In the analyses presented above, the assumption of the availability of perfect CSI at all APs was made to simplify the analysis to obtain some key insights to system performance. However, in practical scenarios, the estimation of CSI may not be perfect.Since imperfect CSI leads to non-optimal ordering of UEs within clusters and adds an additional self-interference component at the receiver output, this can degrade the system performance significantly. We will asymptotically evaluate the effect of imperfect CSI on overall outage performance under the assumption of optimal UE ordering at each cluster.

Corollary 3 states the effect of imperfect CSI on the asymptotic probability of outage.

Corollary 3. *Under full-order clustering, with imperfect CSI and i.i.d. channel gains, in the high SINR regime ($\bar{\gamma} \to \infty$), the probability of outage for GCoMP-NOMA can be approximated as*

$$P_{out}^{(M)} \approx \left(\left(\frac{M\Gamma(M)}{K(\gamma_{th}^{'})^{M}} \right)^{1/M} \frac{\chi^2 N}{\sum_{m=1}^{M} (1-\chi^2)\,\hat{N} + N} \bar{\gamma} \right)^{-M}, \tag{3.18}$$

where $\chi \in [0,1]$ is a measure of the accuracy of the channel estimation and $\hat{N}$ is the PSD of a complex Gaussian random variable that expresses the estimation error.

Proof. Under imperfect CSI, the estimated channel between the m-th APs and the k-th UE can be modeled as [79]

$$\hat{h}_{km} = \chi_m h_{km} + \sqrt{1-\chi_m^2}\mathcal{I}_{km}, \tag{3.19}$$

where $\chi_m \in [0,1]$ is a measure of the accuracy of the channel estimation at the m-th AP and $\mathcal{I}_{km}$ is a complex Gaussian random variable independent of h_{km} having zero-mean and variance $\hat{N}_m$ with $|\sqrt{1-\chi^2}\mathcal{I}_{km}| \sim \text{Exp}\left(0, (1-\chi^2)\,\hat{N}_m\right)$. At every AP, channel estimation errors $\{\sqrt{1-\chi^2}\mathcal{I}_{km}\}_{k=1,\ldots,K}^{m=1,\ldots,M}$ will be treated as a self-interference component at the output

of the receiver decoder. Accordingly, (3.2) can be rewritten as

$$\gamma_k = \frac{\sum_{i=1}^{M} \Lambda_i^k \chi_m^2 |h_{ki}|^2}{\sum_{j=1}^{K} \left(\Delta_j^k \chi_j^2 |h_{kj}|^2 + (1 - \chi_j^2) \hat{N}_j \right) + N_k}, \tag{3.20}$$

Then, after following the same derivation steps that led to Corollary 2, and assuming that all APs has identical channel estimation accuracy (i.e, $\chi_m = \chi, m = 1, \dots M$), Eq. (3.20) is confirmed. ■

3.6 Power Allocation for GCoMP-Enabled NOMA

In this section, a practical clustering method for GCoMP-NOMA scheme is presented and an optimal transmission power allocation model for the proposed system is developed.

3.6.1 A Low-Complexity Full-Order Clustering Scheme

The exact model discussed in **Section 3.3** is considered as the optimal scheme for GCoMP-NOMA. However, this model assumes that any UE may have a different ordering w.r.t. different serving APs. This will result in a relatively high complexity at a NOMA receiver, since in the worst case scenario (i.e, different ordering for every UE at different APs), the set of constraints for successful SIC operation will increase significantly which will be reflected negatively on the SIC unit design. Another issue of the optimal GCoMP-enabled NOMA is that it requires a complicated scheduling algorithm that first assigns the set of signals to be decoded at every power level and then allocates power fractions for these signals. This should be conducted considering the received signals of all UEs.

To solve this problem, we propose that every UE has the same ordering over all related clusters. This can be achieved by defining a global channel-quality-based metric for every UE that takes all links between every UE and all connected APs into consideration. For simplicity of the following analysis, we will focus on the model of full-order clustering (i.e,

$n = M$) in which all APs cooperate to serve a set of K UEs simultaneously (Cell-free scenario). **Algorithm 2** shows the proposed sub-optimal clustering protocol that finds a single cluster which contains the K UEs served by M APs.

Algorithm 2 : Sub-optimal M-th order clustering protocol.

Require: $\mathbf{H} \in \mathbb{C}^{K \times M}, M, K$

1: $h_{\text{norm}}(k) = \sum_{m=1}^{M} |H(k,m)|^2, \forall\ k = 1, \ldots, K$

2: $\Psi_M = \Phi$

3: **for** $i = 1 : K$ **do**

4: $\quad k^* = \min_k \left(h_{\text{norm}}(k)\right)$

5: $\quad \Psi_M = \Psi_M \bigcup \{k^*\}$ and $h_{\text{norm}}(k^*) = \infty$

6: **end for**

The main idea of this method is to produce a 'global' cluster vector that contains the ordered indices of the entire set of K UEs. Specifically, the norm of the gain vector of any arbitrary UE and all connected APs is utilized as the global ordering metric to find the cluster members. Similar to **Algorithm 1**, the complexity of **Algorithm 2** will be linearly dependent on the network size i.e, M and K. It is important to mention that using the exact model under n-th order clustering, the number of SIC operations required at the k-th UE falls within the integer interval $[0,\ \sum_{i=1}^{n} |\Psi_{n,m_i}| + 2n - 3]$, where m_i are the indices of APs associated with the k-th UE. On the other hand, using the low-complexity scheme, the number of SIC operations required at the k-th UE falls within the integer interval $[0,\ |\Psi_n| - 1]$. Note that $\sum_{i=1}^{n} |\Psi_{n,m_i}| + 2n - 3 \geq |\Psi_n| - 1, \forall n$ with equality holding when every UE have identical ordering w.r.t. every serving AP. Note that the price paid for reducing the number of required SIC operations per UE is a reduction on the average per-UE performance. This can be inferred by the fact that using the norm of UE channel gain w.r.t. all serving APs reduces the performance of NOMA scheme, since some APs may consider UEs with

strong channel gain as a weak UEs and vise versa[4].

After finding Ψ_M, the goal now is to formulate and solve the optimization problem to determine the transmission power coefficients ($w_{km}, \forall\ k = 1, \dots K$ and $m = 1, \dots M$) for the transmitted power from all cooperating APs to the set of UEs in Ψ_M. Based on the proposed system model, the norm metric of the UEs will be ordered such that $||\mathbf{H}(1, [1, \dots, M])||_2 \leq ||\mathbf{H}(2, [1, \dots, M])||_2 \leq \dots \leq ||\mathbf{H}(K, [1, \dots, M])||_2$. Note that when lower order clustering is used (i.e, $n < M$), the link quality of the k-th UE should be divided by the $\bar{I}^k_{\text{INCI}}$. Accordingly, the optimization problem can be formulated as

$$
\begin{aligned}
\mathbf{P}_1 : \max_{w_{km}} \quad & \sum_{k=1}^{K} \log_2 \left(1 + \frac{\sum_{m=1}^{M} w_{km}\gamma_{km}}{\sum_{m=1}^{M} \left(\sum_{j=k+1}^{K} w_{jm} \right) \gamma_{km} + 1} \right) \\
& \text{Subject to:} \\
\mathbf{C}_1 : & \log_2 \left(1 + \frac{\sum_{m=1}^{M} w_{km}\gamma_{km}}{\sum_{m=1}^{M} \left(\sum_{j=k+1}^{K} w_{jm} \right) \gamma_{km} + 1} \right) \geq R_k, \\
\mathbf{C}_2 : & \sum_{m=1}^{M} \left(w_{\delta_l m} - \sum_{i=\delta_l+1}^{l} w_{i,m} \right) \gamma_{lm} \geq P_s, \\
\mathbf{C}_3 : & \sum_{k=1}^{K} w_{km} \leq 1,\ \forall\ m = 1, \dots, M, \\
& \forall k, \forall\ \delta_l = 1, \dots, l-1 \text{ and } l = 2, \dots, K,
\end{aligned}
\tag{3.21}
$$

where $\gamma_{km} = P_M |h_{km}|^2 / N_k$, R_k is the minimum required normalized transmission rate for a UE and is represented by the condition $\mathbf{C}_1$, $\mathbf{C}_2$ refers to the set of $\sum_{l=2}^{K}(l-1) = \frac{K(K-1)}{2}$ conditions required for successful SIC operation with receiver sensitivity of P_s, and $\mathbf{C}_3$ represents the set of K conditions related to the maximum power budget per AP. The optimization problem $\mathbf{P}_1$ in (4.15) is a multiuser sum-rate maximization problem in an interference-limited environment. In general, these type of problems are non-convex due to the existence of dependent variables at the denominator of the SINR, which creates a random convex-concave

[4]Note that an exact evaluation of the performance reduction will require the derivation of optimal power allocation scheme for the exact model discussed previously.

oscillation of the objective function[5]. However, by the introduction of NOMA condition to these type of problems (e.g. set of constraints given by $\mathbf{C}_2$), the convexity status of the overall objective function changes. **Lemma 1** states the convexity status of the problem in (4.15).

Lemma 1. *Given the proposed GCoMP-enabled NOMA scheme, the optimization problem* $\boldsymbol{P}_1$ *formulated in (4.15), which maximizes the normalized sum-rate of K NOMA UEs per cluster, is a convex problem.*

Proof. See **Appendix C**. ■

Due to the convexity of problem (4.15), a closed-form expression for optimal transmission power fractions $\{w_{km}\}_{k=1,\ldots,K}^{m=1,\ldots,M}$ can be derived using the Lagrangian multipliers method as follows. For the simplicity of analysis, we will illustrate the derivation of problem with relatively small system parameters (e.g. $M = 2$ and $K = 3$); however, the same procedure can be used to generalize the solution for any set of parameters. The Lagrangian function of problem $\mathbf{P}_1$ in (4.15) can then be written as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{a},\boldsymbol{\eta},\boldsymbol{\mu},\boldsymbol{\tau}\right) = &\sum_{k=1}^{3}\log_2\left(1+\frac{\sum_{m=1}^{2}w_{km}\gamma_{km}}{\sum_{m=1}^{2}\left(\sum_{j=k+1}^{3}w_{jm}\right)\gamma_{km}+1}\right)\\ &+\sum_{k=1}^{3}\eta_k\left[\sum_{m=1}^{2}\left(w_{km}-\gamma_{\text{th}}^{k}\sum_{j=k+1}^{3}w_{jm}\right)\gamma_{km}-\gamma_{\text{th}}^{k}\right]+\mu_1\left[P_s-\sum_{m=1}^{2}\left(w_{1m}-w_{2m}\right)\gamma_{2m}\right]\\ &+\sum_{i=2}^{3}\mu_i\left[P_s-\sum_{m=1}^{2}\left(w_{i-1m}-\sum_{j=i}^{3}w_{jm}\right)\gamma_{3m}\right]+\sum_{m=1}^{2}\tau_m\left[1-\sum_{k=1}^{3}w_{km}\right],\end{aligned} \tag{3.22}$$

where $\boldsymbol{a} = \{w_{ij}\}_{i=1,2,3}^{j=1,2}$, and $\eta \geq 0$, $\mu \geq 0$ and $\tau \geq 0$ are the Lagrange multipliers corresponding to $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$, respectively. Further discussion on the solution of Problem $\mathbf{P}_1$ in (4.15) is given in **Appendix D**.

[5]The convex-concave oscillation is caused by the fact that the sum-rate equation with limited-interference is equivalent to the difference between two independently changing convex and concave functions.

Finally, it is important to mention that for the proposed GCoMP-enabled NOMA scheme, in practical scenarios, the set of cooperating APs may not be able to provide every UE within their cluster with its minimum rate requirement due to the power budget limitations on different APs (infeasible problem). In such cases, the UE with the minimum norm metric (w.r.t. all connected APs) will be removed from its current NOMA cluster and join another cluster that uses a different sub-band.

3.7 Numerical Results

In this section, we provide some numerical results to discuss the performance of the proposed scheme under different system parameters, and then present illustrative results on the proposed NOMA transmission power allocation scheme. Each value is obtained via 2×10^6 Monte-Carlo simulation runs. For simplicity, we study only the case where all channel gains are i.i.d. Table 5.2 presents the main system parameters used to obtain the simulation and analytical results. These values were carefully selected to comply with practical values used

Table 3.3 Simulation parameters

Parameter	Value
AWGN PSD per UE	-169 dBm/Hz
Transmit power budget at a AP, P	Variable
SIC sensitivity, P_s	1 dBm
SINR threshold per UE, γ_{th}	15 dBm
Target SINR outage probability, ϵ	10^{-5}

in the industry.

3.7.1 Outage and Capacity Analysis

In this section, we present results on the outage and capacity performance of the proposed scheme. We first start by evaluating the performance gain of the proposed GCoMP scheme compared to that of conventional CoMP system (considering both orthogonal multiple access [OMA] and NOMA paradigms). Fig. 3.4 shows the average spectral efficiency per UE with different cooperation and multiple access scenarios. It can be noticed that, with OMA,

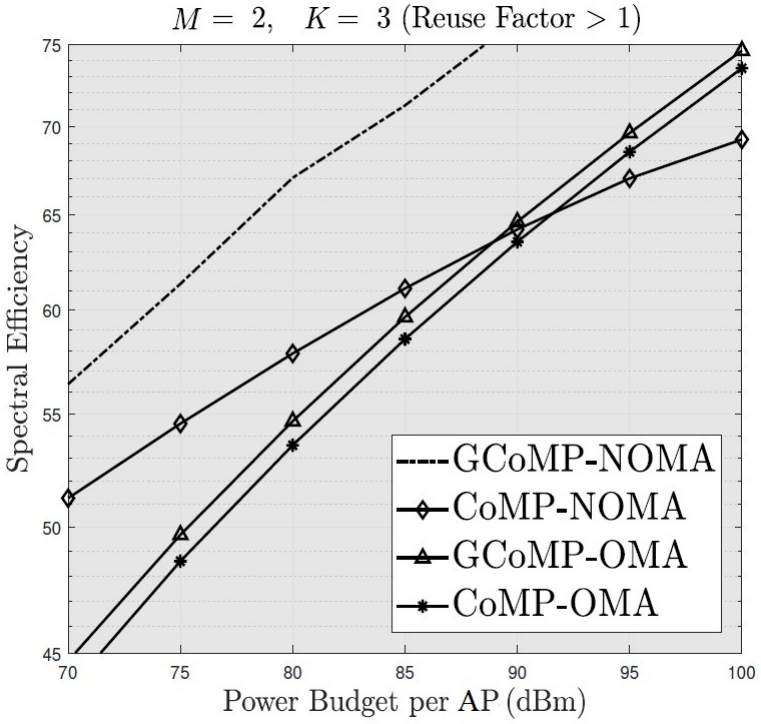

Figure 3.3 Average spectral efficiency with different cooperation and MA scenarios.

a slight enhancement can be achieved when moving from CoMP to its generalized version (GCoMP), when $M = 2$. However, more significant enhancements can be achieved for higher values of M, as shown in **3.4**. Additionally, the GCoMP-NOMA system achieves the best performance for a wide range of power budget while the performance of CoMP-NOMA scheme lies between those of GCoMP-NOMA and GCoMP-OMA for relatively low power levels. The performance of CoMP-NOMA degrades significantly as the maximum transmission power budget per AP increases due to interference caused by the non-cooperating APs

to the cell-centre UEs in other cells. Note that, to simulate the CoMP-NOMA scheme and compare it with the other schemes in a fair manner, we assume that the interference power from the non-cooperating APs to be 10^{-3} times the overall transmission power which takes into consideration the high distances between cell-centre UEs and other cells in a typical JT-CoMP scheme. To evaluate the n-th order clustering scheme, Fig. 3.4 shows the probability of outage of the proposed system under different clustering levels versus the maximum transmission power budget per AP. It can be noticed from Fig. 3.4 that the outage performance

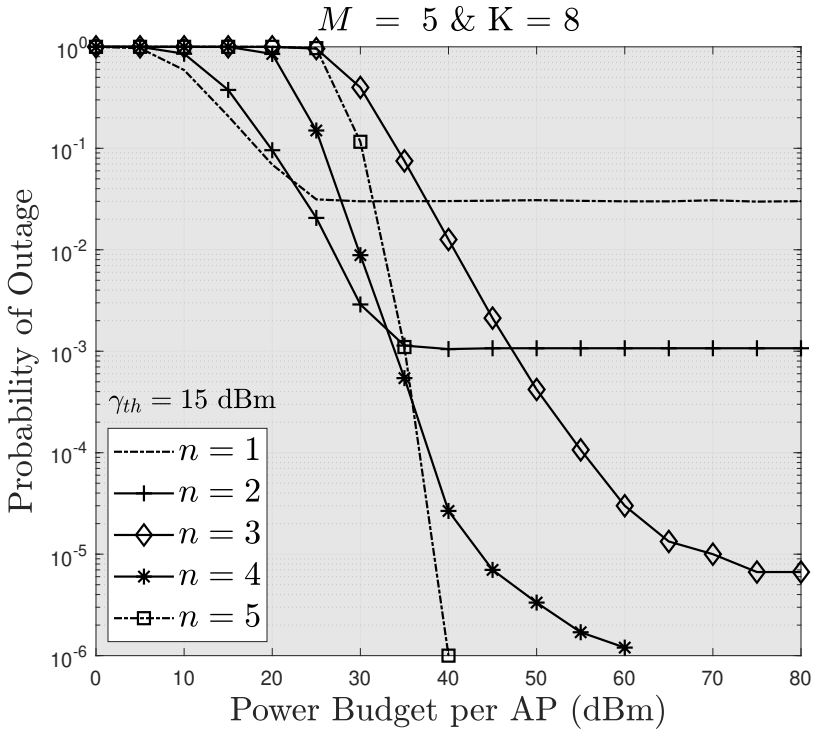

Figure 3.4 $\mathrm{P}_{\mathrm{out}}^{(n)}$ versus P for different clustering order n.

improves significantly when the clustering order increases. When many APs, e.g. $M = 5$ are transmitting in the same cluster sub-band while the clustering order is low, the outage performance deteriorates. For example, a total service blockage occurs at $M = 5$ and $n = 1$. However, when the clustering order n is relatively close to the number of cooperating APs (M), the interference on different UEs caused by non-serving APs becomes tolerable.

It is also important to investigate the effect of increasing the number of UEs per single

NOMA cluster on the overall outage performance. Fig. 3.5 (a) shows the outage probability for a UE versus the maximum transmission power budget per AP under full-order clustering. It can be noticed that when only one AP is serving a set of NOMA UEs, increasing the

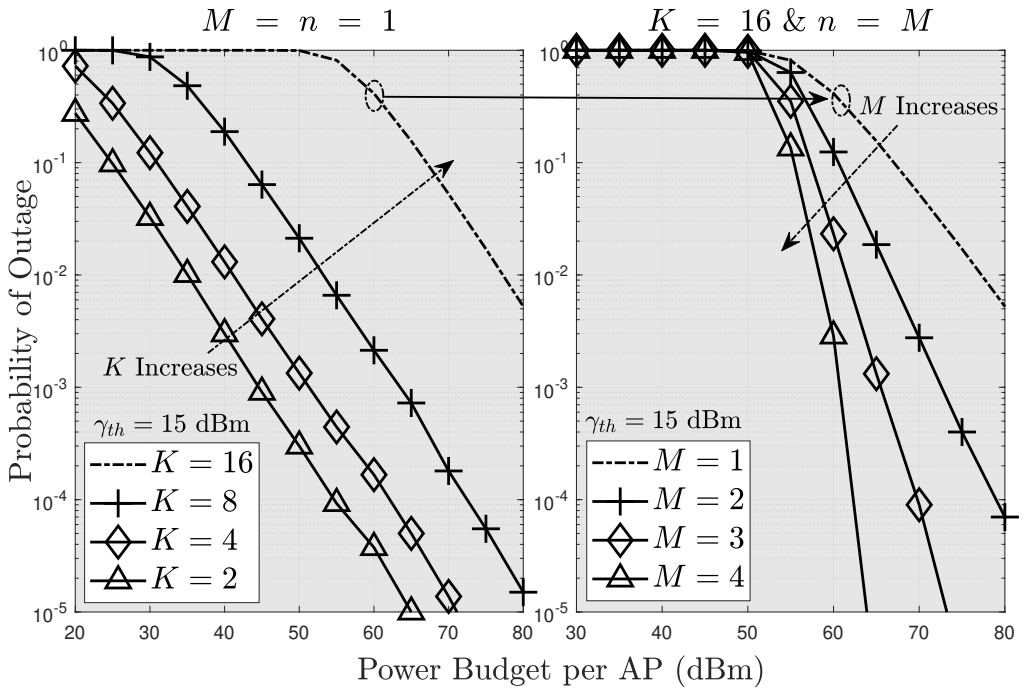

Figure 3.5 $\mathrm{P}_{\mathrm{out}}^{(n)}$ versus P for different numbers of UEs (K) and APs (M).

NOMA cluster size will cause a significant degradation in system coding gain (i.e, the outage performance curve shifts to the right). This is because, as K increases, the average number of interfering signal components with lower weights than the desired signal will increase. This observation discourages the potential use of NOMA in its conventional form with one serving AP at the massive scale.

To compensate for the performance degradation caused by large NOMA cluster size, the number of serving APs per cluster can be increased as in the proposed n-th order clustering scheme. Fig. 3.5 (b) shows the outage probability for a UE versus the maximum power budget per AP under different number of serving APs and with a relatively large cluster size ($K = 16$). It can be noticed that a large number of NOMA clusters can co-exist in the same spectrum band when the number of cooperating APs (M) increases.

Finally, Fig. 3.6 shows the ϵ-outage capacity (C_ϵ) versus transmission power P for different number of cooperating APs. As was concluded from (3.17), increasing M exponentially

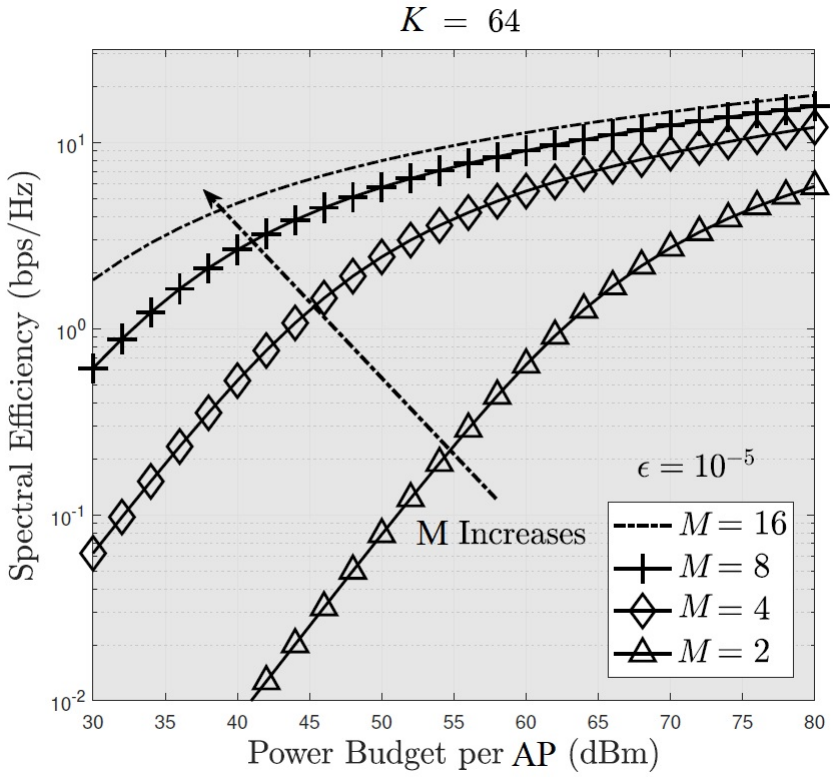

Figure 3.6 C_ϵ versus P for different number of APs (M).

annihilates the negative effect caused by the large number of NOMA UEs using the same sub-band.

3.7.2 Imperfect CSI

In this section, the effect of imperfect channel on the performance of the proposed system is investigated. Fig. 3.7 (a) shows the effect of different values of channel estimation accuracy at the APs (we assume that $\chi_i = \chi, \forall i$). It can be noticed from this figure that an approximate performance loss of 1.7 dB occurs for every 5% decrease in channel estimation accuracy. It is worth mentioning that the overall effect of imperfect CSI can be decreased by increasing the number of cooperative APs per NOMA cluster, as can be noticed from Fig. 3.7 (b).

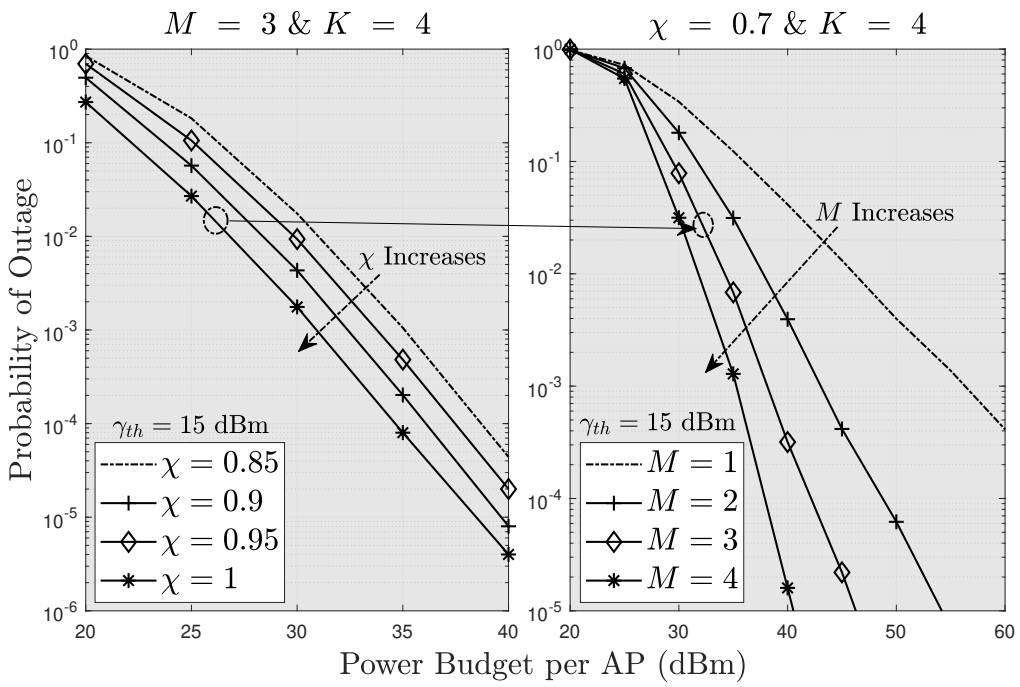

Figure 3.7 Sum-rate versus P under imperfect CSI.

However, such an enhancement is only achieved when the estimation accuracy is sufficiently large (e.g. 70.71%, with equal χ for all APs such that the inequality $M\chi^2 \geq M(1-\chi^2)$ holds). In other words, when the CSI estimation error is severe, increasing the number of cooperative APs will improve the system performance.

3.7.3 Power Allocation

In this section, we study the spectral efficiency of the simplified GCoMP-enabled NOMA scheme with optimal transmission power allocation. Additionally, the performance of the proposed GCoMP-enabled NOMA is compared with its OMA counterparts. The optimal solution is obtained by solving the KKT conditions derived in Eqs. C.1 in **Appendix D** for every channel realization and then selecting the set of feasible solutions (when transmission power budget is adequate to fulfill all the constraints of the optimization problem $\mathbf{P}_1$ in (4.15). The channel realization here is assumed to follow i.i.d. fading distribution with unity

variance. For a given transmission power budget value, the overall spectral efficiency is calculated by averaging the accumulated sum-rate produced from realizations with feasible solutions for that power budget. Additionally, the minimum power rate for every UE is considered as the achievable rate for the same UE using OMA with GCoMP.

Fig. 3.8 (a) shows the optimized spectral efficiency (sum-rate) versus the maximum transmission power budget of GCoMP network layout with both multiple access schemes (NOMA and OMA). As can be noticed, the sum-rate of the GCoMP-enabled NOMA scheme

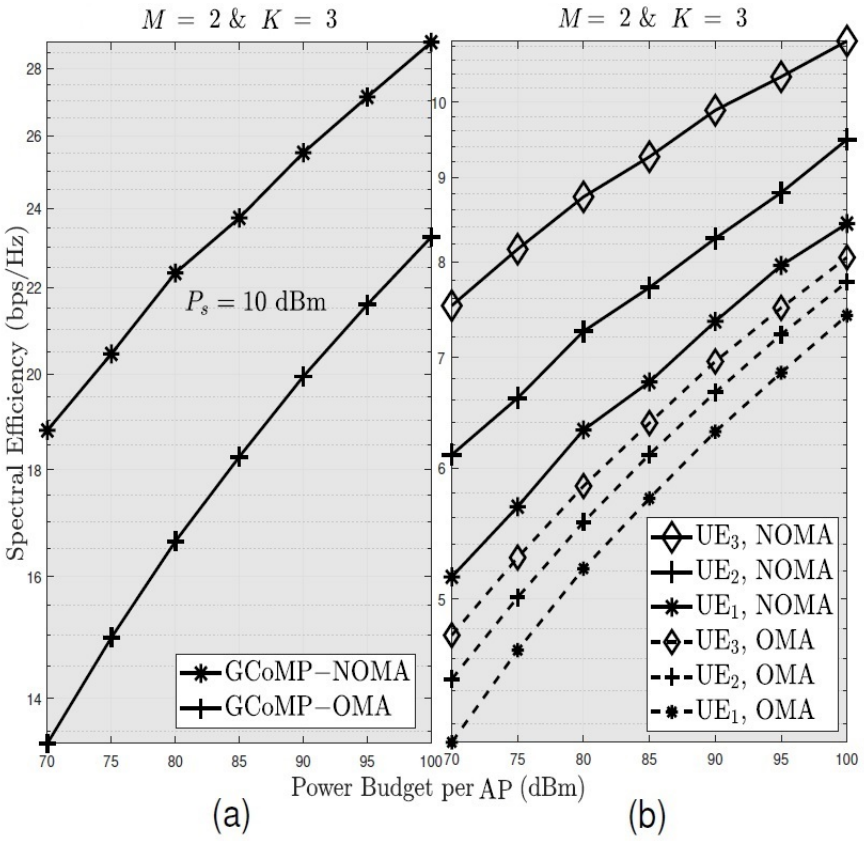

Figure 3.8 Sum-rate versus P under optimal transmission power allocation.

with $M = 2$ and $K = 3$ is superior to that of GCoMP with OMA scheme. However, the spectral efficiency enhancement decreases as the maximum transmission power budget increases.

To study the performance of UEs individually, Fig. 3.8 (b) shows the spectral efficiency

per UE under optimal transmission power allocation. Note that the UEs are ordered in an ascending mode from the least norm UE to the highest norm UE. It can be noticed that a UE with a higher norm always yields a larger spectral efficiency compared to the one with a lower norm. This is because power allocation for NOMA is based on inverse water-filling method until all UEs are provided with minimum rate requirements, and then water-filling is utilized. In particular, we notice that, the smaller the power budget per AP, the larger is the number of APs which apply power allocation to the cluster members for NOMA, and the higher the power budget, the smaller is the number of APs which apply power allocation to meet the requirements of NOMA (i.e, most of the APs then use the ordinary water-filling).

3.8 Conclusion

In this chapter, a novel generalized CoMP-enabled NOMA scheme has been proposed and evaluated. In particular, the traditional joint-transmission CoMP scheme has been generalized to be applied for all UEs (i.e, cell-centre as well as cell-edge UEs) within the coverage area of a wireless network. Furthermore, every base station has been assumed to apply a multiuser NOMA scheme for all UEs associated to it. To evaluate the proposed scheme, the closed-form expressions for the probability of outage and outage capacity per-UE with different orders of BS cooperation have been derived. To reduce the complexity of the proposed system, a low-complexity full order clustering protocol has been designed for the generalized CoMP-enabled NOMA system where the optimal transmission power allocation method has been obtained. Findings show that it is possible to deploy NOMA with a large number of UEs per sub-band and tolerable complexity as long as the number of cooperating base stations is comparable to the number of NOMA UEs. One disadvantage of the proposed scheme is that each UE may belong to different NOMA clusters that are run through different APs. Despite its optimality, implementing the proposed protocol in practice requires significantly high centralized hardware and computational complexity. In the following two chapters,

we propose two different practical realization-friendly models for large-scale NOMA-enabled cell-free networks.

Chapter Four

Network-Centric AP Clustering: A DRL Design

4.1 Introduction

In the existing literature, performance analysis of cell-free networks has focused only on the derivation of an approximate mathematical limits of the per-UE transmission rate. This is due to the fact that other important performance metrics such as the probability of outage (or alternatively coverage probability) are significantly difficult to be derived in a closed-form. Furthermore, for the cell-free networks, beamforming techniques (uplink decoding and downlink precoding) were basically developed and tested for relatively small number of UEs and APs. The reason is that computing the beamforming vectors for massive numbers of UEs and APs using the conventional algorithms (such as bisection search and gradient descent) gives rise to real-time implementation issues such as convergence time and computational complexity. In **Chapter 3**, we proposed a downlink variable-order UE-centric based clustering scheme that assign an optimized subset of UEs for each AP at which a multi-UEs NOMA power allocation scheme is conducted resulting in every UE being simultaneously served by n number of APs. The proposed scheme has decreased the lengths of the beamforming vectors

for every NOMA cluster by a factor that is linearly proportional to clustering order.

In this chapter, we propose a novel (network-centric)-based dynamic cell-free network architecture suitable for 6G networks with massive connectivity. For a fully centralized cell-free network, we first derive a closed-form expression of the per-UE probability of outage. To reduce the complexity of joint processing of signals from all UEs in a centralized cell-free network, we propose a dynamic clustering scheme for APs. For real-time implementation of both dynamic AP clustering and uplink beamforming, we develop a deep reinforcement learning (DRL) scheme, namely, the hybrid DDPG-DDQN scheme. The rest of this chapter is organized as follows. The general system model and assumptions for the proposed dynamic cell-free network is presented in **Section 4.2**. In **Section 4.3**, a closed-form expressions for the probability of outage of a static cell-free network and that of the proposed dynamic cell-free network with random clustering are derived. In **Section 4.4**, we present an SIC-aided signal detection scheme and a diversity combining scheme. Also, the joint optimization problem of AP clustering and beamforming design for the dynamic cell-free model is formulated in this section. In **Section 4.5**, we propose a novel hybrid DRL method that jointly performs AP clustering and optimization of the beamforming vectors. Numerical and simulation results are presented in **Section 4.6** while the chapter is concluded in **Section 4.7**.

Notations: For a random variable (rv) X, $F_X(x)$ and $f_X(x)$ represent cumulative distribution function (CDF) and probability density function (PDF), respectively. $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ denote probability and expectation, respectively. For a given matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, $\boldsymbol{A}^H$ represents the Hermitian transpose of $\boldsymbol{A}$. A proper circularly symmetric rv $Z = X + jY$ is denoted by $Z \sim \mathcal{CN}(0, \sigma_Z)$ where the PDFs of X and Y are given respectively by $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{\frac{x^2}{2\sigma_X^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{\frac{y^2}{2\sigma_Y^2}}$ while a Gamma rv is denoted by $X \sim \mathcal{G}(\alpha, \beta)$, with PDF as $f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$, where $\beta > 0$, $\alpha > 0$, and $\Gamma(z)$ is the Euler Gamma function and a PDF expression of Nakagami-$\mathcal{M}$ rv is given by $f_X(x) = \frac{2\mathcal{M}^{\mathcal{M}} x^{2\mathcal{M}-1}}{\Gamma(\mathcal{M})\Omega^{\mathcal{M}}} e^{-\frac{\mathcal{M}}{\Omega}x^2}$.

4.2 Cell-Free Network Architecture With Dynamic Clustering

4.2.1 System Model and Assumptions

We consider an uplink network with M single-antenna APs and K single-antenna UEs at fixed locations within a certain coverage area (Fig. 4.1). All APs are connected to each other through backhaul links to form a centralized cell-free network architecture [80]. This network setup enables the distributed APs to collaborate simultaneously to serve all UEs within the network coverage area. The network is equipped with a BBU pool (which is referred to as CPU) that performs the last-stage processing tasks to detect/decode signals from every UE. The channel state information (CSI), which will be required at the CPU for signal detection for individual UEs, is assumed to be estimated by the CPU through training the pilot sequences that are not completely orthogonal. Note that in this network setup, the CSI and the instantaneous clustering information will be only required at the CPU for decoding the signals from the UEs. In the proposed dynamic cell-free network, all the M APs in the network are partitioned among a set of $\mathcal{N}$ subgroups (known as clusters) in which the n-th cluster consists of $\mathcal{D}_n$ APs such that $1 \leq \mathcal{D}_n \leq M - (\mathcal{N} - 1)$. Note that when $\mathcal{N} = M$, every cluster will contain a single AP and this falls back to the centralized cell-free network model. Additionally, the n-th cluster, where $n = 1, \ldots, \mathcal{N}$, acts as an single virtual AP with a $\mathcal{D}_n$-antenna DAS that combines the overall received signals using a predefined combining technique (such as maximal ratio combining [MRC], equal gain combining [EGC], selection combining [SC], etc.) [81]. We assume that channel gain between the k-th UE and the m_n-th antenna (the subscript n denotes the n-th cluster) follows the following distribution:

$$g_{m_n k} = L_{m_n k}^{-\kappa} \mathcal{F}_{m_n k}^{1/2} h_{m_n k}, \tag{4.1}$$

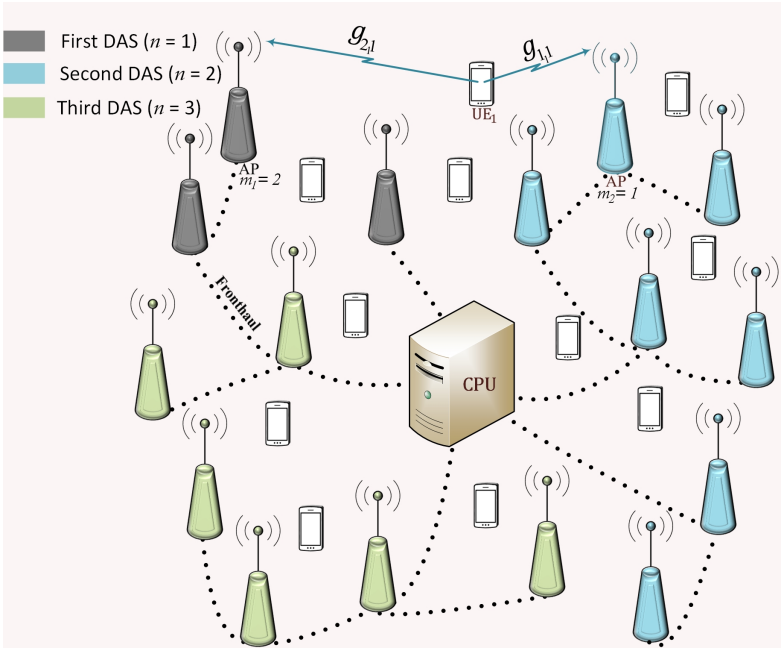

Figure 4.1 Example of a cell-free network with dynamic clustering (for $M = 16$ and $\mathcal{N} = 3$).

where $L_{m_nk} = ||d_{m_nk}||$ is the Euclidean distance between the k-th UE and the m_n-th antenna in the n-th cluster (virtual AP), κ is the path-loss exponent with a value depends on the propagation environment with $\kappa \geq 2$, $\mathcal{F}_{m_nk} = 10^{\frac{\sigma_{m_nk} z_{m_nk}}{10}}$ is the log-normal shadowing correlation factor with standard deviation σ_{m_nk} and $z = \sqrt{\varrho} a_{m_nk} + \sqrt{1-\varrho} b_{m_nk}$, where $a_{m_nk} \sim \mathcal{N}(0,1)$ and $b_{m_nk} \sim \mathcal{N}(0,1)$ are independent and identically distributed (i.i.d) rvs that represent the shadow fading effect at the k-th UE and the m_n-th antenna, respectively, Here ϱ is the transmitter-receiver shadow fading correlation coefficient [82] and h_{m_nk} is the small-scale channel fading between the k-th UE and the m_n-th antenna at the n-th cluster. Furthermore, we assume that h_{m_nk}s follow independent but not identically distributed (i.n.d) Nakagami-$\mathcal{M}_{m_nk}$ distributions with spreading and shape parameters $\mathcal{M}_{m_nk}$ and Ω_{m_nk}, respectively, i.e. $|h_{m_nk}|^2 \sim \mathcal{G}(\alpha_{m_nk}, \beta_{m_nk})$, where $\alpha_{m_nk} = \mathcal{M}_{m_nk}$ is the shape parameter and $\beta_{m_nk} = \frac{\mathcal{M}_{m_nk}}{\Omega_{m_nk}}$ is the rate parameter. For simplicity of analysis, we assume that

$L_{m_nk}^{-\kappa}$ and $\mathcal{F}_{m_nk}$ $\forall$ m, n and k are known[1]. Accordingly, we have $|g_{m_nk}|^2 \sim \mathcal{G}(\alpha_{m_nk}, \beta'_{m_nk})$, where $\beta'_{m_nk} = \frac{\beta_{m_nk} L_{m_nk}^{2\kappa}}{\mathcal{F}_{m_nk}} = \frac{\mathcal{M}_{m_nk} L_{m_nk}^{2\kappa}}{\Omega_{m_nk} \mathcal{F}_{m_nk}}$. Let us denote by $\boldsymbol{\mathcal{C}} = \{\mathcal{C}_1 \dots \mathcal{C}_N\}$ the set of all possible clustering configurations of APs such that every cluster contains at least one AP.

4.2.2 Uplink Network Training (CSI Acquisition)

Under the assumption that $L_{m_nk}^{-\kappa}$ and $\mathcal{F}_{m_nk}$ are known, the aim of network training is to estimate the small-scale fading from the overall channel gain. In order to achieve this, the CPU first assign a pilot sequence to each active UE (denoted by $\boldsymbol{\varphi}_k = \left[\varphi_{k,1} \dots \varphi_{k,\tau_p}\right]^H$ such that $||\boldsymbol{\varphi}_k||^2 = 1$, where $\tau_p \leq \tau_c$ is the length of the pilot training sequence [in samples], which is less than or equal the channel coherence time [τ_c]) which in turn sends that sequence to all APs with constant power. Accordingly, the received pilot vector at the m_n antenna of the n-th DAS is given by

$$\boldsymbol{y}_{p,m_n} = \sum_{k=1}^{K} \sqrt{\tau_p \rho_k} g_{m_nk} \boldsymbol{\varphi}_k + \boldsymbol{\eta}_{m_n}, \tag{4.2}$$

where ρ_k is the normalized transmitted power for each symbol of the k-th UE pilot vector and $\boldsymbol{\eta}_{m_n} \in \mathbb{C}^{\tau_p \times 1}$ is the zero-mean complex additive white Gaussian noise (AWGN) vector related to pilot symbols with i.i.d rvs, i.e. $\eta_{m_n} \sim \mathcal{CN}(0, 1/2)$. To find the best estimate of g_{m_nk} (denoted by $\hat{g}_{m_nk} = L_{m_nk}^{-\kappa} \mathcal{F}_{m_nk}^{1/2} \hat{h}_{m_nk}$) given the vector of observations $\boldsymbol{y}_{p,m_n}$, we first project $\boldsymbol{y}_{p,m_n}$ in (4.2) over $\boldsymbol{\varphi}_k^H$ as follows:

$$\begin{aligned} \dot{y}_{p,m_n} &= \boldsymbol{\varphi}_k^H \boldsymbol{y}_{p,m_n} \\ &= \underbrace{\sqrt{\tau_p \rho_k} g_{m_nk}}_{\text{Desired Value}} + \underbrace{\sum_{k'=1, k' \neq k}^{K} \sqrt{\tau_p \rho_{k'}} g_{m_nk'} \boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_{k'} + \boldsymbol{\varphi}_k^H \boldsymbol{\eta}_{m_n}}_{\text{Noise Value}}. \end{aligned} \tag{4.3}$$

Estimating g_{m_nk} from (4.3) can be optimally achieved by using the maximum *a posteriori* decision rule (MAP). The Bayesian estimator of g_{m_nk} is found to be identical to that of the

[1]The effect of shadow fading correlation model will be investigated in more detail in the section on numerical results later in this chapter.

minimum mean square method (MMSE) [83, 84]. Furthermore, under the condition that $\boldsymbol{\varphi}_k,\ \forall\ k = 1, \ldots, K$ can either be orthogonal or nonorthogonal, $\boldsymbol{\varphi}_k^H \boldsymbol{y}_{p,m_n}$ in (4.3) represents a sufficient statistics for optimal estimation of g_{m_nk} (MMSE). Accordingly, the best estimate of g_{m_nk} can be found by [37]

$$\hat{g}_{m_nk} = \frac{\mathbb{E}\left[\dot{y}^*_{m_nk} g_{m_nk}\right]}{\mathbb{E}\left[|\dot{y}_{m_nk}|^2\right]} \dot{y}_{m_nk} = \mathcal{E}_{m_nk} \dot{y}_{m_nk}. \tag{4.4}$$

Under the assumption that g_{m_nk}s are proper i.n.d complex Gaussian rvs and that $\boldsymbol{\eta}_{m_n}$s are zero-mean i.i.d rvs, $\forall\ n, m$ and k, the MMSE estimation constant $\mathcal{E}_{m_nk}$ can be written as

$$\mathcal{E}_{m_nk} = \frac{\sqrt{\tau_p \rho_k} L^{-\kappa}_{m_nk} \mathcal{F}^{-1/2}_{m_nk} \left(\frac{\alpha_{m_nk}}{\beta^2_{m_nk}}\right)}{\tau_p \sum_{l=1}^{K} \rho_l L^{-\kappa}_{m_nl} \mathcal{F}^{-1/2}_{m_nl} \left(\frac{\alpha_{m_nl}}{\beta^2_{m_nl}}\right) |\boldsymbol{\varphi}^H_{m_nk} \boldsymbol{\varphi}_{m_nl}|^2 + 1}, \tag{4.5}$$

where we use the fact that $\mathbb{E}\left[|h_{m_nk}|^2\right] = \frac{\alpha_{m_nk}}{\beta_{m_nk}}$. Note that, if all UEs assigned a set of mutually orthogonal pilot sequences (i.e. $|\boldsymbol{\varphi}^H_k \boldsymbol{\varphi}_l| = 0,\ \forall\ k \neq l$), the estimated small-scale channel fading in (4.3) reduces to a scaled version of the exact fading gain plus a relatively small AWGN noise portion. However, depending on the applications (e.g. mMTC applications) and due to length limitations of τ_p, non-orthogonal pilot signals have to be used among some active UEs.

4.2.3 Uplink Data Transmission

In the proposed dynamic cell-free network, the composite of signals from all UEs is received by each APs within the network. For the k-th UE, composite signal from the m_n-th antenna is first multiplied by a DAS parameter (denoted by G_{m_nk}) and then the sum of signals from all $\mathcal{D}_n$ antennas of the n-th cluster is then multiplied by a beamforming parameter (denoted by w_{nk}). This combining process is conducted such that the signal component related to the k-th UE is maximized while the remaining interference plus noise component is minimized and it takes place at the baseband level in the network CPU before the detected signal related to the k-th UE is forwarded to its final destination. Accordingly, the signal used to detect

$$
\begin{aligned}
y_k &= \sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} G_{m_n k} \left[\sum_{l=1}^{K} \hat{g}_{m_n l} \sqrt{p_l} x_l + \tilde{\eta}_{m_n} \right] \\
&= \underbrace{\sqrt{\tau_p \rho_k p_k} x_k \sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} \mathcal{E}_{m_n k} G_{m_n k} g_{m_n k}}_{\text{Desired Signal}} + \sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} G_{m_n k} \Bigg[\underbrace{\sum_{l=1, l\neq k}^{K} \sqrt{\tau_p \rho_l p_l} x_l \mathcal{E}_{m_n l} g_{m_n l}}_{\text{Inter-User Interference}} \\
&\quad + \underbrace{\sum_{v=1, v\neq k}^{K} \sqrt{\tau_p \rho_v p_k} x_k \mathcal{E}_{m_n k} |\boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_v| g_{m_n v} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \sqrt{\tau_p \rho_u p_q} x_q \mathcal{E}_{m_n q} |\boldsymbol{\varphi}_q^H \boldsymbol{\varphi}_u| g_{m_n u}}_{\text{non-orthogonal pilot-related estimation error}} \Bigg] \\
&\quad + \sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} G_{m_n k} \Bigg[\underbrace{\sqrt{p_k} \mathcal{E}_{m_n k} x_k |\boldsymbol{\varphi}_k^H \boldsymbol{\eta}_{m_n}| + \sum_{z=1, z\neq k}^{K} \sqrt{p_z} \mathcal{E}_{m_n z} x_z |\boldsymbol{\varphi}_z^H \boldsymbol{\eta}_{m_n}|}_{\text{AWGN-related estimation error}} + \underbrace{\tilde{\eta}_{m_n}}_{\text{AWGN}} \Bigg].
\end{aligned}
\tag{4.6}
$$

$$
\gamma_k = \frac{\sum_{m=1}^{M} |\tilde{g}_{mk}|^2}{\sum_{m=1}^{M} \left[\sum_{l=1, l\neq k}^{K} |\tilde{g}_{ml}|^2 + \sum_{v=1, v\neq k}^{K} |\tilde{g}_{mv}|^2 + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} |\tilde{g}_{mu}|^2 \right] + 1}, \tag{4.7}
$$

---

the k-th UE component at CPU is given by (4.6) at the top of next page, where w_{nk} is the n-th element of the beamforming vector related to the k-th UE such that $0 \leq w_{nk} \leq 1$, $G_{m_n k}$ is the DAS combining parameter of the k-th UE at the m_n-th AP in the n-th cluster, p_k is the uplink transmission power of the k-th UE such that $0 \leq p_k \leq P_k$, where P_k is the power budget of the k-th UE, x_k is the transmitted symbol of the k-th UE such that $\mathbb{E}[|x_k|^2] = 1$ and $\tilde{\eta}_{m_n}$ is the AWGN at the m_n-th AP with $\tilde{\eta}_{m_n} \sim \mathcal{CN}(0, 1/2)$. We assume that $\tilde{\eta}_{m_n}$s are from a set of i.i.d rvs.

4.3 SINR Outage Performance

4.3.1 SINR Outage for a centralized Cell-Free Network

In this section, we will focus on studying the outage behaviour of centralized cell-free networks, i.e. with $\mathcal{N} = M$ and $\mathcal{D}_n = 1, \forall\ n = 1, \ldots, \mathcal{N}$. Note that for the centralized cell-free network, there will be only one clustering configuration for APs which is given by $\mathcal{C}_1 = \{\{\text{AP}_1\}, \ldots, \{\text{AP}_M\}\}$. Accordingly, the second summation in (4.6) will disappear with $n = m$ and $G_{m_n k} = 1,\ \forall\ n, m_n$ and k. The instantaneous signal-to-interference-plus-noise ratio (SINR) at the input of the k-th UE detector is given by 4.7 at the top of this page [85], where $|\tilde{g}_{mk}|^2 \sim \mathcal{G}\left(\tilde{\alpha}_{mk}, \tilde{\beta}_{mk}\right)$, $|\tilde{g}_{ml}|^2 \sim \mathcal{G}\left(\tilde{\alpha}_{ml}, \tilde{\beta}_{ml}\right)$, $|\tilde{g}_{mv}|^2 \sim \mathcal{G}\left(\tilde{\alpha}_{mv}, \tilde{\beta}_{mv}\right)$, $|\tilde{g}_{mq}|^2 \sim \mathcal{G}\left(\tilde{\alpha}_{mq}, \tilde{\beta}_{mq}\right)$, $|\tilde{g}_{mu}|^2 \sim \mathcal{G}\left(\tilde{\alpha}_{mu}, \tilde{\beta}_{mu}\right)$ with $\tilde{\alpha}_{mk} = \mathcal{M}_{mk}$, $\tilde{\beta}_{mk} = \frac{\mathcal{M}_{mk}\dot{\sigma}_{mk}L_{mk}^{2\kappa}}{\Omega_{mk}\mathcal{F}_{mk}w_{mk}^2\tau_p\rho_k p_k \mathcal{E}_{mk}^2}$, $\tilde{\alpha}_{ml} = \mathcal{M}_{ml}$, $\tilde{\beta}_{ml} = \frac{\mathcal{M}_{ml}\dot{\sigma}_{mk}L_{ml}^{2\kappa}}{\Omega_{ml}\mathcal{F}_{ml}w_{mk}^2\tau_p\rho_l p_l \mathcal{E}_{ml}^2}$, $\tilde{\alpha}_{mv} = \mathcal{M}_{mv}$, $\tilde{\beta}_{mv} = \frac{\mathcal{M}_{mk}\dot{\sigma}_{mk}L_{mv}^{2\kappa}}{\Omega_{mk}\mathcal{F}_{mvs}w_{mk}^2\tau_p\rho_v p_k \mathcal{E}_{mk}^2|\boldsymbol{\varphi}_k^H\boldsymbol{\varphi}_v|^2}$, $\tilde{\alpha}_{mu} = \mathcal{M}_{mu}$, $\tilde{\beta}_{mu} = \frac{\mathcal{M}_{mq}\dot{\sigma}_{mk}L_{mu}^{2\kappa}}{\Omega_{mq}\mathcal{F}_{mu}w_{mk}^2\tau_p\rho_u p_q \mathcal{E}_{mq}^2|\boldsymbol{\varphi}_q^H\boldsymbol{\varphi}_u|^2}$, with

$$\dot{\sigma}_{mk} = \sum_{m=1}^{M} w_{mk}^2 \left[\sum_{t=1}^{\tau_p} \left(p_k \mathcal{E}_{mk}^2 \varphi_{k,t}^2 + \sum_{z=1, z\neq k}^{K} p_z \mathcal{E}_{mz}^2 \varphi_{z,t}^2\right) + 1\right].$$

Equation (4.7) is derived utilizing the fact that when both transmitter and receiver has knowledge of the estimated CSI, channel fading parameters can be replaced by their actual values instead of their second moments. for example, the numerator of (4.7) can be written as $\mathbb{E}\left[|\boldsymbol{\mathcal{G}}_k^H \boldsymbol{\mathcal{W}}_k x_k|^2\right] = \boldsymbol{\mathcal{W}}_k^H \boldsymbol{\mathcal{R}}_k \boldsymbol{\mathcal{W}}_k$, where $\boldsymbol{\mathcal{G}}_k = [C_{1k} g_{1k} \ldots C_M g_{\mathcal{D}_1 k}]$, $\boldsymbol{\mathcal{W}}_k = [w_1 \ldots w_M]$, $C_{m_n k} = \sqrt{\tau_p \rho_k p_k}\mathcal{E}_{m_n k} G_{m_n k}$ and $\boldsymbol{\mathcal{R}}_k = \mathbb{E}\left[|\boldsymbol{\mathcal{G}}_k^H \boldsymbol{\mathcal{G}}_k|^2\right]$ represent the auto-correlation matrix of k-th UE signal and is defined as $\boldsymbol{\mathcal{R}}_k = \bar{\boldsymbol{\mathcal{G}}}_k \bar{\boldsymbol{\mathcal{G}}}_k^H + \boldsymbol{C}_{\boldsymbol{\mathcal{G}}_k}$, where $\bar{\boldsymbol{\mathcal{G}}}_k$ and $\boldsymbol{C}_{\boldsymbol{\mathcal{G}}_k}$ are the mean and covariance matrices of $\boldsymbol{\mathcal{G}}_k$, respectively. If the instantaneous CSI is known for both transmitter and receiver, $\boldsymbol{\mathcal{R}}_k$ can be then defined as $\boldsymbol{\mathcal{R}}_k = \boldsymbol{\mathcal{G}}_k \boldsymbol{\mathcal{G}}_k^H$ [85]. Using a similar procedure, the interference power component of (4.7) can be proven. Additionally, the power

of AWGN component is derived utilizing the fact that all noise samples are i.i.d circularly symmetric Gaussian rvs with zero-mean and constant variance $\sigma^2_{m_n} = \sigma^2 = 1/2,\ \forall\ m, n.$ A receiver is in an outage if the SINR of the received signal falls bellow a certain predefined threshold value (denoted by γ_{th}). **Theorem 3** below gives an accurate approximation for the average probability of outage of the k-th UE.

Theorem 3. *If* $|\tilde{g}_{mk}|^2 \sim \mathcal{G}(\tilde{\alpha}_{mk},\ \tilde{\beta}_{mk})$, $|\tilde{g}_{ml}|^2 \sim \mathcal{G}(\tilde{\alpha}_{ml},\ \tilde{\beta}_{ml})$, $|\tilde{g}_{mu}|^2 \sim \mathcal{G}(\tilde{\alpha}_{mu},\ \tilde{\beta}_{mu})$ *and* $|\tilde{g}_{mv}|^2 \sim \mathcal{G}(\tilde{\alpha}_{mv},\ \tilde{\beta}_{mv})$ *where* $|\tilde{g}_{mk}|^2$, $|\tilde{g}_{ml}|^2$, $|\tilde{g}_{mu}|^2$ *and* $|\tilde{g}_{mv}|^2$ *are independent rvs with* $m = 1, \ldots, M$ *and* $k = 1, \ldots, K$. *The probability of outage of the k-th UE in a centralized cell-free network is*

$$P_{out}^{(k)} = \frac{\dot{\alpha}_{mk}^{-1} \dot{\beta}_{mk'}^{\dot{\alpha}_{mk'}} \left(\gamma_{th} \dot{\beta}_{mk}\right)^{\dot{\alpha}_{mk}} \Gamma\left(\dot{\alpha}_{mk} + \dot{\alpha}_{mk'}\right)}{\Gamma\left(\dot{\alpha}_{mk}\right) \Gamma\left(\dot{\alpha}_{mk'}\right) \left(\gamma_{th} \dot{\beta}_{mk} + \dot{\beta}_{mk'}\right)^{\dot{\alpha}_{mk} + \dot{\alpha}_{mk'}}} \times {}_2F_1\left(1, \dot{\alpha}_{mk} + \dot{\alpha}_{mk'}; 1 + \dot{\alpha}_{mk}; \frac{\gamma_{th} \dot{\beta}_{mk}}{\gamma_{th} \dot{\beta}_{mk} + \dot{\beta}_{mk'}}\right), \tag{4.8}$$

where γ_{th} *is the SINR threshold value and* ${}_2F_1(.)$ *is the Gauss hypergeometric function function [86, Eq. 9.11.1].* $\dot{\alpha}_{mk}$, $\dot{\beta}_{mk}$, $\dot{\alpha}_{mk'}$ *and* $\dot{\beta}_{mk'}$ *are defined in Appendix A*

Proof. See **Appendix ??**. ■

To achieve the best performance in a centralized cell-free network, the beamforming vectors for all UEs throughout all APs will need to be optimized at the CPU. This will require huge processing and incur significant delay, especially when a very large number of UEs are using the same time-frequency resources.

4.3.2 SINR Outage for a Dynamic Cell-Free Network

In a dynamic cell-free network, M APs will be divided into $\mathcal{N}$ clusters where each cluster represents a virtual AP in a DAS. Furthermore, every possible AP clustering configuration

$$\gamma_k^{\{\mathcal{C}_j\}} = \frac{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} |\tilde{g}_{m_n k}|^2}{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1, l\neq k}^{K} |\tilde{g}_{m_n l}|^2 + \sum_{v=1, v\neq k}^{K} |\tilde{g}_{m_n v}|^2 + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} |\tilde{g}_{m_n u}|^2 \right] + 1}, \tag{4.9}$$

$$\dot{\alpha}_{m_n k} = \frac{\left(\sum_{n=1}^{\mathcal{N}} \sum_{m_n}^{\mathcal{D}_n} \tilde{\alpha}_{m_n k} / \tilde{\beta}_{m_n k} \right)^2}{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \tilde{\alpha}_{m_n k} / \tilde{\beta}^2_{m_n k}}, \qquad \dot{\beta}_{m_n k} = \frac{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \tilde{\alpha}_{m_n k} / \tilde{\beta}^2_{m_n k}}{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \tilde{\alpha}_{m_n k} / \tilde{\beta}_{m_n k}},$$

$$\dot{\alpha}_{m_n k'} = \frac{\left(\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{m_n l}}{\tilde{\beta}_{m_n l}} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{m_n u}}{\tilde{\beta}_{m_n u}} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{m_n v}}{\tilde{\beta}_{m_n v}} \right] \right)^2}{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{m_n l}}{\tilde{\beta}^2_{m_n l}} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{m_n u}}{\tilde{\beta}^2_{m_n u}} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{m_n v}}{\tilde{\beta}^2_{m_n v}} \right]},$$

$$\dot{\beta}_{m_n k'} = \frac{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{m_n l}}{\tilde{\beta}^2_{m_n l}} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{m_n u}}{\tilde{\beta}^2_{m_n u}} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{m_n v}}{\tilde{\beta}^2_{m_n v}} \right]}{\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{m_n l}}{\tilde{\beta}_{m_n l}} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{m_n u}}{\tilde{\beta}_{m_n u}} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{m_n v}}{\tilde{\beta}_{m_n v}} \right]}.$$

$\mathcal{C}_j$ will result in different values of SINR for UEs (denoted $\gamma_k^{\{\mathcal{C}_j\}}$). Following the same derivation steps used to derive (4.7), $\gamma_k^{\{\mathcal{C}_j\}}$ of the k-th UE related to y_k in (4.6) and under the j-th AP clustering configuration can be expressed by 4.9 at the top of next page, where $|\tilde{g}_{m_n k}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_n k}, \tilde{\beta}_{m_n k})$, $|\tilde{g}_{m_n l}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_n l}, \tilde{\beta}_{m_n l})$, $|\tilde{g}_{m_n v}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_n v}, \tilde{\beta}_{m_n v})$ and $|\tilde{g}_{m_n u}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_n u}, \tilde{\beta}_{m_n u})$ with $\tilde{\alpha}_{m_n k} = \mathcal{M}_{m_n k}$, $\tilde{\beta}_{m_n k} = \frac{\mathcal{M}_{m_n k} \dot{\sigma}_{m_n k} L^{2\kappa}_{m_n k}}{\Omega_{m_n k} \mathcal{F}_{m_n k} w^2_{nk} G^2_{m_n k} \tau_p \rho_k p_k \mathcal{E}^2_{m_n k}}$, $\tilde{\alpha}_{m_n l} = \mathcal{M}_{m_n l}$, $\tilde{\beta}_{m_n l} = \frac{\mathcal{M}_{m_n l} \dot{\sigma}_{m_n k} L^{2\kappa}_{m_n l}}{\Omega_{m_n l} \mathcal{F}_{m_n l} w^2_{nk} G^2_{m_n k} \tau_p \rho_l p_l \mathcal{E}^2_{m_n l}}$, $\tilde{\alpha}_{m_n v} = \mathcal{M}_{m_n v}$, $\tilde{\beta}_{m_n v} = \frac{\mathcal{M}_{m_n v} \dot{\sigma}_{m_n k} L^{2\kappa}_{m_n v}}{\Omega_{m_n k} \mathcal{F}_{m_n v} w^2_{nk} G^2_{m_n k} \tau_p \rho_v p_k \mathcal{E}^2_{m_n k} |\boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_v|^2}$, $\tilde{\alpha}_{m_n u} = \mathcal{M}_{m_n u}$, $\tilde{\beta}_{m_n u} = \frac{\mathcal{M}_{m_n q} \dot{\sigma}_{m_n k} L^{2\kappa}_{m_n u}}{\Omega_{m_n q} \mathcal{F}_{m_n u} w^2_{nk} G^2_{m_n k} \tau_p \rho_u p_q \mathcal{E}^2_{m_n q} |\boldsymbol{\varphi}_q^H \boldsymbol{\varphi}_u|^2}$, with

$$\dot{\sigma}_{m_n k} = \sum_{n=1}^{\mathcal{N}} w^2_{nk} \sum_{m_n=1}^{\mathcal{D}_n} G^2_{m_n k} \left[\sum_{t=1}^{\tau_p} \left(p_k \mathcal{E}^2_{m_n k} \varphi_{k,t} + \sum_{l=1, l\neq k}^{K} p_l \mathcal{E}^2_{m_n l} \varphi_{l,t} \right) + 1 \right]. \tag{4.10}$$

Generally, the performance of a dynamic cell-free network will depend on the clustering algorithm that assigns APs to clusters. **Corollary 4** below states the outage performance of a dynamic cell-free network under a *random clustering scheme*. This *random AP clustering* refers to the case where the APs are allocated to different clusters randomly. That is, in a cell-free network with M APs and $\mathcal{N} \leq M$ clusters, APs are assigned to the $\mathcal{N}$ clusters

randomly such that every cluster contains at least one AP.

Note that when the APs are partitioned into clusters based on some other criterion (e.g. based on relative channel gains of the UEs to the different APs), the distribution of the SINR will need to be obtained by using the theory of order statistics [16].

Corollary 4. *If* $|\tilde{g}_{m_nk}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_nk}, \tilde{\beta}_{m_nk})$, $|\tilde{g}_{m_nl}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_nl}, \tilde{\beta}_{m_nl})$, $|\tilde{g}_{m_nu}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_nu}, \tilde{\beta}_{m_nu})$ *and* $|\tilde{g}_{m_nv}|^2 \sim \mathcal{G}(\tilde{\alpha}_{m_nv}, \tilde{\beta}_{m_nv})$ *where* $|\tilde{g}_{m_nk}|^2$, $|\tilde{g}_{m_nl}|^2$, $|\tilde{g}_{m_nu}|^2$ *and* $|\tilde{g}_{m_nv}|^2$ *are independent rvs with* $n = 1, \ldots, \mathcal{N}$, $m_n = 1, \ldots, \mathcal{D}_n$ *such that* $1 \leq \mathcal{D}_n \leq M - (\mathcal{N} - 1)$ *and* $\{k, l, u, v\} = 1, \ldots, K$ *such that* $k \neq l \neq u \neq v$. *The probability of outage* $P_{out}^{(k)}$ *of the k-th UE in the proposed dynamic cell-fee network under random AP clustering is identical to that given in* ***Theorem 1****, with* $\dot{\alpha}_{m_nk}$, $\dot{\beta}_{m_nk}$, $\dot{\alpha}_{m_nk'}$, $\dot{\beta}_{m_nk'}$ *defined after (4.9) at the top of next page.*

Proof. This follows due to the fact that equations (4.7) and (4.9) are only differ by the scaling parameters of the rvs. ■

In order to be able to use the derived expressions of probability of outage (**Theorem 1** and **Corollary 1**), we will need the optimized beamforming vectors of all UEs obtained for SIC-based signal detection and diversity combining of DAS signals. In the next section, we propose a novel DRL-based design that finds an efficient clustering configuration and the related beamforming matrix. The proposed design aims at either maximizing the instantaneous normalized sum rate or the instantaneous normalized minimum UE rate considering an SIC-based signal detection and a DAS diversity combing scheme at the CPU.

4.4 Joint Clustering and Beamforming Design

This section formulates the general problem of jointly performing the clustering of APs and designing the beamforming vectors. However, before presenting the problem formulation,

$$\gamma_k^{\{\mathcal{C}_j\}} = \frac{\sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} |\check{g}_{m_n k}|^2}{\sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} \left[\sum_{l=1}^{k-1} |\check{g}_{m_n l}|^2 + \sum_{v=1, v\neq k}^{K} |\check{g}_{m_n v}|^2 + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} |\check{g}_{m_n u}|^2 \right] + 1}, \tag{4.11}$$

---

we describe the SIC technique for detection of the signals received by the CPU and also the diversity technique to combine the signals from multiple DAS antennas.

4.4.1 SIC-Enabled Signal Detection

The main idea of power-domain SIC is to use power control such that signals from different UEs have distinct power levels[2]. To detect signal from a certain UE, the SIC unit first detects signals of UEs with higher power levels, subtracts their contributions from the overall received signal, and then detects the intended signal. Specifically, for the proposed dynamic cell-free network, when detecting the signal from the k-th UE, we first arrange the received signal components from the UEs in an ascending order as $\sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} |\check{g}_{m_n 1}|^2 \leq \ldots \leq \sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} |\check{g}_{m_n k}|^2 \leq \ldots \leq \sum_{n=1}^{\mathcal{N}} \sum_{m_n=1}^{\mathcal{D}_n} |\check{g}_{m_n K}|^2$. Then, the beamforming vector related to the k-th UE (denoted by $\boldsymbol{w}_k = [w_{1k} \ldots w_{\mathcal{N}k}]$) is found such that a certain objective function of $\gamma_k^{\{\mathcal{C}_j\}}$, $\forall\, k = 1, \ldots, K$ is maximized, where $\gamma_k^{\{\mathcal{C}_j\}}$ in (4.9) is modified as (4.11) at the top of next page, in which $|\check{g}_{m_n k}|^2 \sim \mathcal{G}(\check{\alpha}_{m_n k}, \check{\beta}_{m_n k})$, $|\check{g}_{m_n l}|^2 \sim \mathcal{G}(\check{\alpha}_{m_n l}, \check{\beta}_{m_n l})$, $|\check{g}_{m_n v}|^2 \sim \mathcal{G}(\check{\alpha}_{m_n v}, \check{\beta}_{m_n v})$ and

$|\check{g}_{m_n u}|^2 \sim \mathcal{G}(\check{\alpha}_{m_n u}, \check{\beta}_{m_n u})$ with $\check{\alpha}_{m_n k} = \mathcal{M}_{m_n k}$, $\check{\beta}_{m_n k} = \frac{\mathcal{M}_{m_n k} \dot{\sigma}_{m_n k} L_{m_n k}^{2\kappa}}{\Omega_{m_n k} \mathcal{F}_{m_n k} G_{m_n k}^2 \tau_p \rho_k p_k \mathcal{E}_{m_n k}^2}$, $\check{\alpha}_{m_n l} = \mathcal{M}_{m_n l}$, $\check{\beta}_{m_n l} = \frac{\mathcal{M}_{m_n l} \dot{\sigma}_{m_n k} L_{m_n l}^{2\kappa}}{\Omega_{m_n l} \mathcal{F}_{m_n l} G_{m_n k}^2 \tau_p \rho_l p_l \mathcal{E}_{m_n l}^2}$, $\check{\alpha}_{m_n v} = \mathcal{M}_{m_n v}$, $\check{\beta}_{m_n v} = \frac{\mathcal{M}_{m_n k} \dot{\sigma}_{m_n k} L_{m_n v}^{2\kappa}}{\Omega_{m_n k} \mathcal{F}_{m_n v} G_{m_n k}^2 \tau_p \rho_v p_k \mathcal{E}_{m_n k}^2 |\boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_v|^2}$, $\check{\alpha}_{m_n u} = \mathcal{M}_{m_n u}$, $\check{\beta}_{m_n u} = \frac{\mathcal{M}_{m_n q} \dot{\sigma}_{m_n k} L_{m_n u}^{2\kappa}}{\Omega_{m_n q} \mathcal{F}_{m_n u} G_{m_n k}^2 \tau_p \rho_u p_q \mathcal{E}_{m_n q}^2 |\boldsymbol{\varphi}_q^H \boldsymbol{\varphi}_u|^2}$, with

$$\dot{\sigma}_{m_n k} = \sum_{n=1}^{\mathcal{N}} w_{nk} \sum_{m_n=1}^{\mathcal{D}_n} G_{m_n k}^2 \left[\sum_{t=1}^{\tau_p} \left(p_k \mathcal{E}_{m_n k}^2 \varphi_{k,t} + \sum_{l=1, l\neq k}^{K} p_l \mathcal{E}_{m_n l}^2 \varphi_{l,t} \right) + 1 \right]. \tag{4.12}$$

[2]This is similar to traditional power-domain non-orthogonal multiple access (PD-NOMA).

We can notice from (4.11) that the CSI estimation error contains portions from all UEs including those with higher power levels than that of the k-th UE. This is because, the SIC operation is also affected by channel estimation errors [87].

4.4.2 Diversity Combining Scheme

Several diversity combining techniques can be used at multi-antenna receivers to combat multipath fading. However, MRC and SC are the most common diversity combining techniques [81]. In this work, we propose the use of a modified version of *Wiener-Hopf* multiple antennas combining scheme under the existence of co-channel interference [88]. Accordingly, the gain factor $\boldsymbol{G}_{nk} = [G_{1_n k} \dots G_{m_n k} \dots G_{\mathcal{D}_n k}]^H$ that is used in combining signals at the n-th virtual AP for the purpose of detection of the k-th UE signal is defined as [88]

$$\boldsymbol{G}_{nk} = \boldsymbol{R}_{nk}^{-1} \hat{\boldsymbol{g}}_{nk}, \tag{4.13}$$

where $\boldsymbol{R}_{nk} = \text{Cov}\left(\sum_{l=1, l\neq k}^{K} \sqrt{p_l}\hat{\boldsymbol{g}}_{nl} + \tilde{\boldsymbol{\eta}}_n\right)$, $\hat{\boldsymbol{g}}_{nl} = [\hat{g}_{1_n l} \dots \hat{g}_{\mathcal{D}_n l}]^H$, and $\tilde{\boldsymbol{\eta}}_n = [\tilde{\eta}_{1_n} \dots \tilde{\eta}_{\mathcal{D}_n}]^H$. Furthermore, the covariance matrix $\boldsymbol{R}_{nk}$ conditioned on instantaneous CSI can be written as

$$\boldsymbol{R}_{nk} = \sum_{l=1, l\neq k}^{K} p_l \hat{\boldsymbol{g}}_{nl} \hat{\boldsymbol{g}}_{nl}^H + \tilde{\boldsymbol{\Sigma}}_n, \tag{4.14}$$

where $\tilde{\boldsymbol{\Sigma}}_n \in \mathbb{C}^{\mathcal{D}_n \times \mathcal{D}_n}$ is a diagonal matrix with m_n-th diagonal element given by $\tilde{\Sigma}_{m_n} = \tilde{\sigma}_{m_n}^2/2$. Note that, when $K = 1$, we have $G_{m_n k} = 2\hat{g}_{m_n k}/\tilde{\sigma}_{m_n}$ which is identical to the MRC scheme that represents the optimal combining scheme under no IUI. Additionally, note that we use the estimated values of CSI $\hat{g}_{m_n k}$ without considering the CSI estimation error. Several works in the literature investigated on optimal combining under imperfect CSI estimation [89, 90]. For brevity, we do not consider this in this chapter.

4.4.3 Optimization Problem Formulation

AP clustering and beamforming should be performed jointly in order to achieve the best possible performance. For this, a general optimization problem can be formulated and solved that jointly finds the optimal cluster set and the corresponding beamforming vectors given an objective function[3]. Accordingly, the general problem of AP clustering and beamforming design can be formulated as

$$\mathbf{P}_1 : \max_{j\in[1,\ldots,\Theta(M,N)],\boldsymbol{W}\in[0\ 1]^{N\times K}} \quad \sum_{k=1}^{K}\log_2\left(1+\gamma_k^{\{\mathcal{C}_j\}}\right)$$

Subject to:

$$\mathbf{C}_1 : \sum_{n=1}^{\mathcal{N}}\left(w_{n\delta_l}^2 - \sum_{i=\delta_l+1}^{l} w_{ni}^2\right)\bar{\gamma}_{m_n l} \geq P_{\mathrm{s}}, \tag{4.15}$$

$$\mathbf{C}_2 : ||\boldsymbol{w}_k||^2 \leq 1,$$

$$\forall\ k, \forall\ \delta_l = 1,\ldots,l-1 \text{ and } l = 2,\ldots,K,$$

where $\Theta(M,\mathcal{N})$ is the maximum number of possible AP clustering configurations which is a function of M and $\mathcal{N}$ (see Sec. 4.4.4), $\boldsymbol{W} \in [0\ 1]^{\mathcal{N}\times K}$ is the beamforming matrix in which the k-th column is used in detecting the received signal component related to the k-th UE, $\bar{\gamma}_{m_n l} = \frac{2p_l G_{m_n l}}{\tilde{\sigma}_{m_n}^2}|\hat{g}_{m_n l}|^2$, $\boldsymbol{w}_k = [w_{1k}\ldots w_{\mathcal{N}k}]$ and $\gamma_k^{\{\mathcal{C}_j\}}$ is given by (4.11). The constraints $\mathbf{C}_1$ refer to the set of $\sum_{l=2}^{K}(l-1) = \frac{K(K-1)}{2}$ conditions required for successful SIC operation with receiver sensitivity of P_{s}. For maximizing the instantaneous minimum normalized rate, the objective function of problem $\mathbf{P}_1$ in (4.15) will be replaced by $\min_k \log_2\left(1+\gamma_k^{\{\mathcal{C}_j\}}\right)$ and the maximization will be with respect to the k-th column of $\boldsymbol{W}$. Note that the receiver deals only with measured channel values (estimated) which include the estimation error as well as the AWGN part. However, the achieved SINR value after the SIC procedure will be decreased by pilot contamination components (as can be noticed from Eq. (4.11) where

[3]In this chapter, the system performance is evaluated for two objective functions: (i) instantaneous normalized sum rate or instantaneous spectral efficiency and (ii) instantaneous normalized minimum per-UE rate.

pilot contamination of signals 1 through $k-1$ was not removed by the SIC process). To optimally solve $\mathbf{P}_1$ in (4.15), we need to simultaneously solve for $\mathcal{C}_j$ and $\boldsymbol{W}$. Specifically, for every possible clustering configuration, there is a related optimal beamforming vector that maximizes the objective function. *The globally optimal solution is the one that gives the best performance among all clustering configurations and the corresponding optimized beamforming vectors ($\boldsymbol{W}$).*

4.4.4 Complexity Analysis

With M APs and K UEs, there will be a beamforming vector for each UE, $\boldsymbol{w}_k = [w_{1k}, \ldots, w_{Mk}]$, to decode its transmitted signal at the CPU. With a specified precision, the beamforming vectors can be obtained by an exhaustive search in which $w_{mk} \in [0\ 1]\,,\ \forall\ m,k$ is partitioned into a set of possible values separated by a certain step size Δ. Accordingly, the optimization problem related to finding optimal values of beamforming vectors of all UEs will have a complexity of $O\left(\left(\frac{1}{\Delta}\right)^{M+K}\right)$. When the cell-free network is clustered into $\mathcal{N} \leq M$ clusters, optimization of the beamforming vectors will have a complexity of $O\left(\left(\frac{1}{\Delta}\right)^{\mathcal{N}+K}\right)$ resulting in a reduction of complexity of $O((\frac{1}{\Delta})^{M-\mathcal{N}})$. Clustering of APs will need to be performed in a time-slot basis based on the instantaneous CSI in the network. For the proposed dynamic cell-free network with M APs and $\mathcal{N}$ clusters, there will be $\Theta\left(M,\mathcal{N}\right) = \mathcal{N}!\,S\left(M,\mathcal{N}\right)$ possible configurations for AP clusters, where $S\left(M,\mathcal{N}\right)$ is the Stirling number, which can be calculated by [91]

$$S\left(M,\mathcal{N}\right) = \frac{1}{\mathcal{N}!}\sum_{i=0}^{\mathcal{N}}(-1)^i\binom{\mathcal{N}}{i}\left(\mathcal{N}-i\right)^M. \tag{4.16}$$

Accordingly, solving Problem $\mathbf{P}_1$ by an exhaustive search will require us going through all possible AP clustering configurations for every given CSI. Therefore, for practical implementation of a dynamic cell-free network, we design a novel deep reinforcement learning (DRL) system based on hybrid DDPG-DDQN model to jointly generate the clustering configura-

tion and the beamforming vectors with an objective to solving the optimization problem in (4.15).

To illustrate the complexity reduction due to dynamic clustering, let us consider an example with $M = 8$, $\mathcal{N} = 3$ and any arbitrary number of UEs K, one possible AP clustering is

$$\mathcal{C}_j = \{\{\overbrace{\mathrm{AP}_1}^{m_{n=1}=1}, \mathrm{AP}_4, \mathrm{AP}_5, \mathrm{AP}_7\}, \{\mathrm{AP}_2\}, \{\mathrm{AP}_3, \mathrm{AP}_6, \mathrm{AP}_8\}\}.$$
$$\underbrace{\qquad\qquad}_{\mathcal{D}_{n=1}=4} \quad \underbrace{\qquad}_{\mathcal{D}_{n=2}=1} \quad \underbrace{\qquad\qquad}_{\mathcal{D}_{n=3}=3}$$

For this example, Fig. 4.2 illustrates the reduction in computational complexity for optimizing the beamforming vectors at the CPU. Here the size of each of the beamforming vectors reduces from 8 to 3. Accordingly, for a precision of $\Delta = 0.01$, the complexity reduction is of $O\left(10^{10}\right)$.

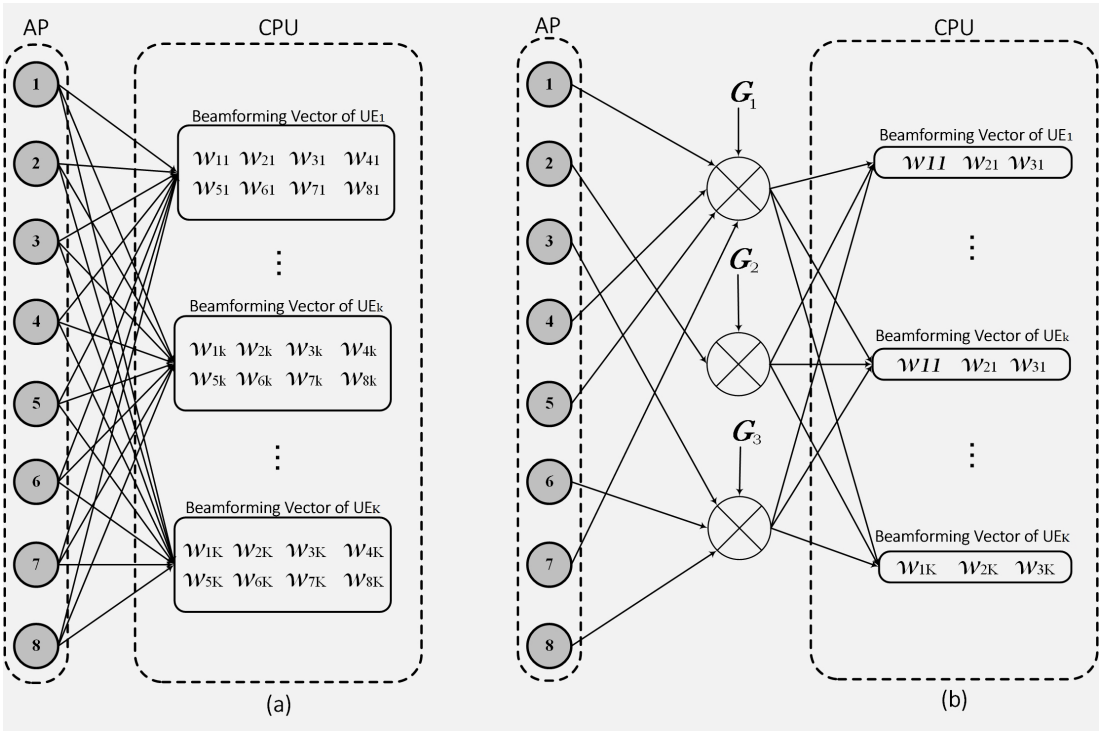

Figure 4.2 Complexity reduction for beamforming optimization in a cell-free network (with $M = 8$ and $\mathcal{N} = 3$): (a) Centralized scenario, (b) Dynamic clustering scenario.

4.5 DRL-Based AP Clustering and Beamforming Design

In this section, we introduce a novel DRL model that solves the general optimization problem (i.e. problem $\mathbf{P}_1$ in (4.15) of jointly optimizing the AP clustering and the beamforming vectors for the UEs.

4.5.1 Theoretical Preliminaries

The concept of reinforcement learning (RL) refers to the learning process of an agent interacting with its environment after receiving certain observations. The environment provides a reward to the agent for every interaction and the RL agent aims to select the right action for the next interaction in order to maximize the discounted reward over a time horizon. This problem can be formulated as a *Markov decision process* (MDP). An MDP is a tuple ($\mathcal{S}$, $\mathcal{A}$, $\mathcal{P}$, $\mathcal{R}$, ζ), where $\mathcal{S}$ represents a K-dimensional (with each state at time t denoted by $\boldsymbol{s}_t$), $\mathcal{A}$ is the action space that contains a finite set of actions from which the agent can choose, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is a transition probability in which $\mathcal{P}(\boldsymbol{s}, a, \boldsymbol{s}')$ defines the probability of observing state $\boldsymbol{s}'$ after executing action a in the state $\boldsymbol{s}$, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the expected reward after being in state $\boldsymbol{s}$ and taking action a, and $\zeta \in [0, 1)$ is the discount factor. To solve the MDP, RL algorithms have been developed to learn and find a discrete value function or a "policy". Such a discretization can lead to lack of generalization and significantly increase the problem dimensionality. Therefore, DRL algorithms based on function approximation by deep neural networks (DNNs) have been proposed.

DRL algorithms can be classified into three types: (i) *value-based* methods such as deep Q-learning (DQL) and SARSA which only learn the so-called value function to find a policy, (ii) *policy-based* methods which learn the policy directly by following the gradient with respect to the policy, and (iii) *actor-critic* methods which are a hybrid of the value-based for the critic and policy-based methods for the actor.

The standard DQL method is the most popular algorithm in the literature due to the availability of a mature theory. For an agent with parameters θ^Q at time t, after its takes action a_t in state $\boldsymbol{s}_t$ and observes the immediate reward r_{t+1} with the resulting state $\boldsymbol{s}_{t+1}$, the DQL update equation is:

$$\begin{aligned} Q(\boldsymbol{s}, a \,|\, \theta^Q_{t+1}) &= Q(\boldsymbol{s}, a \,|\, \theta^Q_t) + \nu \Big[r_{t+1} + \zeta \max_{a'} Q(\boldsymbol{s}_{t+1}, a' \,|\, \theta^Q_t) - Q(\boldsymbol{s}_t, a_t \,|\, \theta^Q_t) \Big] \\ &= Q(\boldsymbol{s}, a \,|\, \theta^Q_t) + \nu \Big[r_{t+1} + \zeta \max_{a'} Q(\boldsymbol{s}_{t+1}, \operatorname*{argmax}_{a'} Q(\boldsymbol{s}_{t+1}, a' \,|\, \theta^Q_t) \,|\, \theta^Q_t) \\ &\quad - Q(\boldsymbol{s}_t, a_t \,|\, \theta^Q_t) \Big], \end{aligned} \tag{4.17}$$

where ν is the learning rate. Computing the term $\max_{a'} \; Q(\boldsymbol{s}_{t+1}, \operatorname*{argmax}_{a'} Q(\boldsymbol{s}_{t+1}, a' \,|\, \theta^Q_t) \,|\, \theta^Q_t)$ introduces a systematic overestimation of the Q-values during the learning that is accentuated by the use of bootstrapping, i.e. learning estimates from estimates. The Q-learning update in (4.17) uses the same Q-network $Q(\boldsymbol{s}, a \,|\, \theta^Q_t)$ both to select and to evaluate an action. After highlighting the overestimation bias in experiments across different Atari game environments, Hasselt et al. [71] decoupled the action selection and evaluation by introducing two deep Q-networks, a Q network and a target network Q' with different parameters θ^Q and $\theta^{Q'}$, respectively, to avoid the maximization bias. The Q' network is used for action selection while the Q network is used for action evaluation. This is known as *deep double Q-learning algorithm (DDQL)*. The DDQL update equation can be expressed as:

$$\begin{aligned} Q(\boldsymbol{s}, a \,|\, \theta^Q_{t+1}) = Q(\boldsymbol{s}, a \,|\, \theta^Q_t) + \nu \Big[r_{t+1} + \zeta \max_{a'} Q(\boldsymbol{s}_{t+1}, \operatorname*{argmax}_{a'} Q'(\boldsymbol{s}_{t+1}, a' \,|\, \theta^{Q'}_t) \,|\, \theta^Q_t) \\ - Q(\boldsymbol{s}_t, a_t \,|\, \theta^Q_t) \Big]. \end{aligned} \tag{4.18}$$

The parameters $\theta^{Q'}$ of the Q' network periodically hard-copy the parameters θ^Q of Q network after t_0 time steps using the Polyak averaging method with parameter $\tau \in [0, 1]$:

$$\theta^{Q'}_{t+t_0} = (1 - \tau)\, \theta^{Q'}_t + \tau\, \theta^Q_t. \tag{4.19}$$

DDQL shows a better performance than standard DQL [71]; however, due to the discretization requirements of the DNN outputs (the action space $\mathcal{A}$), it results in a huge expansion of the action space dimensionality when used in the optimization of an objective function of continuous dependent variables. *This dimensionality issue makes it unattractive for solving the beamforming problem under massive number of UEs and APs.* However, it is a relevant candidate for the clustering problem of the APs since it avoids the need for the exhaustive search method. This motivates the utilization of the "DDPG" policy in beamforming design problem.

DDPG belongs to the class of actor-critic algorithms. It concurrently learns a Q-function network approximation $Q(\boldsymbol{s}, a|\theta^Q)$ called the critic, and a policy network approximation $\mu(\boldsymbol{s}|\theta^\mu)$ called the actor. The Q-function network is trained using the Bellman equation, while the policy network is learnt using the Q-function. Unlike the DQL policies which output the probability distribution $\pi(a|\boldsymbol{s})$ across a discrete action space $\mathcal{A}$, the policy network of DDPG directly maps states to actions. Specifically, at every time step t, it maximizes its loss function defined as:

$$J(\theta) = \mathbb{E}\Big[Q(\boldsymbol{s}, a) \mid \mathcal{S} = \boldsymbol{s}_t, a = \pi(a|\boldsymbol{s}_t)\Big] \tag{4.20}$$

and updates its weights θ by following the gradient of (4.20):

$$\nabla J_{\theta^\mu}(\theta) \approx \nabla_a Q(\boldsymbol{s}, a)\, \nabla \mu(s|\theta^\mu). \tag{4.21}$$

This update rule represents the deterministic policy gradient (DPG) theorem, rigorously proved by Silver et al. in the supplementary material of [72]. The term $\nabla_a Q(\boldsymbol{s}, a)$ is obtained from a Q-network $Q(\boldsymbol{s}, a|\theta^Q)$ called the critic by backpropagating its output w.r.t. the action input $\mu(\boldsymbol{s}|\theta^\mu)$. When the number of actions is very large, this actor-critic training procedure solves the intractability problem of DQN [73] by using the following approximation:

$$\max_a Q(\boldsymbol{s}, a) \approx Q(\boldsymbol{s}, a|\theta^Q)|_{a=\mu(\boldsymbol{s}|\theta^\mu)}. \tag{4.22}$$

Similar to DQN, two tricks are employed to stabilize the training of the DDPG actor-critic architecture, namely, 1) the experience replay buffer R to train the critic, and 2) target networks for both the actor and the critic which are updated using the Polyak averaging in the same way as in (4.19). Now that we have provided a brief DRL background on the two methods used by our proposed algorithm (DDQL and DDPG), a detailed description of the proposed DRL-based AP clustering and beamforming design for a dynamic cell-free network will be presented in the following subsection.

4.5.2 Agent Design for AP Clustering and Beamforming

Our goal is to design a DRL system that jointly optimizes the clustering of APs and the beamforming vectors given a certain CSI matrix $\boldsymbol{H} = [\hat{\boldsymbol{g}}_1 \dots \hat{\boldsymbol{g}}_K]$. In this context, we develop a hybrid DDPG-DDQL DRL system that simultaneously learns the best clustering subsets-beamforming vector pairs given a certain CSI matrix $\boldsymbol{H}$ (or any other metric related to $\boldsymbol{H}$). Fig. 4.3 shows a schematic block diagram of the developed hybrid DDPG-DDQL DRL scheme. For a certain AP clustering configuration, we opt for the actor-critic DDPG algorithm [92] to find the optimized beamforming vector $\boldsymbol{\omega} = \text{Reshape}\,(\boldsymbol{W})$, where $\boldsymbol{\omega} \in \mathbb{R}^{\mathcal{N}K\times 1}$ and $\boldsymbol{W} \in \mathbb{R}^{\mathcal{N}\times K}$. Note that all elements of $\boldsymbol{\omega}$ are continuous within the range [0 1] which motivates us to use the DDPG learning algorithm that has the property of dealing with continuous action space rather than discrete action space with probabilistic outcomes. For the AP clustering problem, we use the well-known DDQL algorithm [71] to find the best AP clustering configuration since the possible number of clustering configurations is finite and the index of each possible configuration is integer.

As shown in Fig. 4.3, the two algorithms interact with a simulated cell-free network environment to solve the optimization problem $\mathbf{P}_1$ in (4.15). The design of the cell-free network environment involves the specification of the environment state $\boldsymbol{s}$ and the definition of immediate reward function r required by the DRL algorithms to approximate the

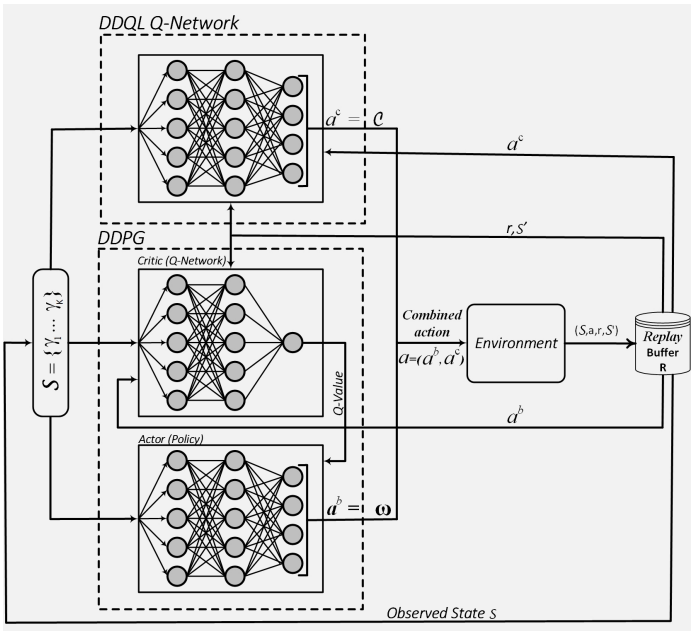

Figure 4.3 Hybrid DDPG-DDQL model for AP clustering and uplink beamforming design.

policies and the Q-values. The state of the environment is a vector of the SINR values of UEs, where the state s_k of the k-th UE is chosen to be the SINR value $\gamma_k^{\{\mathcal{C}_j\}}$ which varies as a function of the instantaneous CSI matrix. Under this setup, the action space $\mathcal{A}$ is the pair of actions $a = (a^b, a^c) = (\boldsymbol{\omega}, \mathcal{C}_j)$. The vector $\boldsymbol{\omega}$ represents the continuous action of the DDPG algorithm and the discrete clustering configuration $\mathcal{C}_j$ is the output of the DDQL algorithm. The superscripts b and c refer to "*beamforming*" and "*clustering*", respectively. After receiving the environment state $\boldsymbol{s}$, the DDPG algorithm outputs the action $a^b = \boldsymbol{\omega}$ and the DDQL outputs the action a^c representing the partition $\mathcal{C}_j$. Table 5.2 summarizes the environment design by specifying the additional problem parameters. *Note that, when the number of clusters equals to the total number of APs ($\mathcal{N} = M$), the network will reduce to the centralized cell-free network with $\mathcal{C}_j = \mathcal{C} = \{\{\text{AP}_1\}, \ldots, \{\text{AP}_M\}\}$.* Accordingly, only the DDPG model will be required to optimize the beamforming vectors ($\boldsymbol{\omega}$).

Our TensorFlow/Keras implementation of the actor and critic networks (including the corresponding target networks) have two hidden layers of 256 and 128 neurons, respectively.

Table 4.1 DRL agent design

Environment Variables	System Equivalence
State $\boldsymbol{s} = \{s_1, \ldots, s_K\}$	$\{\gamma_1^{\mathcal{C}_j}, \ldots, \gamma_K^{\mathcal{C}_j}\}$
Reward r	$\sum_{k=1}^{K} \log\left(1 + \gamma_k^{\mathcal{C}_j}\right)$
Action $\mathcal{A}$	$(a^b, a^c) = (\boldsymbol{\omega}, \mathcal{C}_j)$
$\mathcal{C}_j$	j-th clustering configuration
K	Number of active UEs

The DDQN networks have two fully-connected layers of 64 neurons followed with an activation function *relu* in each, and a final linear fully-connected layer. We use a discount factor $\zeta = 0.99$, a learning rate $\nu = 5 \cdot 10^{-5}$, a Polyak averaging parameter $\tau = 10^{-3}$, and an experience replay buffer of size $R = 20000$. The critic optimizer is *Adam* with its default hyper-parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We train all the networks on 2500 episodes, with 500 time step each.

The offline time-complexity of training a fully connected network is dominated by the matrix multiplications during forward and backward passes. For any given layer i, let m_i, n_i, and b denote the size of the input, the output and the batch. Assuming an $\mathcal{O}(n^3)$ matrix multiplication algorithm, the complexity of forward and backward passes is of $\sum_i \mathcal{O}(m_i\, n_i\, b)$. The order of time-complexity of online inference is $\sum_i \mathcal{O}(m_i\, n_i)$ since $b = 1$ for the RL setting. The offline space-complexity is linear in the number of the input neurons only since the size of the hidden layers is constant.

4.5.3 Description of the Hybrid DDPG-DDQL Algorithm

For the centralized cell-free network, only the DDPG algorithm is trained. However, for the dynamic cell-free network, both the DDPG and DDQL networks are trained.

- DDPG network:
  - *The actor network* $\mu(\boldsymbol{s}|\theta^{\mu})$ maps the SINR values of UEs to the beamforming matrix $\boldsymbol{\omega}$. The output of the network is a^b, i.e. a flat list of all the elements of $\boldsymbol{\omega} \in [0, 1]$.
  - *The target actor network* $\mu'(\boldsymbol{s}|\theta^{\mu'})$: time-delayed copy of the actor network $\mu(\boldsymbol{s}|\theta^{\mu})$.
  - *The critic network* $Q(\boldsymbol{s}, a^b|\theta^{Q})$: maps the SINR values and the output action of $\mu(\boldsymbol{s}|\theta^{\mu})$ to their corresponding Q-value.
  - *The target critic network* $Q'(\boldsymbol{s}, a^b|\theta^{Q'})$: time-delayed copy of the critic network $Q(\boldsymbol{s}, a^b|\theta^{Q})$.
- DDQL network:
  - *The* Q_c*-network* $Q_c(\boldsymbol{s}, a^c|\theta^{Q_c})$ maps the SINR values of UEs to the the Q-values of the state and all the clustering partitions.
  - *The target* Q_c*-network* $Q'_c(\boldsymbol{s}, a^c|\theta^{Q'_c})$: time-delayed copy of the Q_c-network $Q(\boldsymbol{s}, a^c|\theta^{Q_c})$.

We describe all of the training steps of our algorithm in **Algorithm 3**. We start by initializing all neural networks and their targets for the beamforming and clustering problems as well as a replay buffer R (lines 1–5). For every episode, we initialize the environment with N access points and K UEs by setting the initial state $\boldsymbol{s}_0$ to a random vector of SINR values of size K (line 7). At every time step t of the episode, the DDQL and DDPG agents pick an action a_t^c and a_t^b, respectively (lines 9–10). The combined action $a_t = (a_t^b, a_t^c)$ is sent to the cell-free network environment which will transit to a new state $\boldsymbol{s}_{t+1}$, and this new state will be returned together with the immediate reward r_t (lines 11–12). After storing the transition tuple $(\boldsymbol{s}_t, a_t, r_t, s_{t+1})$ in the replay buffer R (line 13), we randomly sample from the experience replay buffer N transitions to train the DDPG and DDQL networks (line 17). We start the DDPG training in line (16) by computing the temporal-difference (TD) target

for the Q-network $Q(\boldsymbol{s}, a|\theta^Q)$ using the target Q-network $Q'(\boldsymbol{s}, a|\theta^{Q'})$. We update the critic $Q(\boldsymbol{s}, a|\theta^Q)$ parameters θ^Q in line 17 using the gradient of the MSE of the loss function of the TD target and the output of the critic. The update of the actor parameters θ^μ uses the Monte Carlo approximation of gradient in line 4.21. The target critic and target policy networks are updated slowly every P iterations (lines 19). Finally, we update the parameters θ^{Q_c} of the DDQL Q-network using the Bellman equation in line (21) after selecting the action using the target Q-network $Q'_c(\boldsymbol{s}, a|\theta^{Q'_c})$ in line 20. Similar to the DDPG target network, we update in line 21 the DDQL target Q-network every P iterations. Note that the DRL model can be easily extended to optimize also the uplink transmission power for UEs $P_k, \forall k = 1, \ldots, K$. However, the power control information will then need to be transmitted to the UEs.

4.6 Numerical Results and Discussions

This section contains a detailed simulations of the proposed scheme and the derived expressions along with insights and technical notes.

4.6.1 Parameters and Assumptions

Table 4.2 Simulation parameters

Parameter	Value
AWGN PSD per UE	-169 dBm/Hz
Path-loss exponent, κ	2
Nakagami fading parameters, $(\mathcal{M}, \Omega)$	$(1, 1)$
Training sequence length, τ_p	K Samples
Pilot transmission power, ρ_k	100 mW, $\forall\ k$
SIC sensitivity, P_s	1 dBm

Algorithm 3 Hybrid DDPG-DDQL algorithm for uplink beamforming and clustering

Randomly initialize the critic $Q(\boldsymbol{s}, a|\theta^Q)$ and the actor $\mu(\boldsymbol{s}|\theta^\mu)$ with weights θ^Q and θ^μ

Initialize target network Q' and μ' with weights: $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$

Randomly initialize the Q_c-network $Q_c(\boldsymbol{s}, a|\theta^{Q_c})$

Initialize the target network $Q'_c(\boldsymbol{s}, a|\theta^{Q'_c})$ with weights: $\theta^{Q'_c} \leftarrow \theta^{Q_c}$

Initialize replay buffer R

for *episode* $= 1, \ldots, E$ **do**

  Receive initial observation state $\boldsymbol{s}_1$ after initializing the environment

  for $t = 1, \ldots, T$ **do**

    Select the beamforming action $a_t^b = \mu\left(\boldsymbol{s}_t|\theta^\mu\right)$

    Select the clustering action $a_t^c = \arg\max_{a^c} Q_c\left(\boldsymbol{s}_t, a^c\right)$

    Define $a_t = (a_t^b, a_t^c)$

    Execute action a_t and observe reward r_t and observe new state $\boldsymbol{s}_{t+1}$

    Store transition $(\boldsymbol{s}_t, a_t, r_t, \boldsymbol{s}_{t+1})$ in R

    Sample a random mini-batch of $\mathcal{L}$ transitions $(\boldsymbol{s}_i, a_i, r_i, \boldsymbol{s}_{i+1})$ from R

    Get a_i^b and a_i^c from a_i

    ▷ **Training the DDPG networks**

    Compute the TD target $y_i = r_i + \zeta\, Q'\left(\boldsymbol{s}_{i+1}, \mu'\left(\boldsymbol{s}_{i+1}|\theta^{\mu'}\right)|\theta^{Q'}\right)$

    Update the critic $Q(\boldsymbol{s}, a|\theta^Q)$ by minimizing the loss: $L = \frac{1}{\mathcal{L}} \sum_i \left(y_i - Q\left(\boldsymbol{s}_i, a_i^b|\theta^Q\right)\right)^2$

    Update the actor policy $\mu(\boldsymbol{s}|\theta^\mu)$ using a Monte-Carlo approximation of (4.21):

$\nabla_{\theta^\mu} J \approx \frac{1}{\mathcal{L}} \sum_i \nabla_a Q\left(\boldsymbol{s}, a|\theta^Q\right)\big|_{\mathcal{S}=\boldsymbol{s}_i, a=\mu(\boldsymbol{s}_i)} \nabla_{\theta^\mu} \mu\left(\boldsymbol{s}|\theta^\mu\right)\Big|_{\mathcal{S}=\boldsymbol{s}_i}$

  Update the DDPG target networks Q' and μ' **if** $mod(t, P) = 0$:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \qquad \text{and} \qquad \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

    ▷ **Training the DDQL networks**

    select $a^* = \arg\max_a Q'_c\left(\boldsymbol{s}_{i+1}, a|\theta^{Q'_c}\right)$

    Update the Q_c using:

$$Q_c(\boldsymbol{s}_i, a_i^c|\theta^{Q_c}) \leftarrow Q_c(\boldsymbol{s}_i, a_i^c|\theta^{Q_c}) + \nu\,\left(r_i + \zeta\, Q_c\left(s_{i+1}, a^*|\theta^{Q_c}\right) - Q_c(\boldsymbol{s}_i, a_i^c|\theta^{Q_c})\right)$$

    Update the DDQL target networks Q'_c **if** $mod(t, P) = 0$: $\theta^{Q'_1} \leftarrow \tau\theta^{Q_1} + (1-\tau)\theta^{Q'_1}$

  end for

end for

Table 4.2 presents the main system parameters used to obtain simulation and analytical results. To simplify simulation and analysis and to concentrate on the most insightful conclusions, we establish some operating assumptions related to channel training and CSI fading models. Additionally, we assume that all channel fading gains $h_{m_n k}$ are i.i.d rvs with $\mathcal{M}_{m_n k} = \mathcal{M} = 1$ and $\Omega_{m_n k} = \Omega = 1$. Similarly, we assume that all AWGN values belong to a set of i.i.d rvs with PSD $\sigma_{m_n}/2 = -169$ dBm/Hz. For large-scale fading, we assume that all APs and UEs are uniformly distributed over a disc of radius 18 m (corresponding to a network total coverage area of 1km^2. Furthermore, to benchmark the proposed DRL scheme, we simulate the proposed system model using *Matlab* and utilize the "*fmincon*" function. This function solves optimization problems with continuous and differentiable multi-dimensional objective functions using the "*bisection method*" for higher dimensions [**?**]. Note that the bisection method is guaranteed to converge to one of the minimum values that lies within the selected initial interval that can be either a local or a global minimum.

4.6.2 Accuracy of *Welch-Satterthwaite* Approximation

We check the accuracy of *Welch-Satterthwaite* approximation of sum of gamma rvs (Fig. 4.4). It can be noticed from Fig. 4.4 that a satisfactory accuracy for the sum of non-identically distributed (different β) gamma rvs can be achieved by *Welch-Satterthwaite* approximation for small, medium, and high numbers of random variables. This justifies the use of *Welch-Satterthwaite* approximation instead the central limit theorem which requires a relatively large number of rvs.

4.6.3 Shadowing Correlation and Channel Estimation Error

Here, we investigate the effect of two important factors that affect the performance of cell-free networks, namely, (a) correlated shadow fading and (b) estimation error resulting from non-

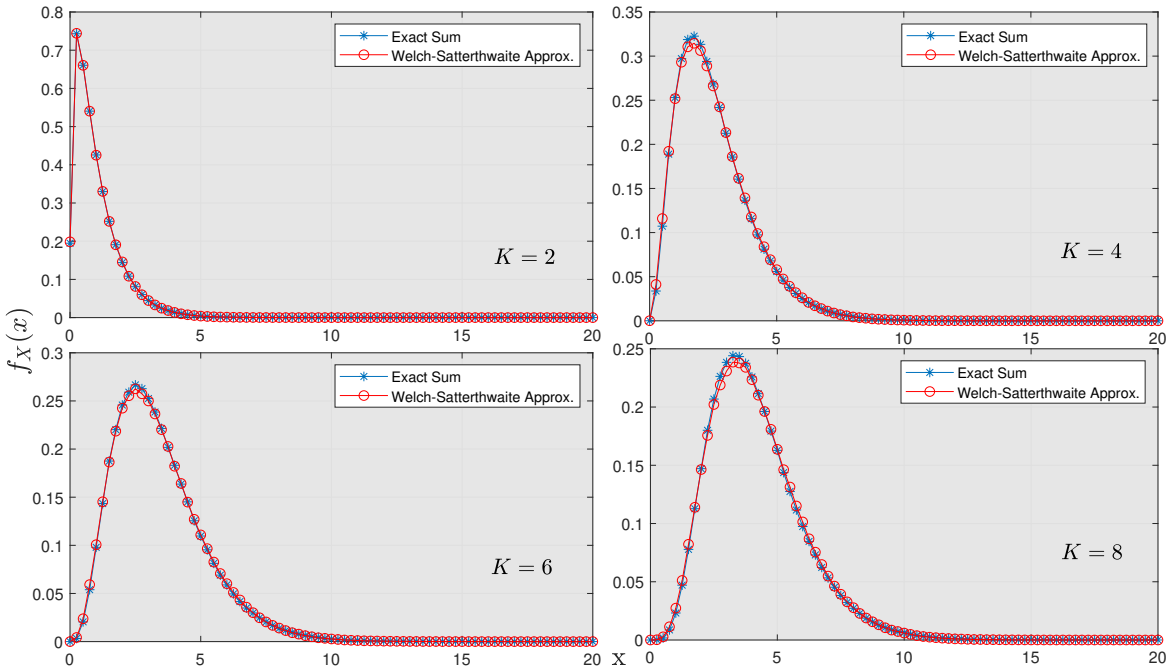

Figure 4.4 An illustrative example of *Welch-Satterthwaite* approximation with different number of rvs (K).

orthogonal pilots. Fig. 4.5(a) shows the effect of transmitter-receiver shadowing correlation under different shadow fading variance $\sigma_{m_n k} = \sigma_{\text{Sh}}$, $\forall\ m, n$ and k. It can be noticed from this figure that when σ_{Sh} is relatively large (8 dB), the effect of shadow fading correlation is relatively small and does not appear for large number of APs (within the region $M \geq 17$). On the other hand, when $\sigma_{\text{Sh}}$is relatively small (4 dB), the effect of shadow fading correlation on per-UE transmission rate becomes more significant and occurs even for small to moderate number of APs (within the region $M \geq 12$).

Fig. 4.5(b) shows the effect of non-orthogonal pilot sequences on the instantaneous per-UE transmission rate under random pilot assignment scheme. It can be noticed from this figure that the per-UE transmission rate decreases as the number of non-orthogonal pilots reduces, with severe performance degradation occurring when the number of non-orthogonal pilots drops bellow $K/2$. This is because every UE will have to deal with more than one non-orthogonal pilot component assigned for other active UEs. Beside, it can be noticed that for large number of non-orthogonal pilot signals (such as $\tau_p \leq K/2$), no noticeable

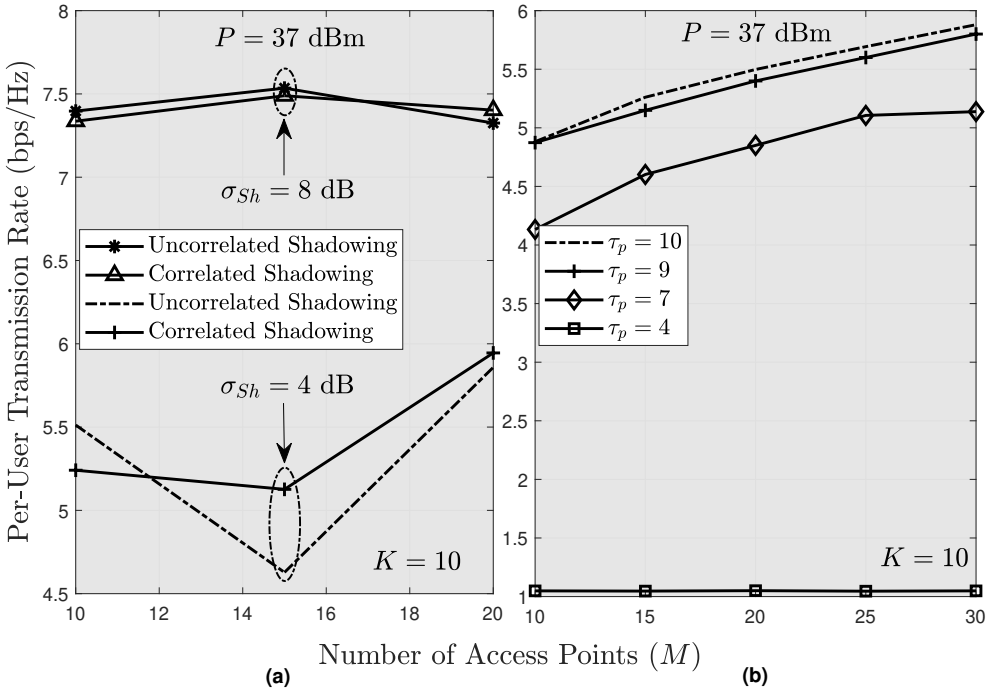

Figure 4.5 Effect of: (a) correlated shadowing and (b) non-orthogonal pilots.

enhancement in per-UE performance can be achieved by increasing the number of APs.

4.6.4 Outage Performance

This section discusses the outage performance of cell-free network in a massive communication regime. Each simulation value is obtained via 2×10^6 Monte-Carlo simulation runs. In Fig. 4.6(a), we evaluate the outage probability for a UE in a centralized cell-free network under different values of K. Fig. 4.6(a) shows that an excellent outage performance can be achieved when the number of APs (M) is much grater than the number of UEs K, i.e. $M >> K$. However, when K becomes comparable to M, the outage performance deteriorates significantly with almost total blockage at $K \approx 0.5M$. Note that this figure is produced assuming no precoding scheme, i.e. $w_{mk} = c, \forall m, k$. A better performance can be achieved by first solving problem $\mathbf{P}_1$ to obtain $\boldsymbol{W}$ and then substituting $\boldsymbol{W}$ into (4.8).

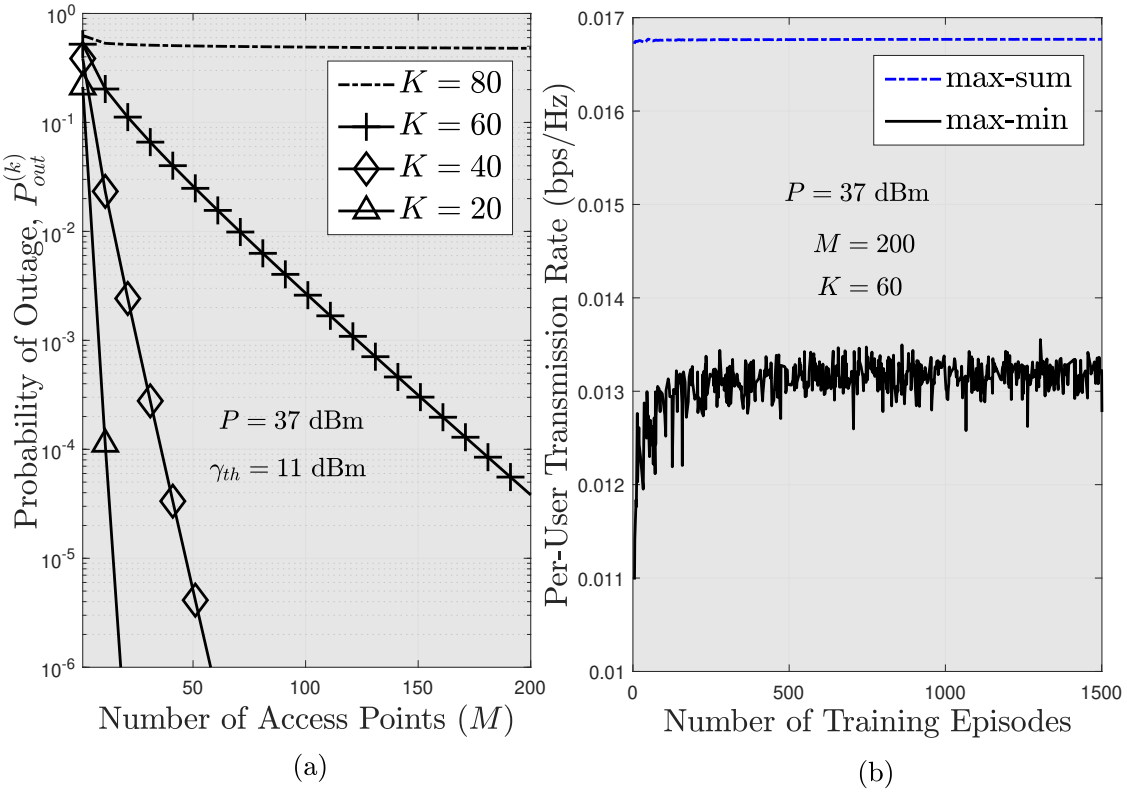

Figure 4.6 UE performance: (a) Probability of outage, (b) Transmission rate per UE.

4.6.5 Performance of DRL Model

In this section, we study the performance of using the proposed DRL model in solving problem $\mathbf{P}_1$. Fig. 4.6(b) shows the per-UE transmission rate in the cell-free network with the proposed DRL method. We can notice that the performance reaches a certain level after around 500 training episodes and then stays around that level. Compared with the max-min objective, we notice a significant performance gain with the max-sum objective. This is due to the fact that, with the max-sum objective, we optimize $\boldsymbol{w}_k, \forall k = 1, \ldots, K$ while with max-min objective, we maximize only $\boldsymbol{w}_{k'}$ and use $\boldsymbol{w}_k, k \neq k'$ & $k = 1, \ldots, K$, where k', denoting the index of the UE with minimum achievable rate, that is found from previous iterations.

To evaluate the performance of the proposed dynamic cell-free network with SIC de-

tection, Fig. 4.7(a) shows the per-UE transmission rate for both centralized and dynamic cell-free networks. It can be noticed that utilizing SIC at the receiver side significantly compensates for the performance loss caused by clustering of the APs. Note that the per-UE transmission rate achieved without SIC in presence of AP clustering is significantly lower than that in a centralized cell-free network.

Additionally, in Fig. 4.7(b), we compare the performance of the proposed DRL-based solution approach compared to traditional methods. We can notice that DRL can achieve

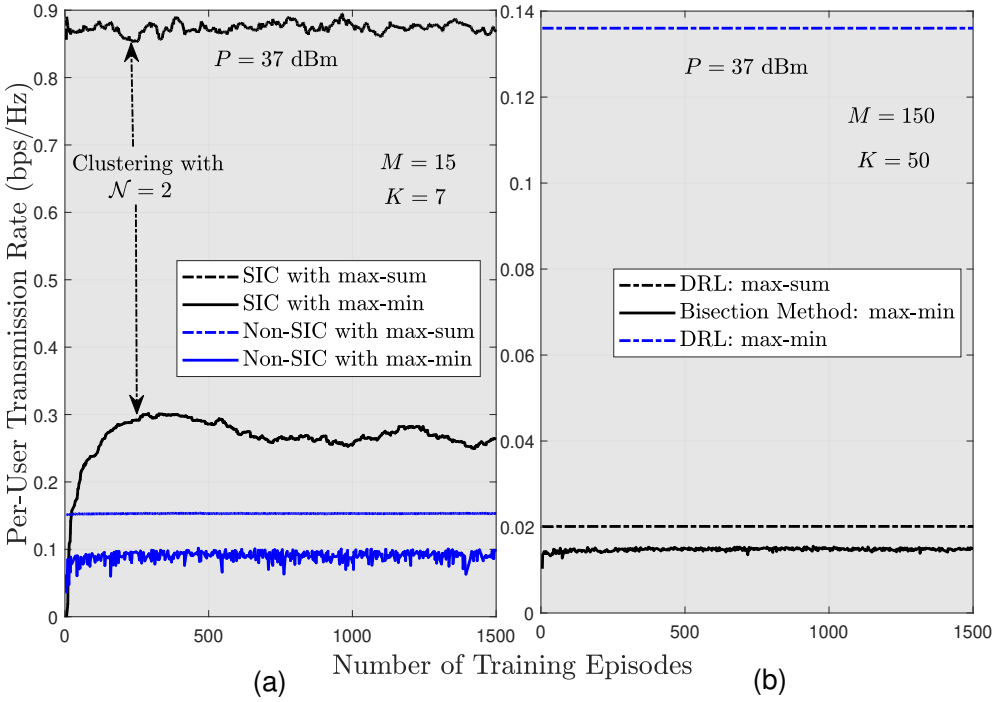

Figure 4.7 Per-UE transmission rate: (a) with SIC, (b) optimal vs. DRL.

per UE transmission rate of around 78% of that with conventional bisection method. This performance degradation comes with a remarkable decrease in the computational complexity. With the proposed DRL-based design, the beamforming vectors can be obtained in an online manner for practical implementation of non-orthogonal multiple access in cell-free networks.

Figs. 4.8(a) and (b) evaluate the convergence rate of the proposed DRL algorithm for two different network setups. Specifically, we plot the absolute value of the difference between

the normalized rate values generated by the DRL model and that by the bisection method (using the Matlab optimization toolkit). It can be noticed that the proposed DRL model

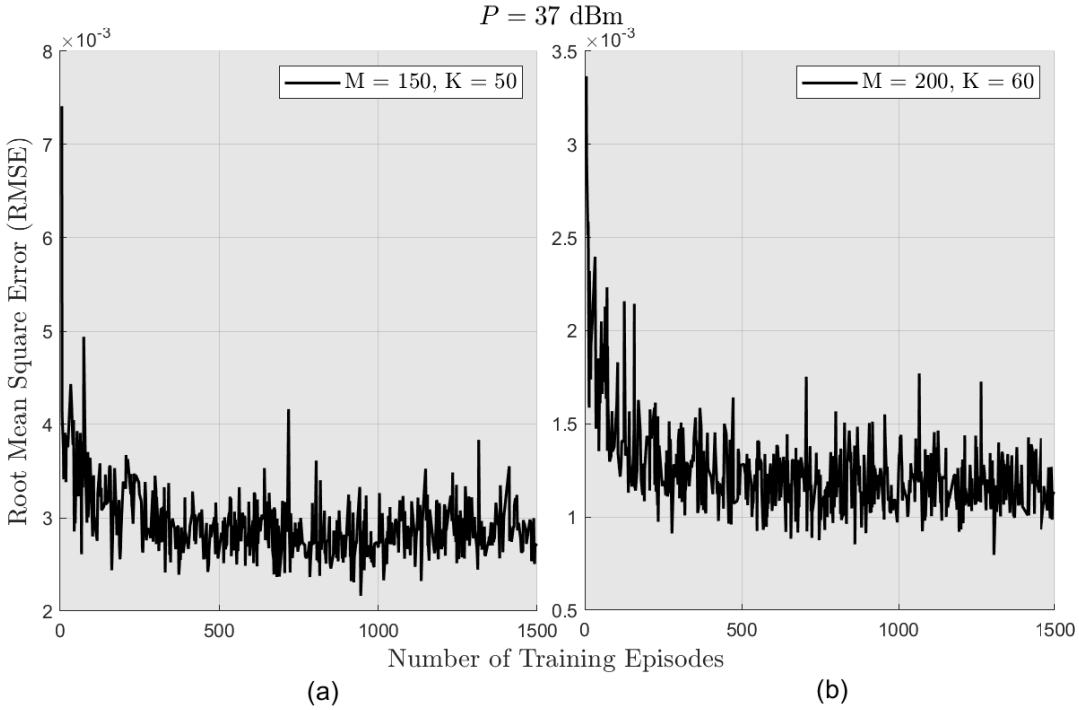

Figure 4.8 DRL convergence rate: (a) Medium-scale scenario, (b) large-scale scenario.

shows a better convergence rate than the bisection method under medium-scale scenario with $M = 150$ and $K = 50$ (in Fig. 4.8(a)) when compared to that of large-scale scenario with $M = 200$ and $K = 60$ (in Fig. 4.8(b)). However, the proposed DRL model under large-scale scenario is found to achieve a better performance by achieving an approximate of 92% of the normalized rate achieved by search method compared to only 78% under medium-scale scenario.

Finally, to evaluate the performance of the cell-free network under different values of per-UE transmission power at the large-scale regime, Fig. 4.9 illustrates the variation in per-UE transmission rate in a centralized large-scale cell-free network for different transmit power values. It can be noticed that rate enhancement becomes smaller when the transmission power is increased in a high P regime (e.g. from $P = 37$ dBm to $P = 50$ dBm). However,

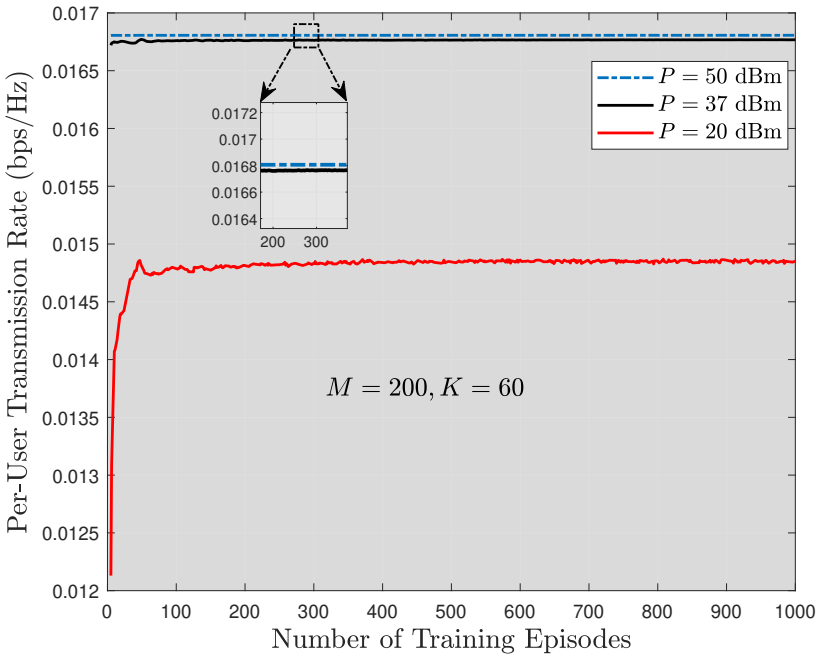

Figure 4.9 Per-UE transmission rate vs. P.

a good rate enhancement still can be achieved when P is changed from a relatively small value to a significantly larger value (e.g. from $P = 20$ dBm to $P = 37$ or $P = 50$ dBm). In addition, the DRL model is found to show a faster convergence with larger P.

4.7 Conclusion

We have first derived closed-form expressions for the probability of outage for an uplink UE in a centralized cell-free network. Next, we have proposed a novel dynamic cell-free network in which, for detection of uplink signals from each UE, the distributed APs are partitioned into a set of subgroups with each subgroup acting as a virtual AP in a distributed antenna system (DAS). Such a clustering is performed based on the current channel state information (CSI) among the APs and all UEs within the network coverage area. The dynamic cell-free network model can reduce the complexity of joint signal processing. For the dynamic cell-free model, we have formulated the general optimization problem of clustering the APs and designing

the beamforming vectors. To solve this optimization problem, we have proposed a hybrid deep deterministic policy gradients (DDPG)-double deep Q-network(DDQN) scheme that jointly selects the optimal network clustering scheme with its optimal beamforming vector values. Possible extensions of this work would include the design and evaluation of more comprehensive DRL models that jointly estimate CSI, select the best clustering configuration for APs, and optimize different beamforming vectors. Also, benchmarking different DRL-based algorithms for optimizing resource allocation in different cell-free network architectures will be valuable for designing future-generation ultra-dense and distributed cell-free wireless networks.

Chapter Five

Self-Organizing mmWave MIMO Cell-Free Network: A Hierarchical DRL Design

we have investigated the utilization of dynamic cell allocation over centralized cell-free networks. We also showed that how decreasing the number of APs by allocating them into subgroups of DAS systems has significantly decreased the processing complexity and convergence rate required to decode UEs signals. However, fully connected cell-free network requires that every AP to be aware of all active UEs within the network coverage area. Such a requirements poses some sort of redundancy since, in massive cell-free network with varying environments, some links between APs and UEs may suffers from severe link degradation and still be treated as viable links and processed accordingly. In this **Chapter**, we propose a variable order AP clustering scheme that allows a variable number of APs within the network to serve each UE based on instantaneous CSI of the network.

5.1 Introduction

5.1.1 Background and Related Work

Cell-free (or cell-less) network architectures are envisioned to increased coverage and transmission rates in future generation wireless systems [37]. In a cell-free wireless network, a large number of UEs in a geographical area will be served simultaneously by a large number of distributed access points (APs) based on non-orthogonal multiple access. The distributed APs in a cell-free system coordinate/cooperate with each other through a centralized processing pool [10] for estimating CSI [23, 40], uplink (downlink) decoding (beamforming) [41–43], and improving transmission performance [10, 16, 48]. The majority of the works on cell-free systems in the literature tackle the following major technical challenges: i) pilot contamination, ii) high computational and hardware complexity of centralized processing, and iii) traffic/signaling overhead. For instance, in [40], the authors designed a joint uplink/downlink pilot training scheme that uses orthogonal subsets of pilots in the downlink instead of using channel reciprocity concept. In [50], the authors developed a semi-blind channel estimation of uplink cell-free massive MIMO network utilizing an enhanced K-means clustering algorithm. In [41], the authors proposed a downlink conjugate beamforming and zero-forcing (ZF) precoding scheme for a fully centralized downlink cell-free network. It was shown that the ZF technique outperforms the conjugate beamforming method at the expense of increased computational complexity. However, when the number of UEs and/or APs increases, the complexity of using ZF beamforming increases significantly due to the requirement of matrix inversion. Accordingly, in [43], a modified conjugate beamforming technique was proposed that uses CSI coordination among the distributed APs. Different machine learning (ML) techniques were used for CSI estimation [23] and beamforming [10] in cell-free networks. For instance, the authors in [23] developed a channel estimation technique for millimeter wave (mmWave)-enabled massive cell-free network using a supervised learning-based denoising

convolutional neural network. The authors in [10] formulated and solved a joint problem for AP clustering and uplink beamforming in a massive cell-free network using DRL techniques.

To reduce the complexity of centralized data processing in a cell-free network, the authors in [16] proposed a UE-centric partitioning method[1]. The proposed method also uses multi-level SIC at each receiver. Another low-complexity design for cell-free networks was proposed in [10]. The core idea is to reduce the dimensionality of beamforming matrices by dynamic clustering of APs. Each cluster then represents a single multi-antenna AP.
All of the aforementioned low-complexity design, however, sacrifice the performance gain of centralized processing. The complexity of solving the beamforming problem in a centralized manner (e.g. to obtain the beamforming vectors at a centralized processing unit) can however be reduced by using a distributed learning and/or processing approach while the detection of the transmitted data is still performed at the central unit. Such a solution has been recently investigated in [25, 56]. Specifically, in [25], the authors utilized supervised learning to solve the beamforming problem in a cell-free network by using a neural network optimizer in each AP. Every AP then obtains the local CSI by estimating only the large-scale fading while considering the small-scale fading as a constant.

Along with cell-free network architectures, mmWave transmissions can be used to improve network capacity [19, 59]. Interestingly, cell-free networks were found to provide an efficient solution for the poor scattering and high path-loss problem of mmWave transmissions [59]. Due to the propagation issues related to mmWave channels, massive multiple-input multiple-output (MIMO) systems are usually used for mmWave-supported UEs [61]. A mmWave and MIMO-based cell-free network, however, requires very significant amount of computational capabilities at the central unit, especially when the number of UEs within the network increases.

[1]We use the terms clustering and partitioning interchangeably.

5.1.2 Motivation and Contributions

For a cell-free network, the complexity of solving the beamforming problem in a centralized manner can be reduced by partitioning the network into a group of cell-free subnetworks, each with independent set of APs and UEs. However, fixed partitioning will not be performance-efficient under fast-varying channel conditions and varying number of UEs per unit area. Therefore, dynamic partitioning into subnetworks based on current network and channel status will be desirable, and for practical feasibility, low-complexity solutions will be required. This motivates us to design a novel mmWave MIMO cell-free network architecture based on dynamic partitioning (or clustering) along with a hybrid analog-digital downlink beamforming method by using DRL techniques. The proposed design provides us with efficient and implementation-friendly solutions.

The main contributions of this chapter can be summarized as follows:

- For a mmWave MIMO cell-free network, we design a self-organizing network architecture that dynamically partitions the network into a group of subnetworks, each acting as an independent cell-free architecture.

- To simultaneously mitigate inter-subnetwork interference (ISNI) and IUI while maximizing the per-UE transmission rate, we develop an innovative hybrid analog beamsteering-digital beamforming method for the proposed mmWave MIMO cell-free network.

- The problem of joint network partitioning, analog beamsteering, and digital beamforming is solved through a novel DRL-cum-convex optimization model. Specifically, the model consists of two interacting networks: i) one DRL model with discrete-action subspace for UE and AP clustering, ii) and another DRL model with continuous-action subspace used for analog beamsteering, the first step of the proposed hybrid beamforming method. The second step of digital beamforming is formulated and solved as a convex optimization problem within the environment of the DRL agent for analog

beamsteering.

- For network partitioning and beamforming, we benchmark several DRL algorithms and evaluate their performances under different system parameters.

Table 5.1 presents the definitions of major system model parameters. The rest of the chapter is organized as follows. **Sec. II** presents the proposed dynamic mmWave MIMO cell-free network architecture and the related modeling assumptions. Also, the general problem formulation to design the proposed system (e.g. clustering and beamforming design) is presented in this section. **Sec. III** presents the beamforming scheme developed for the proposed system. The hierarchical DRL model for joint clustering and beamforming is presented in **Sec. IV**. In **Sec. V**, the complexity analysis of the proposed models and algorithms is presented. **Sec. VI** present the simulation results. **Sec. VII** concludes the chapter.

Table 5.1 Definitions of major system model parameters

Parameter	Definition	Parameter	Definition
eAP	Enhanced Access Point	$\mathcal{N}$	Number of possible subnetwork (or cluster) configurations
ECP	Edge Cloud Processor	$\mathcal{A}_{m_n}$	Beamsteering matrix of eAP m_n
ISNI	Inter-Subnetwork Interference	$\mathcal{D}^{\mathrm{A}}_{n,j}$	Number of eAPs in the n-th subnetwork
IUI	Intra-UE Interference	$\mathcal{D}^{\mathrm{U}}_{n,j}$	Number of UEs in the n-th Subnetwork
NCC	Network Cloud Controller	$\boldsymbol{w}_{k_n m_n}$	Beamforming vector for the link $m_n \rightarrow k_n$
$\boldsymbol{B}^*$	Hermitian transpose of a matrix $\boldsymbol{B}$	m_n	m-th eAP in the n-th subnetwork
$\boldsymbol{B}^T$	Transpose of a matrix $\boldsymbol{B}$	k_n	k-th UE in the n-th subnetwork
$\boldsymbol{H}_{k_n m_n}$	CSI for the link $m_n \rightarrow k_n$	$\boldsymbol{\mathcal{H}}_{k_n m_n}$	Equivalent CSI for the $m_n \rightarrow k_n$ link
M	Number of eAPs	$\mathcal{S}$	State space of a DRL model
a	Number of antennas per eAP	$\boldsymbol{s}$	State vector at time t
K	Number of UEs	$\boldsymbol{s}'$	State vector at time $t+1$
u	Number of antennas per UE	r	Immediate reward of a DRL agent
N	Number of cell-free subnetworks	$\boldsymbol{A}$	Action space of DRL model
$\mathcal{L}$	Number of mmWave paths	$\boldsymbol{a}$	Action vector at time t
$\boldsymbol{\mathcal{C}}_j$	j-th clustering configuration	$\boldsymbol{a}'$	Action vector at time $t+1$

5.2 System Model and Assumptions

5.2.1 Network Architecture

We consider a downlink network with M APs and K UEs (**Fig. 5.1**). Each of the APs and UEs is assumed to be equipped with a and u antennas, respectively. To enable multiuser transmission, each AP is assumed to be equipped with L RF chains. All of the APs are

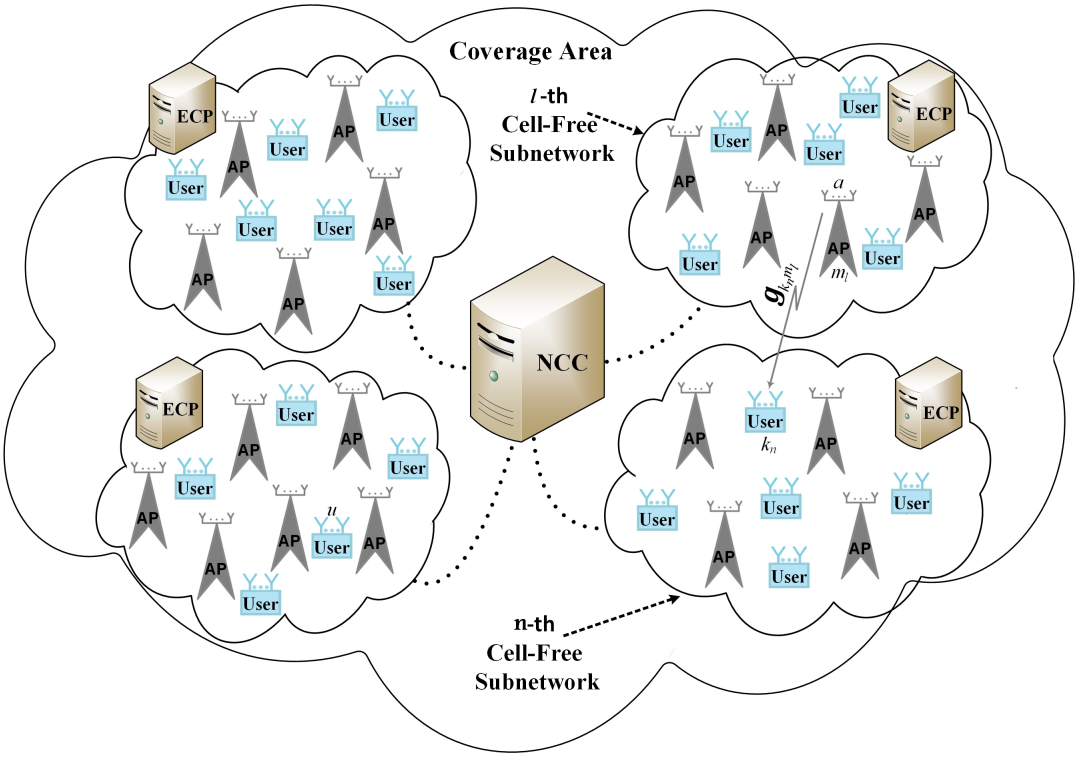

Figure 5.1 An example scenario for the proposed network architecture ($M = 20, K = 21$, and $N = 4$).

connected to each other through fronthaul/backhaul links to form a cell-free network architecture [80]. This enables the distributed APs to simultaneously serve all UEs within the network coverage area. Specifically, the APs collaborate through a network cloud controller (NCC). Each of the APs is assumed to be equipped with a baseband processor that is capable of performing operations related to uplink channel training and downlink beamforming of signals transmitted to different UEs. Such an AP is referred to as an "enhanced-AP" (eAP)

to distinguish it from conventional APs with passband transmission/reception functionalities. In the proposed network architecture, it is assumed that the all of the eAPs and UEs are partitioned into a set of N, where $1 \leq N \leq M$, non-overlapping clusters (i.e. cell-free subnetworks) on a time-slot basis. All of the UEs of a certain subnetwork are served by all of the eAPs of that subnetwork using the same time-frequency resources. Accordingly, the number of RF chains required at each eAP will be equal to the maximum allowable number of UEs per cell-free subnetwork, i.e. $L = K - N + 1$[2]. The group of all baseband processors of eAPs within each cluster can be coordinated to form an virtual edge cloud processor (ECP) unit that is responsible for performing multiuser downlink beamforming within each subnetwork, considering signals from other clusters as Inter-subnetwork interference (ISNI) components. Each eAP may act as an ECP for its subnetwork, or all eAPs of a single subnetwork may form a virtual ECP. Furthermore, the clustering of the cell-free network into a group of non-overlapping cell-free subnetworks is assumed to be performed centrally at the NCC. These two operations of network clustering and subnetwork beamforming are performed either at each time-slot or every several time-slots, based on current CSI and time-varying propagation characteristics of the network. Note that, when $N = 1$, all eAPs and UEs of the network will belong to the same subnetwork which will form a fully centralized cell-free network. On the other hand, when $N = M$, the overall architecture will act as a conventional wireless cellular network with a reuse factor of 1.

Let us denote by $\boldsymbol{\mathcal{C}} = \{\{\boldsymbol{\mathcal{C}}_1^{\mathrm{A}}, \boldsymbol{\mathcal{C}}_1^{\mathrm{U}}\} \dots \{\boldsymbol{\mathcal{C}}_j^{\mathrm{A}}, \boldsymbol{\mathcal{C}}_j^{\mathrm{U}}\} \dots \{\boldsymbol{\mathcal{C}}_{\mathcal{N}}^{\mathrm{A}}, \boldsymbol{\mathcal{C}}_{\mathcal{N}}^{\mathrm{U}}\}\}$ the set of all possible AP-UE clustering configurations such that every cluster contains at least one AP and one UE. $\mathcal{N}$ is the total number of possible clustering configurations which is a function of M, K, and N, i.e. $\mathcal{N} = \Theta(M, K, N)$ (to be defined in subsequent sections). As an example, with

[2]The maximum number of UEs per subnetwork may be defined based on the hardware cost/complexity of the eAPs. This will have a direct impact on the number of RF chains required per eAP and the average amount of energy consumption [93].

$M = 4$, $K = 3$, and $N = 2$, one possible set is

$$\boldsymbol{\mathcal{C}}_j = \{\underbrace{\{\{\mathrm{AP}_1, \mathrm{AP}_3, \mathrm{AP}_4\}}_{\boldsymbol{\mathcal{C}}^{\mathrm{A}}_{1,j}}, \underbrace{\{\mathrm{UE}_2\}}_{\boldsymbol{\mathcal{C}}^{\mathrm{U}}_{1,j}}\}, \{\underbrace{\{\mathrm{AP}_2\}}_{\boldsymbol{\mathcal{C}}^{\mathrm{A}}_{2,j}}, \underbrace{\{\mathrm{UE}_1, \mathrm{UE}_3\}}_{\boldsymbol{\mathcal{C}}^{\mathrm{U}}_{2,j}}\}\}.$$

Let $\mathcal{D}^{\mathrm{A}}_{n,j}$ and $\mathcal{D}^{\mathrm{U}}_{n,j}$ represent, respectively, the number of eAPs and UEs at the n-th subnetwork of the j-th possible configuration, where $\mathcal{D}^{\mathrm{A}}_{n,j} = \text{Cardinality}\left\{\boldsymbol{\mathcal{C}}^{\mathrm{A}}_{n,j}\right\}$ and $\mathcal{D}^{\mathrm{U}}_{n,j} = \text{Cardinality}\left\{\boldsymbol{\mathcal{C}}^{\mathrm{U}}_{n,j}\right\}, n = 1, \ldots, N$ and $j = 1, \ldots, \Theta\left(M, K, N\right)$. For this model, we assume that the m_n-th eAP sends a weighted sum of signals of all UEs within the n-th subnetwork. Accordingly, for a given cell-free network clustering configuration, $\boldsymbol{\mathcal{C}}_j$, the antennas of the m-th eAP at the n-th cluster (denoted by m_n) will have at least $\mathcal{D}^{\mathrm{U}}_{n,j}$ streams[3]. For simplicity, we assume that at each time instant, the m_n-th eAP will use only $\mathcal{D}^{\mathrm{U}}_{n,j} \leq L$ RF chains at a time.

The NCC and the ECP will be responsible for the entire communication process. At the beginning of each time slot, the NCC will first estimate the CSI values for the UEs with respect to all serving eAPs. Then the processes of eAP clustering and per-subnetwork downlink beamforming will be performed jointly by the NCC and the virtual ECP.

5.2.2 Channel and Antenna Model

The communications between the eAPs and the distributed UEs occur in the 24-39 GHz mmWave bands in which transmissions suffer from limited scattering and spatial selectivity. The asymptotic orthogonality assumption among different mmWave channels does not apply to highly correlated mmWave MIMO channels [58]. Accordingly, we adopt the well-known three-dimensional clustered model [95]. We consider a uniform planner array (UPA)[4] at the m_n-th eAP and k_n-th UE with $a = L_{m,1}L_{m,2}$ and $u = L_{k,1}L_{k,2}$ for which $L_{m,1}(L_{k,1})$ and

[3]This assumption will require that the number of RF chains at each eAP does not fall below $\mathcal{D}^{\mathrm{U}}_{n,j}$ [94].

[4]UPA is suitable for mmWave beamforming due to smaller array dimensions, ability to perform 3D beamforming (at the elevation domain), and possibility of packing many antenna elements in a small space [96, 97].

$L_{m,2}(L_{k,2})$ represent the number of columns and rows of antenna elements, respectively. The downlink channel gain matrix for the $m_n \rightarrow k_n$ link (denoted by $\boldsymbol{H}_{k_n m_n} \in \mathbb{C}^{u \times a}$) can be then expressed as [95, 98]

$$\begin{aligned} \boldsymbol{H}_{k_n m_n} &= \sum_{l=1}^{\mathcal{L}} h_{k_n m_n, l} \boldsymbol{b}_{\mathrm{U}} \left(\vartheta_{k_n m_n, l}, \varphi_{k_n m_n, l}\right) \boldsymbol{b}_{\mathrm{A}}^{*} \left(\theta_{k_n m_n, l}, \phi_{k_n m_n, l}\right) \\ &= \sum_{l=1}^{\mathcal{L}} h_{k_n m_n, l} \mathcal{B} \left(\boldsymbol{\theta}_{k_n m_n, l}, \boldsymbol{\phi}_{k_n m_n, l}\right), \end{aligned} \tag{5.1}$$

where $\mathcal{L}$ is the number of paths for the $m_n \rightarrow k_n$ link, $h_{k_n m_n, l} = \sqrt{\frac{1}{\kappa + \mathcal{L} - 1}} \alpha_{k_n m_n, l}$ is the complex channel gain at the l-th path in the $m_n \rightarrow k_n$ link with $\alpha_{k_n m_n, l} \sim \mathcal{CN}\left(0, \sigma_{k_n m_n, l}\right)$, in which $\sigma_{k_n m_n, 1} = \kappa$ (the ratio of the line-of-sight [LoS] path power to non-line-of-sight [NLoS] path power), and $\sigma_{k_n m_n, l} = 1, l = 2, \ldots, \mathcal{L}$. Also, $\mathcal{B}\left(.\right) = \boldsymbol{b}_{\mathrm{U}}\left(.\right) \boldsymbol{b}_{\mathrm{A}}^{*}\left(.\right)$ with $\boldsymbol{\theta}_{k_n m_n} = [\vartheta_{k_n m_n, l}, \theta_{k_n m_n, l},]$ and $\boldsymbol{\phi}_{k_n m_n} = [\varphi_{k_n m_n, l}, \phi_{k_n m_n, l}]$. In (5.1), $\boldsymbol{b}_{\mathrm{U}}\left(\vartheta_{k_n m_n, l}, \varphi_{k_n m_n, l}\right) \in \mathbb{C}^{u \times 1}$ and $\boldsymbol{b}_{\mathrm{A}}\left(\theta_{k_n m_n, l}, \phi_{k_n m_n, l}\right) \in \mathbb{C}^{a \times 1}$ are the antenna array responses at the k_n-th UE and the m_n-th eAP, respectively. The antenna array response at the m_n-th eAP and k_n-th UE, respectively, can be defined as

$$\begin{aligned} &\boldsymbol{b}_{\mathrm{A}}\left(\theta_{k_n m_n, l}, \phi_{k_n m_n, l}\right) \\ &= \Bigg[e^{j2\pi \frac{d\left(0 \sin \theta_{k_n m_n, l} \cos \phi_{k_n m_n, l} + 0 \sin \phi_{k_n m_n, l}\right)}{\lambda}}, \ldots, e^{j2\pi \frac{d\left(w \sin \theta_{k_n m_n, l} \cos \phi_{k_n m_n, l} + z \sin \phi_{k_n m_n, l}\right)}{\lambda}}, \ldots, \\ &e^{j2\pi \frac{d\left(\left(L_{m,1}-1\right) \sin \theta_{k_n m_n, l} \cos \phi_{k_n m_n, l} + \left(L_{m,2}-1\right) \sin \phi_{k_n m_n, l}\right)}{\lambda}} \Bigg]^{T}, \end{aligned} \tag{5.2}$$

$$\begin{aligned} &\boldsymbol{b}_{\mathrm{U}}\left(\vartheta_{k_n m_n, l}, \varphi_{k_n m_n, l}\right) \\ &= \Bigg[e^{j2\pi \frac{d\left(0 \sin \vartheta_{k_n m_n, l} \cos \varphi_{k_n m_n, l} + 0 \sin \varphi_{k_n m_n, l}\right)}{\lambda}}, \ldots, e^{j2\pi \frac{d\left(w \sin \vartheta_{k_n m_n, l} \cos \varphi_{k_n m_n, l} + z \sin \varphi_{k_n m_n, l}\right)}{\lambda}}, \\ &\ldots, e^{j2\pi \frac{d\left(\left(L_{k,1}-1\right) \sin \vartheta_{k_n m_n, l} \cos \varphi_{k_n m_n, l} + \left(L_{k,2}-1\right) \sin \varphi_{k_n m_n, l}\right)}{\lambda}} \Bigg]^{T}, \end{aligned} \tag{5.3}$$

where $\theta_{k_n m_n,l}$ and $\vartheta_{k_n m_n,l}$ are the elevation angles at the m_n-th eAP and k_n-th UE, respectively, $\phi_{k_n m_n,l}$ and $\varphi_{k_n m_n,l}$ are the azimuth angles at the m_n-th eAP and k_n-th UE related to the l-th path in the $m_n \to k_n$ link, respectively, d is the antenna spacing of eAPs and UEs, and λ is the carrier wavelength.

5.2.3 Downlink Data Transmission

We assume that downlink transmission is performed based on two types of beamforming schemes, namely, analog RF beamsteering and baseband digital beamforming. Assuming that a certain cell-free network partitioning configuration (say $\boldsymbol{\mathcal{C}}_j$) is selected by the NCC, the received combined signal at the i_n-th UE can be expressed as

$$\begin{aligned} y_{i_n} &= \boldsymbol{\delta}_{i_n}^{\mathrm{T}} \sum_{l=1}^{N} \sum_{m_l=1}^{\mathcal{D}_{l,j}^{\mathrm{A}}} \boldsymbol{H}_{i_n m_l} \sum_{k_l=1}^{\mathcal{D}_{n,j}^{\mathrm{U}}} \boldsymbol{\mathcal{A}}_{m_l} \boldsymbol{w}_{k_l m_l} x_{k_l} + \boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{\eta}_{i_n} \\ &= \underbrace{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}} \boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_n} \boldsymbol{\mathcal{A}}_{m_n} \boldsymbol{w}_{i_n m_n} x_{i_n}}_{\text{Desired Signal}} + \underbrace{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}} \boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_n} \boldsymbol{\mathcal{A}}_{m_n} \boldsymbol{w}_{k_n m_n} \sum_{k_n=1, k_n \neq i_n}^{\mathcal{D}_{n,j}^{\mathrm{U}}} x_{k_n}}_{\text{IUI}} \\ &\quad + \underbrace{\sum_{l=1, l\neq n}^{N} \sum_{m_l=1}^{\mathcal{D}_{l,j}^{\mathrm{A}}} \boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_l} \sum_{k_l=1}^{\mathcal{D}_{l,j}^{\mathrm{U}}} \boldsymbol{\mathcal{A}}_{m_l} \boldsymbol{w}_{k_l m_l} x_{k_l}}_{\text{ISNI}} + \underbrace{\boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{\eta}_{i_n}}_{\text{AWGN}}, \end{aligned} \tag{5.4}$$

where $\boldsymbol{H}_{i_n m_l} \in \mathbb{C}^{u \times a}$ is the channel gain matrix for the $m_l \to i_n$ link, $\boldsymbol{\mathcal{A}}_{m_l} \in \mathbb{C}^{a \times \mathcal{D}_{l,j}^{\mathrm{U}}}$ is the analog RF beamsteering matrix at the m_l-th eAP, $\boldsymbol{w}_{k_l m_l} \in \mathbb{R}^{\mathcal{D}_{l,j}^{\mathrm{U}} \times 1}$ is the digital baseband beamforming vector related to the $m_l \to k_l$ link, $\boldsymbol{\delta}_{m_l} \in \mathbb{C}^{u \times 1}$ is the analog beamsteering/combining vector at the k_l-th UE, x_{k_l} is the transmitted symbol related to the k_l-th UE such that $\mathbb{E}\left[|x_{k_l}|^2\right] = P/\mathcal{D}_{l,j}^{\mathrm{U}}$ where P is the transmission power budget at each eAP, and $\boldsymbol{\eta}_{i_n}$ is the additive white Gaussian noise (AWGN) vector at the input of the i_n-th UE where $\boldsymbol{\eta}_{i_n} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{i_n} \boldsymbol{I}\right), \forall i_n = 1, 2, \ldots, \mathcal{D}_{n,j}^{\mathrm{U}}, n = 1, \ldots, N$. The instantaneous signal-to-interference-plus-noise ratio (SINR) at the input of the i_n-th UE under clustering configura-

tion $\boldsymbol{\mathcal{C}}_j$ can be expressed as

$$\gamma_{i_n}^{\{\boldsymbol{\mathcal{C}}_j\}} = \frac{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}} |\boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_n} \boldsymbol{\mathcal{A}}_{m_n} \boldsymbol{w}_{i_n m_n}|^2}{\left(\begin{array}{c} \sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}} \sum_{k_n=1, k_n \neq i_n}^{\mathcal{D}_{n,j}^{\mathrm{U}}} |\boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_n} \boldsymbol{\mathcal{A}}_{m_n} \boldsymbol{w}_{k_n m_n}|^2 + \tilde{\sigma}_{i_n} \sum_{j=1}^{u} \delta_{i_n j}^2 \\ + \sum_{l=1, l \neq n}^{N} \left(\frac{\mathcal{D}_{n,j}^{\mathrm{U}}}{\mathcal{D}_{l,j}^{\mathrm{U}}} \right)^2 \sum_{m_l=1}^{\mathcal{D}_{l,j}^{\mathrm{A}}} \sum_{k_l=1}^{\mathcal{D}_{l,j}^{\mathrm{U}}} |\boldsymbol{\delta}_{i_n}^{\mathrm{T}} \boldsymbol{H}_{i_n m_l} \boldsymbol{\mathcal{A}}_{m_l} \boldsymbol{w}_{k_l m_l}|^2 \end{array} \right)}, \tag{5.5}$$

where $\tilde{\sigma}_{i_n} = \left(\frac{\sigma_{i_n} \mathcal{D}_{n,j}^{\mathrm{U}}}{2P} \right)^2$. Note that equation (5.5) is derived based on the assumption that both transmitter and receiver have a full knowledge of CSI of the corresponding link.

5.2.4 General Problem Formulation

To achieve the best performance with the proposed cell-free architecture, the operations of network partitioning, analog beamsteering, and digital beamforming must be jointly optimized (e.g. by solving an optimization problem globally). The objective of this problem can be, for example, maximization of the sum-rate of all UEs (i.e. *max-sum* objective), or maximization of the minimum rate of the UEs (i.e. *max-min* objective to achieve fairness). The general problem formulation can be stated as follows:

$$\begin{aligned} \mathsf{P}_1 : \quad & \max_{j, \left\{ \boldsymbol{A}_{m_n}, \boldsymbol{\Delta}_{m_n}, \boldsymbol{W}_{m_n} \right\}_{m_n=1,\ldots,D_{n,j}^{\mathrm{A}}}^{n=1,\ldots,N}} \quad f\left(\left\{ \gamma_{i_n}^{\{\boldsymbol{\mathcal{C}}_j\}} \right\}_{i_n=1,\ldots,\mathcal{D}_{n,j}^{\mathrm{U}}}^{n=1,\ldots,N} \right) \\ & \text{Subject to:} \\ & \mathbf{C}_1 : |\boldsymbol{\mathcal{A}}_{m_n}(q,z)|^2 = 1, \forall \;\; q = 1, \ldots, a, \text{ and } z = 1, \ldots, \mathcal{D}_{n,j}^{\mathrm{U}}, \\ & \mathbf{C}_2 : |\boldsymbol{\delta}_{k_n}(q)|^2 = 1, \forall \;\; k_n = 1, \ldots, \mathcal{D}_{n,j}^{\mathrm{U}}, \text{ and } q = 1, \ldots, u, \\ & \mathbf{C}_3 : ||\boldsymbol{W}_{m_n}\left(\left[1 \ldots \mathcal{D}_{n,j}^{\mathrm{U}} \right], z \right)||^2 \leq 1, \forall \;\; z = 1, \ldots, \mathcal{D}_{n,j}^{\mathrm{U}}, \end{aligned} \tag{5.6}$$

where $\boldsymbol{\mathcal{A}}_{m_n} \in \mathbb{C}^{a \times \mathcal{D}_{n,j}^{\mathrm{U}}}$ is the analog beamsteering matrix at the m_n-th eAP, $\boldsymbol{W}_{m_n} \in \mathbb{R}^{\mathcal{D}_{n,j}^{\mathrm{U}} \times \mathcal{D}_{n,j}^{\mathrm{U}}}$ is the digital beamforming matrix at the m_n-th eAP with $\boldsymbol{W}_{m_n} = \left[\boldsymbol{w}_{1_n m_n}, \ldots, \boldsymbol{w}_{\mathcal{D}_{n,j}^{\mathrm{U}} m_n} \right]$, where $\boldsymbol{w}_{k_n m_n} = \left[\sqrt{w_{k_n m_n, 1}}, \ldots, \sqrt{w_{k_n m_n, \mathcal{D}_{n,j}^{\mathrm{U}}}} \right]^{\mathrm{T}}$ is the digital baseband beamforming vector related to the $m_n \rightarrow k_n$ link, $\boldsymbol{\Delta}_{m_n} \in \mathbb{R}^{u \times \mathcal{D}_{n,j}^{\mathrm{U}}}$ is the analog beamsteering/combining matrix

related to all UEs of the m_n-th subnetwork, in which $\boldsymbol{\Delta}_{m_n} = \left[\boldsymbol{\delta}_{1_n}, \ldots, \boldsymbol{\delta}_{\mathcal{D}^{\mathrm{U}}_{n,j}}\right]$, where $\boldsymbol{\delta}_{k_n} = [\delta_{k_n 1}, \ldots, \delta_{k_n u}]^{\mathrm{T}}$ is the analog beamsteering/combining vector at the k_n-th UE. Furthermore, the index $j \in [1, \ldots, \Theta(M, K, N)]$ refers to the selected cell-free partitioning configuration.

$\mathtt{P}_1$ is a combinatorial optimization problem which is characterized by: i) non-convexity of the objective function $f(.)$ (discrete j), ii) discrete nature of optimization variable j, iii) non-affine nature of the constraints $\mathbf{C}_1$ and $\mathbf{C}_2$. To solve $\mathtt{P}_1$ optimally, a simultaneous optimization for j, $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\Delta}_{m_n}$, and $\boldsymbol{W}_{m_n}$, $\forall\ n = 1, \ldots, N$, $m_n = 1, \ldots, \mathcal{D}^{\mathrm{A}}_{n,j}$, and $k_n = 1, \ldots, \mathcal{D}^{\mathrm{U}}_{n,j}$ will be required. This is achieved by going through every possible clustering configuration of the cell-free network ($\boldsymbol{\mathcal{C}}_j, j = 1, \ldots, \Theta(M, K, N)$), and for each $\boldsymbol{\mathcal{C}}_j$, we will need to find the corresponding optimal analog beamsteering and digital beamforming matrices (i.e. $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\Delta}_{m_n}$, and $\boldsymbol{W}_{m_n}$). *The globally optimal solution is then the one that gives the best performance among all possible clustering configurations and the corresponding matrices* $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\Delta}_{m_n}$, *and* $\boldsymbol{W}_{m_n}$. The solution will have a combinatorial computational complexity in terms of the network parameters such as M, K, N, a_{m_n}, and u_{k_n} (see Section 5.5 for further discussions).

5.3 mmWave Hybrid Beamforming Design

As has been mentioned in the last section, the problem $\mathtt{P}_1$ in (5.6) is a combinatorial optimization problem with four overlapping feasible spaces (spaces of j, $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\Delta}_{m_n}$, and $\boldsymbol{W}_{m_n}$). Such a problem can be solved by global optimization techniques such as deterministic methods (e.g. inner and outer approximation and cutting-plane methods), stochastic methods (e.g. direct Monte-Carlo sampling and stochastic tunneling), and heuristic methods (e.g. genetic algorithms and swarm-based optimization algorithms) [99]. However, generating an efficient solution of $\mathtt{P}_1$ with reasonable computational complexity and short computing time becomes very challenging as the number of eAPs and/or UEs increases. In this section, we

develop an efficient low-complexity mixed DRL-cum-convex optimization-based solution of $\mathtt{P}_1$. In the following, we will first discuss the problem of downlink beamforming at each cell-free subnetwork and then develop a hierarchical DRL-based scheme that jointly performs network clustering and per-subnetwork hybrid beamforming.

5.3.1 Hybrid Beamforming for Cell-Free MIMO

Partitioning the overall cell-free network architecture (i.e. eAPs and UEs) into a set of computationally independent cell-free subnetworks introduces ISNI to the received signal. Furthermore, simultaneous in-band transmission will cause IUI to all UEs that belong to the same subnetwork. We develop a novel hybrid analog-digital beamforming scheme that efficiently mitigates the effects of ISNI and IUI. Specifically, in the proposed method, downlink beamforming at the multi-antenna eAPs within each cluster is performed in two consecutive stages (**Fig. 5.2**). Under a certain network partitioning configuration (e.g. $\boldsymbol{\mathcal{C}}_j$ for some j),

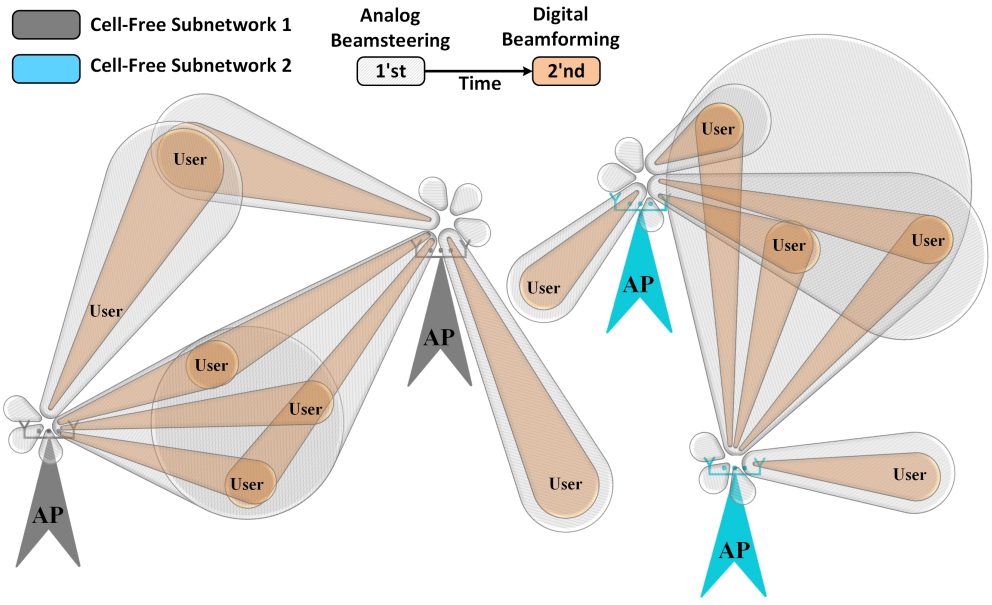

Figure 5.2 Hybrid Beamforming: Example scenario.

each cell-free subnetwork first performs an analog beamsteering for all eAPs such that the ISNI from the nearby clusters is minimized. This is achieved by directing the main beams

of eAPs (i.e. the main lobes of antennas at each eAP) to the UEs belonging to the same cluster and setting the beam directions of annihilated side lobes to the UEs located outside the intended subnetwork (ISNI minimization). Once the beams of different subnetworks are steered to their desired coverage areas, digital beamforming (**Fig. 5.3**) is performed at each eAP to maximize the overall performance and mitigate IUI for the UEs that are located within the intended cell-free subnetwork. Digital beamforming is performed using the overall effective channel after applying the analog beamsteering phase matrices to the original CSI matrices.

For the proposed beamforming scheme, each eAP is assumed to be equipped with L RF chains. Each UE within each subnetwork is assigned to one communication stream by each eAP in that subnetwork[5]. To achieve this, the number of UEs at each cell-free subnetwork must not exceed the number of RF chains at each eAP (i.e. $\mathcal{D}^{\mathrm{U}}_{n,j} \leq L$). **Fig. 5.3** illustrates the functional block diagram of the m_n-th eAP transmitter/beamformer and the k_n-th UE. At the UE side, we assume that signal from different antennas are combined through a low-

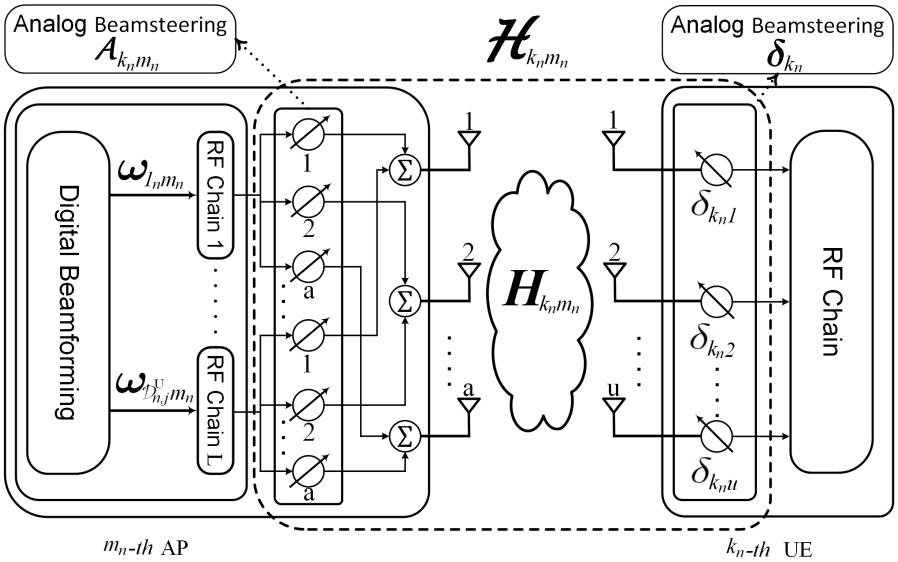

Figure 5.3 Block diagram of the proposed hybrid beamforming scheme.

complexity analog beamsteering/combining scheme using the analog combining vector $\boldsymbol{\delta}_{k_n}$.

[5]A single baseband communication stream is handled by a single RF chain.

After analog beamsteering at all eAPs and UEs, digital beamforming takes place considering the effective CSI obtained after applying analog beamsteering at both eAPs and UEs (i.e. after applying $\boldsymbol{\mathcal{A}}_{m_n}$ and $\boldsymbol{\delta}_{k_n},\ \forall\ n, m, k$).

5.3.2 Analog Beamsteering Subsystem: ISNI Mitigation

The main beam (also referred to as the main lobe) of an antenna element contains the largest portion of the field strength (either radiated or absorbed). The direction of the main beam of a single antenna can be adjusted to match the direction of arrival of the transmitted signal (elevation and azimuth angles) [100]. We propose an analog beamsteering technique to be used in the first-stage of downlink signal transmission at each cell-free subnetwork. This is achieved by utilizing the spatial signatures between UEs of the overall cell-free network and those of eAPs of the intended subnetwork. Specifically, analog beamsteering in a cell-free subnetwork is used to minimize the ISNI caused to the UEs from outside the intended subnetwork. This is achieved by directing the main lobes of the eAPs of each subnetwork toward the UEs belonging to the same subnetwork and setting the directions of the weakest beam side lobes to those outside the intended subnetwork.

Designing the beamsteering objective function: In order to mitigate the ISNI components, analog beamsteering matrices $\boldsymbol{\mathcal{A}}_{m_n}, \forall\ m_n = 1, \ldots, \mathcal{D}^{\mathrm{U}}_{n,j}$ and $n = 1, \ldots, N$ at the m_n-th eAP have to be designed such that they "*zero-force*" the communication links between all eAPs of the n-th subnetwork with UEs outside the intended subnetwork. At the same time, the communication links between all eAPs of the n-th subnetwork and UEs inside the intended subnetwork are optimized. To achieve this, let us first define the "*null space*" of an arbitrary $m_x \rightarrow k_y$ MIMO link using the following axiom. Let $\boldsymbol{H}_{k_y m_x} \in \mathbb{C}^{u \times a}$ be an arbitrary mmwave MIMO channel matrix. If the singular value decomposition (SVD) of

$\boldsymbol{H}_{k_y m_x}$ is given by

$$\begin{aligned}\boldsymbol{H}_{k_y m_x} &= \boldsymbol{U}_{k_y m_x} \boldsymbol{\Sigma}_{k_y m_x} \boldsymbol{V}^*_{k_y m_x} \\ &= \left[\boldsymbol{U}^{(1)}_{k_y m_x} \boldsymbol{U}^{(0)}_{k_y m_x}\right] \boldsymbol{\Sigma}_{k_y m_x} \left[\boldsymbol{V}^{(1)}_{k_y m_x}, \boldsymbol{V}^{(0)}_{k_y m_x}\right]^*,\end{aligned} \tag{5.7}$$

then the left null space of $\boldsymbol{H}_{k_y m_x}$ is given by

$$\text{Null}_L\left(\boldsymbol{H}_{k_y m_x}\right) = \boldsymbol{U}^{(0)}_{k_y m_x}. \tag{5.8}$$

Furthermore, the right null space of $\boldsymbol{H}_{\boldsymbol{k_y m_x}}$ is given by

$$\text{Null}_R\left(\boldsymbol{H}_{k_y m_x}\right) = \boldsymbol{V}^{(0)}_{k_y m_x}, \tag{5.9}$$

where $\boldsymbol{U}_{k_y m_x} \in \mathbb{C}^{u \times u}$ and $\boldsymbol{V}_{k_y m_y} \in \mathbb{C}^{a \times a}$ are unitary matrices, and $\boldsymbol{\Sigma}_{k_y m_x} \in \mathbb{R}^{u \times a}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{H}_{k_y m_x}$. $\boldsymbol{U}^{(1)}_{k_y m_x} \in \mathbb{C}^{u \times r}$ and $\boldsymbol{V}^{(1)}_{k_y m_x} \in \mathbb{C}^{a \times r}$ are the matrices with columns from $\boldsymbol{U}_{k_y m_x}$ and $\boldsymbol{V}_{k_y m_x}$, respectively, corresponding to the non-zero eigenvalues of $\boldsymbol{H}_{k_y m_x}$, and $\boldsymbol{U}^{(0)}_{k_y m_x} \in \mathbb{C}^{u \times (u-r)}$ $\boldsymbol{V}^{(0)}_{k_y m_x} \in \mathbb{C}^{a \times (a-r)}$ with columns from $\boldsymbol{U}_{k_y m_x}$ and $\boldsymbol{V}_{k_y m_x}$, respectively, corresponding to the zero eigenvalues of $\boldsymbol{H}_{k_y m_x}$, where $r = \text{rank}\left(\boldsymbol{H}_{k_y m_x}\right)$.

Remark: To guarantee the existence of a null space for any arbitrary $\boldsymbol{H}_{k_y m_x}$, the number of antennas at each eAP must exceed that of UEs served by that eAP. This condition complies with the fact that mmWave networks use massive MIMO systems at all distributed eAPs. Given the left and right null spaces of $\boldsymbol{H}_{k_y m_x}$, the projection of a complex vector $\boldsymbol{\delta}_{k_y}$ into $\boldsymbol{U}^{(0)}_{k_y m_x}$ can be given by

$$\boldsymbol{\delta}^{\perp}_{k_y m_x} = \boldsymbol{\delta}^{\text{T}}_{ky} \boldsymbol{U}^{(0)}_{k_y m_x} \left(\boldsymbol{U}^{(0)}_{k_y m_x}\right)^*. \tag{5.10}$$

Furthermore, the projection of a matrix $\boldsymbol{\mathcal{A}}_{m_x}$ into $\boldsymbol{V}^{(0)}_{k_y m_x}$ can be given by

$$\boldsymbol{\mathcal{A}}^{\perp}_{k_y m_x} = \boldsymbol{V}^{(0)}_{k_y m_x} \left(\boldsymbol{V}^{(0)}_{k_y m_x}\right)^* \boldsymbol{\mathcal{A}}_{m_x}. \tag{5.11}$$

(5.10) and (5.11) above can be derived using the fact that $\left(\boldsymbol{\delta}_{k_y} - \boldsymbol{\delta}^{\perp}_{k_y m_x}\right)^{\text{T}} \boldsymbol{U}^{(0)}_{k_y m_x} = \boldsymbol{0}$, $\left(\boldsymbol{V}^{(0)}_{k_y m_x}\right)^* \left(\boldsymbol{\mathcal{A}}_{m_x} - \boldsymbol{\mathcal{A}}^{\perp}_{k_y m_x}\right) = \boldsymbol{0}$ and $\boldsymbol{U}^{(0)}_{k_y m_x} \left(\boldsymbol{U}^{(0)}_{k_y m_x}\right)^* = \boldsymbol{I}_u$, $\boldsymbol{V}^{(0)}_{k_y m_x} \left(\boldsymbol{V}^{(0)}_{k_y m_x}\right)^* = \boldsymbol{I}_a$, where

$\boldsymbol{I}_u \in \mathbb{R}^{u\times u}$, $\boldsymbol{I}_a \in \mathbb{R}^{a\times a}$ are identity matrices. Using a similar procedure, the projection of $\boldsymbol{\delta}_{k_y}$ and $\boldsymbol{\mathcal{A}}_{m_x}$ on the left and right "*non-annihilating*" subspaces of $\boldsymbol{H}_{k_y m_x}$ can be given, respectively, by [6]

$$\boldsymbol{\delta}^{\not\perp}_{k_y m_x} = \boldsymbol{\delta}^{\mathrm{T}}_{k_y} \boldsymbol{U}^{(1)}_{k_y m_x} \left(\boldsymbol{U}^{(1)}_{k_y m_x}\right)^*, \text{ and} \tag{5.12}$$

$$\boldsymbol{\mathcal{A}}^{\not\perp}_{k_y m_x} = \boldsymbol{V}^{(1)}_{k_y m_x} \left(\boldsymbol{V}^{(1)}_{k_y m_x}\right)^* \boldsymbol{\mathcal{A}}_{m_x}. \tag{5.13}$$

The objective of analog beamsteering is to reduce ISNI within each cell-free subnetwork. However, focusing only on ZF technique to remove the interfering beams (i.e. ISNI) between adjacent subnetworks may result in misalignment of the main beams of the eAP antennas with those of the UEs within the same subnetwork, and hence reduced/inappropriate antenna directivity. Therefore, we propose a novel analog beamsteering scheme based on maximizing the so-called "*secrecy sum power gains*" at each subnetwork. This is done at the n-th subnetwork by maximizing the sum of powers of two channel projections: i) the non-annihilating projections of $\boldsymbol{\delta}_{k_n}$ and $\boldsymbol{\mathcal{A}}_{m_n}$ on $\boldsymbol{H}_{k_n m_n}$ (i.e, $\boldsymbol{\delta}^{\not\perp}_{k_n m_n}$ and $\boldsymbol{\mathcal{A}}^{\not\perp}_{k_n m_n}$, $\forall\ k_n = 1, \ldots, \mathcal{D}^{\mathrm{U}}_{n,j}$ & $m_n = 1, \ldots, \mathcal{D}^{\mathrm{A}}_{n,j}$), and ii) the annihilating projections of $\boldsymbol{\mathcal{A}}_{m_n}$ on $\boldsymbol{H}_{k_l m_n}$ (i.e. $\boldsymbol{\mathcal{A}}^{\perp}_{k_l m_n}$, $\forall\ l \neq n$).

We are now ready to formulate the beamsteering optimization problem $\mathtt{P}_2$ to mitigate ISNI, which will need to be solved for the n-th cell-free subnetwork, as follows:

$$\mathtt{P}_2: \max_{\left\{\boldsymbol{\delta}_{k_n}, \boldsymbol{\mathcal{A}}_{m_n}\right\}_{k_n=1,\ldots,D^{\mathrm{U}}_{n,j}}^{m_n=1,\ldots,D^{\mathrm{A}}_{n,j}}} \sum_{m_n=1}^{\mathcal{D}^{\mathrm{A}}_{n,j}} \left(\sum_{k_n=1}^{\mathcal{D}^{\mathrm{U}}_{n,j}} ||\boldsymbol{\delta}^{\not\perp}_{k_n m_n} \boldsymbol{\Sigma}_{k_n m_n} \boldsymbol{\mathcal{A}}^{\not\perp}_{k_n m_n}||^2 + \sum_{l=1, l\neq n}^{N} \sum_{k_l=1}^{\mathcal{D}^{\mathrm{U}}_{l,j}} ||\boldsymbol{\delta}^{\perp}_{k_l m_n} (t-1)\, \boldsymbol{\Sigma}_{k_l m_n} \boldsymbol{\mathcal{A}}^{\perp}_{k_l m_n}||^2 \right), \tag{5.14}$$

Subject to:

$\mathbf{C}_1: |\boldsymbol{\mathcal{A}}_{m_n}(q,z)|^2 = 1, \forall\ q = 1, \ldots, a$ and $z = 1, \ldots, u$,

$\mathbf{C}_2: |\boldsymbol{\delta}_{k_n}(q)|^2 = 1, \forall\ \ q = 1, \ldots, u$.

[6] Here, the non-annihilating subspace refers to the subspace of $\boldsymbol{H}_{k_y m_x}$ after its null subspace has been removed.

In the objective function of $\mathtt{P}_2$ in (5.14), the variable matrices $\boldsymbol{\mathcal{A}}^{\not\vee}_{k_n m_n}$, $\boldsymbol{\mathcal{A}}^{\not\vee}_{k_n m_n}$, and $\boldsymbol{\delta}^{\not\vee}_{k_n m_n}$ are non-linear functions of $\boldsymbol{\delta}_{k_y}$ and $\boldsymbol{\mathcal{A}}_{m_n}$. This relationship can be inferred from (5.12) and (5.13) as a non-linear truncation of unitary matrices of the SVD of $\boldsymbol{\delta}_{k_y}$ and $\boldsymbol{\mathcal{A}}_{m_n}$. Accordingly, $\mathtt{P}_2$ is a non-convex combinatorial optimization problem.

5.3.3 Digital Beamforming Subsystem: Transmission Rate Maximization

After the analog beamsteering at all MIMO transmitters (eAPs) and receivers (UEs) has been performed, the actual CSI matrices (i.e. $\boldsymbol{H}_{k_n m_n},\ \forall\ n, m_n$, and k_n) will be multiplied by the beamsteering matrices (from the right side) and the analog combining/beamsteering vectors (from the left side). Accordingly, the effective channel gain at the $m_n \rightarrow k_n$ link (denoted by $\boldsymbol{\mathcal{H}}_{k_n m_n} \in \mathbb{C}^{1\times \mathcal{D}^{\mathrm{U}}_{n,j}}$) will be given by

$$\boldsymbol{\mathcal{H}}_{k_n m_n} = \boldsymbol{\delta}^{\mathrm{T}}_{k_n} \boldsymbol{H}_{k_n m_n} \boldsymbol{\mathcal{A}}_{m_n}. \tag{5.15}$$

Note that, the i_n-th element in $\boldsymbol{\mathcal{H}}_{k_n m_n}$ ($i_n = 1_n, \ldots, \mathcal{D}^{\mathrm{U}}_{n,j}$) corresponds to the signal radiated from the beam steered at UE i_n within the m_n-th subnetwork. However, each element in $\boldsymbol{\mathcal{H}}_{k_n m_n}$ will contain portions of signals send to all UEs of the n-th subnetwork.

UE ordering and SIC decoding: In downlink multiuser single-input single-output (SISO) wireless networks, SIC-based UEs (usually referred to as non-orthogonal multiple access [NOMA] UEs) are first ordered based on their instantaneous channel gains. Then, the UEs with lower link gains are allocated higher transmission power compared to those with better communication link. At the receiver side, multi-level SIC operations are conducted such that the interfering signals related to the UEs with lower channel gains are decoded and then subtracted [101]. When the UEs of a downlink NOMA system are served by a single eAP with multiple antennas or by multiple single-antenna APs, the UEs can be ordered based on the norm of their channel quality vector [16]. Such a channel quality metric is denoted as the

"effective channel gain". In our considered network model, however, all the eAPs and the UEs in each cluster are assumed to be equipped with multiple antennas. Therefore, the channel gain between each eAP and any arbitrary NOMA UE is represented by a complex matrix. We use the squared norm of the effective channel gain vectors (i.e. $\boldsymbol{\mathcal{H}}_{k_n m_n}$) as the NOMA effective channel gain. Accordingly, we assume that the UEs within the n-th cluster are arranged in an ascending order as $\frac{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}}|\boldsymbol{\mathcal{H}}_{1_n m_n}|^2}{\sum_{l=1,l\neq n}^{N}\sum_{m_l=1}^{\mathcal{D}_l^{\mathrm{A}}}\sum_{k_l=1}^{\mathcal{D}_l^{\mathrm{U}}}|\boldsymbol{\mathcal{H}}_{k_l m_l}|^2} \leq \ldots \leq \frac{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}}|\boldsymbol{\mathcal{H}}_{i_n m_n}|^2}{\sum_{l=1,l\neq n}^{N}\sum_{m_l=1}^{\mathcal{D}_l^{\mathrm{A}}}\sum_{k_l=1}^{\mathcal{D}_l^{\mathrm{U}}}|\boldsymbol{\mathcal{H}}_{k_l m_l}|^2} \leq \ldots \leq \frac{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}}|\boldsymbol{\mathcal{H}}_{\mathcal{D}_{n,j}^{\mathrm{U}} m_n}|^2}{\sum_{l=1,l\neq n}^{N}\sum_{m_l=1}^{\mathcal{D}_l^{\mathrm{A}}}\sum_{k_l=1}^{\mathcal{D}_l^{\mathrm{U}}}|\boldsymbol{\mathcal{H}}_{k_l m_l}|^2}$. Note that, for the UE ordering, we divide the effective channel metric of each UE by the corresponding ISNI assuming the availability of channel gain matrices of the overall cell-free network at each eAP. With proper power allocation for the UEs, at the receiver side, the i_n-th UE located in the n-th cluster will be able to remove interference components from $i_n - 1$ UEs with higher overall power gain. Accordingly, $\gamma_{i_n}^{\{\boldsymbol{c}_j\}}$ can be rewritten as

$$\gamma_{i_n}^{\{\boldsymbol{c}_j\}} = \frac{\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}}|\boldsymbol{\mathcal{H}}_{i_n m_n}\boldsymbol{w}_{i_n m_n}|^2}{\left(\begin{array}{l}\sum_{m_n=1}^{\mathcal{D}_{n,j}^{\mathrm{A}}}\sum_{k_n=i_n+1}^{\mathcal{D}_{n,j}^{\mathrm{U}}}|\boldsymbol{\mathcal{H}}_{i_n m_n}\boldsymbol{w}_{k_n m_n}|^2+\tilde{\sigma}_{i_n}\sum_{j=1}^{u}\delta_{i_n j}^2 \\ +\sum_{l=1,l\neq n}^{N}\left(\frac{\mathcal{D}_{n,j}^{\mathrm{U}}}{\mathcal{D}_{l,j}^{\mathrm{U}}}\right)^2\sum_{m_l=1}^{\mathcal{D}_{l,j}^{\mathrm{A}}}\sum_{k_l=1}^{\mathcal{D}_{l,j}^{\mathrm{U}}}|\boldsymbol{\mathcal{H}}_{i_n m_l}\boldsymbol{w}_{k_l m_l}|^2\end{array}\right)}, \tag{5.16}$$

where $\tilde{\sigma}_{i_n} = \left(\frac{\sigma_{i_n}\mathcal{D}_{n,j}^{\mathrm{U}}}{2P}\right)^2$.

We propose a beamforming (i.e. precoding) scheme that maximizes the sum-rate of the UEs. The digital beamforming problem at the n-th cell-free subnetwork under a certain

cell-free network partitioning configuration ($\boldsymbol{\mathcal{C}}_j$) can be formulated as

$$\mathtt{P}_3: \max_{\{\boldsymbol{w}_{k_n m_n}\}_{k_n=1,\ldots,D^{\mathrm{U}}_{n,j}}^{m_n=1,\ldots,D^{\mathrm{A}}_{n,j}}} \quad \sum_{n=1}^{N}\sum_{i_n=1}^{\mathcal{D}^{\mathrm{U}}_{n,j}} \log_2\left(1+\gamma_{i_n}^{\{\boldsymbol{\mathcal{C}}_j\}}\right)$$

Subject to:

$$\mathbf{C}_1: \sum_{m_n=1}^{\mathcal{D}^{\mathrm{A}}_{n,j}} \left(||\boldsymbol{\mathcal{H}}_{i_n m_n}\boldsymbol{w}_{\delta_{i_n} m_n}||^2 - \sum_{w=\delta_{i_n}+1}^{i_n} ||\boldsymbol{\mathcal{H}}_{i_n m_n}\boldsymbol{w}_{w m_n}||^2 \right) \geq \epsilon, \tag{5.17}$$

$$\mathbf{C}_2: ||\boldsymbol{w}_{k_n m_n}||^2 \leq 1,$$

$$\forall\ \delta_{i_n} = 1,\ldots,i_n-1, l=2,\ldots,\mathcal{D}^{\mathrm{U}}_{n,j}, m_n=1,\ldots,\mathcal{D}^{\mathrm{A}}_{n,j},$$

$$k_n = 1,\ldots,\mathcal{D}^{\mathrm{U}}_{n,j}, \text{ and } n=1,\ldots,N.$$

In [16, Appendix C], it was shown that problem $\mathtt{P}_3$ in (5.17) represents a convex optimization problem under the assumption of UE ordering and SIC-based decoding. Specifically, it was shown that the objective function of $\mathtt{P}_3$ in (5.17) can be decomposed into a sum of convex and concave functions with the convex function having a more increasing rate than that of the concave one. Furthermore, it is easy to confirm that the constraints $\mathbf{C}_1$ and $\mathbf{C}_2$ represent affine relations of $\boldsymbol{w}_{k_n m_n}$[7]. This convex problem can be easily solved by using the Karush-Kuhn-Tucker (KKT) conditions and utilizing some numerical methods for calculating the first and second differentiation of the Lagrangian function[8].

[7]This can be easily confirmed by rewriting the vector form of $\mathbf{C}_1$ and $\mathbf{C}_2$ in a sum of products format, rather than vector format.

[8]Despite the existence of the second derivative of the objective function, a closed-form expression is difficult to derive due to the multi-dimensional nature of the optimization variables.

5.4 Hierarchical DRL Design: Joint Network Partitioning and Hybrid Beamforming

5.4.1 DRL Techniques for Solving Optimization Problems

DRL techniques have been used to solve optimization problems in wireless communications systems (e.g. for optimization of downlink power control in a multi-cell system [**?**], beamforming optimization in a cell-free network [10]). In these cases, a DRL agent (e.g. a network entity) aims at learning the "optimal" mapping between a system state $\boldsymbol{s}$ and the action $\boldsymbol{a}$ (e.g. a policy function or a value function) in order to maximize its reward discounted reward over a time horizon. Depending on the agent objective, DRL techniques are commonly classified into three categories:

- *Value-based* methods such as deep Q-learning (DQL) and state–action–reward–state–action (SARSA) learn the value function $V(\boldsymbol{s})$ or the state-value function $Q(\boldsymbol{s}, \boldsymbol{a})$ to find a policy.

- *Policy-based* methods obtain the mapping between the system state and the action (i.e. policy) directly. These methods generally suffer from noisy gradients and high variance [68].

- *Actor-critic* methods are a hybrid of the value-based and policy-based methods. Value-based methods are used to reduce the variance of the policy-based methods by estimating the value function or the action-value function (a.k.a. the critic) to improve the performance of the policy (a.k.a. the actor).

5.4.2 Hierarchical DRL Architecture

The proposed solution consists of two-levels of interacting DRL models. The first-level of the proposed system is responsible for network partitioning (i.e. clustering) and it consists of a single DRL model. The agent of the first-level DRL model is located at the NCC and is mainly responsible for partitioning the overall cell-free network into a set of non-overlapping cell-free subnetworks. The second-level of the proposed hierarchical architecture consists of N independent DRL subsystems. Each DRL subsystem is responsible for conducting the hybrid analog beamsteering-digital beamforming process in a single cell-free subnetwork. This is achieved by training the DRL subsystem agent to optimize the analog beamsteering vectors of all eAPs and UEs of the corresponding cell-free subnetwork while the digital beamforming problem for the same subnetwork is modeled and solved as a convex optimization problem inside the environment of the DRL subsystem for analog beamsteering. All of the DRL subsystems are within the environment of the first-level DRL clustering system (**Fig. 5.4**). In terms of the time-scale of operation, the overall cell-free network is assumed to cluster

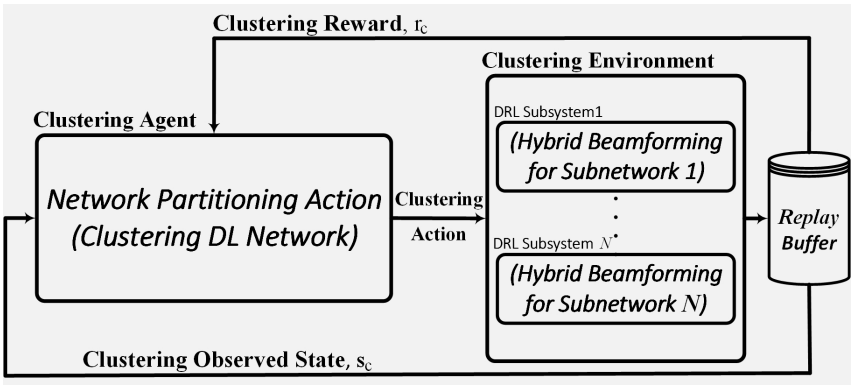

Figure 5.4 Block diagram of the hierarchical DRL clustering system.

(partition) every $\tau \geq 1$ time instants[9]. On the other hand, the hybrid beamforming process is assumed to take place at each time slot. Further details about the DRL model action spaces, rewards, and observed states are described in the following sections.

[9]The value of τ may be considered as a design parameter that can depend, for example, on the time-varying nature of the propagation environment.

5.4.3 Learn to "*Cluster*"

Our objective is to design a self-organizing cell-free network that has the ability to self-partition (self-cluster) into a group of cell-free subnetworks, in a time-slot basis, based on the instantaneous CSI. For the proposed dynamic cell-free network with M eAPs, K UEs, and N subnetworks, there will be $\Theta(M,N) = \left(\frac{N!}{\sqrt{2}}\right)^2 C(M,N)\, C(K,N)$ possible configurations for cell-free subnetworks, where $C(M,N)$ is the Stirling number, which can be calculated as [91]

$$C(M,N) = \left\{ \begin{array}{c} M \\ N \end{array} \right\} = \frac{1}{N!} \sum_{i=0}^{N} (-1)^i \binom{N}{i} (N-i)^M . \tag{5.18}$$

Optimally updating the cell-free network configuration on a time-slot basis requires going through all possible configurations (as discussed in Section 5.2.4) which will be practically infeasible for a massive cell-free network with large numbers of eAPs and UEs[10].

In this section, we design several low-complexity DRL-based methods to efficiently perform network clustering on a time-slot basis. Each of these methods accepts certain network information (e.g. instantaneous CSI values across the entire network) and outputs a certain network partitioning configuration that maximizes a predefined network performance metric. **Table** 5.2 summarizes the environment design for the DRL models in terms of the problem parameters. Note that, in **Table** 5.2, the DRL system state vector ($\boldsymbol{s}_{\mathrm{c}}$) corresponds to a

Table 5.2 DRL model for network partitioning

Clustering Environment Variables	Network Partitioning System Equivalence
State $\boldsymbol{s}_{\mathrm{c}} = \{s_{\mathrm{c},1}, \ldots, s_{\mathrm{c},N}\}$	$\left\{\prod_{t=1}^{\tau} \prod_{i_1=1}^{\mathcal{D}^{\mathrm{U}}_{1,j}} \gamma_{i_1}^{\{\mathcal{C}_j\}}(t), \ldots, \prod_{t=1}^{\tau} \prod_{i_N=1}^{\mathcal{D}^{\mathrm{U}}_{N,j}} \gamma_{i_N}^{\{\mathcal{C}_j\}}(t)\right\}$
Reward r_{c}	$\prod_{t=1}^{\tau} \left(\prod_{n=1}^{N} \left(\sum_{i_n=1}^{\mathcal{D}^{\mathrm{U}}_{n,j}} \log\left(1+\gamma_{i_n}^{\{\mathcal{C}_j\}}(t)\right)\right)\right)$
Action $\boldsymbol{a}_{\mathrm{c}}$	$\boldsymbol{\mathcal{C}}_j = \left\{\left\{\boldsymbol{\mathcal{C}}^{\mathrm{A}}_{1,j}, \boldsymbol{\mathcal{C}}^{\mathrm{U}}_{1,j}\right\}, \ldots, \left\{\boldsymbol{\mathcal{C}}^{\mathrm{A}}_{N,j}, \boldsymbol{\mathcal{C}}^{\mathrm{U}}_{N,j}\right\}\right\}$

[10] As an example, for $M = 100$, $K = 50$, and $N = 10$, there will be approximately $1.28962122 \times 10^{138}$ possible cell-free subnetwork configurations.

clustering configuration with N clusters (i.e. an N-element vector) and the value of each element is the product of the SINR values of all the UEs in the corresponding cluster (or partition). The immediate reward value for a state is given by the product of the sum of the rates of the UEs in each cluster, under the corresponding clustering configuration. Among the different DRL methods, we investigate (i) value-based DRL methods, namely, the deep double Q-network DDQN [66] and State–action–reward–state–action (SARSA) [67], (ii) a policy-based DRL method, namely, the policy gradient (PG) [102] method, and (iii) the actor-critic (AC) [69] method. The performance, complexity, and convergence rate of these methods are then evaluated and compared.

Value-based DRL methods (e.g. DDQN and SARSA): In value-based clustering, each network partitioning configuration is assigned a certain value through a state-value function $V^{\pi}(\mathbf{\Gamma}_N)$, also known as the expected return function when starting at a certain state $\mathbf{\Gamma}_M$, where $\mathbf{\Gamma}_N = (\gamma_1^{\mathrm{P}}, \ldots, \gamma_N^{\mathrm{P}})$ and $\gamma_n^{\mathrm{P}} = \prod_{t=1}^{\tau} \prod_{i_n=1}^{\mathcal{D}_{n,j}^{\mathrm{U}}} \gamma_{i_n}^{\{\boldsymbol{\mathcal{C}}_j\}}(t)$. The state-value function is defined as $V^{\pi}(\mathbf{\Gamma}_N) = \mathbb{E}[r|\mathbf{\Gamma}_N, \pi]$, where r is the immediate reward, π is the followed policy which can be found such that $V^{*}(\mathbf{\Gamma}_N) = \max_{\pi} V^{\pi}(\mathbf{\Gamma}_N)$, $\forall\ \gamma_n^{\mathrm{P}} \in \mathbb{R},\ n = 1, \ldots, N$. Given $V^{*}(\mathbf{\Gamma}_N)$, the optimal policy π^{*} is found by selecting the best cell-free network partitioning configuration that maximizes $\mathbb{E}_{\mathbf{\Gamma}'_N \sim \mathcal{T}\left(\mathbf{\Gamma}'_N | \mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j\right)}[V^{*}(\mathbf{\Gamma}'_N)]$, where $\mathcal{T}(\mathbf{\Gamma}'_N | \mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j)$ is the transition dynamics that is usually unavailable. Hence, the value function is replaced by a quality state-action-value function $Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j)$, which is different from V^{π} due to the fact that a random cell-free network partitioning configuration action $\boldsymbol{\mathcal{C}}_0$ is provided and the policy π is only counted from the succeeding state, i.e. $Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j) = \mathbb{E}[r|\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j, \pi]$. The learning of the Q^{π} network is performed by using the Bellman equation with the recursive form $Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j) = \mathbb{E}_{\mathbf{\Gamma}'_N}[r' + \zeta Q^{\pi}(\mathbf{\Gamma}'_N, \pi(\mathbf{\Gamma}'_N))]$ [65]. This means that the quality function can be improved by bootstrapping (i.e. using current values of Q^{π} to improve our estimate). This modeling is the basis of Q-learning [66] and SARSA [67] algorithms that is defined as $Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j) \leftarrow Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j) + \alpha\delta$, where α is the learning rate and $\delta = Y - Q^{\pi}(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j)$ is the temporal difference error with Y representing a target (as in standard regression problems).

Using the Q-learning cell-free network partitioning agent, the target Y directly approximates Q^* by setting $Y = r + \zeta \max_j Q^\pi\left(\mathbf{\Gamma}'_N, \boldsymbol{\mathcal{C}}_j\right)$ (off-policy agent), where ζ is the discount factor. On the other hand, the SARSA algorithm improves the estimate of Q^π by deriving a behavioural policy from Q^π. This is achieved by setting $Y = r + \zeta Q^\pi\left(\mathbf{\Gamma}'_N, \boldsymbol{\mathcal{C}}'_j\right)$ (on-policy agent).

Policy-based DRL method for clustering: In policy gradient (PG) algorithms, the modeling and optimization of a certain policy is conducted directly through a parameterized function, $\mu_\theta\left(\boldsymbol{\mathcal{C}}_j, \mathbf{\Gamma}_N\right)$. The value of the objective function (the reward) directly depends on the policy. The value of the reward function for PG-based clustering is given by

$$J(\theta) = \sum_{\boldsymbol{s}_\mathrm{c} \in \mathbb{R}^N} d^\mu\left(\mathbf{\Gamma}_N\right) V^\mu\left(\mathbf{\Gamma}_N\right) = \sum_{\boldsymbol{s}_\mathrm{c} \in \mathbb{R}^N} d^\mu\left(\mathbf{\Gamma}_N\right) \sum_{j \in \mathbb{Z}} \mu_\theta\left(\boldsymbol{\mathcal{C}}_j | \mathbf{\Gamma}_N\right) Q^\mu\left(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j\right), \tag{5.19}$$

where $d^\mu\left(\mathbf{\Gamma}_N\right)$ is the stationary state distribution of Markov chain. Note that the gradient of $J\left(\theta\right)$ (denoted by $\nabla_\theta J\left(\theta\right)$) depends both on the selected actions $\boldsymbol{a}_\mathrm{c}$ and the stationary distribution $d^\mu\left(\mathbf{\Gamma}_N\right)$. We also use a PG algorithm that simplifies the computation of the gradient by removing the dependence of $J\left(\theta\right)$ on $d^\mu\left(\mathbf{\Gamma}_N\right)$ as follows [68, Sec. 13.2]:

$$\begin{aligned} \nabla_\theta J\left(\theta\right) &= \nabla_\theta \sum_{\boldsymbol{s}_\mathrm{c} \in \mathbb{R}^N} d^\mu\left(\mathbf{\Gamma}_N\right) \sum_{j \in \mathbb{Z}} \mu_\theta\left(\boldsymbol{\mathcal{C}}_j | \mathbf{\Gamma}_N\right) Q^\mu\left(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j\right) \\ &\propto \sum_{\boldsymbol{s}_\mathrm{c} \in \mathbb{R}^N} d^\mu\left(\mathbf{\Gamma}_N\right) \sum_{j \in \mathbb{Z}} \mu_\theta\left(\boldsymbol{\mathcal{C}}_j | \mathbf{\Gamma}_N\right) \nabla_\theta Q^\mu\left(\mathbf{\Gamma}_N, \boldsymbol{\mathcal{C}}_j\right). \end{aligned} \tag{5.20}$$

The general policy gradient method has a high variance. Accordingly, the following general form is used as a foundation of different PG algorithms:

$$\nabla_\theta J\left(\theta\right) = \mathbb{E}_{\mu_\theta}\left[\sum_{t=0}^{T-1} G_t \nabla_\theta \log \mu_\theta\left(\boldsymbol{\mathcal{C}}_j | \mathbf{\Gamma}_N\right)\right]. \tag{5.21}$$

The PG-based DRL model for network partitioning can be then trained through the following steps:

i. Initialize the actor $\mu\left(\mathbf{\Gamma}_N\right)$ with random weights θ_μ.

ii. For each training episode (i.e. every T training steps), generate the experiences by following $\mu(\mathbf{\Gamma}_N)$: the actor generates the probability values for each possible cell-free partitioning, then the DRL agent randomly selects an action based on a certain probability distribution. This process continues for T steps.

iii. At each step of a certain episode, calculate the return value $G_t = \sum_{l=t}^{T} \zeta^{l-1} r_l$.

iv. Find the cumulative sum of the actor network gradients during one entire learning episode as

$$d\theta_\mu = \sum_{t=1}^{T} G_t \nabla_{\theta_\mu} \ln \mu(\mathbf{\Gamma}_N | \theta_\mu) \tag{5.22}$$

v. Update the actor network using: $\theta_\mu \leftarrow \theta_\mu + \alpha d\theta_\mu$, where α is the learning rate.

(Actor-critic)-based DRL method: In the PG-based partitioning algorithm, the value function G_t is estimated based on a preassigned policy. However, the estimation of G_t for a predefined policy introduces a relatively high variance of the policy gradient which in turn reduces the quality of cell-free network partitioning action. In order to tackle the high variance problem, a second DNN can be used that can accurately learn the value of G_t [70].

5.4.4 Learn to "Beamsteer"

As discussed in Sec. 5.3.1, the downlink beamforming is performed through two consecutive stages, namely, analog beamsteering and digital beamforming. In this section, we develop a mixed DRL-cum-convex optimization subsystem that performs the two-stage beamforming operation for each cell-free subnetwork (**Fig. 5.5**). In the proposed system, the non-convex analog beamsteering (problem $\mathtt{P}_2$, which is a non-convex combinatorial optimization problem) is solved by training a DRL agent that "interacts" with the propagation medium (i.e. DRL environment) on a time-slot basis. The convex digital beamforming problem $\mathtt{P}_3$, which

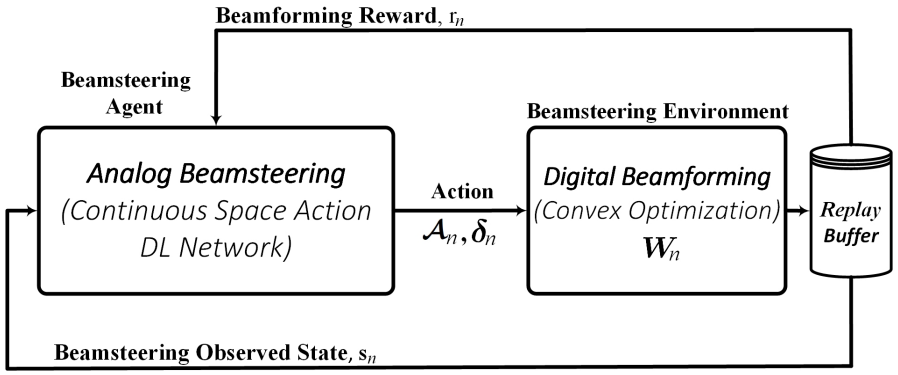

Figure 5.5 Block diagram of the n-th mixed DRL-convex subsystem.

is a strictly convex optimization problem, on the other hand, is solved within the DRL environment by using conventional convex optimization methods (e.g. Newton and Broyden methods[11]). This process of hybrid analog beamsteering-digital beamforming is performed independently at each subnetwork on a time-slot basis.

Table 5.3 shows the main design parameters of the DRL model. Note that the oper-

Table 5.3 DRL model for hybrid beamforming in subnetwork n

Beamforming Environment Variables	Network Beamforming System Equivalence
State $\boldsymbol{s}_{\mathrm{b}} = \left\{ s_{\mathrm{b},1_n}, \ldots, s_{\mathrm{b},\mathcal{D}^{\mathrm{U}}_{n,j}} \right\}$	$\left\{ \gamma_1^{\{\mathcal{C}_j\}}, \ldots, \gamma_{\mathcal{D}^{\mathrm{U}}_{n,j}}^{\{\mathcal{C}_j\}} \right\}$ (Involving Digital Beamforming)
Reward r_{b}	$\sum_{m_n=1}^{\mathcal{D}^{\mathrm{A}}_{n,j}} \left(\sum_{k_n=1}^{\mathcal{D}^{\mathrm{U}}_{n,j}} \lVert \boldsymbol{\delta}_{k_n}^{\nparallel} \boldsymbol{\Sigma}_{k_n m_n} \boldsymbol{\mathcal{A}}_{m_n}^{\nparallel} \rVert^2 + \sum_{l=1, l\neq n}^{N} \sum_{k_l=1}^{\mathcal{D}^{\mathrm{U}}_{l,j}} \lVert \boldsymbol{\delta}_{k_l}^{\perp} (t-1) \boldsymbol{\Sigma}_{k_l m_n} \boldsymbol{\mathcal{A}}_{m_n}^{\perp} \rVert^2 \right)$
Action $\boldsymbol{a}_{\mathrm{b}}$	$\{\boldsymbol{\delta}_{k_n}, \boldsymbol{\mathcal{A}}_{m_n}\}_{k_n=1,\ldots,\mathcal{D}^{\mathrm{U}}_{n,j}}^{m_n=1,\ldots,\mathcal{D}^{\mathrm{A}}_{n,j}}$

ation of digital beamforming is performed as a part of the DRL environment computations that produce the observed states and reward for the analog beamsteering agent. The proposed subsystem can be implemented by using several DRL algorithms. In this chapter, we implement and benchmark two DRL algorithms with continuous action space, namely, the PG algorithm and the Soft Actor-Critic Agents (SAC) algorithm.

Policy-based beamsteering: The PG algorithm used previously for cell-free network partitioning can be also implemented to learn the best beamsteering vectors by solving problem

[11]The Lagrangian function of $\mathtt{P}_3$ is twice differentiable w.r.t all optimization variables [103].

$\mathtt{P}_2$. This can be achieved by optimizing over the discrete action space and then estimating the best continuous beamsteering action. Such an approximation process is relatively slow/inefficient. However, working directly with policies that emit probability distributions can increase the estimation speed of the continuous action space since sampling a well-known distribution is easier than sampling from value functions.

Soft actor-critic beamsteering: On-policy actor-critic algorithms improve the stability of the network by allowing random exploration of experience from actions replay buffers [68]. However, this on-policy training results in a poor sample complexity. On the other hand, off-policy algorithms have been developed to improve the sampling efficiency while maintaining robustness by developing more advanced variance reduction techniques and at the same time incorporating the off-policy samples (e.g. the DDPG family of algorithms) [92]. However, the interaction between the off-policy DDQN value estimator and the deterministic actor setting makes DDPG extremely difficult to stabilize and adjust the hyper-parameter settings. This issue becomes more severe as the size of the cell-free network increases. We propose to utilize the SAC algorithm to solve the beamsteering problem at each cell-free subnetwork [104, Algorithm 1]. The SAC algorithm enables off-policy actor-critic training with a stochastic actor. The main difference between the SAC algorithm and the GP and AC ones is that the SAC algorithm uses a general objective that maximizes entropy along with the cumulative reward [105]. The addition of policy entropy encourages the actor network to explore new experiences. Accordingly, the expected reward in Eq. (5.21) can be modified to [105]

$$\nabla_{\theta_\mu} J\left(\theta_\mu\right) = \nabla_{\theta_\mu} \sum_{t=0}^{T-1} \mathbb{E}_{\left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n\right) \sim \rho_{\theta_\mu}} \left[r\left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n\right) + \alpha \mathcal{H}\left(\mu\left(\boldsymbol{\mathcal{F}}_n | \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}\right)\right)\right], \quad (5.23)$$

where $\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} = \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}$, $\boldsymbol{\mathcal{F}}_n = \boldsymbol{\mathcal{F}}_n$, and $\mathcal{H}\left(.\right)$ is the entropy measure of the policy $\mu\left(\boldsymbol{\mathcal{F}}_n | \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}\right)$, α is a temperature factor that determines the relative importance of the policy entropy against the reward $r\left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n\right)$, and $\rho_\mu\left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}\right)$ and $\rho_\mu\left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n\right)$ are the state and

state-action of the trajectory distribution introduced by $\mu\left(\boldsymbol{\mathcal{F}}_n|\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)$. The soft state-value function of SAC algorithm is given by [104]

$$V\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)=\mathbb{E}_{\boldsymbol{a}_{\mathrm{b}}\sim\boldsymbol{\mu}}\left[Q\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)-\log\mu\left(\boldsymbol{\mathcal{F}}_n|\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)\right]. \tag{5.24}$$

Accordingly, the soft Q-value will be defined as [104]

$$Q\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)=r\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)+\gamma\mathbb{E}_{\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\sim\rho_\mu\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\right)}\left[Q\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime},\boldsymbol{\mathcal{F}}'_n\right)-\log\mu\left(\boldsymbol{\mathcal{F}}'_n|\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\right)\right]. \tag{5.25}$$

The SAC algorithm aims to learn three functions, namely, i) a policy function with parameters θ and π_θ, ii) a soft Q-value function approximated (parameterized) by w and Q_w, and iii) a soft state value function parameterized by ψ and V_ψ. The soft state value is trained to minimize the mean square error with gradient function given as follows [104]:

$$\begin{aligned}\nabla_\psi J_V(\psi)&=\nabla_\psi\mathbb{E}_{\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\sim\mathcal{R}}\left[\frac{1}{2}\left(V_\psi\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)-\mathbb{E}\left[Q_w\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)-\log\pi_\theta\left(\boldsymbol{\mathcal{F}}_n|\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)\right]\right)^2\right]\\&\approx\nabla_\psi V_\psi\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)\left(V_\psi\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)-Q_w\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)+\log\pi_\theta\left(\boldsymbol{\mathcal{F}}_n|\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)\right),\end{aligned} \tag{5.26}$$

where $\mathcal{R}$ is the distribution of previously sampled actions and states (in the replay buffer). Furthermore, the soft Q function is trained to minimize the soft Bellman residual with gradient function given as

$$\begin{aligned}\nabla_w J_Q(w)=\nabla_w\mathbb{E}_{\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)\sim\mathcal{R}}&\left[\frac{1}{2}\left(Q_w\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)-\left(r\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)\right.\right.\right.\\&\left.\left.\left.+\zeta\mathbb{E}_{\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\sim\rho_\pi\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)}\left[V_{\bar{\psi}}\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\right)\right]\right)\right)^2\right]\\\approx\nabla_w Q\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)&\left(Q_w\left(\boldsymbol{\mathcal{F}}_n,\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}}\right)-r\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}},\boldsymbol{\mathcal{F}}_n\right)-\zeta V_{\bar{\psi}}\left(\boldsymbol{\Gamma}_n^{\{\mathcal{C}_j\}\prime}\right)\right),\end{aligned} \tag{5.27}$$

where $\bar{\psi}$ is an exponentially moving average target function. The desired policy is then trained using the information projection that is defined in terms of Kullback-Leibler (KL)-

divergence [106]. Accordingly, the policy is updated according to

$$\pi_{\text{new}} = \underset{\pi' \in \Pi}{\text{argmax}}\, D_{\text{KL}} \left(\pi' \left(. | \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right) || \frac{\exp\left\{ Q^{\pi_{\text{old}}} \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, . \right) \right\}}{Z^{\pi^{\text{old}}} \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right)} \right), \tag{5.28}$$

where Π denotes a set of potential policies that π must restricted to. $Z^{\pi^{\text{old}}} \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right)$ is a partitioning function that is used for normalizing the distribution. The objective update function of the policy π_θ is [104]

$$\begin{aligned} \nabla_\theta J_\pi(\theta) &= D_{\text{KL}} \left(\pi_\theta \left(. | \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right) || \exp\left\{ Q_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n \right) - \log Z_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right) \right\} \right) \\ &= \mathbb{E}_{\boldsymbol{\mathcal{F}}_n \sim \pi} \left[-\log \left(\frac{\exp\left\{ Q_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n \right) - \log Z_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right) \right\}}{\pi_\theta \left(\boldsymbol{\mathcal{F}}_n | \mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right)} \right) \right] \\ &= \mathbb{E}_{\boldsymbol{\mathcal{F}}_n \sim \pi} \left[\log \pi_\theta \left(\boldsymbol{\mathcal{F}}_n | \boldsymbol{s}_{\text{b}} \right) - Q_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}}, \boldsymbol{\mathcal{F}}_n \right) + \log Z_w \left(\mathbf{\Gamma}_n^{\{\mathcal{C}_j\}} \right) \right]. \end{aligned} \tag{5.29}$$

We use the SAC algorithm developed in [104, Algorithm 1].

Algorithm 4 shows the sequence of processes performed during network operation, where E_{c} and E_{b} are the number of episodes for the DRL models used for clustering and beamsteering, respectively, T_{c} and T_{b} are the number of training steps in each episode for the DRL models used for clustering and beamsteering, respectively. .

5.5 Complexity Analysis and Signaling Overhead

To solve problem $\mathtt{P}_1$ in (5.6), the following subproblems will need to be solved: the combinatorial problem related to selecting the best network partitioning configuration, the non-convex problem related to finding the best beamsteering matrices (problem $\mathtt{P}_2$ in (5.14)), and a

[12]The type of network will depend on the utilized clustering and beamsteering techniques (**Sec. V.B and Sec. V.C**).

Algorithm 4 Training for network partitioning and beamforming

Input: $\boldsymbol{H_{k_n m_n}}$, M, K, a, u, and N.
Initialize target networks of main DRL system and subsystems in **Figs. 5.4** and **(b)**[12].
Initialize replay buffers of the main DRL system and the subsystems in **Figs. 5.4** and **(b)**.
Initialize $\boldsymbol{\mathcal{C}}_j$, $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\delta}_{k_n}$ and $\boldsymbol{w}_{k_n m_n}$, $\forall\ n, m_n$, and k_n.
for Episodes$_{\text{Clustering}} = 1$ to E_{c} **do**
    for $t_{\text{c}} = 1$ to T_{c} **do** ▷ Training steps for network partitioning
        Update weights of target networks using Algorithms in **Sec. V.B**
    $n = 1$ to N ▷ Simultaneous beamforming in all subnetworks
        for Episodes$_{\text{Beamsteering}} = 1$ to E_{b} **do**
            for $t_b = 1$ to T_b **do**
                Update weights of target networks using the Algorithms in **Sec. V.C**.
                Compute $\boldsymbol{\mathcal{H}}_{k_n m_n},\ \forall\ k_n = 1, \ldots, \mathcal{D}^{\text{U}}_{n,j}$ and $m_n = 1, \ldots, \mathcal{D}^{\text{A}}_{n,j}$.
                Solve problem $\mathtt{P}_3$ using a convex optimizer.
                Update $\boldsymbol{\mathcal{A}}_{m_n}$, $\boldsymbol{\delta}_{k_n}$, and $\boldsymbol{w}_{k_n m_n}$, $\forall\ n, m_n$, and k_n.
            end for
        end for
        Update $\boldsymbol{\mathcal{C}}_j$.
    end for
end for

convex optimization problem related to finding the optimal digital beamforming at each eAP.

In **Sec. 5.4.3**, we discussed how the complexity of finding best network partitions grows exponentially with increasing values of M, K, and N, as can be seen from (5.18). Furthermore, solving $\mathtt{P}_2$ in (5.14) through an exhaustive search with a step size Δ will have a complexity of order $O\left(\prod_{n=1}^{N}\left(\frac{1}{\Delta}\right)^{(a\times\mathcal{D}_{n,j}^{\mathrm{A}})\times(u\times\mathcal{D}_{n,j}^{\mathrm{U}})}\right)$. Since the problem $\mathtt{P}_3$ in (5.17) is strictly convex, the solution for this problem has a computational complexity of $O\left(\left(\sum_{n=1}^{N}\mathcal{D}_{n,j}^{\mathrm{U}}\times\mathcal{D}_{n,j}^{\mathrm{A}}\right)^{3}\right)$.

To evaluate the time-complexity of a deep neural network used in a DRL model, the conventional measure is the *floating-point operations per second* (FLOPs). For any fully connected layer $\mathbb{L}_i$ of input size I_i and output size O_i, the number of FLOPs is given by $\mathrm{FLOPS}(\mathbb{L}_i) = 2\,I_i\,O_i$. The policy network has two hidden layers of size 256 and 128. Thus, for the DRL models, the total number of FLOPS during the inference is

$$\mathrm{FLOPs}_{\mathrm{DRL}} = \sum_{i=1}^{3}\mathrm{FLOPs}(\mathbb{L}_i) \quad = 2\cdot\Big(256\cdot|\mathcal{S}| + 128\cdot|\boldsymbol{A}|{+}32768\Big), \tag{5.30}$$

where $|\mathcal{S}|$ and $|\boldsymbol{A}|$ are the dimensions of the state space and action space, respectively. **Table 5.4** compares the FLOPS for inference as well as the convergence rate for the DRL algorithms used in this chapter. Note that for network clustering, the dimensionality of the action

Table 5.4 Complexity of different DRL models for clustering

DRL Agent	**Inference FLOPS**	**Convergence**
Conventional solution	$\Theta\,(M, K, N)$	Linear convergence
SARSA	$32768 + 256\cdot K + 128$	Slow
DDQN	$32768 + 256\cdot K + 128$	Geometric
PG	$32768 + 256\cdot K + 128$	Sub-linear
Actor-Critic	$2\,(32768 + 256\cdot K + 128)$	Fast

space is $|\boldsymbol{A}|= 1$. Similarly, the complexity and convergence properties for the considered beamsteering agents are summarized in **Table 5.5**[13]. $\mathcal{M}_n$ in **Table 5.5** represents the

[13]The ellipsoid method requires a total of $O\left(\left[\left(\left(\mathcal{D}_{n,j}^{\mathrm{A}}\times\mathcal{D}_{n,j}^{\mathrm{U}}\right)\times(a\times u)\right]^{4}q\right)$ operations, where q is the length of binary coding of the input.

Table 5.5 Complexity of DRL models for beamsteering in the n-th subnetwork

DRL Agent	**Inference FLOPS**	**Convergence**
Conventional solution	$O\left(\left[\left(\mathcal{D}_{n,j}^{\mathrm{A}} \times \mathcal{D}_{n,j}^{\mathrm{U}}\right) \times (a \times u)\right]^2 q\right)$	Linear convergence
PG	$32768 + 256 \cdot K + 128 \cdot \mathcal{M}_n$	Sub-linear
DDPG	$32768 + 256 \cdot K + 128 \cdot \mathcal{M}_n$	Unknown
SAC	$32768 + 256 \cdot K + 128 \cdot \mathcal{M}_n$	Unknown

dimensionality of the analog beamsteering problem and is given by $\mathcal{M}_n = \left(\mathcal{D}_{n,j}^{\mathrm{A}} \times \mathcal{D}_{n,j}^{\mathrm{U}}\right) \times (a \times u)$.

In terms of signaling overhead, in the proposed methods, the NCC first will have to collect the estimated CSI matrices from distributed eAPs and send full CSI to the ECP of each cell-free subnetwork. Next, the NCC will collect the performance metric (e.g. sum-rate) from the ECP of each subnetwork and use it to decide on the new network partitioning configuration.

5.6 Numerical Results

5.6.1 Parameters and Assumptions

Table 5.6 presents the values of different parameters used in generating the simulation results. All the results for the conventional methods are produced using Mont-Carlo simulations each with 10^6 runs. Additionally, we assume that all channel small-scale fading gains $h_{k_n m_n}$ are drawn from a set of i.i.d random variables. We assume that all APs and UEs are uniformly distributed over a disc of radius 18 m (corresponding to a network total coverage area of $1\mathrm{km}^2$).

Table 5.6 Simulation parameters

Parameter	Value
AWGN PSD at UE	-169 dBm/Hz
Path-loss exponent	2 (outdoor)
mmWave carrier frequency, $\frac{3\times10^8}{\lambda}$	24 GHz (unless specified otherwise)
mmWave paths, $\mathcal{L}$	3 (unless specified otherwise)
SIC sensitivity, P_s	1 dBm
# of training episodes	$\{2000, 4000\}$
# of training steps/episode	200
Discount factor, ζ	0.01
Learning rate, α	0.001

5.6.2 Hybrid Beamforming Scheme

We start this section by evaluating the performance of the proposed network architecture under the designed hybrid beamforming system (**Fig. 5.6(a)**). It can be noticed from this figure that the designed hybrid analog beamsteering-digital beamforming scheme significantly outperforms that of conventional (all digital) beamforming scheme. For example, a gain of 2.3 bps/Hz and 1.5 bps/Hz are achieved at 35 dBm with $N = 1$ and $N = 2$, respectively (upper graph in **Fig. 5.6(a)**). Interestingly, even without network partitioning (i.e. for $N = 1$), our proposed hybrid beamforming scheme shows significant increase in performance compared to its conventional beamforming counterpart. The reason is that our designed objective function for beamsteering (see **Problem** $\mathtt{P}_2$ in (5.14)) aims to enhance beams of desired UEs and "*zero-null*" beams to undesired UEs at the same time. Accordingly, with $N = 1$, beamsteering will focus on optimally directing the antenna main lobes of APs and UEs to each other. However, this performance gain is observed to decrease as the network scales up (see lower graph of **Fig. 5.6(a)**). In order to study the effect of multiple antennas on

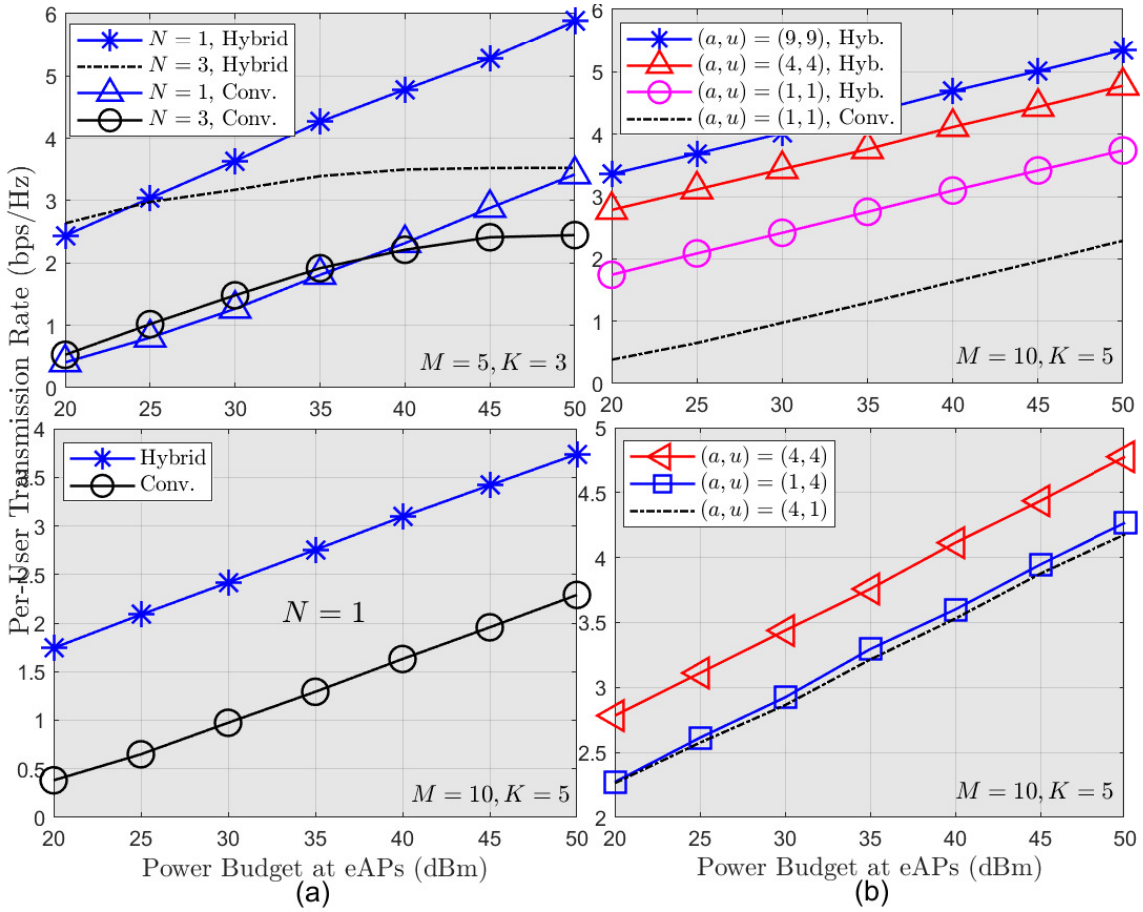

Figure 5.6 Hybrid vs. conventional beamforming techniques.

the per-UE rate performance, **Fig. 5.6(b)** shows per-UE transmission rate versus different MIMO layouts. It can be noticed that a significant increase in per-UE rate performance can be achieved by increasing the number of antennas at the UEs and eAPs (upper graph of **Fig. 5.6(b)**). This rate enhancement, however, decreases as the values of a and/or b increase (due to increased interference levels). Furthermore, the system performance enhances better as the antenna order at the UEs increases more than that at the eAPs (lower graph of **Fig. 5.6(b)**).

5.6.3 Evaluation and Benchmarking of Hierarchical DRL Models

The performances of the DRL models for network partitioning and analog beamsteering are investigated separately. This is done by first training the different DRL clustering agents while using conventional methods for hybrid analog beamsteering-digital beamforming operations. On the other hand, the DRL-based beamforming methods are evaluated while clustering is performed through the trained DRL agents in the inference mode. This separate evaluation enables us to extract more insights on the performances of that DRL models under discrete and continuous action spaces.

We start by evaluating the performances of the DRL-based clustering algorithms for the proposed self-partitioning cell-free network architecture (**Fig. 5.7(a)**). We use two training modes for each of the studied DRL algorithms. The first training mode considers a fixed CSI (i.e. constant $\boldsymbol{H}$ matrix), while in the second mode, we use different CSI realizations at every training step of each episode. In **Fig. 5.7(a)**, we train four DRL agents using PG, DDQN, SARSA, and AC algorithms for network partitioning for a single CSI realization. As can be observed, the PG algorithm provides the best clustering performance in terms of stability and convergence, while the DDQN algorithm comes second, and the SARSA algorithm comes last with significantly high covariance in the Q-values per episode and slower convergence rate (upper graph in **Fig. 5.7(a)**). In terms of per-UE rate performance, even though all of the three algorithms show relatively similar performance levels, however, with a closer look, one can find out that the PG algorithm provides the highest per-UE transmission rate.

Fig. 5.7(b) evaluates the effect of training DRL agents during the real-time operation of the cell-free network. Specifically, we assume that a training step is performed during one time slot. This means that, state transitions of the DRL model will result from both current action $\boldsymbol{a}_c$ and the instantaneous CSI $\boldsymbol{H}$. It can be observed from **Fig. 5.7(b)** that changing $\boldsymbol{H}$ during training of the clustering agents has a negative impact on both the convergence rate and the per-UE rate performance. This can be observed clearly by the

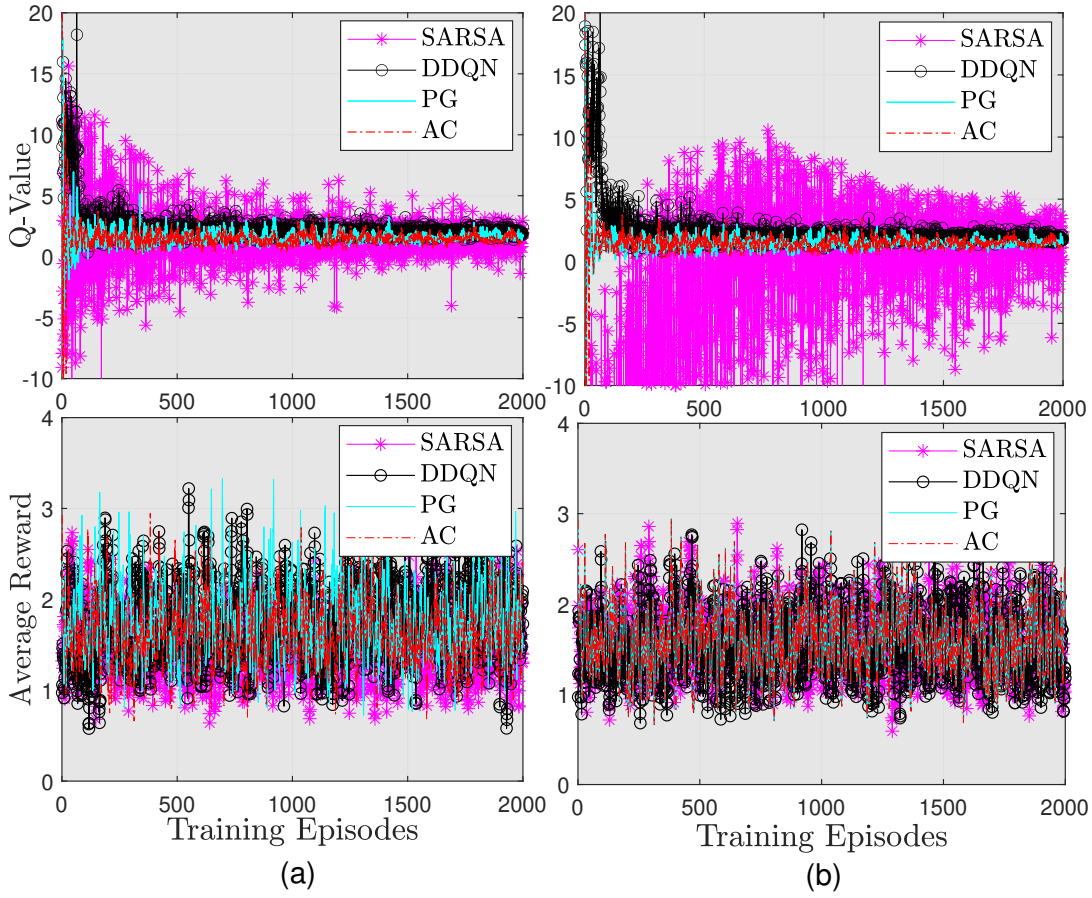

Figure 5.7 Performance of clustering agents: (a) fixed CSI, and (b) varying CSI.

significant increase in the variance of the Q-values in the upper graph of **Fig. 5.7(b)**. It can also be observed that the convergence of the SARSA-based clustering is the worst. For SARSA, to tackle the weak stability issue, we double the number of training episodes from 2000 to 4000 episodes and retrain the SARSA agent under varying CSI conditions (**Fig. 5.8(a)**). As can be observed from this figure, increasing the number of episodes improves the stability of the SARSA algorithm significantly. However, with more training episodes the per-UE rate performance does not improve (with average reward of around 1.5408 bps/Hz). More numerical results on the performances of the DRL-based clustering schemes are given in **Table 5.7**. It can be observed that the off-policy algorithm (i.e. DDQN) gives the worst performance under varying CSI conditions. This is due to the fact that the DDQN agent selects the action related to the highest Q-value in a deterministic fashion, without any

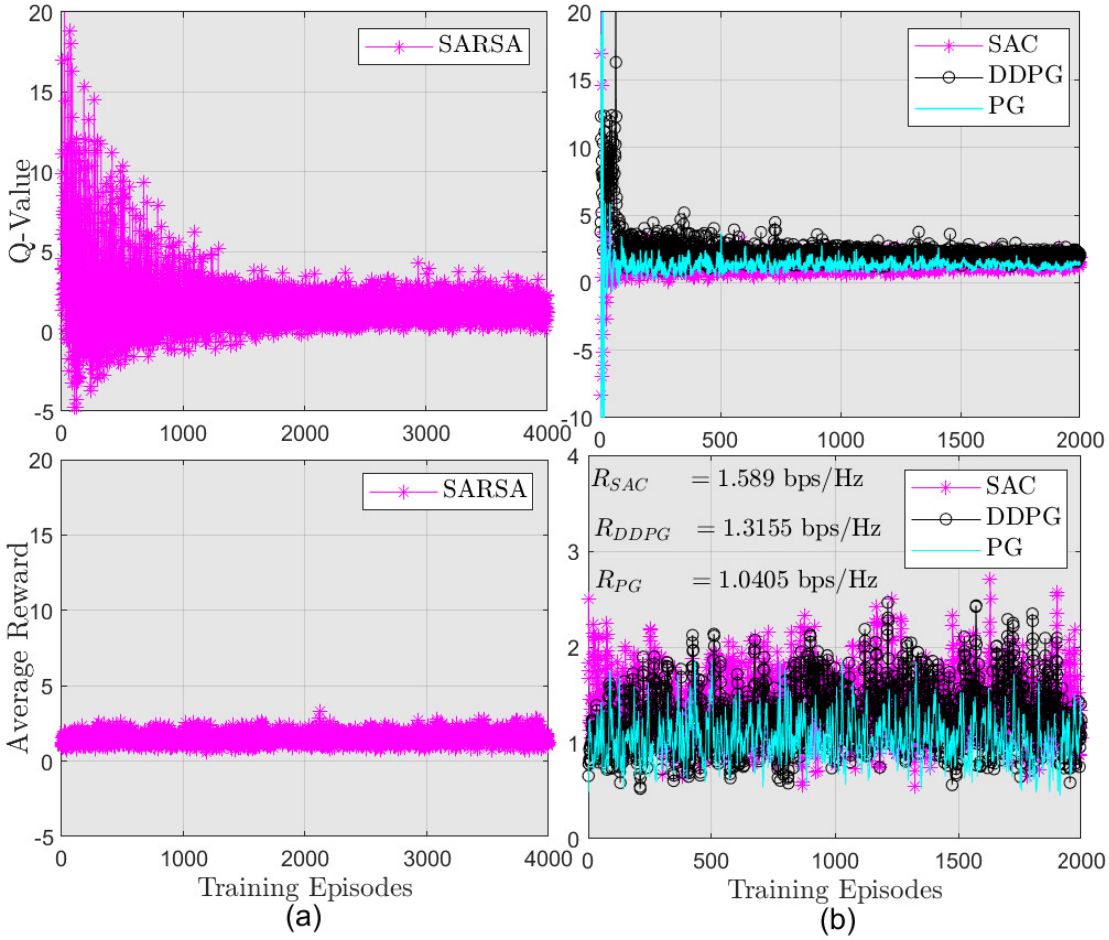

Figure 5.8 Performance of different clustering agents with different number of training episodes.

exploration. This action-selection strategy will prevent the network from learning/sensing the stochastic variations of the states (or alternatively the CSI matrix $\boldsymbol{H}$). On the other hand, all of the on-policy based algorithms (PG, SARSA, and AC) show good performance under varying CSI, and we observe a noticeable enhancement on per-UE rate performance due to the AC algorithm. The reason is that the AC algorithm allows the DRL agent to learn the stochastic properties of the state (or alternatively, the CSI matrix $\boldsymbol{H}$). Finally, **Fig. 5.8(b)** evaluates and compares the performance of several DRL-based beamsteering methods. For these simulations, we use the inference mode of the DDQN algorithm to solve the network partitioning problem. It can be noticed that the AC algorithm shows the best per-UE rate performance compared to the DDPG and PG algorithms. Furthermore, when

Table 5.7 Numerical results on the performances of different clustering schemes.

$(M, K) = (5, 3)$ and $(a, u) = (1, 1)$ with optimal performance: 2.03786 pbs/Hz				
Agent	Average reward	Inference mode Fixed $\boldsymbol{H}$	Inference mode Variable $\boldsymbol{H}$	Training duration (2000 episodes)
PG trained by fixed $\boldsymbol{H}$	1.7607 bps/Hz	1.5626 bps/Hz	1.4841 bps/Hz	15.4072 Mins
PG trained by varying $\boldsymbol{H}$	1.6090 bps/Hz	1.5592 bps/Hz	1.5332 bps/Hz	15.0098 Mins
DDQN trained by fixed $\boldsymbol{H}$	1.7308 bps/Hz	1.5802 bps/Hz	1.5355 bps/Hz	16.5467 Mins
DDQN trained by varying $\boldsymbol{H}$	1.5579 bps/Hz	1.5321 bps/Hz	1.4654 bps/Hz	14.0558 Mins
SARSA trained by fixed $\boldsymbol{H}$	1.4733 bps/Hz	1.4376 bps/Hz	1.5153 bps/Hz	18.2317 Mins
SARSA trained by varying $\boldsymbol{H}$	1.5862 bps/Hz	1.5047 bps/Hz	1.5072 bps/Hz	16.3313 Mins.
AC trained by fixed $\boldsymbol{H}$	1.4186 bps/Hz	1.5876 bps/Hz	1.4991 bps/Hz	15.3618 Mins
AC trained by varying $\boldsymbol{H}$	1.6084 bps/Hz	1.5664 bps/Hz	1.4919 bps/Hz	15.5362 Mins

both network clustering and analog beamsteering are implemented through the DRL agents, the per-UE rate performance of the SAC algorithm drops to around 70% of the optimal performance, and for the PG algorithm, it is 51% of the optimal performance.

5.7 Conclusion

A novel self-partitioning MIMO cell-free network architecture has been proposed in which a cell-free network is partitioned into a set of independent cell-free subnetworks. To efficiently solve the problem of network partitioning for a large-scale network, we have proposed, evaluated, and benchmarked several state-of-the-art DRL methods with discrete action space. Furthermore, to reduce the interference between adjacent cell-free subnetworks, we have designed a novel downlink hybrid analog beamsteering-digital beamforming scheme. We have also evaluated several state-of-the-art DRL methods with continuous action space to solve the combinatorial problem of analog beamsteering while the digital beamforming problem has been solved as a strictly convex optimization problem. Results have showed a significant rate enhancement and complexity reduction due to the proposed hybrid beamforming

scheme compared to its conventional all-digital counterpart. It has been observed that online training of different DRL agents is only slightly affected by changing the CSI in the network. However, changing the CSI can significantly affect the variance and convergence rate of the DRL algorithms such as the SARSA algorithm. Furthermore, it has also been noticed that all DRL methods for network clustering and beamsteering have almost the same per-UE rate performance with a slight superiority of the PG and AC algorithms when used for network clustering and analog beamsteering, respectively. A potential extension of this work is to enable distributed beamforming at each cell-free subnetwork. This may be achieved by utilizing a multiple agent algorithm with continuous action space. Another extension of this work is to solve the problem of pilot assignment using distributed multiple agent DRL modeling.

Chapter Six

Conclusion

6.1 Summary of the Thesis

In this thesis, we have developed several enabling techniques for massive multiple access in future wireless networks. To achieve this goal, we first proposed a novel partially overlapping NOMA clusters scheme, the D-OMA scheme, where different adjacent NOMA clusters are allowed to overlap by certain percentage with each other (**Chapter 2**). The spectral efficiency gained because of spectrum overlapping process is then used to compensate for a potential decrease in the NOMA clusters sizes. Such an enabled reduction on NOMA cluster results in a significant reduction in hardware and computational complexity, especially for terminal wireless UEs with low budgets. In order to enhance the per-UE spectral efficiency and quality of service throughout all network coverage area, we have developed several novel cell-free wireless network architectures that enables distributed APs to simultaneously serve all UEs within their coverage area, using the same time/frequency resources. To reduce the hardware and computational complexity raised by the centralized beamforming of the proposed cell-free network, we first proposed a UE-centric approach that allow each UE to be simultaneously served by the n-th best APs through multi-level NOMA operations (**Chapter 3**). The proposed scheme showed significant improvement in the overall performance and a

significant complexity reduction for terminal wireless UEs. However, as different UEs may belong to different NOMA clusters at different APs, some kind of centralized processing will still be required to optimizes signals of each UEs by all connected APs. To solve this issue, we have designed a scalable network-centric uplink cell-free network architecture that cluster different APs into subgroups with each subgroup act as a low-complexity DAS receiver (**Chapter 4**). Instantaneous clustering of APs at a time slot-bases, however, requires going through all possible clustering configurations and then select the one that results in a best performance. To solve this issue, we have designed a novel hybrid discrete-continuous DRL model that jointly learn the best AP clustering configuration and beamforming matrices from instantaneous CSI of the overall network. The proposed system showed a great potentials to support massive number of UEs with low hardware and computational complexities, while maintaining a satisfactorily per-UE performance compared to the fully centralized cell-free network. To completely remove the requirements related to centralized processing, we also have developed a novel self-partitioning/self-organizing cell-free network that, at a time slot-bases, partition the whole cell-free network into a set of independent cell-free subnetworks (**Chapter 5**). In order to decrease the interference among adjacent cell-free subnetworks, we have designed a novel hybrid analog beamsteering-digital beamforming model for each subnetwork that maximizes the antenna beams over UEs from the same subnetwork and zero-null beams of each APs toward UEs from other cell-free subnetworks. Once beams are optimized through analog beamsteering operations, a digital beamforming scheme is conducted to maximize a joint objective function of UEs at each cell-free subnetwork. Finally, to efficiently conduct network partitioning, analog beamsteering and digital beamforming, we have proposed, implemented and benchmarked several DRL algorithms for both cell-free network partitioning and beamsteering.

6.2 Research Outcomes: Publication List

Most of the technical work of this thesis have been published or submitted for possible publication. Below is a list of published/submitted technical papers related to this thesis work:

Refereed Journal Papers (Published/Submitted)

[6] Y. Al-Eryani and E. Hossain, "Self-Organizing mmWave MIMO Cell-Free Networks With Hybrid Beamforming: A Hierarchical DRL-Based Design," submitted to the ***IEEE Trans. on Commun.***, 2021.

[5] F. Fredj, Y. Al-Eryani, S. Maghsudi, M. Akrout, and E. Hossain, "Distributed Uplink Beamforming in Cell-free Networks Using Deep Reinforcement Learning," submitted to the ***IEEE Trans. on Mobile Computing***, 2021.

[4] Y. Al-Eryani, M. Akrout and E. Hossain, "Antenna Clustering for Simultaneous Wireless Information and Power Transfer in a MIMO Full-Duplex System: A Deep Reinforcement Learning-Based Design," ***IEEE Trans. on Commun.***, 2021 (*Early Access*).

[3] 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 Sel. Areas in Commun.***, 2020 (*Early Access*).

[2] Y. Al-Eryani, E. Hossain and D. I. Kim, "Generalized Coordinated Multipoint (GCoMP)-Enabled NOMA: Outage, Capacity, and Power Allocation," ***IEEE Trans. on Commun.***, vol. 67, no. 11, pp. 7923–7936, Nov. 2019, doi: 10.1109/TCOMM.2019.2931971.

[1] Y. Al-Eryani and E. Hossain, "The D-OMA Method for Massive Multiple Access in 6G: Performance, Security, and Challenges," ***IEEE Veh. Technol. Mag.***, vol. 14, no. 3, pp. 92–99, Sept. 2019.

Refereed Invited Papers:

[IP2] Y. Al-Eryani, E. Hossain, "A Dynamic Cell-Less Architecture for Ultra-Dense Wireless Networks," ***IEEE COMSOC Technical Committees Newsletter, TCN***, Nov. 2019.

[IP1] E. Hossain and Y. Al-Eryani, "Large-scale NOMA: Promises for Massive Machine-type Communication," ***IEEE COMSOC TCCN Newsletter***, Feb. 2019.

6.3 Future Research Directions

As the requirements for more transmission rates, lower latency, more reliability and fairness increases, the development of new techniques and strategies for wireless networks must continue in an increasing rate. In this section, we provide some future research directions and trends based on our achieved thesis work.

- Developing a virtually centralized processing with distributed functionalities: In this research direction, different tasks of massive wireless networks have to be optimally allocated among network resources in a distributed way while guaranteeing the coordination/cooperation with other network entities. Quantum computing and federated learning may play a great deal in guaranteeing privacy and security of transmitted data, which will significantly encourage cooperation among distributed network entities.

- Developing low-complexity algorithms for CSI estimation. This may be achieved through designing distributed pilot assignment algorithms that decreases pilot contamination over massive MIMO cell-free networks.

References

[1] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. on Sel. Areas in Commun.*, pp. 1–1, 2020.

[2] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Veh. Techol. Mag.*, vol. 13, no. 3, pp. 72–80, Sep. 2018.

[3] P. Wang, J. Xiao, and L. P, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," *IEEE Veh. Techol. Mag.*, vol. 1, no. 3, pp. 4–11, Sep. 2006.

[4] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Survey & Tutorials*, vol. 20, no. 1, pp. 629–646, 2018.

[5] R16-38812, "Technical specification group radio access network: Study on non-orthogonal multiple access (NOMA) for NR," in *3GPP Technical Report*, Nov. 2018.

[6] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[7] S. Islam, N. Avazov, O. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Survey & Tutorials*, vol. 19, no. 2, pp. 721–742, May 2017.

[8] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 5, pp. 3444–3458, May 2018.

[9] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, 2017.

[10] 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 J. Select. Areas Commun.*, pp. 1–1, 2020.

[11] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, 2017.

[12] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, 2016.

[13] M. D. Renzo, A. Zappone, M. Debbah, M. Alouini, C. Yuen, J. D. Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," *IEEE J. on Select. Areas in Commun.*, pp. 1–1, 2020.

[14] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, 2018.

[15] 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 J. Select. Areas Commun.*, pp. 1–1, 2020.

[16] Y. Al-Eryani, E. Hossain, and D. I. Kim, "Generalized coordinated multipoint (GCoMP)-enabled NOMA: Outage, capacity, and power allocation," *IEEE Trans. on Wireless Commun.*, vol. 67, no. 11, pp. 7923–7936, Nov. 2019.

[17] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *IEEE Int'l Conf. on Commun. (ICC'19)*, 2019, pp. 1–6.

[18] G. Interdonato, E. Björnson, H. Quoc Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. on Wireless Commun. and Net.*, vol. 2019, no. 1, p. 197, Aug. 2019.

[19] X. Zhang, J. Wang, and H. V. Poor, "Statistical delay and error-rate bounded QoS provisioning over mmWave cell-free mMIMO and FBC-HARQ-IR based 6G wireless networks," *IEEE J. on Select. Areas in Commun.*, vol. 38, no. 8, pp. 1661–1677, 2020.

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

[21] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2017.

[22] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.

[23] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmWave massive MIMO through deep learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10 325–10 329, 2019.

[24] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46 317–46 350, 2019.

[25] S. Chakraborty, E. Björnson, and L. Sanguinetti, "Centralized and distributed power allocation for max-min fairness in cell-free massive MIMO," in *2019 53rd Asilomar Conf. on Sig., Sys., and Computers*, 2019, pp. 576–580.

[26] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised-learning power control for cell-free wireless systems," in *IEEE 30-th Annual Int'l Symposium on Pers., Indoor and Mobile Radio Commun. (PIMRC'19)*, 2019, pp. 1–5.

[27] F. Fredj, Y. Al-Eryani, S. Maghsudi, M. Akrout, and E. Hossain, "Distributed uplink beamforming in cell-free networks using deep reinforcement learning," 2020.

[28] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, no. 12, pp. 126–133, 1997.

[29] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Veh' Technol. Conf. (VTC Spring)*, 2013, pp. 1–5.

[30] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[31] Y. Al-Eryani and E. Hossain, "The D-OMA method for massive multiple access in 6G: Performance, security, and challenges," *IEEE Veh. Techol. Mag.*, vol. 14, no. 3, pp. 92–99, Sep. 2019.

[32] Y. Sun, Z. Ding, X. Dai, and O. A. Dobre, "On the performance of network NOMA in uplink CoMP systems: A stochastic geometry approach," *IEEE Trans. on Wireless Commun.*, vol. 67, no. 7, pp. 5084–5098, July 2019.

[33] Y. Tian, A. R. Nix, and M. Beach, "On the performance of opportunistic NOMA in downlink CoMP networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 998–1001, May 2016.

[34] Y. Li and G. A. Aruma Baduge, "NOMA-aided cell-free massive MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 6, pp. 950–953, Dec. 2018.

[35] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, L. Hanzo, and P. Xiao, "On the performance of cell-free massive MIMO relying on adaptive NOMA/OMA mode-switching," *IEEE Trans. on Wireless Commun.*, pp. 1–1, 2019.

[36] Y. Zhang, H. Cao, M. Zhou, and L. Yang, "Spectral efficiency maximization for uplink cell-free massive MIMO-NOMA networks," in *IEEE Int'l Conf. on Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.

[37] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, March 2017.

[38] A. Abdallah and M. M. Mansour, "Angle-based multipath estimation and beamforming for FDD cell-free massive MIMO," in *IEEE 20'th Inter'l Workshop on Signal Proc. Advances in Wireless Commun. (SPAWC'19)*, 2019, pp. 1–5.

[39] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 264–11 268, 2018.

[40] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, 2019.

[41] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, 2017.

[42] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive mimo at millimeter wave," *IEEE Trans. on Green Commun. and Networking*, vol. 3, no. 3, pp. 651–663, 2019.

[43] M. Attarifar, A. Abbasfar, and A. Lozano, "Modified conjugate beamforming for cell-free massive MIMO," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 2, pp. 616–619, April 2019.

[44] A. Zhou, J. Wu, E. G. Larsson, and P. Fan, "Max-min optimal beamforming for cell-free massive MIMO," pp. 1–1, 2020.

[45] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44 596–44 612, 2019.

[46] ——, "Fronthaul-constrained cell-free massive MIMO with low resolution ADCs," *IEEE Access*, vol. 8, pp. 116 195–116 215, 2020.

[47] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1038–1053, 2020.

[48] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 68, no. 7, pp. 4247–4261, 2020.

[49] A. Liu and V. K. N. Lau, "Joint bs-user association, power allocation, and user-side interference cancellation in cell-free heterogeneous networks," *IEEE Trans. Signal Processing*, vol. 65, no. 2, pp. 335–345, 2017.

[50] X. Huang, X. Zhu, Y. Jiang, and Y. Liu, "Efficient enhanced k-means clustering for semi-blind channel estimation of cell-free massive MIMO," in *IEEE Inter'l Conf. on Commun. (ICC'20)*, 2020, pp. 1–6.

[51] M. Bashar, A. Akbari, K. Cumanan, H. Q. Ngo, A. G. Burr, P. Xiao, M. Debbah, and J. Kittler, "Exploiting deep learning in limited-fronthaul cell-free massive MIMO uplink," *IEEE J. on Sel. Areas in Commun.*, pp. 1–1, 2020.

[52] C. D'Andrea, A. Zappone, S. Buzzi, and M. Debbah, "Uplink power control in cell-free massive MIMO via deep learning," in *2019 IEEE 8'th Int'l Workshop on Comput. Advances in Multi-Sensor Adaptive Proc. (CAMSAP'19)*, 2019, pp. 554–558.

[53] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[54] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 706–709, 2017.

[55] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive mimo approach," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, 2020.

[56] F. Fredj, Y. Al-Eryani, S. Maghsudi, M. Akrout, and E. Hossain, "Distributed uplink beamforming in cell-free networks using deep reinforcement learning," 2020.

[57] O. Semiari, W. Saad, M. Bennis, and M. Debbah, "Integrated millimeter wave and sub-6 GHz wireless networks: A roadmap for joint mobile broadband and ultra-reliable

low-latency communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 109–115, 2019.

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

[59] J. García-Morales, G. Femenias, and F. Riera-Palou, "Energy-efficient access-point sleep-mode techniques for cell-free mmWave massive MIMO networks with non-uniform spatial traffic density," vol. 8, 2020, pp. 137 587–137 605.

[60] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44 596–44 612, 2019.

[61] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. of Select. Topics in Sig. Process.*, vol. 10, no. 3, pp. 436–453, 2016.

[62] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE J. on Sel. Areas in Commun.*, vol. 34, no. 3, pp. 510–527, March 2016.

[63] E. Castañeda, A. Silva, A. Gameiro, and M. Kountouris, "An overview on resource allocation techniques for multi-user MIMO systems," *IEEE Commun. Survey & Tutorials*, vol. 19, no. 1, pp. 239–284, Feb. 2017.

[64] M. Kountouris, D. Gesbert, and T. Sälzer, "Enhanced multiuser random beamforming: dealing with the not so large number of users case," *IEEE J. on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1536–1545, Oct. 2008.

[65] R. Bellman, "On the theory of dynamic programming," *Proceedings of the National Academy of Sciences*, vol. 38, no. 8, pp. 716–719, 1952. [Online]. Available: https://www.pnas.org/content/38/8/716

[66] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013.

[67] G. Rummery, M. Niranjan, and U. of Cambridge. Engineering Department, *On-line Q-learning Using Connectionist Systems*, ser. CUED/F-INFENG/TR. University of Cambridge, Department of Engineering, 1994. [Online]. Available: https://books.google.ca/books?id=JdyRPgAACAAJ

[68] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press Cambridge, Massachusetts London, England., 2017.

[69] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," 2016.

[70] P. Christodoulou, "Soft actor-critic for discrete action settings," 2019.

[71] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Thirtieth AAAI conf. on artificial intelligence*, 2016.

[72] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," 2014.

[73] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

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

[75] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[76] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics)*. Dover Publications, 1965.

[77] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. on Wireless Commun.*, vol. 51, no. 8, pp. 1389–1398, 2003.

[78] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005, pp. i–vi.

[79] K. S. Ahn and R. W. Heath, "Performance analysis of maximum ratio combining with imperfect channel estimation in the presence of cochannel interferences," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 3, pp. 1080–1085, March 2009.

[80] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Mag. Commun.*, vol. 41, no. 3, pp. 108–113, March 2003.

[81] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. New York: Wiley, 2005.

[82] 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 Int'l Workshop on Signal Proc. Advances in Wireless Commun. (SPAWC)*, 2015, pp. 201–205.

[83] R. M. Gray and L. D. Davisson, *An Introduction to Statistical Signal Processing*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.

[84] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. on Sel. Areas in Commun.*, vol. 31, no. 2, pp. 264–273, Fe. 2013.

[85] W. Liao, T. Chang, W. Ma, and C. Chi, "QoS-based transmit beamforming in the presence of eavesdroppers: An optimized artificial-noise-aided approach," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1202–1216, 2011.

[86] S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. Academic Press, Inc., 2000.

[87] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. on Sel. Areas in Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.

[88] R. A. Monzingo.

[89] R. You, Hong Li, and Y. Bar-Ness, "Diversity combining with imperfect channel estimation," *IEEE Trans. on Wireless Commun.*, vol. 53, no. 10, pp. 1655–1662, Oct. 2005.

[90] R. K. Mallik, "Optimized diversity combining with imperfect channel estimation," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1176–1184, March 2006.

[91] O. P. Ronald L. Graham, Donald E. Knuth, *Concrete Mathematics*. Addison-Wesley, Reading MA, 1988.

[92] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[93] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "Rf chain reduction for MIMO systems: A hardware prototype," *IEEE Systems Journal*, pp. 1–12, 2020.

[94] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, 2015.

[95] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 345–364, 2011.

[96] A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, 2007.

[97] W. Tan, S. D. Assimonis, M. Matthaiou, Y. Han, X. Li, and S. Jin, "Analysis of different planar antenna arrays for mmwave massive MIMO systems," in *IEEE Veh. Technol. Conf. (VTC'85 Spring)*, 2017, pp. 1–5.

[98] B. Ai, K. Guan, G. Li, and S. Mumtaz, "Chapter 8–mmwave massive MIMO channel modeling," in *mmWave Massive MIMO*, S. Mumtaz, J. Rodriguez, and L. Dai, Eds. Academic Press, 2017, pp. 169–194.

[99] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[100] C. A. Balanis, *Antenna theory: analysis and design*, 4th ed. Hoboken, NJ: John Wiley, 2016.

[101] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[102] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," in *Machine Learning*, 1992, pp. 229–256.

[103] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," 1965.

[104] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," *arXiv e-prints*, p. arXiv:1801.01290, Jan. 2018.

[105] B. D. Ziebart, "Modeling purposeful adaptive behavior with the principle of maximum causal entropy," Ph.D. dissertation, USA, 2010, aAI3438449.

[106] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[107] P. V. Csukhatme, "Tests of significance for samples of the χ^2-population with two degrees of freedom," *Annals of Eugenics*, vol. 8, no. 1, pp. 52–56, Aug. 2012.

[108] H. A. David and H. N. Nagaraja, *Order Statistics, Third Edition.* John Wiley & Sons, Inc., 2005.

[109] L. P. Qian, Y. J. Zhang, and J. Huang, "Mapel: Achieving global optimality for a non-convex wireless power control problem," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 3, pp. 1553–1563, March 2009.

[110] F. E. Satterthwaite, "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, vol. 2, no. 6, p. 110, Dec. 1946.

APPENDIX

Appendix A

Proof of Theorem 1

First, let us define $I^k_{\text{INCI}} = \sum_{l=n+1}^{M} \Phi_l |h_{k,l}|^2 = \sum_{l=n+1}^{M} y_{k,l}$, where $Y_{k,l} \sim \text{Exp}\,(\lambda_{k,l})$ and $\lambda_{k,l} = 1/2\Phi_w \sigma^2_{k,l}$. Using the theory of order statistics, the PDF of I^k_{ICI} can be defined as

$$\begin{aligned} f_{I^k_{\text{ICI}}}(y_{k,n+1}, \ldots, y_{k,M}) &= \sum_{i_{n+1},\ldots,i_M}^{\{1,2,\ldots,M\}} f_{Y_{k,i_{n+1}}}(y_{k,n+1}) \ldots f_{Y_{k,i_M}}(y_{k,M}) \times \prod_{j=1}^{n} \Big(1 - F_{Y_{k,i_j}}(y_{k,n+1})\Big) \\ &= \sum_{i_{n+1},\ldots,i_M}^{\{1,2,\ldots,M\}} \prod_{l=n+1}^{M} \lambda_{k,i_l} e^{-\lambda_{k,i_l} y_{k,l}} \prod_{j=1}^{n} e^{-\lambda_{k,i_j} y_{k,n+1}}, \end{aligned} \tag{A.1}$$

where $y_{k,M} \leq \ldots \leq y_{k,n+1}$ and $\{i_1, \ldots, i_M\}$ are distinct indices that take values from $\{1, \ldots, M\}$. Accordingly, the value of $\bar{I}_{\text{INCI}}$ can be defined as

$$\begin{aligned} \bar{I}^m_{\text{INCI}} = &\sum_{i_{n+1},\ldots,i_M}^{\{1,2,\ldots,M\}} \int_0^\infty \int_{y_{k,M}}^\infty \ldots \int_{y_{k,n+2}}^\infty \sum_{i=n+1}^{M} y_{k,i} \\ &\times \prod_{l=n+1}^{M} \lambda_{k,i_l} e^{-\lambda_{k,i_l} y_l} \prod_{j=1}^{n} \lambda_{k,i_l} e^{-\lambda_{k,i_j} y_{k,n+1}} d \ldots dy_{k,m}. \end{aligned} \tag{A.2}$$

Due to the dependence among $y'_i s$, this integral cannot be changed into a product of independent integrals. Therefore, $\bar{I}^k_{\text{INCI}}$ can be rewritten as

$$\bar{I}^k_{\text{INCI}} = \sum_{i_{n+1},\ldots,i_M}^{\{1,2,\ldots,M\}} \int_0^\infty \ldots \int_0^\infty \sum_{w=1}^{M-n} w x_w \times \left(\prod_{j=n+1}^{M} \lambda_{i_j} e^{\left(\sum_{q=n+1}^{j} \lambda_{k,i_q} + \sum_{m=1}^{n} \lambda_{k,i_m}\right) x_j} \right) \ldots dx_{M-n}, \tag{A.3}$$

where we have used Sukhatme transformation of rvs such that $x_i = y_{k,n+i} - y_{k,n+i+1}$ and $x_n = y_{k,M}$ represent an independent random variables [107]. Due to the independence among xs, (A.3) can be easily changed into a sum of a product of one-dimensional integral. Hence, we obtain (3.11).

Appendix B

Proof of Theorem 2

In (3.12), we have two types of ordering to be considered. The first one is the ordering of BSs w.r.t the k-th UE. The second ordering is the ordering of the k-th UE w.r.t all clusters it belongs to. First, we will consider the ordering of the BSs connected to the k-th UE. Accordingly, the PDF of z_k can be defined as

$$\begin{aligned} f_{Z_k}(z_{k,1},\ldots,z_{k,n}) &= \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} f_{Z_{k,i_1}}(z_{k,1})\ldots f_{Z_{k,i_n}}(z_{k,n}) \prod_{j=n+1}^{M} F_{Z_{k,i_j}}(y_n) \\ &= \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} \prod_{l=1}^{n} \alpha_{k,i_l} e^{-\alpha_{k,i_l} z_{k,l}} \prod_{j=n+1}^{M} \left(1 - e^{-\alpha_{k,i_j} z_{k,n}}\right). \end{aligned} \tag{B.1}$$

Utilizing the Sukhatme transformation, the MGF of Z_k is given by

$$M_{Z_k}(s) = \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} \left[\prod_{q=1}^{n} \frac{\alpha_{k,i_q}}{\sum_{r=1}^{q} \alpha_{k,i_r} - qs} - \sum_{h_1=1}^{M-n} (-1)^{h_1} \sum_{j_1 \le \ldots \le j_{h_1}}^{\{n,\ldots,M\}} \prod_{d=1}^{n} \frac{\alpha_{k,i_d}}{C_d^{k(h_1,i,j)} - ds} \right], \tag{B.2}$$

where $\{j_1,\ldots,j_{h_1}\}$ are distinct ordered indices taking values from $\{n,\ldots,M\}$ and $C_d^{k(h_1,i,j)}$ is defined as

$$C_{d(h_1,i,j)}^{k} = \begin{cases} \sum_{r=1}^{d} \alpha_{k,i_r} & d \in [1,\cdots,n-1] \\ \sum_{r=1}^{d} \alpha_{k,i_r} + \sum_{l=1}^{h_1} \alpha_{k,j_l} & d = n \end{cases}$$

. To obtain the PDF $f_{Z_k}(z)$, it is convenient to express (B.2) as a partial fraction expression. Specifically,

$$M_{Z_k}(s) = \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} \left(\prod_{q=1}^{n} \frac{\alpha_{k,i_q}}{q} \right) \left[\sum_{t_1=1}^{n} \frac{\eta_{t_1}^{k}}{\rho_{t_1}^{k} - s} - \sum_{h_1=1}^{M-n} (-1)^{h_1} \sum_{j_1 \le \ldots \le j_{h_1}}^{\{n,\ldots,M\}} \sum_{t_2=1}^{n} \frac{\eta_{t_2}^{k}}{\rho_{t_2}^{k} - s} \right], \tag{B.3}$$

where $\rho_{t_1}^k = \sum_{r=1}^{t_1} \alpha_{k,i_r}/t_1$, $\rho_{t_2}^k = C_{t_2}^k(h_1,i,j)/t_2$, $\eta_{t_1}^k = \prod_{m=1,m\neq t_1}^{n} \left(\rho_m^k - \rho_{t_1}^k\right)^{-1}$ and $\eta_{t_2}^k = \prod_{m=1,m\neq t_2}^{n} \left(\rho_m^k - \rho_{t_2}^k\right)^{-1}$. Note that (B.3) is valid only under the assumption that all $\rho_{t_1}^k$(and $\rho_{t_2}^k$) are distinct (the i.n.d case). Upon finding the Laplace transform $L_{Z_k}(x) = M_{Z_k}(-x)$ and using the theory of inverse Laplace transform, the PDF of Z_k (denoted by $f_{Z_k}(z)$) is then given by

$$f_{Z_k}(z) = \sum_{i_1,\ldots,i_n}^{\{1,2,\ldots,M\}} J_1(k,i) \left[\sum_{t_1=1}^{n} \eta_{t_1}^k e^{-\rho_{t_1}^k z} - \sum_{h_1=1}^{M-n} (-1)^{h_1} \sum_{j_1\leq\ldots\leq j_{h_1}}^{\{n,\ldots,M\}} \sum_{t_2=1}^{n} \eta_{t_2}^k e^{-\rho_{t_2}^k z} \right], \quad \text{(B.4)}$$

where $J_1(k,i) = \left(\prod_{q=1}^{n} \frac{\alpha_{k,i_q}}{q}\right)$. Now, we consider the ordering of the k-th UE within the cluster of its best serving BS. By utilizing the CDF expression of the k-th order statistics for the set of i.n.d rvs given in [108, Eq. 5.2.1], the outage probability of the proposed system is given as in **2**.

Appendix C

Proof of Lemma 1

To proof the convexity of the problem in (4.15), we need to first prove that the objective function is concave and then show that all constraints represent an affine transformation of dependent variables. It is apparent that the objective function (let us denote it as R_s) is twice differentiable for all dependent variables (w_{km}). However, due to the two-dimensional nature of dependent variables (functions of m and k), finding the Hessian matrix of the partial second derivative even for the most simplified model ($M = 2$ and $K = 2$) would be very lengthy and tedious. Therefore, we use an intuitive method to proof the concavity of R_s.

The objective function at (4.15) can be rewritten as

$$R_s = \sum_{k=1}^{K} \left[\underbrace{\log_2 \left(\theta_1(k) + 1\right)}_{f_1(k)} + \underbrace{\log_2 \left(\frac{1}{\theta_2(k) + 1} \right)}_{f_2(k)} \right], \tag{C.1}$$

where $\theta_1(k) = \sum_{m=1}^{M} \left(w_{km} + \sum_{i=k+1}^{K} w_{im} \right) \gamma_{km}$ and $\theta_2(k) = \sum_{m=1}^{M} \left(\sum_{i=k+1}^{K} w_{im} \right) \gamma_{km}$. It is apparent that $f_1(f_2)$ is a monotonically increasing (decreasing) function of $\theta_1(k)\,(\theta_2(k))$ with f_1 being a strictly concave function and f_2 being a strictly convex function (logarithmic functions). Additionally, due to the set of constraints $\mathbf{C}_2$, the overall value of $\theta_1(k)$ will be always strictly greater than that of $\theta_2(k)$ and any increase in $\theta_2(k)$ will result in a higher increase in $\theta_1(k)$. Accordingly, the degree of convexity of $f_2(k)$ will be always greater that the degree of concavity of $f_1(k)$ which will make the summation $f_1(k) + f_2(k)$ to be always

strictly convex (for all $w_{ij}, i = 1, \ldots, K$ and $j = 1, \ldots, M$). Finally, since R_s represents a positive linear sum of convex functions, R_s is a convex function as well. Note that without the constraints $\mathbf{C}_2$, the problem in (4.15) will be neither convex nor concave. Nevertheless, a global optimal point for such non-convex problem can be found using the 'MAPEL' algorithm [109].

Now, we need to prove that the constraints $\mathbf{C}_1$ through $\mathbf{C}_3$ represent affine constraints. It is apparent that the constraints $\mathbf{C}_2$ and $\mathbf{C}_3$ represent affine functions for all w_{ij}. Additionally, the constraints $\mathbf{C}_1$ can be rewritten as $\sum_{m=1}^{M} w_{km}\gamma_{km} - \gamma_{\text{th}}^{k}\left(1 + \sum_{m=1}^{M}\left(\sum_{i=k+1}^{K} w_{im}\right)\gamma_{km}\right) \geq 0$, where $\gamma_{\text{th}}^{k} = 2^{R_k} - 1$, which represents an affine function as well. Hence, **Lemma 1** is proved.

Appendix D

Derivation of problem P_1 in (4.15)

Since the problem **J** in (4.15) is convex with affine constraints, the Karush-Kuhn-Tucker (KKT) conditions can be given by taking the partial derivative of (3.22) w.r.t. $\{w_{km}\}_{k=1,2,3}^{m=1,2}$, $\{\eta_k\}_{k=1,2,3}$, μ_1, $\{\mu_i\}_{i=1,2}$ and $\{\tau_m\}_{m=1,2}$ as follows:

$$\frac{\partial \mathcal{L}}{\partial w_{11}} = \frac{\gamma_{11}}{\sum_{m=1}^{2}\left(\sum_{i=1}^{3} w_{im}^{*}\right)\gamma_{1m}+1} + \eta_1^{*}\gamma_{11} - \mu_1^{*}\gamma_{21} - \mu_2^{*}\gamma_{31} - \tau_1^{*} = 0, \tag{C.1a}$$

$$\frac{\partial \mathcal{L}}{\partial w_{12}} = \frac{\gamma_{12}}{\sum_{m=1}^{2}\left(\sum_{i=1}^{3} w_{im}^{*}\right)\gamma_{1m}+1} + \eta_1^{*}\gamma_{12} - \mu_1^{*}\gamma_{22} - \mu_2^{*}\gamma_{32} - \tau_2^{*} = 0, \tag{C.1b}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_{21}} = {} & \frac{-\gamma_{11}\left(\sum_{m=1}^{2} w_{1m}^{*}\gamma_{1m}\right)}{\left(\sum_{m=1}^{2}\left[\sum_{i=1}^{3} w_{ik}^{*}\right]\gamma_{1m}+1\right)\left(\sum_{m=1}^{2}\left[\sum_{i=2}^{3} w_{im}^{*}\gamma_{1m}\right]+1\right)} \\ & + \frac{\gamma_{21}}{\sum_{m=1}^{2}\left(\sum_{i=2}^{3} w_{jm}^{*}\right)\gamma_{2m}+1} - \eta_1^{*}\gamma_{\mathrm{th}}^{1}\gamma_{11} \\ & + \eta_2^{*}\gamma_{21} + \mu_1^{*}\gamma_{21} + \mu_2^{*}\gamma_{31} - \mu_3^{*}\gamma_{31} - \tau_1^{*} = 0,\end{aligned} \tag{C.1c}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial w_{22}} = {} & \frac{-\gamma_{12}\left(\sum_{m=1}^{2} w_{1m}^{*}\gamma_{1m}\right)}{\left(\sum_{m=1}^{2}\left[\sum_{i=1}^{3} w_{im}^{*}\right]\gamma_{1m}+1\right)\left(\sum_{m=1}^{2}\left[\sum_{j=2}^{3} w_{jm}^{*}\gamma_{1m}\right]+1\right)} \\ & + \frac{\gamma_{2,2}}{\sum_{m=1}^{2}\left(\sum_{j=2}^{3} w_{jm}^{*}\right)\gamma_{2m}+1} - \eta_1^{*}\gamma_{\mathrm{th}}^{1}\gamma_{12} \\ & + \eta_2^{*}\gamma_{22} + \mu_1^{*}\gamma_{22} + \mu_2^{*}\gamma_{32} - \mu_3^{*}\gamma_{32} - \tau_2^{*} = 0,\end{aligned} \tag{C.1d}$$

$$\frac{\partial \mathcal{L}}{\partial w_{31}} = \frac{-\gamma_{11}\left(\sum_{m=1}^{2} w_{1m}^{*}\gamma_{1m}\right)}{\left(\sum_{m=1}^{2}\left[\sum_{i=1}^{3} w_{im}^{*}\right]\gamma_{1m}+1\right)\left(\sum_{m=1}^{2}\left[\sum_{j=2}^{3} w_{jm}^{*}\gamma_{1m}\right]+1\right)} + \frac{-\gamma_{21}\left(\sum_{m=1}^{2} w_{2m}^{*}\gamma_{2m}\right)}{\left(\sum_{m=2}^{2}\left[\sum_{j=2}^{3} w_{jm}^{*}\right]\gamma_{2m}+1\right)\left(\sum_{m=1}^{2} w_{3m}^{*}\gamma_{2m}+1\right)} + \frac{\gamma_{31}}{\sum_{m=1}^{2} w_{3m}^{*}\gamma_{3m}+1} - \eta_{2}^{*}\gamma_{21} + \eta_{3}^{*}\gamma_{31} - \mu_{3}^{*}\gamma_{31} - \tau_{1}^{*} = 0, \tag{C.1e}$$

$$\frac{\partial \mathcal{L}}{\partial w_{32}} = \frac{-\gamma_{12}\left(\sum_{m=1}^{2} w_{1m}^{*}\gamma_{1m}\right)}{\left(\sum_{m=1}^{2}\left[\sum_{i=1}^{3} w_{im}^{*}\right]\gamma_{1m}+1\right)\left(\sum_{m=1}^{2}\left[\sum_{j=2}^{3} w_{jm}^{*}\gamma_{1m}\right]+1\right)} + \frac{-\gamma_{22}\left(\sum_{m=1}^{2} w_{2m}^{*}\gamma_{2m}\right)}{\left(\sum_{m=2}^{2}\left[\sum_{j=2}^{3} w_{jm}^{*}\right]\gamma_{2m}+1\right)\left(\sum_{m=1}^{2} w_{3m}^{*}\gamma_{2m}+1\right)} + \frac{\gamma_{32}}{\sum_{m=1}^{2} w_{3m}^{*}\gamma_{3m}+1} - \eta_{2}^{*}\gamma_{22} + \eta_{3}^{*}\gamma_{32} - \mu_{3}^{*}\gamma_{32} - \tau_{2}^{*} = 0, \tag{C.1f}$$

$$\frac{\partial \mathcal{L}}{\partial \eta_{q}} = \sum_{m=1}^{2}\left(w_{qm}^{*} - \gamma_{\text{th}}^{q}\sum_{j=q+1}^{3} w_{jm}^{*}\right)\gamma_{qm} - \gamma_{\text{th}}^{q} \geq 0, \tag{C.1g}$$

$$\frac{\partial \mathcal{L}}{\partial \mu_{1}} = P_{s} - \sum_{m=1}^{2}\left(w_{1m}^{*} - w_{2m}^{*}\right)\gamma_{2m} = 0, \tag{C.1h}$$

$$\frac{\partial \mathcal{L}}{\partial \mu_{d}} = P_{s} - \sum_{m=1}^{2}\left(w_{(d-1)m}^{*} - \sum_{l=d}^{3} w_{lm}^{*}\right)\gamma_{3m} = 0, \tag{C.1i}$$

$$\frac{\partial \mathcal{L}}{\partial \tau_{v}} = 1 - \sum_{k=1}^{3} w_{kv}^{*} = 0, \tag{C.1j}$$

where $q = 1, 2, 3$, $d = 2, 3$ and $v = 1, 2$. The set of points $\{w_{km}^{*}\}_{k=1,2,3}^{m=1,2} \geq 0$, $\{\eta_{m}^{*}\}_{m=1,2,3} \geq 0$, $\mu_{1}^{*} \geq 0$, $\{\mu_{i}^{*}\}_{i=1,2} \geq 0$ and $\{\tau_{m}^{*}\}_{m=1,2} \geq 0$ that satisfy conditions (C.1) are both the primal and dual optimal solutions for problem $\mathbf{J}$ in (4.15). In addition, the primal feasibility conditions, i.e. conditions $\mathbf{C}_{1} - \mathbf{C}_{3}$ has to be check for every feasible solutions. Since the equations (C.1) are differentiable, we may utilize one of the numerical methods used for solving a set of differentiable non-linear equations such as Newton or Broyden methods [103].

Appendix E

Proof of Theorem 3

First, let us define $X = \sum_{m=1}^{M} |\tilde{g}_{mk}|^2$ and

$$Y = \sum_{m=1}^{M} \left[\sum_{l=1, l\neq k}^{K} |\tilde{g}_{ml}|^2 + \sum_{v=1, v\neq k}^{K} |\tilde{g}_{mv}|^2 + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} |\tilde{g}_{mu}|^2 \right]. \tag{D.1}$$

Let $W = Y + 1 \approx Y$. (This assumption is valid for large values of K and M.) The derivation of closed-form expression for X and Y is very complicated due to the non-equal rate parameters ($\tilde{\beta}_{m,k}$ and $\tilde{\beta}_{m,l}$) which makes the MGF method unusable to derive $f_X(x)$ and $f_Y(y)$. Here, we use an accurate approximation of the sum of independent Gamma rvs with different shape and rate parameters. This is inspired by the well-known *Welch-Satterthwaite* approximation [110] which is used to approximate the sum of independent Chi-square distributions. Here, we use *moment matching* method to match the mean and variance of the sum of i.n.d gamma rvs (different shape and rate parameters) with those of a single gamma rv. Accordingly, we have $X \sim \mathcal{G}\left(\dot{\alpha}_{mk}, \dot{\beta}_{mk}\right)$ and $Y \sim \mathcal{G}\left(\dot{\alpha}_{mk'}, \dot{\beta}_{mk'}\right)$, where

$$\dot{\alpha}_{mk} = \frac{\left(\sum_{m=1}^{M} \tilde{\alpha}_{mk}/\tilde{\beta}_{mk}\right)^2}{\sum_{m=1}^{M} \tilde{\alpha}_{mk}/\tilde{\beta}_{mk}^2}, \tag{D.2}$$

$$\dot{\beta}_{mk} = \frac{\sum_{m=1}^{M} \tilde{\alpha}_{mk}/\tilde{\beta}_{mk}^2}{\sum_{m=1}^{M} \tilde{\alpha}_{mk}/\tilde{\beta}_{mk}}, \tag{D.3}$$

$$\dot{\alpha}_{mk'} = \frac{\left(\sum_{m=1}^{M} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{ml}}{\tilde{\beta}_{ml}} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{mv}}{\tilde{\beta}_{mv}} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{mu}}{\tilde{\beta}_{mu}}\right]\right)^2}{\sum_{m=1}^{M} \left[\sum_{l=1, l\neq k}^{K} \frac{\tilde{\alpha}_{ml}}{\tilde{\beta}_{ml}^2} + \sum_{v=1, v\neq k}^{K} \frac{\tilde{\alpha}_{mv}}{\tilde{\beta}_{mv}^2} + \sum_{q=1, q\neq k}^{K} \sum_{u=1, u\neq q}^{K} \frac{\tilde{\alpha}_{mu}}{\tilde{\beta}_{mu}^2}\right]}, \tag{D.4}$$

$$\dot{\beta}_{mk'} = \frac{\sum_{m=1}^{M}\left[\sum_{l=1,l\neq k}^{K}\frac{\tilde{\alpha}_{ml}}{\tilde{\beta}_{ml}^2}+\sum_{v=1,v\neq k}^{K}\frac{\tilde{\alpha}_{mv}}{\tilde{\beta}_{mv}^2}+\sum_{q=1,q\neq k}^{K}\sum_{u=1,u\neq q}^{K}\frac{\tilde{\alpha}_{mu}}{\tilde{\beta}_{mu}^2}\right]}{\sum_{m=1}^{M}\left[\sum_{l=1,l\neq k}^{K}\frac{\tilde{\alpha}_{ml}}{\tilde{\beta}_{ml}}+\sum_{v=1,v\neq k}^{K}\frac{\tilde{\alpha}_{mv}}{\tilde{\beta}_{mv}}+\sum_{q=1,q\neq k}^{K}\sum_{u=1,u\neq q}^{K}\frac{\tilde{\alpha}_{mu}}{\tilde{\beta}_{mu}}\right]}. \tag{D.5}$$

Note that this approximation becomes very close to the exact expression even for small to moderate number of APs and/or UEs. This makes it useful for modeling large-scale cell-free architectures. The CDF of the ratio distribution ($Z = \frac{X}{W}$) can be expressed as

$$F_Z(z) = \mathbb{P}\left(x \leq zW\right) = \int_0^{\infty} \mathbb{P}\left(x \leq zw\right) f_Y(w) dw. \tag{D.6}$$

By substituting $f_Y(w)$ and $\mathbb{P}\left(x \leq zw\right) = \gamma\left(\dot{\alpha}_{mk}, \dot{\beta}_{mk} zw\right)/\Gamma\left(\dot{\alpha}_{mk}\right)$, where $\gamma(.)$ is the upper incomplete gamma function [76, Eq. 6.5.3] in (**A**.1), we have

$$\begin{aligned} F_Z(z) = &\frac{\dot{\beta}_{mk'}^{\dot{\alpha}_{mk'}}}{\Gamma(\dot{\alpha}_{mk})\Gamma(\dot{\alpha}_{mk'})} \\ &\times \int_0^{\infty} x^{\dot{\alpha}_{mk'}-1} e^{-\dot{\beta}_{mk'}x} \gamma\left(\dot{\alpha}_{mk}, \dot{\beta}_{mk} zw\right) dw. \end{aligned} \tag{D.7}$$

This integral can be solved using the following relation: [86, Eq. 6.455.2]

$$\begin{aligned} \int_0^{\infty} x^{\mu-1} e^{-\beta x} \gamma\left(\nu, \alpha x\right) dx = &\frac{\alpha^{\nu}\Gamma\left(\mu+\nu\right)}{\nu\left(\alpha+\beta\right)^{\mu+\nu}} \\ &\times {}_2F_1\left(1, \mu+\nu; \nu+1; \frac{\alpha}{\alpha+\beta}\right), \end{aligned} \tag{D.8}$$

where $\mathrm{Re}\left(\alpha+\beta\right) > 0, \mathrm{Re}\beta > 0$ and $\mathrm{Re}\left(\mu+\nu\right) > 0$. Not that practical values of $\alpha_{mk'}, \beta_{mk'}, \alpha_{mk}, \beta_{mk}$ and z lies within the range of these integral conditions. By substituting (D.8) in (D.7) and utilizing the relationship $P_{\text{out}}^{(k)} = F_Z\left(\gamma_{\text{th}}\right)$, **Theorem 3** is proved.