Cell-Free Massive MIMO for Next-Generation Communication and Sensing Systems

by

Umut Demirhan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2024 by the
Graduate Supervisory Committee:

Ahmed Alkhateeb, Chair
Gautam Dasarathy
Georgios Trichopoulos
Nicolò Michelusi

ARIZONA STATE UNIVERSITY

May 2024

ABSTRACT

With the significant advancements of wireless communication systems that aim to meet exponentially increasing data rate demands, two promising concepts have appeared: (i) Cell-free massive MIMO, which entails the joint transmission and processing of the signals allowing the removal of classical cell boundaries, and (ii) integrated sensing and communication (ISAC), unifying communication and sensing in a single framework. This dissertation aims to take steps toward overcoming the key challenges in each concept and eventually merge them for efficient future communication and sensing networks.

Cell-free massive MIMO is a distributed MIMO concept that eliminates classical cell boundaries and provides a robust performance. A significant challenge in realizing the cell-free massive MIMO in practice is its deployment complexity. In particular, connecting its many distributed access points with the central processing unit through wired fronthaul is an expensive and time-consuming approach. To eliminate this problem and enhance scalability, in this dissertation, a cell-free massive MIMO architecture adopting a wireless fronthaul is proposed, and the optimization of achievable rates for the end-to-end system is carried out. The evaluation has shown the strong potential of employing wireless fronthaul in cell-free massive MIMO systems.

ISAC merges radar and communication systems, allowing effective sharing of resources, including bandwidth and hardware. The ISAC framework also enables sensing to aid communications, which shows a significant potential in mobile communication applications. Specifically, radar sensing data can address challenges like beamforming overhead and blockages associated with higher frequency, large antenna arrays, and narrow beams. To that end, this dissertation develops radar-aided beamforming and blockage prediction approaches using low-cost radar devices and evaluates them in real-world systems to verify their potential.

At the intersection of these two paradigms, the integration of sensing into cell-free

massive MIMO systems emerges as an intriguing prospect for future technologies. This integration, however, presents the challenge of considering both sensing and communication objectives within a distributed system. With the motivation of overcoming this challenge, this dissertation investigates diverse beamforming and power allocation solutions. Comprehensive evaluations have shown that the incorporation of sensing objectives into joint beamforming designs offers substantial capabilities for next-generation wireless communication and sensing systems.

To my family.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

With the significant advances in technology in the last several decades, such as cellular phones, higher-quality entertainment systems, and smart wearable devices, wireless communication has become an integral part of daily life. Consequently, the demand for higher data rates and more capable wireless communication systems has been ever-increasing. To meet this demand in recent years, developments in communication systems have focused on the use of larger antenna arrays for directionality gains within multiple-input multiple-output (MIMO) systems, the use of higher-frequency bandwidth due to the more extensive availability, and the densification of the networks. These developments resulted in challenges, specifically, accurately directing the narrow beams, coordination between basestations, and interference management. This has also brought out various opportunities for further advancement, notably (i) cell-free massive MIMO and (ii) integrated sensing and communications.

Cell-free massive MIMO is the concept where the classical cellular systems with cell boundaries, which caused interference and limitations in seamless connection, are replaced with a cell-less approach with the coordination among the access points. In cell-free Massive MIMO, the access points serve all the users in the area jointly. This way, it takes advantage of densification and also provides diversity advantages with the utilization of multiple access points for each user. Further, with the common adoption of channel estimation with channel reciprocity in cell-free massive MIMO systems, its overhead was significantly reduced. Thanks to these advantages, cell-free massive MIMO has become an important theme in research for future communication systems.

Integrating the sensing and communications systems, under the umbrella of *integrated*

sensing and communications (ISAC), has also attracted significant interest both from academia and industry. In essence, ISAC aims to merge radar sensing and communication operations, both of which use electromagnetic waves for different purposes, by providing a more comprehensive framework that can allow the sharing of resources that includes the available bandwidth and hardware. This framework mainly covers three fundamental approaches, given as (a) joint sensing and communications, where the design of the signals is carried out to support both sensing and communications, (b) sensing aided by communication, where the communication signals are utilized for sensing, (c) communication aided by sensing, where the sensing information is utilized to enhance the communication performance.

In the ISAC framework, *sensing-aided communication* (c) has been of particular importance since the significant bottlenecks in the current communication systems can be mitigated with radar sensing. Radar sensing has long been utilized to track mobile targets for both military and civil purposes. The tracking information of the users and other targets can be crucial in overcoming the challenges in communication due to mobility. For instance, accurate beamforming toward highly mobile users is a difficult objective because of the large number of narrow beams that can efficiently take advantage of MIMO antenna arrays. This problem, however, can be resolved with the aid of radar sensing, which can provide efficient localization/tracking of mobile users, and this tracking can aid communications to provide seamless high data rates.

Moreover, the integration of joint sensing and communication (a) also holds significant promise, offering the potential to support communication and various sensing applications. The dense deployment and collaborative processing capabilities inherent in cell-free massive MIMO can amplify the effectiveness of ISAC in systems employing a multi-static sensing design. In this setup, signals transmitted by diverse APs for sensing can be received and processed individually at different APs and/or collectively at a central unit. Exploring the full realization of this potential feature presents a compelling avenue with substantial

prospects for future communication systems.

Although these advances provide an important perspective for resolving the problems in actual deployments, there are still challenges to be overcome. In particular, (i) the coordination in cell-free massive MIMO and its deployment brings significant costs due to the wired fronthaul, (ii) the gap between theoretical and practical real-world applications in radar-aided communications, in particular beamforming and blockage prediction/handover, is still too wide, (iii) employment of ISAC in densely deployed cell-free massive MIMO and its system design. In the following sections, the details of these challenges are presented.

1.1 Fronthaul Limitations in Cell-free Massive MIMO

A critical challenge with the current cell-free massive MIMO systems is its reliance on optical fiber-based fronthaul for distributing data and synchronization signals from the central processing unit (CPU) to the access points (APs). This highly increases its infrastructure cost and installation time and limits the flexibility and scalability of the cell-free massive MIMO deployment. Specifically, to realize cell-free massive MIMO, it is interesting to place a large number of dense access points, however, the deployment of access points requires a high-capacity fronthaul which results in significant costs. In the literature, the fronthaul cost and limitations have partially been taken into account with the design of the fronthaul limited systems in many studies [1–4]. These studies mainly aimed to optimize the performance of cell-free massive MIMO with fronthaul rate and quantization constraints. With the optimization and evaluation of the fronthaul limited systems, it becomes possible to use more inexpensive alternatives to fiber optic cables. The work in the literature, however, has not eliminated the wired fronthaul requirement, whose deployment can still render significant costs due to placement difficulties. These motivate the elimination of wire-link, which could also enable a flexible installation, and the design and

investigation of cell-free massive MIMO systems with wireless fronthaul.

1.2 Realization of Radar-aided Communication in Real-World

Millimeter wave (mmWave) and terahertz (THz) communications systems rely on the beamforming gains of the narrow beams to achieve sufficient received signal power. Finding the best narrow beam (or beam pair), however, requires high beam training overhead, which makes it hard for these systems to support highly mobile applications such as vehicular, drone, or augmented/virtual reality communications [5]. One important observation here is that the beam selection problem highly relies on the transmitter/receiver locations and the geometry/characteristics of the surrounding environment. This means that acquiring some awareness about the surrounding environment and the transmitter/receiver locations could potentially help the mmWave beam selection problem. An efficient way to acquire this awareness is the use of radar sensors, which have been successfully utilized to locate and track mobile targets. The work in the literature has mainly focused on joint sensing and communication systems [6], which require fundamental changes in the current communication systems to be fully realized. Resolving this, off-the-shelf low-cost radar sensors can easily be attached to the current systems to aid the beamforming [7]. Although initial investigations toward these directions have been carried out, taking these approaches to the real world and utilizing them in actual systems (i) require a deeper investigation into the problem, (ii) taking the practical concerns into account in the solution design, and (iii) implementation of radar-aided communication systems in the real world. Further, with the availability of real-world data, machine learning, which has not been considered in the literature, could be efficiently utilized. These motivate the need for practical applications and real-world evaluation of radar-aided beamforming.

Future wireless networks also attempt to meet the increasing demand for low latency and high reliability. With the use of the higher frequency bands, millimeter-wave (mmWave)

and sub-terahertz (sub-THz), the propagation characteristics at these frequencies also make the communication highly susceptible to the blockages. Specifically, there are two important characteristics of mmWave/sub-THz communication systems: (i) These systems rely mainly on line-of-sight (LOS) links to guarantee sufficient receive signal power, and (ii) this dependency on LOS links coupled with the high penetration loss at mmWave/THz bands. These make the communication systems very sensitive to blockages. If these links are blocked, for example by moving objects, this could cause sudden performance degradation or even a link disconnection, which highly challenges the reliability and latency of these networks. The blockage problem could also be mitigated by the use of sensing systems. In particular, the environment information obtained from the sensing can aid the prediction of the blockages ahead of time. The information of the future blockages can be utilized to eliminate the latency and reliability issues, e.g., through proactive handover. With this motivation, it is interesting to develop radar-aided blockage prediction methods and test them in the real-world to evaluate the potential.

1.3 Joint Sensing and Communication in Cell-free Massive MIMO

With the shift towards densely deployed APs in future communication systems, it becomes natural to extend the ISAC functionality into the distributed deployments. This extension presents the opportunity to process the sensing information in the distributed system to develop solutions that can take advantage of radar signals' observation from multiple points. Taking advantage of this opportunity, however, brings the additional difficulty of the design, where the beamforming of the cell-free massive MIMO system needs to be able to communicate users while sensing the target efficiently by taking multi-static sensing into account. This motivates to develop beamforming and resource allocation solutions for the cell-free ISAC MIMO systems.

1.4 Overview of Contributions

The critical challenges discussed in Section 1.1, Section 1.2, and Section 1.3 are addressed in this dissertation. The primary contributions can be summarized as follows.

- In Chapter 2, the cell-free massive MIMO architecture with the wireless fronthaul is proposed. Specifically, the chapter first presents the model of the proposed architecture, problem formulation and solution development for its rate optimization. Further in this chapter, the proposed system is extended with a mixed-fronthaul approach, where the fronthaul is a mixture of wired and wireless fronthaul links. Accordingly, the earlier solution is extended to apply to the mixed-fronthaul scenario. Finally, the proposed architectures and solutions are evaluated, and the potential of the wireless fronthaul in cell-free massive MIMO systems is revealed.

- In Chapter 3, the radar-aided beamforming problem is formulated for an off-the-band low-cost radar. Specifically, two different beamforming problems are formulated for this radar: (i) Beam prediction, where a single radar measurement is utilized to predict the immediate beam, and (ii) beam tracking, where a sequence of radar measurements is utilized to predict immediate or future beams. For each of these problems, two solutions adopting the classical signal processing and deep learning approaches are developed. Further, a real-world dataset is built for the evaluation of these systems, and the performances of the radar-aided beamforming solutions are presented. The results highlight the potential of the radar for the communication systems in real-world.

- In Chapter 4, in a similar direction, the radar-aided proactive blockage prediction problem is formulated for an off-the-band low-cost radar. Given a sequence of radar measurements, the problem aims to predict the future blockages. For this problem,

again, two solutions adopting the classical signal processing and deep learning approaches are developed. Further, a real-world dataset is built for the evaluation of these systems, and the performances of the radar-aided blockage prediction solutions are presented. These results highlight the further potential of the radar for the communication systems in real-world.

- In Chapter 5, the joint sensing and communication is introduced to cell-free massive MIMO systems, integrating communications and sensing for distributed MIMO. In particular, a multi-static sensing metric is derived for the sensing objective, and joint optimization problems for beam optimization and power allocation optimization are proposed. The solutions for these problems are derived using semi-definite programming techniques. These proposed solutions are then compared with the classical approaches, namely, sensing-prioritized communication and communication-prioritized sensing beamforming. Our results indicated that the beamforming optimization can provide the communication data rates of communication-prioritized sensing, while achieving a better sensing rate than the sensing-prioritized communication. Further, the direct beam optimization solution provides significant gain over the power allocation with regularized zero-forcing beams, highlighting the advantages of joint sensing and communication beamforming design in cell-free massive MIMO systems.

1.5 Notation

The notation used throughout the thesis is given as follows. $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, $\mathcal{A}$ is a set of scalars, and $\boldsymbol{\mathcal{A}}$ is a set of vectors or sets. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ are transpose, Hermitian (conjugate transpose), conjugate, inverse, and pseudo-inverse of $\mathbf{A}$, respectively. $\|\mathbf{a}\|$ is the l_2-norm of $\mathbf{a}$ and $\|\mathbf{A}\|_{\mathrm{F}}$ is the Frobenius norms of $\mathbf{A}$. $\mathbf{I}$ and $\mathbf{1}$

are the identity matrix and the vector of ones. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathcal{U}[a, b]$ is a uniform random variable in $[a, b]$. $\mathbb{E}[\cdot]$ and $\otimes$ denote expectation and Kronecker product.

Chapter 2

CELL-FREE MASSIVE MIMO WITH WIRELESS FRONTHAUL

2.1 Overview

Cell-free massive MIMO systems have promising data rates and coverage gains. These systems, however, typically rely on fiber-based fronthaul for the communication between the central processing unit and the distributed access points (APs), which increases the infrastructure cost and installation complexity. To address these challenges, this section proposes two architectures for cell-free massive MIMO systems based on wireless fronthaul that is operating at a higher band compared to the access links. These dual-band architectures ensure a high data rate fronthaul while reducing the infrastructure cost and enhancing the deployment flexibility and adaptability. To investigate the achievable data rates with the proposed architectures, we formulate the end-to-end data rate optimization problem accounting for the various practical aspects of the fronthaul and access links. Then, we develop a low-complexity yet efficient joint beamforming and resource allocation solution for the proposed architectures based on user-centric AP grouping. With this solution, we show that the proposed architectures can achieve comparable data rates to those obtained with optical fiber-based fronthaul under realistic assumptions on the fronthaul bandwidth, hardware constraints, and deployment scenarios. This highlights a promising path for realizing the cell-free massive MIMO gains in practice while reducing the infrastructure and deployment overhead.

2.2 Introduction

Can we realize the cell-free massive MIMO gains with wireless fronthaul? This section attempts to answer this important question. Cell-free massive MIMO systems are promising uniform coverage and high data rate gains, even for scenarios with very dense users [8, 9]. A critical challenge, however, with the current cell-free massive MIMO systems is its reliance on optical fiber-based fronthaul for distributing data and synchronization signals from the central processing unit (CPU) to the access points (APs). This highly increases its infrastructure cost and installation time and limits the flexibility and scalability of the cell-free massive MIMO deployment. To address these challenges, the chapter proposes an alternative cell-free massive MIMO architecture with higher-band fronthaul, e.g., a millimeter wave (mmWave) or terahertz (THz) fronthaul for a sub-6GHz cell-free massive MIMO system. The use of higher frequency band fronthaul has two key advantages: (i) The large bandwidth available at the higher frequency bands provides a high data rate fronthaul, and (ii) the higher-band fronthaul signals (lower wavelength) have a high capability to synchronize the sub-6GHz AP transceivers. For this proposed architecture, the chapter investigates whether it can achieve comparable achievable rates to those obtained by classical fiber-fronthaul based cell-free massive MIMO systems.

2.2.1 Prior Work

Motivated by its potential to increase the data rate and manage multi-user interference, the idea of having distributed antennas simultaneously serving the same users has been previously investigated in the network multiple-input multiple-output (MIMO) [10], distributed MIMO [11], and coordinated multipoint with joint transmission (CoMP-JT) [12, 13]. These earlier approaches did not scale well with the increasing number of users and a large number of access points (APs) due to the channel estimation and feedback over-

head. To overcome this limitation, in [8, 14–17], the authors investigated cell-free massive MIMO, where reciprocity of the time-division duplexing (TDD) is leveraged to estimate the downlink channels directly from the joint uplink pilot transmissions. In the proposed cell-free optimization framework, the imperfection of channel state information (CSI) was taken into account, and only the long-term channel coefficients were utilized for the power allocation. Despite the interesting data rate and coverage gains, the realization of cell-free massive MIMO networks suffers from the limitations and high infrastructure cost of the CPU-AP fiber links. To reduce cost, cheaper wire-link alternatives with limited capacity were considered. For example, [18] investigated cell-free performance with quantized fronthaul transmissions. However, the approaches in [2, 3, 18, 19] did not eliminate the wired connection requirement, which is still associated with high infrastructure cost, high installation time, and limited deployment flexibility. To clarify, the joint consideration of the wireless fronthaul and cell-free massive MIMO networks has not been studied in the literature.

Another relevant line of work is presented in the different network architectures that adopted wireless fronthaul/backhaul such as small-cells [20, 21], ultra-dense networks [22], heterogeneous networks [23], and the cloud radio access networks (C-RAN) [24–26]. In [20–23], wireless fronthaul was adopted to show the potential advantages of the mmWave backhaul. This was, however, limited to scenarios with no cooperation between basestations. In C-RAN architectures, multiple (normally a few) base stations coordinate to serve the users. For example, in [24], the time-frequency resources of an OFDMA-based C-RAN system consisting of a few base stations were optimized to maximize the weighted sum rate of a large number of users. The solution in [24] assumed the fronthaul transmission to a single base station at a time. To overcome this limitation, the work in [25] proposed a multicast beamforming for the downlink ultra-dense C-RAN and designed user-centric clusters of the base stations for the transmissions. The work in [24, 25], however, was limited to

architectures with a few base stations. This is partially due to the large channel estimation/feedback overhead associated with the solutions in [24, 25], which limit them from scaling to a large number of antennas. This motivates the research for new approaches to enable the potential gains of cell-free massive MIMO systems with the adoption of the wireless fronthaul.

2.2.2 Contributions

In this section, we propose a new architecture for cell-free massive MIMO systems where the fronthaul is implemented using high-frequency (e.g., mmWave) wireless links to serve access links at lower frequencies (e.g., at a sub-6GHz band). The proposed architecture has two key motivations: (i) The larger available bandwidth at the high-frequency band ensures high data rates for the fronthaul and (ii) the small wavelength of the high-frequency signals provides high synchronization accuracy for the access points that are operating at a lower frequency band. Further, the wireless fronthaul enables a modular, flexible, and scalable architecture with low infrastructure cost and low installation time. With all these potential gains, an important question is whether this architecture is capable of achieving comparable data rate gains to those achieved with optical fiber-based cell-free massive MIMO systems. To answer this question, we develop an efficient communication scheme, analyze its performance, and draw important insights about the proposed architecture and data rate optimization approaches. The main contributions of the chapter can be summarized as follows:

- **Proposing an efficient architecture for cell-free massive MIMO systems based on higher-frequency wireless fronthaul**. The proposed architecture has the potential of ensuring higher data rate fronthaul and high synchronization accuracy for the distributed APs. Further, it requires low infrastructure cost and low installation time and provides interesting flexibility and adaptability gains for cell-free massive

MIMO systems.

- **Developing an efficient communication model for the wireless fronthaul-based cell-free massive MIMO system**. This model assigns a group of APs to each user and optimizes the multicast beamforming at the central processing unit to simultaneously serve the AP group of each user. The adopted system model accounts for the practical constraints on the higher-frequency band (mmWave) beamforming architectures [27]. Specifically, the CPU applies analog beamforming. Given the spatial multiplexing constraints of the analog beamforming, we use time-division multiple-access (TDMA) to multicast the message of each user-centric group.

- **Formulating the end-to-end data rate maximization problem for the proposed architecture and communication model**. The optimization problem takes the wireless fronthaul, access channels, and AP grouping into account. In particular, it aims to determine the user-centric AP group selection, fronthaul beamforming vectors, TDMA schedule, and AP power coefficients to maximize the end-to-end data rates.

- **Developing near-optimal end-to-end data rate maximization solution for the proposed architecture**. The solution adopts an iterative group selection algorithm, which is coupled with the fronthaul and access channel data rate maximization subproblems. Specifically, in each iteration, the group size and AP selection are determined based on the channel estimates, and then the fronthaul/access rates are optimized for the given groups.

- **Extending the architecture and developed solutions to the mixed wireless/wired fronthaul case.** A mixed-fronthaul architecture with wire-connected AP clusters (for example, through a radio stripe [9]) is proposed. In this architecture, only the leader AP in each wire-connected cluster has a wireless fronthaul link with the CPU. The

proposed data rate maximization solutions are generalized to account for the mixed-fronthaul.

The proposed solutions are extensively evaluated using numerical simulations, which draw important insights into the performance of the proposed cell-free architectures. Based on the results, the high-frequency wireless fronthaul can provide sufficient data rates for the cell-free massive MIMO by taking advantage of the larger bandwidth availability. With the fully wireless architecture, it is possible to achieve data rates with only $10 - 20\%$ degradation compared to the fiber fronthaul-based cell-free massive MIMO architectures. Further, the mixed-fronthaul architecture significantly reduces the bandwidth requirements and improves the data rates. In our simulations, the mixed-fronthaul architecture enabled data rates very similar to the fiber-fronthaul based solutions with reasonable fronthaul bandwidth requirements.

2.2.3 Key Idea

Cell-free massive MIMO is a promising enabler for the high data rate and coverage requirements in future wireless communications systems. Current cell-free massive MIMO architectures, however, assume that all the antennas are connected to the central processing unit with wired fronthaul, which is typically optical fiber [28]. This complicates the deployment process of these systems and increases their cost and installation time. To overcome these challenges, we propose an alternative architecture for cell-free massive MIMO systems that relies on a wireless fronthaul operating at a much higher frequency band compared to the access band, as depicted in Figure 2.1. For example, a mmWave fronthaul channel with sub-6GHz access channels or a terahertz fronthaul supporting mmWave access channel. Thanks to the wireless fronthaul, **the proposed architecture is modular, flexible, scalable, with low cost and installation time**. In particular, replacing the wired-fronthaul with a wireless connection provides modularity and flexibility in the deployment.

This way, the costs and installation time can be reduced as the APs can be directly deployed, potentially without any (or with minimal) engineering assistance. Consequently, this allows easier deployment of the APs and potential deployment scalability of the system [1] . While the wireless fronthaul provides these advantages, it may have challenges for the data rates and synchronization, which motivates the use of a higher-frequency fronthaul. The use of a higher frequency band in the fronthaul compared to the access band has the following advantages (In the rest of the chapter, for ease of exposition, we will assume a mmWave fronthaul and a sub-6GHz access channel):

- **High Data-Rate Fronthaul:** The availability of the large bandwidth at the mmWave frequency band enables the central processing unit (equipped with a mmWave transceiver) to support high data-rate fronthaul. This way, the crucial large bandwidth availability can be utilized for maintaining the high data rate gains of cell-free massive MIMO systems, as will be shown in Section 2.7.

- **High Synchronization Accuracy:** Using a mmWave fronthaul (where signals have a small wavelength) to synchronize the sub-6GHz APs (where signals have a much higher wavelength) has the potential of ensuring precise clock synchronization among the APs, which is essential for the operation of the cell-free massive MIMO systems. Specifically, the synchronization of the APs can be achieved over the mmWave fronthaul with the aid of the CPU via master-slave type algorithms (e.g., [30]) or via network-centric solutions (e.g., [31]). The APs use two clock signals, one for the mmWave fronthaul and the other for the sub-6GHz access channel. The APs can synchronize their mmWave clocks and time by referring to the CPU. These signals can then be utilized for more accurate sub-6GHz synchronization by converting the

[1]Note that a scalability notion for cell-free massive MIMO was formally defined in[29], which aims to keep the complexity of the channel estimation, signal processing for beamforming, fronthaul signaling, and power control optimization limited for asymptotically increasing number of users. It is important to note that this is different than our use of scalability, that is for the deployment of the APs.

mmWave (smaller wavelength) signal to a large wavelength clock [2] .

Further, with regards to the mmWave fronthaul, we would like to highlight two points: (i) The cost of mmWave transceivers may currently be high. However, it is expected to decrease over time with mass adoption and production. Further, in Section Section 2.6, we propose an extension of the architecture that requires a much smaller number of mmWave transceivers. (ii) The potential coverage limitations of the mmWave fronthaul can be reduced with the employment of a large antenna array at the CPU, as will be considered in the following sections. In general, we envision that the proposed architecture will be promising for the scenarios where the CPU serves APs distributed within a few hundred meters.

Now, with the interesting gains of having wireless fronthaul-based cell-free massive MIMO systems, and while mmWave fronthaul may have relatively high data rates compared to other wireless solutions, it is important to answer the question: **Can this wireless fronthaul based architecture achieve the same data rates of the fiber fronthaul based cell-free massive MIMO architectures?** In this chapter, this question is targeted. Towards this objective, we first describe the system model of the proposed mmWave fronthaul-based cell-free massive MIMO architecture in Section 2.3 and formulate the end-to-end (from the CPU to the users) achievable rate optimization problem in Section 2.4. Then, we develop an efficient transmission strategy for the proposed architecture in Section 2.5 and discuss some important extensions in Section 2.6. Finally, the achievable rate of the proposed architecture and comparisons with the fiber-based solutions are provided in Section 2.7.

[2]For instance, if two mmWave clocks are synchronized with a Δf frequency offset between them, the sub-6GHz clocks obtained from the these (or from the same source) can translate into a smaller frequency difference, possibly at the ratio of the carrier frequencies (e.g., for 30 GHz and 3 GHz carrier frequencies, it could be possible to obtain the clocks with frequency difference $\frac{\Delta f}{10}$). A similar gain is expected if a THz-based fronthaul is leveraged to synchronize a mmWave access channel.

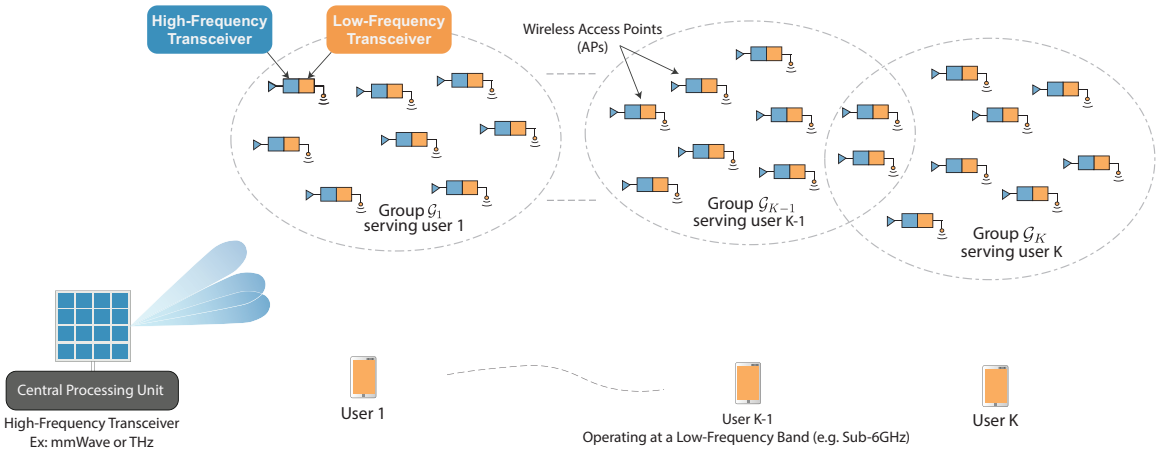


Figure 2.1: An Illustration of the Proposed Architecture Where a mmWave CPU Provides Fronthaul to the Different Sets of User-Centric AP Groups via Different Beamforming Vectors. The Wireless APs Jointly Serve Their Users Through a Cell-Free Massive MIMO Access Channel at Sub-6GHz.

2.3 System Model

We consider the distributed massive MIMO system in Figure 2.1, where a CPU communicates with M wireless-APs over a high-frequency (e.g., mmWave) wireless fronthaul and the wireless-APs serve K user equipment (UEs) over a low-frequency (e.g., sub-6GHz) channel. We will use $\mathcal{M} = \{1, \ldots, M\}$ and $\mathcal{K} = \{1, \ldots, K\}$ to denote the sets of M APs and K UEs. For the rest of the chapter, we will refer to the downlink channel from the CPU to the APs as the *fronthaul channel* and the downlink channel from APs to the UEs as the *access channel*. Further, for ease of exposition and without loss of generality, we will assume that the fronthaul channel is operating over a mmWave band, with a bandwidth B^{fh}, and the access channel is adopting a sub-6GHz band, with a bandwidth B^{ac}. The algorithms and results of the chapter, however, can be applied to other dual-band architectures, such as terahertz fronthaul with a mmWave access channel.

To beamform the signal to the APs, the CPU employs an antenna array of N elements, while the APs and UEs are, for simplicity, assumed to have single antennas. We consider a user-centric grouping approach where the message of UE k is jointly transmitted by a

subset of APs, $\mathcal{G}_k \subseteq \mathcal{M}$. We denote the set of the user-centric groups as $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_1, \ldots, \mathcal{G}_K\}$. Note that different groups may include the same APs since multiple APs are utilized in the transmission to every UE. For that, we will also define $\mathcal{U}_m \subseteq \mathcal{M}$ for $m \in \mathcal{M}$ as the set of users that are being served by the AP m. For the distances between the CPU and m-th AP and between the m-th AP and k-th user, we will use d_m^{fh} and d_{mk}^{ac} to denote them. It is important to mention here that we do not assume any knowledge about the positions or distances between the CPU, APs, or the users.

At the Fronthaul Link: The mmWave CPU adopts TDMA to serve the K user-centric groups. In each TDMA slot, the CPU beamforms the signal toward the APs that serve one user. The duration of the TDMA slot allocated for serving APs of UE k is denoted by t_k. Let $\mathbf{h}_m \in \mathbb{C}^N$ denote the channel between the CPU and the mth AP. If the message intended for user k is represented by $q_k^{\text{fh}} \in \mathbb{C}$, with $\mathbb{E}[\left|q_k^{\text{fh}}\right|^2] = 1$, then the received signal at the AP m can be written as

$$y_m^{\text{fh}} = \sqrt{\rho^{\text{fh}}} \mathbf{h}_m^H \mathbf{f}_k q_k^{\text{fh}} + w_m^{\text{fh}}, \tag{2.1}$$

where ρ^{fh} is the normalized fronthaul transmission power and $w_m^{\text{fh}} \sim \mathcal{CN}(0, 1)$ is the receive noise at the mth AP. The vector $\mathbf{f}_k \in \mathbb{C}^N$ is the CPU beamforming vector intended to focus the signal to the APs that serve user k. To satisfy the practical mmWave hardware constraints, we assume that the CPU adopts analog-only beamforming implemented by a network of quantized phase shifters [27]. This means that the beamforming vector $\mathbf{f}_k$ can only be selected from a certain set of vectors, that we define by the codebook $\boldsymbol{\mathcal{F}}$. If each phase shifter has q bits, i.e., 2^q possible phase shift values defined by the set $\mathcal{Q} \in \{0, \frac{\pi}{2^q}, \frac{2\pi}{2^q}, ..., \frac{(2^q-1)\pi}{2^q}\}$, then we can write

$$\boldsymbol{\mathcal{F}} = \left\{ \frac{1}{\sqrt{N}} [e^{j\phi_1}, \ldots, e^{j\phi_N}]^T : \phi_n \in \mathcal{Q}, \forall n \in \{1, \ldots, N\} \right\}, \tag{2.2}$$

and $\mathbf{f}_k \in \boldsymbol{\mathcal{F}}$. We assume that the channels between the CPU and APs are available at the

CPU. This is motivated by the stationarity of the CPU and APs and the channel reciprocity assumption.

At the Access Link: Each AP decodes the CPU signal received by its mmWave receiver and prepares it for transmission over the sub-6GHz access channel. Similar to [8], the APs transmit the weighted sum of the users' messages, where each message is multiplied by a different beamforming and power control coefficient. Since each AP m is assumed to contribute in serving a set of users $\mathcal{U}_m$, the transmitted signal from the mth AP, x_m, can be written as

$$x_m = \sqrt{\rho^{\text{ac}}} \sum_{k \in \mathcal{U}_m} \sqrt{p_{mk}} f_{mk}^{\text{ac}} q_k^{\text{ac}}, \tag{2.3}$$

where ρ^{ac} is the shared APs coefficient for the transmit power coefficient of the APs, f_{mk}^{ac} and p_{mk} denote the beamforming and power control coefficients of the m-th AP for the k-th UE, and q_k^{ac} represents the intended message for the k-th UE which satisfies $\mathbb{E}\left[|q_k^{\text{ac}}|^2\right] = 1$. After all the APs prepare their messages, they transmit them to their users. Note that all the APs are assumed to maintain sufficient clock synchronization at the sub-6GHz access channel, which is further facilitated by the adoption of mmWave-based synchronization, as briefly highlighted in Section 2.2.3. Now, if g_{mk} denotes the access channel coefficient between the m-th AP and k-th UE, the received signal at the UE k can be expressed as

$$y_k^{\text{ac}} = \sum_{m=1}^{M} x_m g_{mk} + w_k, \tag{2.4}$$

where $w_k \sim \mathcal{CN}(0, 1)$ is the receive noise at user k.

Channel Models: For the fronthaul channel, we adopt a geometric channel model [17, 32]. In this model, there are V propagation paths between the CPU and each AP. We denote the large- and small-scale fading and azimuth angle of departure of the path v by $\beta_{m,v}^{\text{fh}}$, $\alpha_{m,v}^{\text{fh}}$ and $\theta_{m,v}$, respectively. With this notation, we can write the channel as

$$\mathbf{h}_m = \sum_{v=1}^{V} \sqrt{\beta_{m,v}^{\text{fh}}} \alpha_{m,v}^{\text{fh}} \mathbf{a}(\theta_{m,v}), \tag{2.5}$$

where $\mathbf{a}(\cdot) \in \mathbb{C}^N$ is the array response vector function. For the access channel, we define the channel coefficient between AP m and UE k as $g_{mk} = \sqrt{\beta_{mk}^{\text{ac}}}\alpha_{mk}^{\text{ac}}$, where β_{mk}^{ac} and α_{mk}^{ac} represent the large- and small-scale fading coefficients.

2.4 Problem Formulation

In this section, we formulate the optimization problem of the end-to-end achievable rate of the proposed wireless fronthaul-based cell-free massive MIMO system. We first present the achievable rates of the access and fronthaul channels. Then, we present a transmission schedule and formulate the end-to-end rate optimization problem with this schedule. In the following, we omit the superscripts denoting fronthaul and access channel variables as they are easily distinguished from the context.

2.4.1 Achievable Rate of the Fronthaul

Considering the system model in Section 2.3, the achievable rate of AP m, that is part of the kth user group $\mathcal{G}_k$, can be written as

$$R_m^{\text{fh}}(\mathbf{f}_k) = \log\left(1 + \rho^{\text{fh}}\left|\mathbf{h}_m^H \mathbf{f}_k\right|^2\right), \tag{2.6}$$

where $\mathbf{f}_k$ is the beamforming vector used by the CPU to serve the set of APs in the group $\mathcal{G}_k$. Now, note that the message of the kth user is simultaneously transmitted to all the APs in $\mathcal{G}_k$, and that all the APs in the group should finish receiving the message before they start transmitting it to the user. To incorporate that, we define the effective rate of the group k as the minimum rate of the APs in the $\mathcal{G}_k$'s group. Mathematically, we write

$$R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k) = \min_{m \in \mathcal{G}_k} \left\{R_m^{\text{fh}}(\mathbf{f}_k)\right\}. \tag{2.7}$$

Moreover, due to the TDMA schedule of the AP groups, the group rates will be further scaled by the TDMA time fractions. Let t_k denote the fraction of TDMA time allocated to

group $\mathcal{G}_k$. Then, the time-scaled fronthaul rate of group k can be written as $t_k R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k)$. We note that these time fractions satisfy the constraints $0 < t_k < 1$ and $\sum_{k=1}^{K} t_k = 1$.

2.4.2 Achievable Rate of the Access Channel

The access channel adopts the main assumptions of the cell-free massive MIMO architectures in [8]. Specifically, we assume the following: (i) The APs synchronously serve the UEs without any cell boundaries, i.e., each AP can serve any UE. (ii) Time-division duplexing (TDD) is adopted for the transmissions, which facilitates the estimation of the downlink access channel coefficients through the uplink pilots. (iii) Only the large-scale fading coefficients are available at the CPU for joint power allocation. Further, we assume that the APs adopt conjugate beamforming for the downlink transmission to the users. More specifically, the uplink pilots are used to estimate the uplink channels (which are also used to construct the downlink channels leveraging channel reciprocity). Then, the information about the large-scale fading coefficients is frequently transmitted to the CPU. The CPU uses this large-scale fading information to determine the access channel power coefficients. The APs adopt this power allocation while jointly serving their users. Next, we first describe the adopted channel estimation procedure in detail before formulating the achievable rates of the access link.

Channel Estimation: In the adopted protocol, the UEs transmit orthogonal pilot sequences of length L_p, $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K \in \mathbb{C}^{L_p}$, simultaneously to be received by all the APs. If ρ_t denotes the power level selected for the pilot transmissions, then the received signal at AP m can be written as

$$\mathbf{y}_m^t = \sqrt{\rho_t L_p} \sum_{k=1}^{K} g_{mk} \boldsymbol{\psi}_k + \mathbf{n}_m, \tag{2.8}$$

where $\mathbf{n}_m \sim \mathcal{CN}(0, \mathbf{I}_{L_p})$ is the receive noise. With this received signal, the MMSE estima-

tor for g_{mk} can be given by

$$\hat{g}_{mk} = \frac{\sqrt{\rho_t L_p}\beta_{mk}}{1 + \rho_t L_p \beta_{mk}} \boldsymbol{\psi}_k^H \mathbf{y}_m^t. \tag{2.9}$$

We define the channel estimation error $\tilde{g}_{mk} = g_{mk} - \hat{g}_{mk}$, and note that the estimation $\hat{g}_{mk}$ and the error $\tilde{g}_{mk}$ are uncorrelated thanks to the estimator. Therefore, the distributions of the estimated channel coefficients and the error can be written as $\hat{g}_{mk} \sim \mathcal{CN}(0, \hat{\beta}_{mk}), \quad \tilde{g}_{mk} \sim \mathcal{CN}(0, \beta_{mk} - \hat{\beta}_{mk})$ with the variance of the estimator is defined as

$$\hat{\beta}_{mk} = \frac{\rho_t L_p \beta_{mk}^2}{1 + \rho_t L_p \beta_{mk}}. \tag{2.10}$$

Achievable Rate: With the adopted conjugate beamforming, the coefficients f_{mn}^{ac} in the received signal equations (2.3)-(2.4) can be replaced by $\hat{g}_{mn}^*$. Further, with the described changes on the signal model, the capacity lower bound for the UEs given in [15] becomes valid, and we can express the achievable rate of user k as

$$R_k^{\text{ac}} = \log_2\Big(1 + \text{SINR}_k\Big), \tag{2.11}$$

$$\text{SINR}_k = \frac{\rho^{\text{ac}} \big(\sum_{m=1}^{M} \sqrt{p_{mk}} \hat{\beta}_{mk} \big)^2}{\rho^{\text{ac}} \sum_{m=1}^{M} \beta_{mk} \sum_{k'=1}^{K} p_{mk'} \hat{\beta}_{mk'} + 1}, \tag{2.12}$$

where we note here that the power coefficient p_{mk} for the AP m and UE k is set to 0 if the AP is not in the group of that UE, i.e. if $m \notin \mathcal{G}_k$. These power coefficients for each AP m also satisfy

$$\sum_{k=1}^{K} p_{mk} \hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M}, \tag{2.13}$$

which captures the total power constraint of AP m.

2.4.3 Transmission Schedule

To achieve the high data rates with the proposed transmission scheme, i.e., cell-free massive MIMO access channel and TDMA-based fronthaul transmission scheme, a fronthaul transmission schedule needs to be designed. On the one hand, the access channel

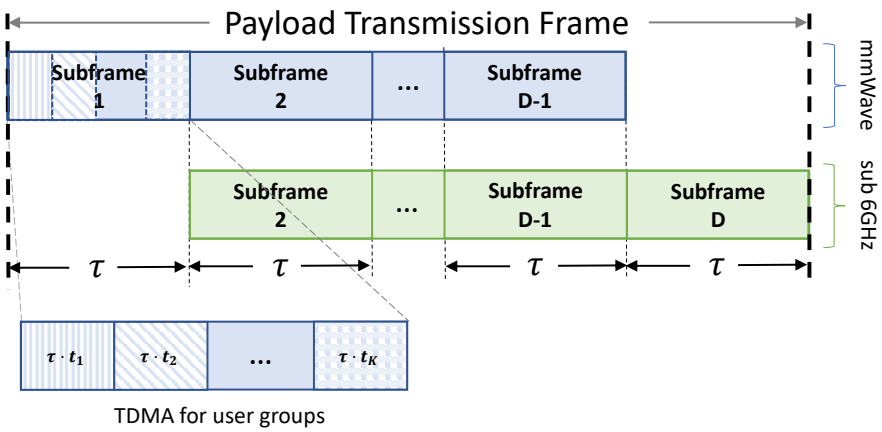


Figure 2.2: The Timing of the Fronthaul and Access Channel Transmissions With the Frame Structure.

requires the message of each UE to be available at all of the APs in its groups. On the other hand, the fronthaul transmissions are proposed in a TDMA manner, and a set of full transmissions needs to be completed before the access channel transmissions. To that end, we propose a frame/subframe structure that allows almost concurrent data rates. Specifically, we consider a frame of downlink data transmission period within the coherence time, that is split into D subframes. Let the duration of this subframe be τ. Then, in the first subframe of a frame, only the fronthaul transmission is carried out. Then, in the following subframes, the APs transmit the data received in the previous subframe, while receiving the data for the access channel transmissions in the next subframe. With a large number of subframes and smaller subframe durations, almost concurrent transmission data rates can be achieved. We illustrate the adopted transmission schedule in Figure 2.2. Further, with this model, the additional transmission delay of the system can be determined as τ, as the access channel transmissions follow the fronthaul transmissions from the previous subframe. Although this may introduce some delay, the small values can be achieved with a careful fronthaul communication design. Next, we formulate the end-to-end data rate problem with the proposed system.

2.4.4 End-to-End Achievable Rate Optimization

Based on the presented transmission schedule and achievable rates for the fronthaul and access channels, we now derive the end-to-end achievable rate of the system. With the adopted system model in Section 2.3 and the schedule given in Section 2.4.3, the access channel rate of UE k during the payload transmission frame can be written as $\frac{D-1}{D}B^{\mathrm{ac}}R_k^{\mathrm{ac}}$, and the fronthaul rate of UE k can be written as $\frac{D-1}{D}t_kB^{\mathrm{fh}}R_k^{\mathrm{fh}}$. As the AP operation is essentially a relaying operation between the fronthaul and access channels, the minimum of these rates provides the end-to-end achievable rate [3] . By dropping $\frac{D-1}{D}$ as $D \to \infty$, we express end-to-end rate of UE k by

$$R_k = \min\left\{B^{\mathrm{ac}}R_k^{\mathrm{ac}},\ t_kB^{\mathrm{fh}}R_k^{\mathrm{fh}}\left(\mathbf{f}_k\right)\right\}, \tag{2.14}$$

Note that the end-to-end channel rate accounts for the fronthaul and access bandwidths, which could be significant since the fronthaul bandwidth in our mmWave-based fronthaul is expected to be much larger than the bandwidth of the sub-6GHz access channel. Now, to optimize the end-to-end rate, we adopt the following formulation of the joint max-min fair rate optimization problem

$$\max_{\boldsymbol{\mathcal{G}},\{\mathbf{f}_k\},\{t_k\},\{p_{mk}\}} \min_k \ R_k \tag{2.15a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} p_{mk}\hat{\beta}_{mk} \le 1, \quad \forall m \in \mathcal{M} \tag{2.15b}$$

$$p_{mk} = 0, \quad \forall m \notin \mathcal{G}_k,\ \forall k \in \mathcal{K} \tag{2.15c}$$

$$0 < t_k < 1, \quad \forall k \in \mathcal{K} \tag{2.15d}$$

$$\sum_{k\in\mathcal{K}} t_k = 1 \tag{2.15e}$$

$$\mathbf{f}_k \in \boldsymbol{\mathcal{F}}, \quad \forall k \in \mathcal{K} \tag{2.15f}$$

[3] This objective is commonly adopted in the relay data rate optimization (e.g., [33]).

which aims to jointly optimize the AP groups, the fronthaul beamforming vectors, the fronthaul time allocation for different groups, and the access channel power coefficients. The problem is non-convex and challenging, especially due to the AP grouping and the fronthaul analog-only beamforming. It is worth mentioning here that the AP grouping and analog beamforming can be optimally designed via an exhaustive search over all the possible groups and candidate beam codewords, but this will require prohibitive complexity. To reduce this complexity, we propose an iterative sub-optimal solution in the following section.

2.5 Proposed Solution

In this section, we develop a suboptimal yet efficient solution for the end-to-end data rate maximization problem of the proposed cell-free massive MIMO architecture. We will then show in Section 2.7 that this developed solution (for the wireless mmWave-fronthauled cell-free massive MIMO architecture) achieves close performance to the upper bound which is given by the fiber-fronthaul based cell-free massive MIMO architecture. To start, we note that the optimization problem in (2.15) can be written in the equivalent form

$$\begin{aligned} \max_{\boldsymbol{\mathcal{G}},\{\mathbf{f}_k\},\{p_{mk}\},\{t_k\}} \ & \min \ \left\{B^{\mathrm{ac}} R_k^{\mathrm{ac}},\, t_k B^{\mathrm{fh}} R_k^{\mathrm{fh}}(\mathbf{f}_k)\right\}_{k=1}^{K} \\ & \text{s.t.} \quad \text{(2.15b)} - \text{(2.15f)} \end{aligned} \tag{2.16}$$

where the objective is to maximize the minimum of all the fronthaul and access channel rates. Before attempting the solution, we highlight the following remark.

Remark 1 *For the given system model and the max-min fairness problem defined in* (2.16)*, the separate optimization of the access channel and fronthaul variables depend on the grouping variable $\boldsymbol{\mathcal{G}}$. Nevertheless, for a fixed grouping $\boldsymbol{\mathcal{G}}$, only the objective function (but not the constraints) retains the variables of both the fronthaul ($\mathbf{f}_k$ and t_k) and access channel (p_{mk}), i.e., each constraint affects either the fronthaul channel or the access channel.*

Therefore, for a given grouping, a two-step approach can be developed with the access channel and fronthaul optimization steps to obtain an optimal solution.

Motivated by Remark 1, instead of the joint optimization of the fronthaul and access rates, we first consider a given grouping structure, and design two separate problems that maximize the fronthaul and access rates: (i) The access channel power coefficients are optimized without any consideration of the fronthaul, and (ii) the beamforming vectors for each group and TDMA time fractions are determined to maximize the fronthaul rate in a fair manner. As this process is conditioned on the grouping structure, an optimization of this grouping is required to maximize the end-to-end data rates. For that, we propose an iterative heuristic algorithm for the AP grouping where each iteration involves solving two sub-problems for the fronthaul and access rates. The next three subsections present the details of this proposed approach.

2.5.1 Access Channel Rate Optimization

As briefly described, for a given AP grouping selection, we first optimize the power coefficients for the access channel, without any fronthaul limitations. This approach allows us to allocate the power of the APs over the access channel, in a similar way to the standard approaches in the cell-free massive MIMO literature. To formulate the access channel optimization problem, we only keep the access channel-related terms and constraints of the original problem defined in (2.16) and write

$$\max_{\{p_{mk}\}} \min_{k} \; B^{\mathrm{ac}} R_k^{\mathrm{ac}} \tag{2.17a}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} p_{mk} \hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M} \tag{2.17b}$$

$$p_{mk} = 0, \quad \forall m \notin \mathcal{G}_k, \, \forall k \in \mathcal{K}, \tag{2.17c}$$

where two further simplifications can be applied to the objective. First, since B^{ac} is a positive constant multiplied by a function of the variables, it can be removed. Second, the objective can be rewritten in terms of the SINR values, instead of the rates of the form $\log(1 + \text{SINR})$. This re-formulation does not change the optimal power coefficients, since $\log(1 + \text{SINR})$ is a non-decreasing function of the SINR values. Thus, we simplify (2.17) and write the grouping sensitive max-min SINR optimization for the access channel as

$$\max_{\{p_{mk}\}} \min_{k} \ \text{SINR}_k \tag{2.18a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_{mk}\hat{\beta}_{mk} \leq 1, \quad \forall m \in \mathcal{M} \tag{2.18b}$$

$$p_{mk} = 0, \quad \forall m \notin \mathcal{G}_k, \ \forall k \in \mathcal{K}, \tag{2.18c}$$

which is in the same form as the power allocation problem of the standard cell-free massive MIMO [8] with the addition of the grouping constraint in (2.18c). When all the APs transmit to all UEs, i.e., $\mathcal{G}_k = \mathcal{M}\ \forall k \in \mathcal{K}$, (2.18c) does not provide any constraints, and the presented optimization of the access channel becomes directly equivalent to the power allocation of cell-free massive MIMO. For the solution of (2.18), we first elaborate on the additional grouping constraint (2.18c), which allows an AP to transmit only to the UEs whose groups include that AP, by restricting the power allocated for the other UEs to 0. It is a linear equality constraint that does not affect the convexity of the problem. Hence, for the problem (2.18), the same solution to the optimal power allocation of the standard cell-free massive MIMO given in [8] can be applied. Further, different sub-optimal solutions with lower complexity and feedback (e.g., [9, 15, 29]) can be applied instead of the optimal solution of the problem.

2.5.2 Fronthaul Channel Rate Optimization

In this subsection, we define an optimization problem to maximize the fronthaul rates of the distributed AP groups in a fair manner. First, by starting with the original formulation given in (2.16), we write a max-min fair fronthaul optimization problem by eliminating the access channel-related constraints and objectives, to get

$$\max_{\{t_k\}\{\mathbf{f}_k\}} \min_k \; t_k R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k) \tag{2.19a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} t_k = 1 \tag{2.19b}$$

$$t_k \geq 0, \quad \forall k \in \mathcal{K} \tag{2.19c}$$

$$\mathbf{f}_k \in \boldsymbol{\mathcal{F}}, \quad \forall k \in \mathcal{K}. \tag{2.19d}$$

For this problem in (2.19), and given the group selections $\{\mathcal{G}_k\}$, we note that the optimal solution can be obtained by first optimizing the rates of all the groups, $R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k)$, $\forall k$, and then optimization the TDMA time allocation. This is because any set of TDMA time fractions, $\{t_k\}$, does not affect the optimization of the beamforming vectors. Therefore, we will first optimize the beamforming vectors to optimize the group rates individually. Then, we can determine the TDMA time fractions for each group to optimize the time-scaled group rates, as described in the remaining part of this subsection.

Beamforming Optimization The beamforming vector of each group needs to be optimized to maximize the group rate. Since this rate is determined by the minimum rate of the APs in the group, we can formulate the beamforming optimization problem of any group $\mathcal{G}_k$ as follows

$$\begin{aligned} \mathbf{f}_k^\star = \arg\max_{\mathbf{f}_k} \quad & \left(R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k) = \min_{m\in\mathcal{G}_k} \left\{ R_m^{\text{fh}}(\mathbf{f}_k) \right\} \right) \\ \text{s.t.} \quad & \mathbf{f}_k \in \boldsymbol{\mathcal{F}}, \end{aligned} \tag{2.20}$$

which coincides with the well-studied multicast beamforming problem with analog phase shifters. The optimal solution to the problem can be obtained by exhaustive search, which

has the complexity increases exponentially with the number of antennas, i.e., $\mathcal{O}(|\mathcal{Q}|^N) = \mathcal{O}(2^{qN})$. To reduce this complexity, several methods have been developed in the literature [34, 35]. In the simulation results in Section 2.7, we adopt the suboptimal iterative solution proposed in [34] due to its low complexity and good performance. With the optimized beamforming vectors, the optimal time fractions of the TDMA can be determined. Next, we will present two approaches to solve this TDMA time fraction optimization problem.

TDMA Optimization. Approach 1: Given the solution of the problem (2.20), $\{\mathbf{f}_k^*\}$, the fronthaul rate of each group k becomes a constant denoted by $R^{\text{fh}}(\mathcal{G}_k) = R^{\text{fh}}(\mathcal{G}_k, \mathbf{f}_k^*)$. With these rate constants, we can simplify the fronthaul rate maximization problem given in (2.19) to be

$$\max_{\{t_k\}} \min_k t_k R^{\text{fh}}(\mathcal{G}_k) \text{ s.t. } \sum_{k\in\mathcal{K}} t_k = 1, \;\; t_k \geq 0 \; \forall k \in \mathcal{K}. \tag{2.21}$$

As the problem is linear, the optimal solution is obtained at the equality of the time-scaled fronthaul rates, i.e., $t_k R_k^{\text{fh}} = t_k' R_{k'}^{\text{fh}} \; \forall k, k' \in \mathcal{K}$. Therefore, the optimal sum of the time-scaled group rates can be given by

$$R^{\text{fh}}(\boldsymbol{\mathcal{G}}) = \text{HM}\{R^{\text{fh}}(\mathcal{G}_k)\}. \tag{2.22}$$

where the derivation is given as follows. We first introduce a slack variable [36], η, to represent the minimum time-scaled data rate of the groups and write the equivalent formulation

$$\begin{aligned} \max_{\{t_k\},\eta} \quad \eta \quad \text{s.t.} \quad & t_k R^{\text{fh}}(\mathcal{G}_k) \geq \eta \; \forall k \in \mathcal{K}, \\ & \sum_{k\in\mathcal{K}} t_k = 1, \;\; t_k \geq 0 \;\; \forall k \in \mathcal{K}. \end{aligned} \tag{2.23}$$

This formulation is a linear program and the solution can be attained by finding maximum η. Let us assume that there exists an optimal η, $\eta^\star$, that satisfies the minimum rate constraint of (2.23), i.e., $t_k R^{\text{fh}}(\mathcal{G}_k) \geq \eta$. Due to the maximization objective, this constraint is satisfied

at the equality, and it is possible to obtain the exact solution thanks to the linear structure of the problem. Hence, for the equality, we can write $\eta^\star = t_k^\star R^{\text{fh}}(\mathcal{G}_k)$, where $t_k^\star$ denotes the optimal time fraction for group k. Then, we have $t_k^\star = \frac{\eta^\star}{R^{\text{fh}}(\mathcal{G}_k)}$. By replacing t_k with $t_k^\star$ in the time constraint of (2.23), $\sum_{k\in\mathcal{K}} t_k = 1$, we obtain

$$\eta^\star = \frac{1}{\sum_{k\in\mathcal{K}} \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}} = \frac{\text{HM}\{R^{\text{fh}}(\boldsymbol{\mathcal{G}})\}}{K} \tag{2.24}$$

which leads to the solution given in (2.22).

Finally, with the obtained access channel and fronthaul rates, we can take the minimum of them to provide an efficient solution to (2.16) for a given grouping structure.

The provided solution so far only aims to maximize the minimum of the rates, which is also the objective of (2.16). However, the optimal solution is not necessarily unique, which motivates the search within this set of solutions for one that could further optimize other objectives. In particular, without decreasing the minimum of the end-to-end rates of the given groups, we may still be able to increase the data rates of the other groups by further optimizing the fronthaul TDMA time allocations of these groups. To detail, the solution given in (2.22) attempts to maximize the minimum of the fronthaul group rates without accounting for the optimized access channel rates. The user rates, however, depend on both the access and fronthaul rates. Therefore, we may be able to further increase the data rates of the users by taking into account the access rates as a restricting constraint. Based on that, we define the following problem, which attempts to optimize the fronthaul rate of each group to meet the access rate of the group in a fair manner.

TDMA Optimization. Approach 2: For a fair allocation, the time-scaled rate of the groups should be equal unless any of the constraints are met. Therefore, the time allocated to each group should be inversely proportional to their rates. To achieve this, we utilize the weighted logarithm function for the objective, i.e., $\sum_k w_k \log(t_k)$, which allows the resources to be allocated fairly, proportional to the weights, w_k. We select the weights as

$w_k = \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}$, to allocate the time fractions inversely proportional to the fronthaul rates of the groups, resulting in equal fronthaul rates. In addition, we upper-bound the fronthaul rate of a user by the access channel rate of that user. With this upper bound, the total time fractions do not necessarily meet the summation equality, $\sum_{k\in\mathcal{K}} t_k = 1$. Hence, we relax the condition by $\sum_{k\in\mathcal{K}} t_k \leq 1$, and write the problem for the time allocation of the TDMA, that maximizes the end-to-end rate in a fair manner, as follows

$$\max_{\{t_k\}} \quad \sum_{k\in\mathcal{K}} \frac{1}{R^{\text{fh}}(\mathcal{G}_k)} \log t_k \tag{2.25a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} t_k \leq 1 \tag{2.25b}$$

$$t_k \leq \frac{B^{\text{ac}} R_k^{\text{ac}}}{B^{\text{fh}} R^{\text{fh}}(\mathcal{G}_k)}, \quad \forall k \in \mathcal{K} \tag{2.25c}$$

$$t_k \geq 0, \quad \forall k \in \mathcal{K}. \tag{2.25d}$$

The problem given in (2.25) is in the form

$$\begin{aligned} \max_{\{t_k\}} \quad & \sum_{k\in\mathcal{K}} w_k \log t_k \\ \text{s.t.} \quad & \sum_{k\in\mathcal{K}} t_k \leq 1,\ t_k \leq \bar{t}_k\ \forall k \in \mathcal{K},\ \ t_k \geq 0\ \forall k \in \mathcal{K} \end{aligned} \tag{2.26}$$

for the constants defined by $w_k = \frac{1}{R^{\text{fh}}(\mathcal{G}_k)}$ and $\bar{t}_k = \frac{B^{\text{ac}} R_k^{\text{ac}}}{B^{\text{fh}} R^{\text{fh}}(\mathcal{G}_k)}$. The problem is concave due to the sum of logarithms in the objective, and it can be solved by the Karush-Kuhn-Tucker (KKT) conditions [36]. For the solution, we first write the Lagrangian function [4] as

$$\mathcal{L}(\mathbf{t}, \boldsymbol{\lambda}, \nu) = \sum_{k\in\mathcal{K}} w_k \log t_k - \nu(\sum_{k\in\mathcal{K}} t_k - 1) - \sum_{k\in\mathcal{K}} \lambda_k (t_k - \bar{t}_k), \tag{2.27}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]$ and ν are the Lagrange multipliers. We note that the partial derivative of the Lagrangian function is $\frac{\partial \mathcal{L}(\mathbf{t},\boldsymbol{\lambda},\nu)}{\partial t_k} = \frac{w_k}{t_k} - \nu - \lambda_k$, and write the KKT

[4] We ignore the constraint $t_k \geq 0$ since the logarithm is only defined in this region, and any solution obtained from the formulation will satisfy the condition.

conditions as

$$\frac{w_k}{t_k^\star} - \nu - \lambda_k = 0, \qquad \text{(Stationarity)}$$

$$\sum_{k\in\mathcal{K}} t_k^\star \le 1, \quad t_k^\star \le \bar{t}_k, \quad \lambda_k, \nu \ge 0, \qquad \text{(Feasibility)}$$

$$\lambda_k(t_k^\star - \bar{t}_k) = 0, \quad \nu(\sum_{k\in\mathcal{K}} t_k^\star - 1) = 0. \qquad \text{(C. Slackness)}$$

From the stationarity condition, we have $\nu = \frac{w_k}{t_k} - \lambda_k$. Combining this with the complementary slackness condition of λ_k, we obtain

$$t_k^\star = \begin{cases} \frac{w_k}{\nu} & \text{if } \lambda = 0, \\ \bar{t}_k & \text{if } \lambda > 0. \end{cases} \tag{2.28}$$

which can be simplified to $t_k^\star = \min\{\frac{w_k}{\nu}, \bar{t}_k\}$. Finally, we define $\eta = \frac{1}{\nu}$ and place $t_k^\star$ in terms of η in the first (primal) feasibility condition, and obtain the solution given as

$$\sum_{k\in\mathcal{K}} t_k^*(\eta) \le 1, \text{ with } t_k^*(\eta) = \min\left\{\frac{B^{\mathrm{ac}} R_k^{\mathrm{ac}}}{B^{\mathrm{fh}} R^{\mathrm{fh}}(\mathcal{G}_k)}, \frac{\eta}{R^{\mathrm{fh}}(\mathcal{G}_k)}\right\}. \tag{2.29}$$

The resulting solution takes a similar form to the water-filling solution with upper bounds. Specifically, the solution can be obtained by finding the maximum individual fronthaul rate (water level), $\eta > 0$, that satisfies (2.29).

For the optimal η value, η^*, we can obtain the optimal time fractions $t_k^* = t_k^*(\eta^*)$. The given solution allocates the fronthaul rates equally among UEs until the satisfaction of individual access channel rates or the use of the total time. In the special case of all the fronthaul rates being smaller than the access channel rates, it allocates the rates equally, coinciding with (2.22).

2.5.3 End-to-End Optimization through Iterative Group Selection

With the proposed fixed-group end-to-end data rate maximization, the efficient selection of the groups becomes crucial in achieving high data rates. However, the group selection problem is a combinatorial problem that has high complexity given the large number of APs and UEs. In particular, there are 2^M possible selections of groups for each UE, leading to a total of 2^{MK} distinct selections for the K users. Moreover, each selection of the groups needs to be utilized with the proposed end-to-end data rate optimization, which results in prohibitive optimization complexity for practical systems. Hence, an efficient design for the group selection is required. To this end, we propose a low-complexity yet efficient solution that is motivated by understanding the end-to-end achievable rate optimization problem and the proposed cell-free massive MIMO architecture. Next, we present the proposed approach, which has two key points, namely the group selection and the group size optimization.

Group Selection: We adopt the following group selection approach: The group of each user is selected as the G APs with the maximum channel gains. Here, it is important to clarify two points: (i) The group size G is assumed to be fixed for the sake of simplicity and a low-complexity solution. Note that the impact of this constraint is expected to be marginal, given the high density of the APs. (ii) The selection of the APs with the maximum channel gains will likely lead to a set of APs that are close to each other, which will lead to efficient beamforming design via more focused beams. Next, we mathematically describe the proposed approach. Let $\{\beta_{1k}^{(o)}, \ldots, \beta_{Mk}^{(o)}\}$ define the ordered set of channel coefficients between user k and the M APs. This set adopts a descending order, i.e., the channel coefficients satisfy $\beta_{mk}^{(o)} \geq \beta_{m'k}^{(o)}$ for any $m \leq m'$. In addition, to formalize the mapping between the original channel coefficients sets $\{\beta_{1k}, ..., \beta_{Mk}\}$ and the ordered channel coefficients set $\{\beta_{1k}^{(o)}, \ldots, \beta_{Mk}^{(o)}\}$, we define the permutation $\varsigma_k(.)$ such that $\beta_{m'k}^{(o)} = \beta_{mk}$ if $\varsigma_k(m') = m$,

i.e., if the mth AP has the m'th highest channel gain. Using these definitions, for a given group size G, the group of UE k can be determined as $\mathcal{G}_k = \bigcup_{m'=1}^{G} \varsigma_k(m')$, which simply selects the G APs with the best channel gains. With this approach, the group selection is reduced to the selection of the parameter G, and it needs to be determined carefully.

Group Size Optimization: The number of APs per group, G, needs to be optimized to maximize the end-to-end rate. Here, it is interesting to highlight the following trade-off: A small group size may lead to more optimized CPU-APs beamforming design and hence high fronthaul rates. At the same time, it may also result in low APs-user beamforming gain and achievable access channel rates. Therefore, to select G, one possible approach is the trial of different group size values from a pre-determined interval (for a given AP structure). Another approach is to start from a certain group size G and then increase/decrease the group size depending on the relation between the access and fronthaul rates. More specifically, if the sum fronthaul rate is larger than the sum access channel rate, i.e., $R^{\text{fh}} \geq R^{\text{acc}}$, the group size is increased. Otherwise, if $R^{\text{fh}} < R^{\text{acc}}$, the group size is decreased. Through the iterations of these updates, the group value can be locally optimized [5] .

2.6 Mixed-Fronthaul Cell-Free Massive MIMO Architecture

In the previous sections, we proposed and designed a cell-free massive MIMO architecture with wireless higher-band fronthaul. While this architecture is flexible, scalable, and has the potential to reduce the installation cost and time of cell-free massive MIMO systems, it has a few drawbacks: (i) Adding a dual-band relay to each remote wireless AP adds an extra cost to the system and (ii) separately powering the distributed APs may require additional infrastructure cost. To address these points while maintaining the same promising

[5] Although the iterative algorithm is applied through the sum data rates, it still achieves some fairness between the users since the group size is fixed.

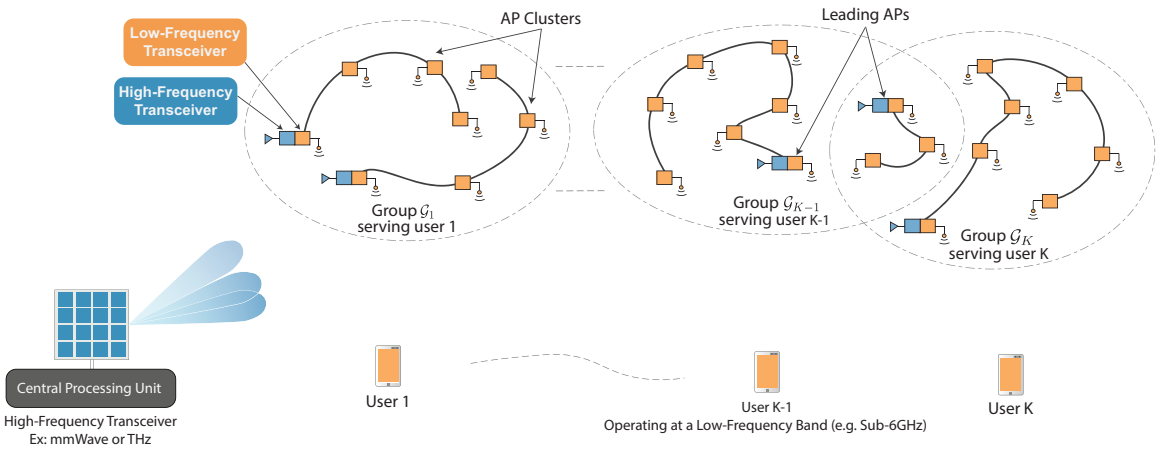


Figure 2.3: An Illustration of the Proposed Mixed-Fronthaul Architecture. In This Architecture, Each Subset of APs Is Connected Together via an Optical Fiber Forming a *cluster*. Each Cluster Has One Leading AP That Is Responsible for the Wireless Fronthaul Communication With the Central Processing Unit. For the Access Link, Each User Is Served by a Group of AP Clusters. This Architecture Has the Potential of Reducing the Fronthaul Cost/Complexity While Maintaining the Data Rate/Coverage Gains As Illustrated in Section 2.7.

gains, we propose a modified architecture with partially connected APs. In this section, we first describe the proposed architecture in Section 2.6.1 and then briefly present its rate optimization approach in Section 2.6.2. The performance evaluation of this architecture is then detailed in Section 2.7.3.

2.6.1 Description of the Proposed Mixed-Fronthul Architecture

To reduce the cost of the distributed APs while maintaining the potential deployment/operational gains of using wireless fronthaul, we propose the alternative cell-free massive MIMO architecture depicted in Figure 2.3. In this architecture, every cluster/set of distributed APs is connected with wired connections, e.g., an optical fiber. Further, each set of APs includes one leading AP that has a wireless-fronthaul to the central unit. This leading AP will be responsible for transmitting/receiving the cluster data to/from the central unit. It is worth noting here that the connection between the APs in each cluster could be realized using radio stripes [9]. With this implementation, the proposed mixed-fronthaul cell-free

massive MIMO architecture combines the gains of the radio stripes and the installation cost/flexibility gains of the higher-band wireless fronthaul, shaping a promising and practical solution for cell-free massive MIMO systems.

From an operation perspective, the proposed mixed-fronthaul architecture reduces the number of APs that are simultaneously using the wireless fronthaul, which relaxes the fronthaul requirements in terms of the equipment cost, the required bandwidth, and the beamforming design. Therefore, this mixed-fronthaul architecture has the potential of further improving the wireless fronthaul rates since better beams can be utilized by the central unit in serving the clusters leading APs. In terms of the end-to-end communication model, the main difference between the mixed-fronthaul architecture and the originally proposed architecture in Section 2.2.3 is that all the APs in one cluster communicate with the central unit through one leading AP. Therefore, and to simplify the end-to-end rate optimization, we assume that all the APs in one cluster will serve the same user. This approach is similar to [37], where the user-centric groups for cell-free massive MIMO are selected from the set of APs connected to the same CPU rather than individual APs. If the number of APs in each cluster is large, however, it could be important to relax this constraint, i.e., to allow any user to be served by only a subset of the cluster APs. In the next subsection, we elaborate more on the proposed rate-optimization approach.

2.6.2 Proposed Rate Optimization Approach

In this subsection, we extend the rate optimization solution presented in Section 2.5 to the modified cell-free massive MIMO architecture with mixed-fronthaul. First, we assume that all the APs in one cluster will be serving the same set of UEs. This is motivated by the negligible data transmission cost within the same cluster. Hence, the group of APs serving each user k will be determined by selecting the set of clusters (instead of selecting individual APs in the separate-AP architecture in Section 2.2.3-Section 2.5). Formally,

assume that there are L distinct AP clusters, such that $\mathcal{C}_l \subseteq \mathcal{M}$, $\forall l \in \{1, \ldots, L\}$ with $\bigcup_{l=1}^{L} \mathcal{C}_l = \mathcal{M}$ and $\mathcal{C}_l \cap \mathcal{C}_{l'} = \emptyset$, $\forall l \neq l'$. Then, the APs group of user k, which we denote as $\mathcal{G}_k^{\text{cluster}}$ to emphasize that they are selected from the available AP *clusters* will satisfy $\mathcal{G}_k^{\text{cluster}} \subseteq \{1, \ldots, L\}$. Further, we let a_l represent the leading AP of the lth cluster. Next, we revisit the fronthaul/access rate optimization and the group selection methodology from the lens of the mixed fronthaul architecture.

Fronthaul Rate Optimization: The initial fronthaul rate optimization was defined in (2.19), which aims to optimize the time and beams of the group $\mathcal{G}_k$. With the mixed architecture, the CPU only needs to transmit the data to the leading APs of the clusters of the k-th UE. To simplify the description, we define the user-centric fronthaul groups $\mathcal{G}_k^{\text{fh}} = \{a_l : \ l \in \mathcal{G}_k^{\text{cluster}}\}$ of the leading APs of the clusters. This notation allows a straightforward use of the fronthaul formulation in (2.19), only by adopting $\mathcal{G}_k^{\text{fh}}$ instead of $\mathcal{G}_k$. With this fronthaul group definition, the rest of the fronthaul optimization methodology can be carried out in the same form.

Access Rate Optimization: Similar to the fronthaul, we let $\mathcal{G}_k^{\text{ac}}$ denote the set of APs serving each user k over the access channel. This can be described as $\mathcal{G}_k^{\text{ac}} = \{\mathcal{C}_l : \ l \in \mathcal{G}_k^{\text{cluster}}\}$. With this notation, we can utilize the original access channel formulation in (2.17) and the adopted rate optimization approach by replacing the AP groups $\{\mathcal{G}_k\}$ with the access channel groups $\{\mathcal{G}_k^{\text{ac}}\}$.

Group Selection: For the selection of the groups, as the user groups are defined in terms of the clusters, a metric for each cluster is needed. For this purpose, we consider the sum of the channel gains of the APs of the clusters, i.e., $\bar{\beta}_{lk} = \sum_{m \in \mathcal{C}_l} \beta_{mk}$, and select the G clusters with the maximum value as the clusters of a UE, $\mathcal{G}_k^{\text{cluster}}$. After determining the active clusters of UEs, the obtained groups of clusters can be converted to the fronthaul and access channel groups of the APs by earlier definitions. Then, the access channel and fronthaul optimization problems can be utilized over the access channel and fronthaul groups,

Table 2.1: The Communication Parameters Used in the Simulations Are Presented.

Parameter	**Fronthaul**	**Access Channel**
Frequency ($f_c^{\text{fh}}, f_c^{\text{ac}}$)	28 GHz	3.5 GHz
Bandwidth (B)	240 MHz	20 MHz
Power ($\tilde{\rho}^{\text{fh}}, \tilde{\rho}^{\text{ac}}$)	30 dBm	10 dBm
Noise Figure (σ^2)	9dB	
Antenna Spacing (d_A)	$\lambda/2$	N/A
CPU Antennas (N)	128	

completing an iteration of end-to-end data rate maximization. The described process will be evaluated in the numerical results section to show the performance of connected APs.

2.7 Numerical Results

In this section, we evaluate the performance of the proposed cell-free massive MIMO architectures and the developed rate optimization algorithms. Further, these proposed solutions are compared with the classical fiber-fronthaul based cell-free massive MIMO systems. The flow of this section is as follows. First, we present the adopted system setup and parameters in Section 2.7.1. Then, we evaluate the performance of the two proposed cell-free massive MIMO architectures, namely with separate APs and connected APs, in Sections 2.7.2 and 2.7.3.

2.7.1 Simulation Setup

In the simulation, we considered the proposed cell-free massive MIMO architectures in Section 2.2.3 (for the separate APs case) and in Section 2.6 (for the connected APs case). For both architectures, we adopt mmWave fronthaul channels at $f^{\text{fh}} = 28$ GHz carrier frequency and access channels at $f^{\text{ac}} = 3.5$ GHz. The bandwidth allocated for the access

channels is $B^{\text{ac}} = 20$ MHz, and various values for the fronthaul bandwidth are considered, as will be discussed in the results. The noise figure is set to 9 dB for both the fronthaul and access links. The central unit is deploying a uniform linear array with $N = 128$ elements and half-wavelength antenna spacing. For the phase-shift set of analog beamforming at the CPU, we use $q = 3$ bit uniform quantizers. The transmit power of the central unit is assumed to be $\tilde{\rho}^{\text{fh}} = 30$ dBm for the fronthaul links while the APs and users transmit powers are set to $\tilde{\rho}^{\text{ac}} = \tilde{\rho}^{t} = 10$ dBm for the access links and pilot transmissions. The noise power is determined based on the bandwidth of the fronthaul and access channels, following $\sigma_n^2 = \sigma^2 \cdot B \cdot k \cdot T$, with the Boltzmann constant k, the channel bandwidth B, the temperature $T = 290K$, and the noise figure σ^2. The normalized power levels of the different channels are then determined by $\rho = \frac{\tilde{\rho}}{\sigma_n^2}$ with the corresponding parameters of the specific channel. These baseline system parameters are summarized in Table 2.1. The values that are selected differently from the baseline parameters in different figures are explicitly stated.

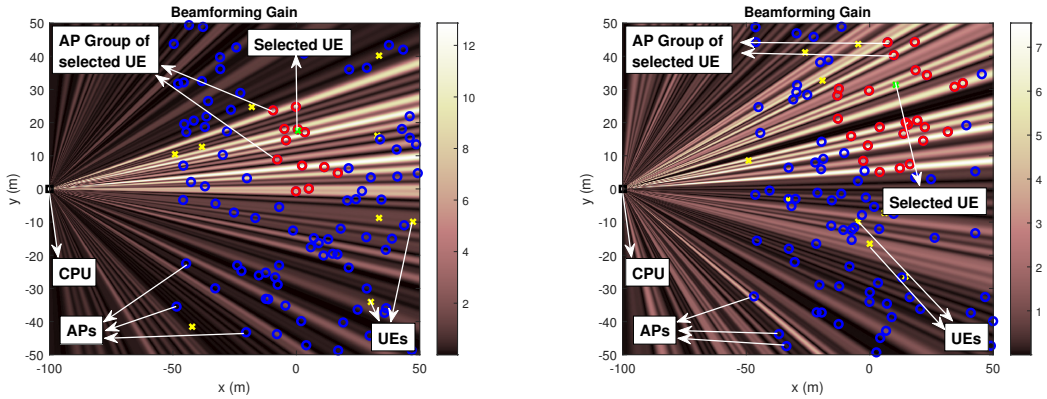
At the fronthaul, we consider a LOS multi-path channel with $V = 1 + V_{\text{NLOS}}$ paths, considering the potential availability of the LOS path with the fixed placement of the CPU and APs. The parameters of the NLOS path are determined randomly, where the number of paths $V_{\text{NLOS}} \sim \mathcal{U}[1, 6]$, angle-of-departure $\theta_{m,v} \sim \mathcal{U}[\pi/2, \pi/2]$ and the small-scale fading $\alpha_{m,v}^{\text{fh}} \sim \mathcal{CN}(0, 1)$ for $v > 1$. The LOS path parameters are determined from the geometry of the simulation model with $\alpha_{m,1}^{\text{fh}} = 1$. For the large-scale channel coefficients of LOS and NLOS paths, we adopt the Urban Micro (UMi) street-canyon model given in 3GPP 38.901 [38] as follows:

$$\text{PL}_{\text{LOS}}(d, f_c) = 32.4 + 21 \log_{10} d + 21 \log_{10} f_c + \mathcal{X}$$

$$\text{PL}_{\text{NLOS}}(d, f_c) = 32.4 + 31.9 \log_{10} d + 21 \log_{10} f_c + \mathcal{X},$$

where f_c is in GHz and d is in meters. $\mathcal{X}$ (dB) is the shadow fading effect determined by a Gaussian random variable with zero mean and standard deviation 4 and 8.2 for LOS and

NLOS paths of the fronthaul. For the access channel, we adopt $\mathrm{PL}_{\mathrm{NLOS}}(d_{mk}^{\mathrm{ac}}, f_c^{\mathrm{ac}})$, however, apply a spatially correlated shadow fading exactly as described in [8] with $\delta = 0.5$ and decorrelation distance 100m. Next, we evaluate the performance of the proposed solutions.



(a) Group of 12 APs

(b) Group of 25 APs

Figure 2.4: The Beamforming Gains of the Optimized Beamforming Vectors for the Groups of (a) 12 APs and (b) 25 APs. The Given Scenario Adopts Randomly Placed $M = 100$ APs and $K = 10$ UEs. The Beamforming Gain of a Group Depends on the Group Size, Decreasing With Larger Groups.

2.7.2 Evaluation of the Wireless Fronthaul Architecture

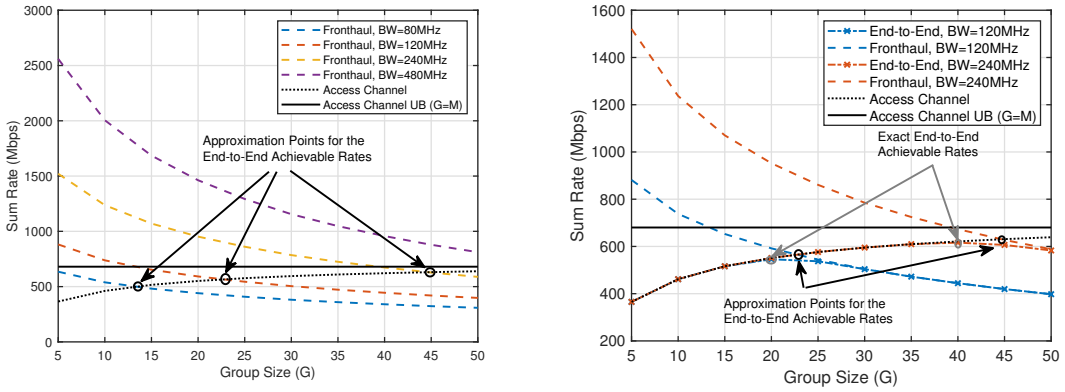
The objective of this subsection is to evaluate the performance of the proposed cell-free massive MIMO architecture with wireless fronthaul (and separate APs). In particular, we want to draw some insights into how this architecture performs if compared with the classical fiber-fronthaul based cell-free massive MIMO architecture and whether it can achieve comparable data rates. To do that, we consider a setup where the UEs and APs are randomly placed over an area of $100m \times 100m$ centered at $(0, 0)$. Unless otherwise stated, the following simulations assume that $K = 10$ UEs and $M = 100$ APs are placed in the considered area. The central unit (CPU) is located at $\mathbf{z}_0 = (x_0, y_0) = (-D, 0)$, i.e., at a distance $D = 100$m away from the center of the square area. The simulations are averaged over 250 realizations. Each realization drops the APs and UEs randomly, adopting a uni-

form probability distribution. With this setup, we study the following important questions.

What is the performance of the fronthaul beamforming? To evaluate the performance of the adopted fronthaul beamforming approach [6] in our wireless-fronthaul based cell-free massive MIMO architecture, we plot the achievable beamforming gain for the considered deployment in Figure 2.4. This figure shows the beamforming gain in x-y coordinates for two different AP group sizes, namely a group of 12 APs in Figure 2.4(a) and a group of 25 APs in Figure 2.4(b). These figures provide a visual verification that the central unit focuses its multi-cast beam toward the APs in the group of interest. In terms of beamforming gain, it varies from 12 (i.e., around 11dB) in the case of 12 APs group to 4 (i.e., around 6dB) in the 25 APs/group case. This highlights the trade-off between the fronthaul rate and access rate as the group sizes increases, as this decreases the fronthaul rate while increasing the access rate. It is worth mentioning here that this fronthaul beamforming can be further improved when adopting hybrid precoding approaches [27, 39]. Also, leveraging machine learning could enable the autonomous design of these beamforming vectors/codebooks [40–42]. This can further reduce the deployment overhead.

What are the achievable fronthaul and access rates? Now, we want to evaluate the fronthaul and access achievable rates using the developed rate optimization solutions in Section 2.5. Adopting the setup described earlier in this section with 10 UEs and 100 APs, we plot the achievable fronthaul/access sum-rates in Figure 2.5(a) versus the AP group size. These achievable sum-rates (sum of the rates of UEs) are obtained from the solutions of the fronthaul (Approach 1) and cell-free massive MIMO solutions, and present the upper bounds on the end-to-end rate. They reflect the behavior of the system within a transmission frame, and their intersection point can be considered as a good approximation

[6] As mentioned in Section 2.5.2, we adopt the suboptimal approach in [34] to optimize the beams. The approach starts from a beam vector and iteratively updates each beamforming coefficient to maximize the rate until the convergence. To ensure good performance, this solution is applied for 100 starting points, and the best-performing beam is adopted in the simulations.



(a) Sum fronthaul and access channel rates

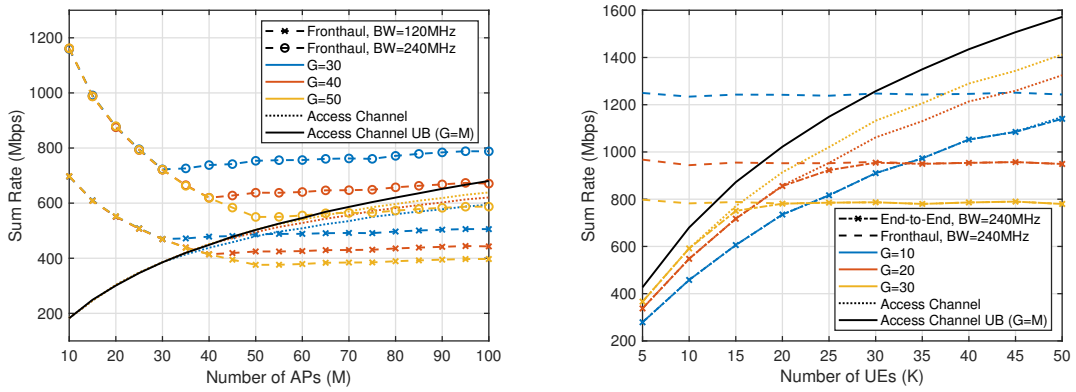
(b) Sum fronthaul, access channel and end-to-end rates

Figure 2.5: The Access Channel and Fronthaul Sum Data Rates of the Proposed Architecture With Different Group Size and Fronthaul Bandwidth Values. (a) Shows the Change in the Fronthaul Rates Based on the Available Bandwidth and Group Size. (b) Includes the Sum of the End-to-End Data Rates Reflecting the Exact Performance.

for the achievable end-to-end sum-rate, as will be discussed shortly in Figure 2.5(b). The group for each user is determined based on the channel gain criterion described in Section 2.5.3, the access rate is optimized based on the solution presented in Section 2.5.1, and the fronthaul time allocation optimization is implemented based on Approach 1 in 2.5.2. As an upper bound on the cell-free massive MIMO system, we show fully activated APs, i.e., $G = M = 100$, without any fronthaul limitations. This curve is abbreviated as Access Channel UB in the figure. As shown in Figure 2.5(a), as the group size increases, the access rate increases because of the higher access channel beamforming gain, and the fronthaul rate decreases because of the multicasting to more APs with a lower beamforming gain. With high enough fronthaul bandwidth, e.g., 240MHz, the proposed rate optimization approach and cell-free massive MIMO architecture intersect at 630Mbps sum-rate. This is very close to the 680Mbps sum-rate achieved with 100 active APs without any fronthaul limitations. These results highlight the promising data rates of the proposed rate optimization solutions.

Can the proposed architecture with wireless fronthaul approach the data rates of the classical fiber-fronthaul based cell-free massive MIMO architecture? This is a key question that we are trying to answer in this chapter. Essentially, the proposed architecture with wireless higher-band fronthaul has clear gains in terms of the installation cost/time/flexibility, but could it also provide comparable achievable rates to the fiber-based architecture? In Figure 2.5(a), we provided some insights into the answer of this question by separately evaluating the achievable fronthaul and access rates. The exact achievable rate, however, is the one with matched fronthaul and access rates. To evaluate this exact rate, we adopt the fronthaul rate optimization (Approach 2) in Section 2.5.2, which ensures that the fronthaul rate of each link does not exceed its corresponding access rate. In Figure 2.5(b), we plot the achievable rates for the access and fronthaul links using Approach 2. This rate then represents the exact end-to-end achievable rates using the proposed architecture. For reference, we also plot the fronthaul rates using Approach 1. As shown in Figure 2.5(b), with sufficient fronthaul bandwidth, the exact achievable end-to-end data rates using the proposed architecture are very comparable to the classical fiber-based cell-free massive MIMO architecture. For example, with 240MHz fronthaul bandwidth, the proposed solution achieves around 620Mbps sum-rate compared to 680Mbps for the fully active APs without any fronthaul limitations. This difference can be further reduced by increasing the fronthaul bandwidth. It is also good to note that the exact rate is very close to the approximation discussed in Figure 2.5(a), i.e., 630.

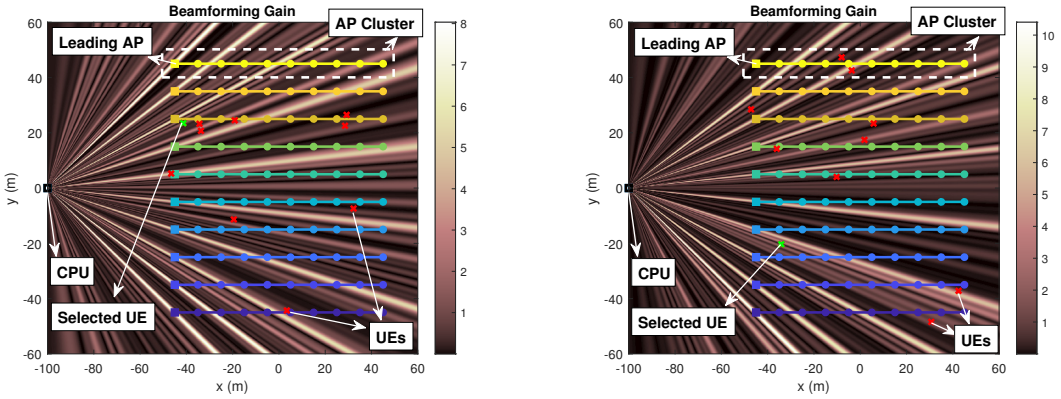
Can the proposed architecture support a large number of distributed APs and UEs? The scalability is a key objective of cell-free massive MIMO systems to be able to support more users/larger areas and to increase the beamforming gains/achievable rates. To draw some insights into the capability of the proposed architecture in supporting large numbers of APs and UEs, we plot the fronthaul/access rates vs the number of APs in Figure 2.6(a), and the number of UEs in Figure 2.6(b). As shown in Figure 2.6(a), for each



(a) Achievable rates with varying number of APs (b) Achievable rates with varying number of UEs

Figure 2.6: The Access Channel and Fronthaul Sum Data Rates of the Proposed Architecture With Different Numbers of (a) APs With $K = 10$ UEs, and (b) UEs With $M = 100$ APs. The Figures Show the Relationship of Group Sizes, G, to Different Numbers of APs and UEs.

group size, the access rates increase with denser APs due to the higher access beamforming gains. At the fronthaul, the rate decreases rapidly for $M \leq G$. In this region, $M \leq G$, the group size is taken as $G = M$. Therefore, every new AP introduced into the system also increases the effective group size by adding APs to the groups. This decreases the fronthaul beamforming gains. For larger values of M, the fronthaul rates slightly increase as the area containing a single AP group becomes smaller, allowing higher beamforming gains. Overall, however, the end-to-end data rates, which are approximately given by the intersections of the fronthaul and access channel rates, are increasing with more APs. This is also very comparable to the increase in data rates experienced by the fiber-based architecture, which highlights the potential of the proposed solution. In Figure 2.6(b), we investigate the effect of the increasing number of UEs. In the figure, the sum fronthaul rate with a fixed group size is almost constant for different numbers of UEs, mainly due to the fronthaul resources and TDMA-based communication design. However, with smaller group size values, the achievable maximum fronthaul rate increases, while less number of APs serve each UE, and the access channel rate decreases. At the access channel, with the scaling of UEs, the



(a) Group size of 10 clusters

(b) Group size of 9 clusters

Figure 2.7: The Beamforming Gains With the Optimized Beamforming Vectors Are Shown for the Mixed-Fronthaul Architecture. In (a), All of the Clusters Are Activated; Hence the Beam Gain Is Maximized Towards All of the Leading APs. In (b), the Bottom 9 Clusters Are Activated, and the Beam Gain Is Not Maximized for the Leading AP of the Top Cluster.

effect of the group size increases. However, even with $K = 50$ APs, the 80% of the data rate with fiber-fronthaul can be obtained while only using 240MHz fronthaul bandwidth.

2.7.3 Evaluation of the Mixed-Fronthaul Fronthaul Architecture

In this subsection, we evaluate the performance of the proposed modified architecture in Section 2.6 with connected APs. As shown in Figure 2.7, we consider a square grid of APs, where the APs are located uniformly in a 10×10 grid with equal distance between the rows and columns. For simplicity, we assume that the 10 APs in each line in the y-axis form a cluster. The leaders of the clusters (that communicate with the central unit through the mmWave fronthaul) are selected as the closest APs to the central unit. Further, we assume that $K = 10$ UEs are randomly located in the $100m \times 100m$ area of interest. The central unit is located at the origin, as depicted in Figure 2.7, and is equipped with a ULA of $N = 128$ antenna elements.

When does the achievable data rate with the proposed architecture approach the optimal solution? To evaluate the achievable data rates with the proposed connected-

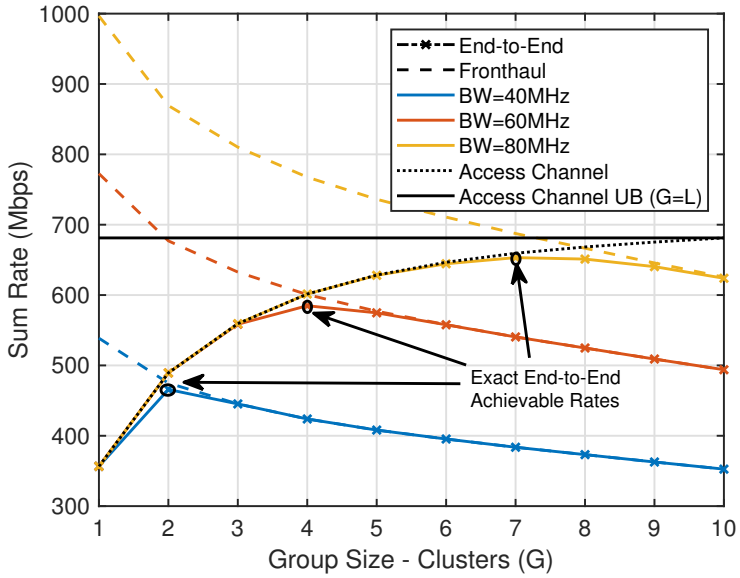


Figure 2.8: For the Mixed-Fronthaul Architecture, the Fronthaul, Access Channel, and End-to-End Sum Rates Are Shown for Different Numbers of Group Sizes (of Clusters). In the Figure, the 10×10 APs Rectangular Grid Scenario Illustrated in APA Figure 2.7 Is Adopted, and Each Cluster Consists of 10 APs.

APs cell-free massive MIMO architecture, we adopt the setup described in the previous paragraph, which is also depicted in Figure 2.7. For this setup, we first investigate the achievable beamforming gains for different numbers of AP clusters per group. To illustrate, we consider one UE and plot the beamforming gains for the clusters serving this UE in two scenarios: (i) When a group of 10 AP clusters (i.e., all the clusters) jointly serve this UE as shown in Figure 2.7(a) and (ii) when a group of the 9 bottom clusters serves the UE, as shown in Figure 2.7(b). As expected, when the number of AP clusters per group increases, the beamforming gain decreases. Next, we calculate the achievable fronthaul and access rates of the same setup in Figure 2.8 for different numbers of AP clusters per group. As shown in this figure, the modified architecture with connected APs has clear data rate gains compared to the separate APs architecture studied in Figure 2.5. With only 80MHz fronthaul bandwidth, the proposed connected AP architecture achieves nearly the same end-to-end rates obtained by the upper bound, which is defined by the classical fiber-

fronthaul based cell-free massive MIMO architecture with fully active APs. This provides a promising solution for flexible, scalable, and efficient cell-free massive MIMO systems.

2.8 Summary

In this chapter, we proposed two wireless-fronthaul based architectures for cell-free massive MIMO systems where the fronthaul link is operating at a higher band (e.g., mmWave) compared to the access band (e.g., sub-6GHz). We formulated the end-to-end data rate optimization problems for these architectures based on user-centric AP grouping. Then, we developed efficient transmission and resource allocation strategies that can achieve near-optimal performance. Simulation results showed that the proposed architectures could achieve data rates that approach the optimal rates obtained with optical fiber-based fronthaul with reasonable assumptions on the number of CPU antennas, coverage area, number of users, etc. The results also showed that these architectures are capable of supporting large numbers of APs under realistic fronthaul bandwidth requirements.

Chapter 3

RADAR-AIDED BEAM PREDICTION AND TRACKING IN REAL-WORLD

3.1 Overview

This chapter presents a machine learning-based real-world demonstration for radar-aided beam prediction in a practical vehicular communication scenario. Leveraging radar sensory data at the communication terminals provides essential awareness about the transmitter/receiver locations and the surrounding environment. This awareness could be utilized to reduce or even eliminate the beam training overhead in millimeter wave (mmWave) and sub-terahertz (THz) MIMO communication systems, which enables a wide range of highly-mobile low-latency applications. In this chapter, we develop deep learning-based radar-aided beam prediction approaches for mmWave/sub-THz systems. The developed solutions leverage domain knowledge for radar signal processing to extract the relevant features fed to the learning models. This optimizes their performance, complexity, and inference time. The proposed radar-aided beam prediction solutions are evaluated using the large-scale real-world dataset DeepSense 6G, which comprises co-existing mmWave beam training and radar measurements. In addition to completely eliminating the radar/communication calibration overhead, the experimental results showed that the proposed algorithms are able to achieve around 90% top-5 beam prediction accuracy while saving 93% of the beam training overhead. This highlights a promising direction for addressing the beam management overhead challenges in mmWave/THz communication systems.

3.2 Introduction

Millimeter wave (mmWave) and terahertz (THz) communications systems rely on the beamforming gains of the narrow beams to achieve sufficient receive signal power. Finding the best narrow beam (or beam pair), however, requires high beam training overhead, which makes it hard for these systems to support highly mobile applications such as vehicular, drone, or augmented/virtual reality communications [5]. One important observation here is that the beam selection problem highly relies on the transmitter/receiver locations and the geometry/characteristics of the surrounding environment. This means that acquiring some awareness about the surrounding environment and the transmitter/receiver locations could potentially help the mmWave beam selection problem. An efficient way to acquire this awareness is by using low-cost radar sensors such as those initially designed for radar applications [43] or by leveraging joint communication-radar systems [44, 45]. *With this motivation, this chapter investigates the potential of leveraging radar sensory data to guide the beam prediction and tracking problems and provides the first machine learning-based real-world demonstration for radar-aided beam prediction and tracking in a practical vehicular communication scenario.*

3.2.1 Prior Work

Leveraging sensory data to guide the mmWave beam selection problem has gained increasing interest in the last few years [7, 46–51]. In [46, 47], the authors proposed to leverage the sub-6GHz channels that are relatively easier to acquire to guide the beam selection problem. Acquiring sub-6GHz channels, however, still requires allocating wireless communication resources and probably additional control signaling. In [48, 49], the user's position information was leveraged by the base station to select the mmWave beam. The position information, though, may not be sufficient to accurately determine the best beam,

which is also a function of the surrounding environment, especially in the non-line-of-sight scenarios. Further, acquiring accurate enough position information to adjust the narrow beams (i) may require expensive positioning systems for the user for the outdoor scenarios, and (ii) is hard to achieve for indoor communication. This motivated leveraging other data modalities for beam selection such as: (i) Vision [48, 50], which could be acquired at low-cost and without consuming any wireless communication/control resources, (ii) Radar [7, 51], which may operate at a different band than that used by the mmWave communication system. The prior work on using radar for beam management, however, relied mainly on classical calibration techniques for the radar and communication systems, which could be expensive and hard to implement in reality. The prior work was also evaluated only using computer simulations and relatively simple scenarios that are different from real-world deployments and practical hardware imperfections.

3.2.2 Contributions

In this work, we develop classical and machine learning-based algorithms for radar-aided mmWave beam prediction and tracking problems. Later we demonstrate their performance using a real-world dataset in a realistic vehicular communication scenario. The main contributions of the chapter can be summarized as follows:

- We formulate the radar-aided beam prediction problem considering practical radar and communication models. For this problem, we leverage classical signal preprocessing approaches for extracting the relevant features such as range-Doppler, range-angle, and range-Doppler-angle maps. Then, we develop a classical solution with object detection and lookup table-based radar angle-to-beam mapping and deep learning algorithms for different types of maps.

- We extend our initial problem formulation and formulate a beam tracking problem,

where the samples from the previous measurements are also utilized with a tracking scheme, and the beam from future instances can also be predicted. With this problem, we develop an extended Kalman filter-based classical solution. Next, we propose a deep learning solution with the long short-term memory (LSTM) architecture to track the beams.

- Leveraging the large-scale real-world dataset, DeepSense 6G [52] that comprises co-existing mmWave beam training and radar measurements, we evaluate and demonstrate the performance of the proposed radar-aided beam prediction and tracking approaches in a realistic vehicular communication scenario.

We also draw important insights about the trade-offs of the various algorithms in terms of beam prediction accuracy, processing time, inference latency, and complexity overhead.

3.2.3 Key Idea

In this chapter, we propose to leverage machine learning to optimize the mapping from the radar measurements to the optimal communication beamforming vectors. The motivation for the proposed approach is based on the following main observations:

- The FMCW radars are designed to collect specific measurements of the environment that are useful for automotive applications. These measurements are generally different from the communication channel state information we typically need to adjust the communication beams. Therefore, a straightforward mapping between the radar and communication channels is non-trivial.

- Using classical approaches such as lookup tables to directly map the radar object detection and positioning information to the best beam may not be efficient. The reason goes back to the imperfections of the practical radar systems that impact the detection and localization performance accuracy. Moreover, the other elements (e.g.,

pedestrians and bikers) in the real-world environments present themselves as a challenge in the detection and localization of the target users. Further, the beam patterns of the practical systems do not cover the field of view in an ideal way, and selecting the best beam may require additional refinement in the angle-beam mapping. Therefore, the practical features of the system and deployment scenarios degrade the performance of potential classical solutions, even in the more ideal LOS scenarios. We will elaborate more on this point in Section 3.8.

- With the recent advances in artificial intelligence, machine learning models have become prevalent in complex mapping and recognition problems. It is mainly thanks to their high capability in extracting the inherent information with significant success rates. Particularly with the FMCW radars, the machine learning-based solutions have significantly improved the object, vehicle, and pedestrian detection and classification problems [53, 54]. Moreover, it enabled more advanced mappings, such as gesture detection with radars [55].

With this motivation, we propose leveraging machine learning, particularly deep learning models, to learn the mapping from the radar measurements to the optimal beamforming vectors. Our solutions will integrate these machine learning models with domain-knowledge-based radar preprocessing techniques to reduce the complexity of the learning problem in realistic environments. The general flow of the proposed machine learning and radar-aided beam prediction approach is illustrated in Figure 3.3. In the next section, we will provide a detailed description of the proposed solution.

3.3 System Model

The considered system in this chapter consists of a base station and a mobile user. The base station employs two main components: (i) A mmWave communication terminal

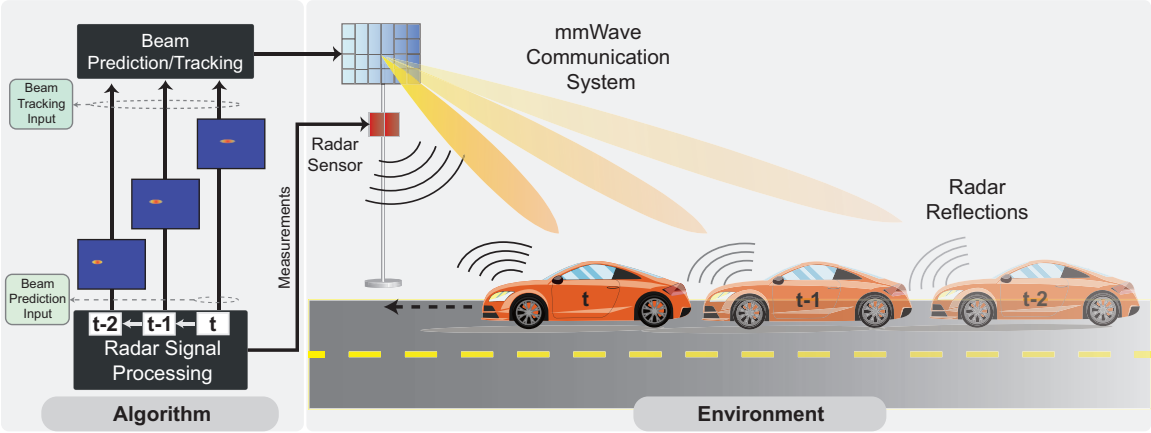


Figure 3.1: The System Model Where the Radar Measurements at the Basestation Are Leveraged to Select the Beam That Serves the Mobile User. The Beam Prediction Only Uses the Current Radar Measurements, While Beam Tracking Leverages the Latest Measurements.

Table 3.1: The Notation Adopted in the Chapter.

Notation	**Description**	**Notation**	**Description**
M_a	# of radar RX antennas	N_a	Angle FFT size
M_r	# of samples per chirp	N_r	Range FFT size
M_c	# of chirps per frame	N_c	Doppler FFT size
μ	Chirp slope	$\mathcal{F}_{\text{2D}(.)}, \mathcal{F}_{\text{3D}(.)}$	2D and 3D FFTs
T_c	Chirp duration	T_f	Frame duration
$\mathbf{X}_l$	Radar data of frame l	$\tilde{\mathbf{X}}_l$	Set of T_o latest radar data
L	# of frames (measurements)	$\mathbf{H}_{\text{RC}}$	Radar cube
B	Beamforming codebook size	$\mathbf{H}_{\text{RA}}$	Range-Angle map
$b_l^\star$	Optimal beam of l-th frame	$\mathbf{H}_{\text{RV}}$	Range-Doppler map

equipped with a phased array that is used to communicate with the mobile user and (ii) an FMCW radar that is leveraged to aid the selection of the mmWave communication beam. The system model is illustrated in Figure 3.1. In the following two subsections, we briefly describe the system and signal models of the communication and radar components. For ease of exposition, we summarize the adopted notation in Table 3.1.

3.3.1 Radar Model

In our system, the base station adopts an FMCW radar. The objective of this radar is to provide sensing observations (perception) about the communication environment. The FMCW radar achieves this objective by transmitting chirp signals and processing their received scattering/reflection from the various objects in the environment. More formally, the FMCW radar transmits a linear chirp signal starting at an initial frequency f_c and linearly ramping by $f_c + \mu t$, which is written as

$$s^{\text{tx}}_{\text{chirp}}(t) = \begin{cases} \sin(2\pi[f_c\, t + \frac{\mu}{2}\, t^2]) & \text{if } 0 \leq t \leq T_c \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\mu = \mathsf{BW}/T_c$ is the slope of the linear chirp signal with BW and T_c representing the bandwidth and duration of the chirp [56].

A single radar measurement frame consists of M_c chirp waves that are transmitted with T_s waiting time between them. After the transmission of the last chirp, no other signals are transmitted until the completion of the frame. Let T_f denote the frame duration; then, we can write the transmitted signal of the radar frame as

$$s^{\text{tx}}_{\text{frame}}(t) = \sqrt{\mathcal{E}_t} \sum_{c=0}^{M_c-1} s_{\text{chirp}}(t - (T_c + T_s) \cdot c), \quad 0 \leq t \leq T_f, \tag{3.2}$$

where $\mathcal{E}_t$ is the transmission power gain. The given transmitted signal is scattered/reflected from the objects in the environment and received back at the radar.

At the receiver, the signal obtained from an antenna is passed through a quadrature mixer that combines the transmit with the received signal resulting in in-phase and quadrature samples. After that, a low-pass filter is applied to the mixed signals. The resulting signal, referred to as the intermediate frequency (IF) signal, reflects the frequency and phase difference between the transmit and receive signals. If a single object exists in the

environment, then the receive IF signal of a single chirp can be written as [57]

$$s_{\text{chirp}}^{\text{rx}}(t) = \sqrt{\mathcal{E}_t \mathcal{E}_r} \exp\left(j2\pi[\mu\tau t + f_c\tau - \frac{\mu}{2}\tau^2]\right), \tag{3.3}$$

where $\mathcal{E}_r$ is the reflection/scattering gain of the object which depends on the radar cross section (RCS) and the path-loss, $\tau = 2d/c$ is the round-trip delay of the reflected signal through the object with d denoting the distance between the object and the radar, and c representing the speed of light.

The receive IF signal, $s_{\text{chirp}}^{\text{rx}}(t)$, is then sampled at the ADC sampling rate, f_s, producing M_r samples for each chirp. The ADC samples from each frame are then collected to be ready for processing. For an FMCW radar with M_a receive antennas, each having the described RF receive chain, the resulting measurements (raw-data) of one frame can be denoted by $\mathbf{X} \in \mathbb{C}^{M_a \times M_r \times M_c}$. For further information on the utilized FMCW radar model and details on the hardware architecture, please refer to [57, 58]. In the following subsection, we describe the communication model.

3.3.2 Communication Model

The considered base station employs a mmWave transceiver with M_a^c antennas to communicate with a single-antenna mobile user. Adopting a geometric channel model with P paths, the channel between the user and the base station can be expressed as [59]

$$\mathbf{h} = \sum_{p=1}^{P} \alpha_p \mathbf{a}(\phi_p, \theta_p), \tag{3.4}$$

where α_p denotes the complex gain and ϕ_p, θ_p represent transmit azimuth and elevation angles of the p-th path at the base station. In the downlink, if the base station transmits the data symbol s_d to the user via the beamforming vector $\mathbf{f} \in \mathbb{C}^{M_a^c}$, then the received signal at the user can be written as

$$y = \sqrt{\mathcal{E}_c}\mathbf{h}^H \mathbf{f} s_d + n, \tag{3.5}$$

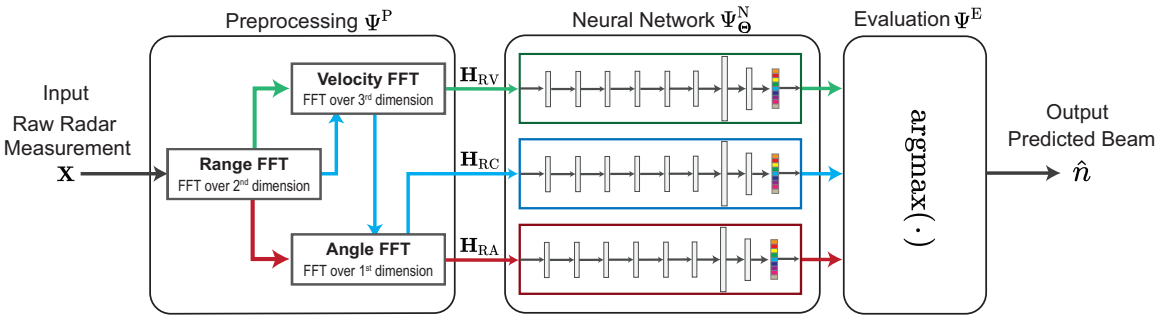


Figure 3.2: The Figure Illustrates the Radar Processing Procedures for the Three Proposed Approaches. The Results of These Processing Procedures, $\mathbf{H}_{\mathrm{RC}}, \mathbf{H}_{\mathrm{RA}}, \mathbf{H}_{\mathrm{RV}}$ Are Then Given to the Neural Networks for the Beam Prediction. Although the Outputs of the Different Preprocessing Approaches' Sizes Are Different, They Utilize the Same Network Architecture With Different Layer Dimensions.

where $n \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise and $\mathcal{E}_c$ is the transmission power gain of the basestation. The beamforming vector f is assumed to be selected from a predefined codebook $\boldsymbol{\mathcal{F}}$ of B beams, $\{\mathbf{f}_b\}$ [39, 42]. With this model, the index of the optimal beam, $b^\star$, becomes the result of the beamforming gain maximization problem given as

$$b^\star = \underset{b}{\operatorname{argmax}} \, |\mathbf{h}^H \mathbf{f}_b|^2 \quad \text{s.t.} \quad \mathbf{f}_b \in \boldsymbol{\mathcal{F}}. \tag{3.6}$$

The optimal solution to this problem can be obtained by an exhaustive search of the possible beamforming vectors.

3.4 Radar Preprocessing

The radar measurements can be processed with the conventional signal processing techniques to extract useful information, such as the range, angle, and Doppler velocity of the various objects in the environment. As will be seen in the following sections, this information could be utilized in the development of efficient radar-aided beam management solutions. With this motivation, this section presents the key radar processing approaches that are adopted in the rest of the chapter. Starting with the radar measurement $\mathbf{X} \in \mathbb{C}^{M_a \times M_r \times M_c}$, three important features that could be extracted are the range, the angle, and the Doppler velocity of the objects. Based on that, we propose three different prepro-

cessing approaches, as illustrated in Figure 3.2. Each approach leverages a certain set of these quantities. To mathematically define these approaches, let us denote the 2D and 3D Fourier transforms by $\mathcal{F}_{\text{2D}(.)}$ and $\mathcal{F}_{\text{3D}}(.)$, respectively. Next, we detail the preprocessing methods used in this chapter.

Range-Angle Maps: The first approach aims at utilizing the range and angle information. For this purpose, first, with an FFT in the direction of the time samples, referred to as the Range FFT, we obtain the chirp signal in the frequency domain. In this domain, the chirp signals are shifted proportionally to the round-trip travel duration of the signal, providing the range information. After that, a clutter removal operation can be applied to obtain cleaner images by a mean removal over the chirp samples. Then, with an another FFT in the direction of the receiver antenna samples, referred to as the Angle FFT, the angular information can be obtained. FFTs of larger sizes can be applied with zero padding to over-sample the angles. Let us denote the size of the range and angle FFTs by N_r and N_a, respectively. Finally, we can construct the final range-angle map by combining the resulting range-angle information for each chirp sample. In a simplified way, the described operation can be mathematically written as

$$\mathbf{H}_{\text{RA}} = \Psi^{\text{P}}_{\text{RA}}(\mathbf{X}) = \sum_{c=1}^{M_c} |\mathcal{F}_{\text{2D}}(\mathbf{X}_{:,:,c})|. \tag{3.7}$$

Range-Doppler Maps: Alternatively, we consider the range-Doppler maps. To construct these maps from the radar measurements, two FFTs through the time samples and chirp samples are applied. Similarly to the previous approach, first, the Range FFT of size N_r is utilized. Differently, the second FFT of size N_c is applied through the chirp samples, referred to as the Doppler FFT. It essentially returns the phase shift over the consecutive chirp samples, mainly caused by the Doppler shift. Finally, again by combining the range-Doppler information of the different receiver antenna samples, we obtain the final

Table 3.2: Complexity and Memory Requirements of Preprocessing.

Network Input	**Preprocessing Complexity**	**Output Size**
$\mathbf{H}_{\text{RC}}$	$\mathcal{O}(N_r N_c N_a \log N_c N_a N_r)$	$N_r N_c N_a$
$\mathbf{H}_{\text{RA}}$	$\mathcal{O}(N_r M_c N_a \log N_r N_a)$	$N_r N_a$
$\mathbf{H}_{\text{RV}}$	$\mathcal{O}(N_r N_c M_a \log N_r N_c)$	$N_r N_c$

range-Doppler map. This operation can be written as

$$\mathbf{H}_{\text{RV}} = \Psi^{\text{P}}_{\text{RV}}(\mathbf{X}) = \sum_{a=1}^{M_a} |\mathcal{F}_{\text{2D}}(\mathbf{X}_{a,:,:})|. \tag{3.8}$$

Radar Cube: The previous approaches combine the angle or Doppler dimensions, reducing the information to a 2D map. Without a dimensionality reduction, we apply the range, Doppler, and angle FTTs, and obtain the radar cube. The resulting radar cube contains all the information of the range, Doppler, and angle of the targets. It can be considered as the stack of range-angle maps of each Doppler value. The operation can be mathematically described as

$$\mathbf{H}_{\text{RC}} = \Psi^{\text{P}}_{\text{RC}}(\mathbf{X}) = |\mathcal{F}_{\text{3D}}(\mathbf{X})|. \tag{3.9}$$

We note that the described radar processing approaches bring different preprocessing complexities and input sizes. In particular, while the radar cube requires a 3D FFT (presenting the most detailed information), it suffers from a high number of dimensions. In contrast, the range-angle and range-Doppler images only require 2D FFTs and provide smaller input sizes. In Table 3.2, we present the complexity of generating the different maps and their output dimensionality. Essentially, the complexity of a k-dimensional FFT can be written as $k \log k$, which is applied through each dimension. As a result, a larger complexity is observed with a larger FFT size and a larger number of FFTs.

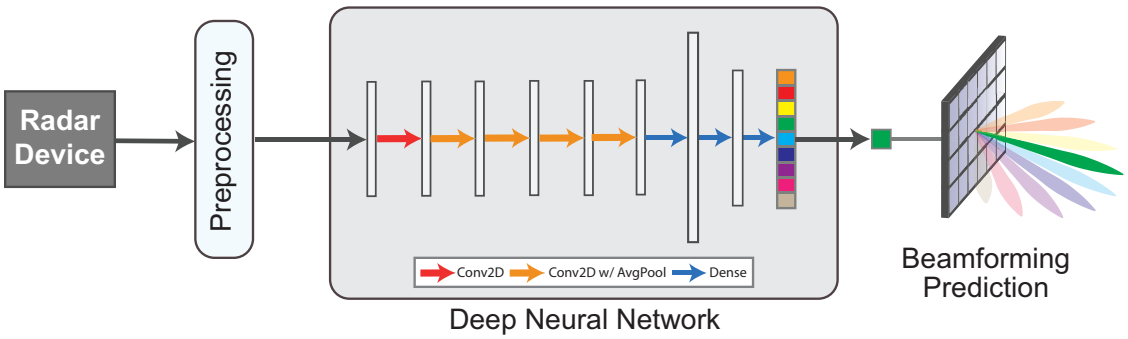


Figure 3.3: The Figure Illustrates the Proposed Approach: (i) the Radar Observations Are Preprocessed to Extract the Useful Features. (ii) These Features Are Then Fed to the Deep Neural Network Model. The Model Returns the Prediction of the Beamforming Vector That Should Be Used at the Basestation for Seamless Communication Performance.

3.5 Radar-aided Beam Prediction

3.5.1 Problem Formulation

In this section, we formally define the radar-aided beam prediction problem building upon the system model described in Section 3.3. Then, we present the problem as a machine learning task.

Problem Definition

We seek to leverage the radar measurements $\mathbf{X}$ in determining the optimal communication beamforming vector $\mathbf{f}_{b^\star}$ (out of the pre-defined codebook $\boldsymbol{\mathcal{F}}$). First, let us introduce the subscript l to indicate the l-th radar frame (which consists of M_c chirps as described in Section II). The radar measurements based on this frame will then be denoted as $\mathbf{X}_l$. Further, we add this subscript l to the beamforming index used in this l-th frame, to be b_l, and the corresponding beamforming vector becomes $\mathbf{f}_{b_l}$. If a single user exists in the line-of-sight (LOS) of the base station, then the radar measurements, $\mathbf{X}_l$, could potentially include useful information about its position/orientation with respect to the base station. This position/orientation information could be leveraged to guide the optimal beam selection. To formulate that, we define the mapping function $\Psi_{\boldsymbol{\Theta}}$ to capture the mapping from the radar

observations to the optimal beamforming index, given by

$$\Psi_{\boldsymbol{\Theta}}(\mathbf{X}_l) = b_l^\star, \tag{3.10}$$

Our objective then is to design the mapping function $\Psi_{\boldsymbol{\Theta}}$ to be able to map the radar measurements to the optimal beam index $b^\star$. Towards this objective, we investigate the possible designs of the mapping function and optimize the set of parameters $\boldsymbol{\Theta}$. Mathematically, we can express this objective by the following optimization problem, which aims at maximizing the accuracy in predicting the optimal beam

$$\Psi_{\boldsymbol{\Theta}^\star}^\star = \arg\max_{\Psi_\Theta} \frac{1}{L} \sum_{l=1}^{L} \mathbf{1}_{\{b_l^* = \Psi_{\boldsymbol{\Theta}}(\mathbf{X}_l)\}}, \tag{3.11}$$

where $\mathbf{1}_E$ is the indicator function of the event E, i.e., $\mathbf{1}_E = 1$ if E occurs, and $\mathbf{1}_E = 0$ otherwise.

Machine Learning Task

We define the machine learning task as follows: Given the l-th radar observation matrix (raw data) $\mathbf{X}_l$, the objective is to design a machine learning model that returns the index of the optimal beam. In other words, the machine learning model aims to return the index of the beam providing the most gain in the beamforming codebook. In a more general sense, the top-K predictions can be utilized. In this case, the model returns the ordered set of the K most likely beam indices. Mathematically, for an observation $\mathbf{X}_l$, the beam prediction task returns the ordered set of the indices of the most promising K beams $\hat{\mathbf{b}}_l^\star = \{\hat{b}_{l,1}^\star, \ldots, \hat{b}_{l,K}^\star\}$ from a codebook of B vectors, where $b_{l,k}^\star$ is the k-th most likely beam. With this notation, the beam prediction task is a multi-class classification problem given by

$$\min \frac{1}{L} \sum_{l=1}^{L} \mathcal{L}\left(\hat{\mathbf{b}}_l^\star, b_l^\star\right), \tag{3.12}$$

where $\mathcal{L}(.,.)$ denotes the loss function. To evaluate the beam prediction accuracy, we adopt the top-K evaluation metric, which is defined as

$$\text{Top-K Accuracy} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{1}_{\{\hat{b}_{l,k}^{\star}=b_l^{\star}\}}. \tag{3.13}$$

Based on this beam prediction problem formulation, we develop two different approaches and present them in the following section.

3.5.2 Proposed Solutions

In this section, we propose two solutions for the beam prediction problem: (i) A classical solution relying on the detection of the object angle based on the range-angle map, and mapping the angle of object to the beam via a lookup table, and (ii) a deep learning solution that utilizes the generated maps to predict the beams.

Conventional Signal Processing Solution

As a classical solution, we consider an object detection method to determine the vehicle position in the range-angle map, and then map it to the communication beam. For this purpose, we first detect the object's position by selecting the maximum reflection amplitude in the range-angle map and then use a lookup table to determine the most likely beam corresponding to the object's angle. We present the details of these steps in the following.

Area of Interest: Different elements in the real-world environments may degrade the detection performance, such as the pedestrians in the field of view. To prevent this, we apply an area of interest to the detection of the object. The area of interest is the area contained in the boundaries that may contain the target object. It can be determined based on prior knowledge or analysis of the scenario. Specifically, in our solution, the range bins in $[\underline{N}_r, \overline{N}_r]$ and the angle bins in $[\underline{N}_a, \overline{N}_a]$ are adopted for the detection of the objects. We note that the lower and upper bounding variables of the angle and range are defined

as $\underline{N}_a, \overline{N}_a \in \{1, \ldots, N_a\}$ and $\underline{N}_r, \overline{N}_r \in \{1, \ldots, N_r\}$. Next, utilizing the defined area of interest, we detect the target object.

Object Detection: In the literature, the objects are usually detected with a constant false alarm rate (CFAR) algorithm. However, CFAR mainly determines the peaks (detection points) in the angle, range, or joint range-angle dimensions. However, in our work, as we only consider a single-vehicle scenario, the detection of multiple points and association to the transmitter is not required. Thus, simplifying the approach, we only detect a single point with the largest amplitude in the angle dimensions. To elaborate, we sum the amplitude in each range bin corresponding to an angle bin and then find the angle bin with the maximum value. Mathematically, we determine the estimated angle bin of the target object, $\tilde{o}^a$, by

$$\tilde{o}^a = \underset{a \in [\underline{N}_a, \overline{N}_a]}{\arg\max} \sum_{r=\underline{N}_r}^{\overline{N}_r} \left[\mathbf{H}_{\mathrm{RA}}\right]_{a,r}. \tag{3.14}$$

With the angle bin of the object determined, it can now be mapped to the beams.

Lookup Table based Beam Mapping: For any sample l, the angle bin of the object is determined by $\tilde{o}_l^a$ and the optimal beam takes the value $b_l^\star \in \{1, \ldots, B\}$. To map the angle bin of the object to the beams, we design a lookup table where each angle bin is matched with the most likely beam index. Specifically, for a given angle bin o, our table contains the conditional probability of the optimal beam index given the object angle estimate, i.e.,

$$\hat{b}_l^\star = \underset{b \in \{1,\ldots,B\}}{\arg\max}\, p(b | \tilde{o}_l^a = o). \tag{3.15}$$

Hence, for each $o \in [\underline{N}_a, \overline{N}_a]$, we obtain the optimal beam via (3.15), to construct our lookup table. The conditional probabilities in the equation can be estimated over the training dataset of L elements by

$$p(b = b | \tilde{o}_l^a = o) = \frac{\sum_{l=1}^{L} \mathbf{1}_{\{b_l^\star = b, \tilde{o}_l^a = o\}}}{\sum_{l=1}^{L} \mathbf{1}_{\{\tilde{o}_l^a = o\}}}. \tag{3.16}$$

This solution essentially maps the angle bin of the object in the radar sample l, $\tilde{o}_l^a$, to the beam that appears most with that angle bin within the training dataset. The complexity

of the lookup table is $\mathcal{O}(N_a)$, which allows a low-complexity beam selection mechanism that can be trained with a comparably low number of samples. Next, we present our deep learning solutions for the beam prediction problem.

Deep Learning Solution

In this subsection, we present our deep learning-based radar-aided beam prediction approach. Our solution integrates radar preprocessing and deep neural networks. This targets reducing the complexity of the learning task and enables efficient training with reasonable dataset sizes. To formalize our approach, we first decompose the radar-to-beam mapping function into three components: (i) The preprocessing function $\Psi^{\mathrm{P}}(.)$, (ii) the neural network function of the parameters $\boldsymbol{\Theta}$, $\Psi^{\mathrm{N}}_{\boldsymbol{\Theta}}(.)$, and (iii) the evaluation function $\Psi^{\mathrm{E}}(.)$. Then, we can write the radar-to-beam mapping function as

$$\Psi_{\boldsymbol{\Theta}}(\mathbf{X}) = \Psi^{\mathrm{E}}(\Psi^{\mathrm{N}}_{\boldsymbol{\Theta}}(\Psi^{\mathrm{P}}(\mathbf{X}))). \tag{3.17}$$

With the decomposition, we can define our solution in terms of the preprocessing, neural networks, and evaluation functions. In the following, we present our approach via the subsections of each function. First, we describe the proposed preprocessing approach.

Preprocessing: As the preprocessing, we separately consider all three alternatives presented in Section 3.4. Therefore, the preprocessing part of the equation is given by any of the defined radar processing approaches in (3.7)-(3.9), e.g., for the radar cube we have $\Psi^{P}(\mathbf{X}) = \Psi^{P}_{\mathrm{RC}}(\mathbf{X}) = \mathbf{H}_{\mathrm{RC}}$. Next, we present the deep neural networks adopted for each modality of the data.

Neural Network Modeling: For the neural networks, to keep the complexity of the approach low, we rely on a comparably simple deep learning model with a design with convolutional and fully-connected (FC) layers. Specifically, the deep neural networks (DNNs) comprise 8 total layers. The first five layers are the convolutional layers with the rectified

linear unit (ReLU) activation functions. In addition, the average pooling is applied after the activation of the convolutional layers to decrease the size of the data. Finally, the output of the fifth convolutional layer is connected to a set of three FC layers, providing N outputs. The each entry of the output indicates a beam.

As the proposed inputs of the neural networks are of different sizes and dimensions, the same network cannot be applied to all types of inputs. Therefore, for different modalities of the radar data, the input, output, and kernel sizes of the DNN layers are adjusted to keep the network size reasonable and similar while providing comparably good performance. Specifically, we adjust the networks for our dataset, which will be described in Section 3.7. In this dataset, the system parameters are given by $M_r = N_r = 256$, $M_c = N_c = 128$, $M_a = 4, N_a \in \{4, 64\}$ and $B = 64$. The designed DNN architectures are summarized in Table 3.3 [1] .

Neural Network Objective: To train the neural networks with the aim of finding the optimal parameters $\mathbf{\Theta}^\star$, we can write the following optimization problem that aims at minimizing the loss between the output of the network and the optimal beam values, $b_l^\star \in \{1, \ldots, B\}$:

$$\mathbf{\Theta}^\star = \arg\min_{\mathbf{\Theta}} \frac{1}{L} \sum_{l=1}^{L} \mathcal{L}\left(\Psi_{\mathbf{\Theta}}^{\mathrm{N}}\big(\Psi^{\mathrm{P}}(\mathbf{X}_l)\big), b_l^\star\right), \tag{3.18}$$

where $\mathcal{L}(.,.)$ denotes the loss function, which should be selected based on the problem type. As our problem is a multi-class classification problem, we utilize the cross-entropy loss given by

$$\mathcal{L}(\hat{\mathbf{n}}, \mathbf{n}) = -\frac{1}{B} \sum_{b=1}^{B} n_b \log(\hat{n}_b), \tag{3.19}$$

where $\mathbf{n} = [n_1, \ldots, n_B]$ is the one-hot encoded vector of the optimal beam $b_l^\star$ and $\hat{\mathbf{n}} = \Psi_{\mathbf{\Theta}}^{\mathrm{N}}(\Psi^{\mathrm{P}}(\mathbf{X}_l))$ is the output of the neural network. The elements of the one-hot encoded

[1] The additional $N_a = 4$ point angle FFT is only applied for the range-angle maps. For the radar cube, only $N_a = 4$ point angle FFT is applied to keep the input size of the different data modalities reasonably similar.

Table 3.3: Deep Neural Network Architectures for Different Input Types.

NN Layers	**Radar Cube** ($\mathrm{H_{RC}}$)	**Range-Doppler** ($\mathrm{H_{RV}}$)	**Range-Angle-64** ($\mathrm{H_{RA}}$)	**Range-Angle-4** ($\mathrm{H_{RA}}$)
Input	$4 \times 256 \times 128$	$1 \times 256 \times 128$	$1 \times 256 \times 64$	$1 \times 256 \times 4$
CNN-1	Output Channels: 8, Kernel: (3, 3), Activation: ReLU			
CNN-2	Output Channels: 16, Kernel: (3, 3), Activation: ReLU			
AvgPool-1	Kernel: (2, 1)		N/A	
CNN-3	Output Channels: 8, Kernel: (3, 3), Activation: ReLU			
AvgPool-2	Kernel: (2, 2)			Kernel: (2, 1)
CNN-4	Output Channels: 4, Kernel: (3, 3), Activation: ReLU			
AvgPool-3	Kernel: (2, 2)			Kernel: (2, 1)
CNN-5	Output Channels: 2, Kernel: (3, 3), Activation: ReLU			
AvgPool-4	Kernel: (2, 2)			Kernel: (2, 1)
FC-1	Input Size: 512, Output Size: 256, Activation: ReLU			
FC-2	Input Size: 256, Output Size: 128, Activation: ReLU			
FC-3	Input Size: 128, Output Size: 64			

vector $\mathbf{n}$ are defined by $n_b = 1$ if $b = b^\star$ and 0 otherwise. Using the defined loss function, the neural network can be trained with back-propagation through the layers. We note that by the construction of proposed neural networks, the DNN models return the soft information, $\hat{\mathbf{n}} \in \mathbb{R}^B$, which needs to be converted to the beam indices.

Evaluation: To evaluate the output of the neural network in terms of the accuracy-based objective function in (3.11), we need to select a single beam from the soft output of the neural network. For this purpose, the maximum of the neural network output can be selected as the prediction of the optimal beamforming vector. This can be mathematically provided by setting $\Psi^{\mathrm{E}}(.) = \arg\max(.)$, completing our solution. In the following section, we describe our dataset adopted in the training and evaluation of the solutions.

3.6 Radar-aided Beam Tracking

3.6.1 Problem Formulation

The beam prediction problem defined in Section 3.5 aims to utilize only the most recent measurement to predict the immediate beam. However, it does not take advantage of the previous measurements and it does not attempt to predict future beams. In this section, we extend the beam prediction problem to track the objects through time and predict their beams in the future time slot, which we call *beam tracking*. This can enable more accurate beam management results due to two reasons: (i) Exploiting the correlation between consecutive measurements in time can make the detection approaches less sensitive to this noise, which improves the prediction accuracy. (ii) Considering sequences instead of separate frames allows the system to extract the object's velocity information (with respect to the global coordinates), which can be utilized for predicting future beams more accurately.

With this motivation, we extend our beam prediction framework to consider a sequence of current/previous radar measurements and to predict not just current but also future beams. Mathematically, we define the beam tracking problem as the mapping from the current/previous measurements to the current or future beams. We use $\boldsymbol{\mathcal{X}}_l = \{\mathbf{X}_{l-T_o+1}, \ldots, \mathbf{X}_{l-1}, \mathbf{X}_l\}$ to denote the available set of radar measurements at time l through an observation interval of $T_o \in \mathbb{Z}^+$ samples.

To generalize the time of the predicted beam, we define the prediction delay $T_p \in \mathbb{Z}$ samples, which captures the time difference between the current beam and the to-be-predicted beam. We define a new mapping function, by generalizing (3.10) to map the latest T_o radar observations to the optimal beamforming index at time $t + T_p$ by

$$\Phi_{\boldsymbol{\Theta}}(\boldsymbol{\mathcal{X}}_l) = b^{\star}_{l+T_p}. \tag{3.20}$$

We note that $T_o = 1$ and $T_p = 0$ correspond to our beam prediction function and problem

formulation in Section 3.5. Now, we revisit the problem definition for the beam tracking problem. Similar to the beam prediction problem defined in Section 3.5, our objective is to design/learn the generalized mapping function $\Phi_{\boldsymbol{\Theta}}$ to be able to map the set of latest T_o radar measurements to the optimal beam index $b^\star \in \{1, \ldots, B\}$ at time $l + T_p$. For this objective, we explore the designs of the mapping function and learn the set of parameters $\boldsymbol{\Theta}$. Mathematically, we can express this objective by

$$\Phi^\star_{\boldsymbol{\Theta}^\star} = \arg\max_{\Phi_\Theta} \frac{1}{L} \sum_{l=1}^{L} \mathbf{1}_{\{b^\star_{l+T_d} = \Phi_{\boldsymbol{\Theta}}(\{\mathbf{X}_{l-T_o+1}, \ldots, \mathbf{X}_l\})\}}. \tag{3.21}$$

With this objective, in the following section, we extend our solutions for the beam prediction for the beam tracking problem.

3.6.2 Proposed Solutions

Building upon our beam prediction solutions given in Section 3.5.2, we propose two new (extended) solutions for the beam tracking framework: (i) An extended Kalman filter-based tracking method using the object detection solution in Section 3.5.2 and a support vector machine (SVM), [60], to predict the beams, and (ii) a long-short term-memory (LSTM) based machine learning solution that is coupled with the previously proposed deep learning models in Section 3.5.2. These two solutions capture the tracking idea (and temporal dependencies) with (i) a Kalman filter and (ii) an LSTM model. In the following two subsections, we present the details of our solutions.

Conventional Signal Processing Solution

First, we develop a radar-aided beam tracking approach based on conventional radar signal processing. Specifically, the proposed framework can be summarized as follows: (i) The target object is detected, and its attributes (range and angle) are determined. (ii) An extended Kalman filter is utilized to track the estimated attributes across the radar measure-

ment sequence and estimate the target's current state (position). (iii) The tracked state of the target is mapped to the beams using an SVM classifier. Next, we present the various stages of the proposed Kalman filtering-based solution in more detail.

Object Detection: For the detection of the object, we adopt the same approach proposed in Section 3.5.2. In particular, we first determine the angle bin of the target within the area of interest by (3.14). Then, given the angle bin of the target, $\tilde{o}_l^a$, we estimate the range bin of the target by finding the maximum amplitude in that angle slice of the range-angle map. It can mathematically be written as

$$\tilde{o}^r = \underset{r \in [\underline{N}_r, \overline{N}_r]}{\arg\max} \left[\mathbf{H}_{\mathrm{RA}}\right]_{\tilde{o}^a, r} . \tag{3.22}$$

With the range and angle of the target bins determined, they can be converted to the corresponding angle and range values, o^a and o^r, based on the FFT size and the radar parameters. Specifically, $o^a = \arcsin(\frac{2\tilde{o}^a}{N_a} - 1)$ and $o^r = \tilde{o}^r d_r$ with $d_r = \frac{c}{2\mathsf{BW}}$ being the range resolution of the radar.

Extended Kalman Filter: The singular observations can be noisy and include other inaccuracies. Thus, their utilization over time can produce more accurate estimations of the states. To that end, the Kalman filter generates a robust estimation of the unknown states given a set of observations over time. For a Kalman filter, two models are required, i.e., the state-transition model and the observation model. In this chapter, we adopt the position and velocity of the object as the state, and the observations are the angle and range of the target object extracted from the radar measurements.

For the state transition model, we adopt the constant velocity model [61] as our system model only considers a straight vehicle movement within a lane. In this constant velocity model, the object's state is given as the Cartesian position and velocity of the object. Mathematically, we define the state of the object by

$$\mathbf{s}_l = \begin{bmatrix} o_l^x & o_l^y & o_l^{v_x} & o_l^{v_y} \end{bmatrix}, \tag{3.23}$$

where o_l^x, o_l^y represent the position of the object in x and y axes, and $o_l^{v_x}, o_l^{v_y}$ denote the velocity in a similar manner. The target's position is updated by the distance covered by the object during a time slot duration, d_t, with the velocity in each corresponding axis. As indicated by the name (constant velocity), the object's velocity is assumed as a constant with the update. Mathematically, the underlying state update model is written as

$$\mathbf{s}_{l+1} = \mathbf{A}\mathbf{s}_l + \mathbf{w}_l, \tag{3.24}$$

where $\mathbf{A}$ is the state update matrix given by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & d_t & 0 \\ 0 & 1 & 0 & d_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.25}$$

and $\mathbf{w}_l$ is the process noise.

For the observation model, we consider the range and angles obtained by the object detection methodology described previously. The measurement at step l can be defined by $\mathbf{z}_l = \begin{bmatrix} o_l^r & o_l^a \end{bmatrix}$. Then, the transformation of the states to the observations (angle and range) is given by

$$o_l^a = \arctan \frac{o_l^x}{o_l^y} \quad \text{and} \quad o_l^r = \sqrt{(o_l^x)^2 + (o_l^y)^2}. \tag{3.26}$$

As these transformation equations are non-linear, we adopt the EKF utilizing the Taylor series approximation of the observation equations. Specifically, the Hessian matrix is given by

$$\mathbf{D}_l = \begin{bmatrix} \frac{o_l^x}{o_l^r} & \frac{o_l^y}{o_l^r} & 0 & 0 \\ \frac{o_l^y}{(o_l^r)^2} & -\frac{o_l^x}{(o_l^r)^2} & 0 & 0 \end{bmatrix}. \tag{3.27}$$

With the described state transition and observation models, the update equations of the EKF are applied as described in Table 3.4 for the estimation of the states.

Table 3.4: Extended Kalman Filter Update Equations.

Prediction Variable	Equation
State Estimate	$\mathbf{s}_{l\|l-1} = \mathbf{A}\mathbf{s}_{l-1}$
State Covariance	$\mathbf{B}_{l\|l-1} = \mathbf{A}\mathbf{B}_{l-1}\mathbf{A}^T + \mathbf{Q}$
Update Variable	**Equation**
Residual	$\mathbf{z}'_l = \mathbf{z}_l - \mathbf{D}_l\mathbf{s}_{l\|l-1}$
Kalman Gain	$G = \mathbf{B}_{l\|l-1}\mathbf{D}_l^T(\mathbf{D}_l\mathbf{B}_{l\|l-1}\mathbf{D}_l^T + \mathbf{R})^{-1}$
State Estimate	$\mathbf{s}_l = \mathbf{s}_{l\|l-1} + G \cdot \mathbf{z}'_l$
State Covariance	$\mathbf{B}_l = (\mathbf{I} - G \cdot \mathbf{D}_l^T)\mathbf{B}_{l\|l-1}$

Support Vector Machine Beam Classification: After the state of the vehicle (position and velocity) is estimated through the EKF, it can be utilized for the beam prediction. For this purpose, we use a support vector machine (SVM) algorithm as the classification solution. SVM aims to obtain the hyperplane separating the data points of different classes with a maximum margin between them. In this solution, the SVM is selected to map the position and velocity to the beam indices for the prediction of the beams. The SVM complements the earlier part of the approach well since, in this approach, a range and velocity region is determined for each beam. Moreover, it can be considered a robust (due to the data-centric approach) extension to the constant velocity model.

Deep Learning Solution

In this subsection, we propose our deep learning solution for the beam tracking problem. Different from the beam prediction task, the neural network needs to utilize a set of measurements rather than a single measurement and also take advantage of the temporal dependency between the samples. For this purpose, we adopt a neural network combining CNN and LSTM. The CNN-LSTM solutions have been successfully utilized in more general

learning, and radar problems [62, 63]. To this end, it is a well-fitting solution to the problem. In this solution, the CNN aims to extract the essential features to be tracked through time via the LSTM cells. The LSTM cells follow the information from the different samples through time. Then, the output of the LSTM cells needs to be converted to the beam prediction, done through a set of linear layers. We detail our approach in the following.

Feature Extraction: As the first step of the deep learning solution, we aim to extract the essential features of the radar maps. The feature extraction is beneficial because (i) the large size of the input data carries a set of irrelevant properties to the beam prediction, and (ii) the complexity of the LSTM can be reduced with the tracking of essential features. For the defined problem of beam tracking, we mainly aim to track the vehicle and corresponding beams through time. The other elements in the radar maps are mainly irrelevant to the problem. Therefore, feature extraction reduces the complexity of tracking the features over time without any extra performance cost.

For the extraction of the features, we utilize the neural networks proposed in Section 3.5.2. In particular, we adopt the deep neural network utilized in the beam prediction problem for the corresponding maps, as given in Table 3.3. However, these neural networks aim to return only B (number of available beams) dimensional output. This output may not be sufficient for the features to be fed to the LSTM for tracking over time. To this end, we remove the last two layers of the architectures presented in Table 3.3 and use them as feature extraction architectures. Then, the output of this network is connected to the LSTM cells, as will be explained next.

LSTM networks: The LSTM architectures are meaningful solutions in extracting dependency among different samples in time or frequency by increasing the performance of the neural networks [64]. In the beam tracking problem, the target vehicles (or the beams corresponding to these vehicles) are aimed to be tracked. Hence, the LSTM presents a natural way for a desirable solution to the beam tracking problem.

The LSTM networks consist of a set of consecutive LSTM cells, each connected to the next one. Each cell takes the input of the corresponding sample (e.g., the second time sample in the input set is given to the second cell) and the output of the previous cell and returns an output. The first cell only takes the input sample, as there are no preceding cells. In LSTMs, the input and output sizes of the cells are the same. As the output of the LSTM architecture, we take the latest cell's output, which utilizes all the information transferred through the previous cells.

Beam classification: After the input is passed through the feature extraction CNN and LSTM layers, the output has the dimensionality of the extracted features. However, this value does not necessarily equal the number of available beams. Therefore, to make the final prediction of the beams, we utilize a set of linear layers. Specifically, we adopt two linear layers, the first returning half the number of features and the second returning an output of size equal to the number of available beams. This is then passed through the evaluation layer (argmax function) as in Section 3.5.2, and the output is taken as the prediction of the beams.

Training: The training procedure is the same as the deep learning solution for the beam prediction problem. We note that the whole network architecture is trained jointly with the same cross-entropy objective function, as in Section 3.5.2, with the change of the aimed beam index belonging to the time $l + T_p$ and input being the set of radar maps.

3.7 Real-World Dataset

To accurately evaluate the performance of the proposed machine learning-based radar-aided beam prediction approaches in a realistic environment, we built a real-world dataset with radar and wireless measurements. In this section, we describe our testbed and present the dataset collection scenario.

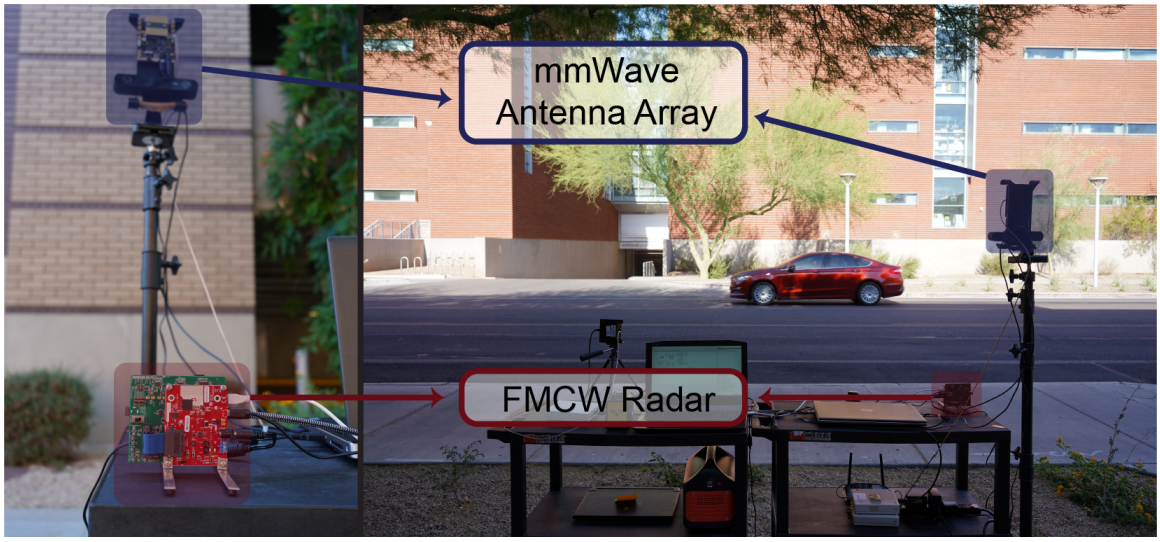


Figure 3.4: The Figure Illustrates the Data Collection Setup, Which Consists of a Base Station Unit and a Mobile Unit (Vehicle). The Base Station Is Equipped With a mmWave FMCW Radar and a mmWave Antenna Array. The Car Is Equipped With an Omni-Directional mmWave Transmitter. The Right Part Shows a Front View, and the Left Part Shows a Back View With a Mobile User (Vehicle) on Sight.

3.7.1 Data Collection Testbed

We adopt the Testbed-1 of the DeepSense 6G dataset[52]. The testbed comprises two units: (Unit 1) The stationary unit, and (Unit 2) the mobile unit. Among other sensors, unit 1 employs an FMCW radar (AWR2243BOOST) which has three transmitter and four receiver antennas, and a mmWave receiver at 60 GHz, which adopts a uniform linear array (ULA) with $M_a^c = 16$ elements. The unit 2 utilizes a 60 GHz quasi-omni antenna, acts as a transmitter, and is always oriented towards the receiver antenna of unit 1. The setup of unit 1 is shown in Figure 3.4 (left), where the receiver antenna array and FMCW radar board are placed in close proximity.

The phased array of unit 1 utilizes an over-sampled beamforming codebook of $B = 64$ vectors, which are designed to cover the field of view. It captures the received power by applying the codebook beams as a combiner. The combiner providing the most power is taken as the optimal beamforming vector. For the radar, we only activated one of the TX an-

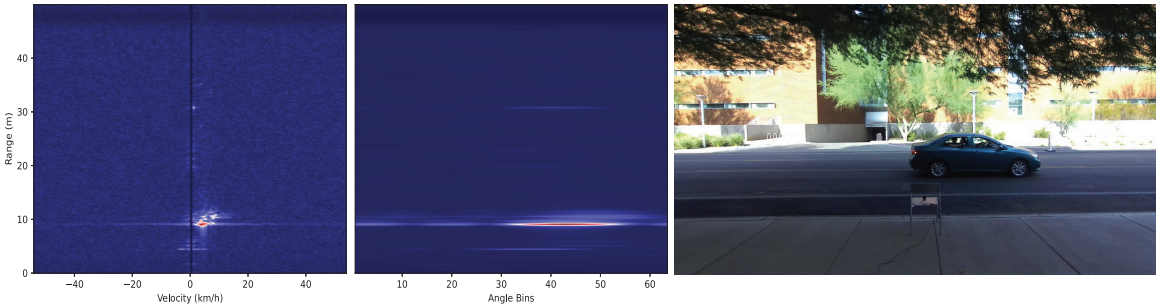


Figure 3.5: A Sample From the Dataset Is Shown With the Current Environment Image (Right) and the Corresponding Range-Angle (Middle) and Range-Velocity (Left) Images. The Car Is on the Right Part of the Camera Image, Moving Away From the Vertical Angle of the Radar Device. The Range-Angle Image Shows the Position at Approximately 9m Distance on the Right-Hand Side, While the Range-Velocity Image Indicates the Increasing Relative Velocity and Range.

tennas, while the data from $M_a = 4$ RX antennas were captured. We adopted a set of radar parameters based on the TIs short range radar (SRR) example, given by ADC sampling rate $f_s = 5$ us, $\mu = 15$ MHz/us, $M_c = 128$ chirps/frame, $M_s = 256$ samples/chirp. This configuration approximately provides [56, 57] the bandwidth $\mathsf{BW} = 768$ MHz, the maximum range of 50m, and the maximum velocity of 54 km/h, which are well-fit for the scenario illustrated in Figure 3.4 (right). Please refer to the data collection testbed description in [52] for further details. Next, we detail the collection scenario and dataset.

3.7.2 Development Dataset

To generate our dataset for the evaluation, we used the testbed described in Section 3.7.1 and adopted Scenario 9 of the DeepSense 6G dataset [52]. In this scenario, a passenger in the back seat of the car holds the transmitter. As shown in Figure 3.5, the car passes by the stationary unit (Unit 1), which collects the radar and beam training measurements. During the data collection, the road was actively used by other cars, pedestrians, and bikers. Our testbed collected the radar measurements and the received power with each communication beam.

In the construction of the dataset, the beam providing the highest power is saved as the

optimal beamforming vector. The data is cleaned by only keeping the samples with the target car in sight. This cleaning operation is performed manually by inspecting the RGB images captured from a camera attached next to the antenna array. The data samples with the other elements (cars on the other lane, pedestrians, and bikers) are also kept to reflect the realistic environment. The final dataset comprises 6319 samples, which are separated with a $70/20/10\%$ split for the training, validation and testing. A sample from the dataset is illustrated in Figure 3.5, with the extracted range-angle and range-Doppler images.

3.8 Results

In this section, we evaluate [2] and compare the performance of the proposed solutions. In particular, we separately evaluate the classical and machine learning-based beam prediction and beam tracking approaches. In this evaluation, we adopt our dataset described in Section 3.7.

3.8.1 Beam Prediction

We first evaluate the beam prediction solutions. In particular, we compare the DNN-based solutions adopting the radar cube, range-velocity, and range-angle maps, respectively, and the classical approach that adopts a lookup solution mapping the beams to the detected angle bins. These different solutions are compared in terms of their prediction accuracy, complexity/inference time, and the required dataset sizes.

Classical Solution: For this approach, we first determine the area of interest. With the given scenario, all the available angle bins are utilized in the problem since the basestation is placed parallel to the road, i.e., $\underline{N}_a = 1, \overline{N}_a = N_a$. However, for the range bins, the target car only passes through the closer lane, and any other object in a different area may cause

[2]The numerical computations are carried out on a server with an Intel Xeon Silver 4216 processor with an Nvidia RTX Titan GPU. The code is written with the standard scientific Python libraries and PyTorch.

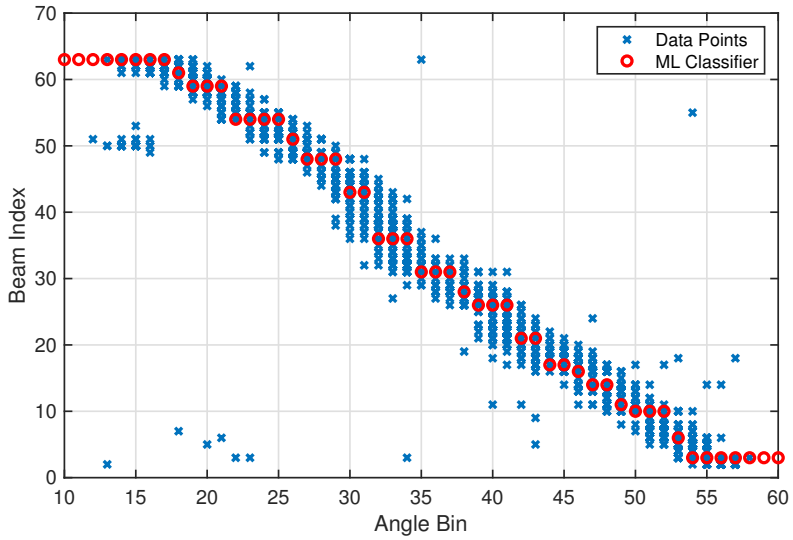


Figure 3.6: The Detected Angle Bin and the Optimal Beam Indices of the Training Samples Are Shown With the Lookup Table (Red) Entries.

the object detection to fail. To that end, the range bins are determined as $\underline{N}_r = 35, \overline{N}_r = 70$, based on the inspection of the detected points on the map. In Figure 3.6, the angle bins and optimal beams of the training data points are shown along with the lookup table. As shown in the figure, there are multiple angle bins corresponding to each beam. This is potentially due to the comparably low angular resolution of the noisy radar measurements.

Deep Learning Training and Evaluation: For the evaluation of the neural networks, we trained the DNN models summarized in Table 3.3 using the Adam algorithm [65] with a learning rate 0.001, batch size 32, and a decay factor $\gamma = 0.1$ applied every 10 epochs. The networks are trained for 40 epochs, and the network parameters showing the best top-1 accuracy over the validation dataset are saved for evaluation. The network training operation is carried out for 5 separate instances, and the average performance is shown in the following results. Next, we compare the beam prediction accuracy of the solutions.

Beam Prediction Accuracy: We first compare the top-1 prediction accuracy of the solutions. As shown in Figure 3.7, the range-angle map-based deep learning solutions over-perform the range-velocity and radar cube solutions. Compared to the deep learning

solutions, the classical solution with an accurately selected range of interest (RoI) shows a slightly inferior performance. However, the simple detection mechanism with the full range-angle maps shows a high degradation in the performance, potentially due to the detection errors with the other objects in the environment. The deep learning solutions achieve better performance without supervision to accurately select the parameters, which shows the robustness and applicability of the solution to the real-world data. Among the deep learning solutions, the range-Doppler maps show comparable performance to the other deep learning solutions. This is potentially due to two reasons: (i) The Doppler velocity is in the velocity towards the basestation, and hence, it contains angular information. (ii) Although the right-to-left and left-to-right movement with respect to the basestation cannot be distinguished in the range-Doppler maps, the traffic flow allows it to be determined. The vehicles moving left-to-right are on the closer lane and separated by distance. Moreover, the scenario only contains the target car moving from left to right as described in Section 3.7. Finally, the radar cube contains the range-angle maps of different Doppler velocity values, however, it cannot perform similarly to the 4-point range-angle solution. This is potentially due to the large size and complexity of the input and the comparative simplicity of the deep learning model. It could be possible to over-perform both solutions with the radar cube adopting more data and higher complexity neural networks, however, it would be computationally prohibitive.

The top-3 and top-5 accuracy of the range-angle images reach up to 79.7% and 93.5%, outperforming the classical solution by a more significant margin. Furthermore, all the approaches show improvements similar to the range-angle-based deep learning solutions with increasing K values of the top-K predictions. Based on the presented results, we conclude that the deep learning solutions show clear potential for real applications, especially with the top-3 and top-5 results. Moreover, the comparison of the range-angle solutions with different angle FFT sizes shows the advantage of the maps generated with higher resolu-

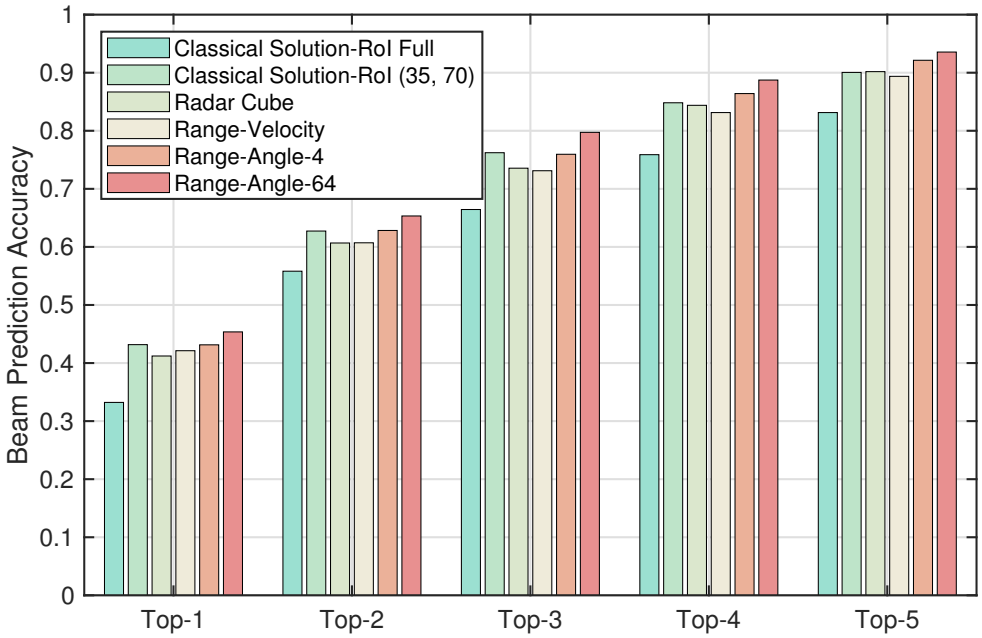


Figure 3.7: The Top-k Test Accuracy of the Proposed Approaches. Range-Angle Map-Based Solutions Outperform the Other Solutions, While the Deep Learning Solutions Slightly Outperform the Classical Approach. The Determination of the Range of Interest Significantly Impacts the Performance, Showing the High Dependence on the Accurate Detection of the Objects.

tion. This is expected to be the case with the baseline solution, however, the behavior is also prevalent in the deep learning solution, which shows the potential advantages of the more accurate preprocessing approaches.

Complexity: In Figure 3.8, we compare the complexity of the DNNs in terms of the number of parameters, preprocessing, and network inference durations [3] . First, in the design of the DNN models, we aimed to keep the number of parameters similar. To that end, the radar cube, range-velocity, and range-angle-based solutions comprise approximately 175k parameters. The classical solutions only require $64 \times K$ parameters for 64-point angle FFTs and top-K evaluation. These values correspond to the top-K beams of each angle

[3]The DNNs are run on the GPU while the preprocessing is applied by the CPU. Therefore, they are not one-to-one comparable with each other. The durations scale based on the corresponding specialized hardware separately.

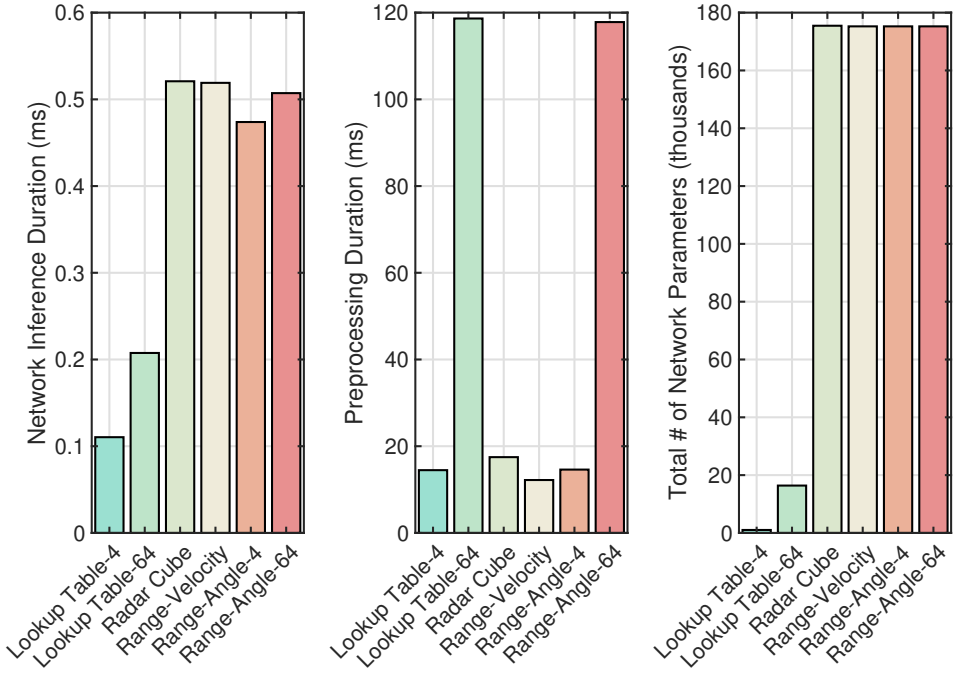


Figure 3.8: The Complexity of the Proposed Radar-Aided Beam Prediction Approaches Are Compared in Terms of the Network Inference Time, the Preprocessing Duration, and the Number of (Neural Network) Parameters. The Generation of High-Resolution Range-Angle Maps Brings Significantly Higher Complexity.

bin. Second, we compare the inference duration of the neural networks as shown on the left part of Figure 3.8. As one can expect, the deep learning solutions show similar inference durations since the number of parameters and network architectures are designed similarly. Third, the central figure in Figure 3.8 illustrates the radar preprocessing durations, where the larger angle FFT adopted in the range-angle maps requires a significant additional duration. This presents a trade-off between the beam prediction accuracy and the complexity of the preprocessing solution. Depending on the hardware availability, one may prefer to design a solution with higher-resolution maps and better accuracy. Without an oversampling of the angle FFT dimension, all the approaches show similar durations, presenting an advantage for better-performing solutions.

Impact of Dataset Size: To draw some insights about the required dataset size, we consider different percentages of the training dataset to train the neural network models. In

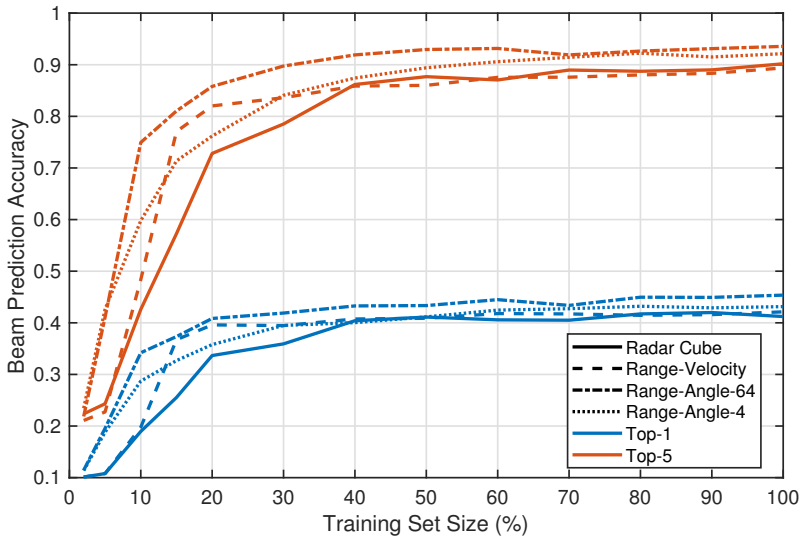


Figure 3.9: Top-1 and Top-5 Beam Prediction Accuracies of the Proposed Radar-Aided Beam Prediction Approaches When a Fixed Percentage of the Training Dataset Is Utilized.

Figure 3.9, we show the average accuracy of the trained networks on the same test samples while only a subset of the training samples is utilized. The figure shows top-1 and top-5 accuracy values. The lines of the same input for different K values show similar behavior with different scaling and accuracy levels. In the figure, the accuracy of the 64-point range-angle map-based solution increases steeper than the others, reaching better accuracy. The radar cube shows the slowest initial increase, possibly due to the larger input size, and eventually reaches similar accuracy to the range-velocity solution. The performance of the 64-point range-angle solution requires $10 - 20\%$ of the data for starting to saturate, while the other solutions need around $20 - 30\%$ of the training data. This might be due to the more straightforward interpretability of the high-resolution range-angle maps for the beam prediction task, potentially requiring less learning and transformation. The radar cube starts to saturate particularly late, and its accuracy may slightly increase with more data, which indicates a potential benefit from a larger dataset or more complex models. Nevertheless, the range-angle and range-Doppler map-based solutions can perform well with the presented low complexity and smaller datasets.

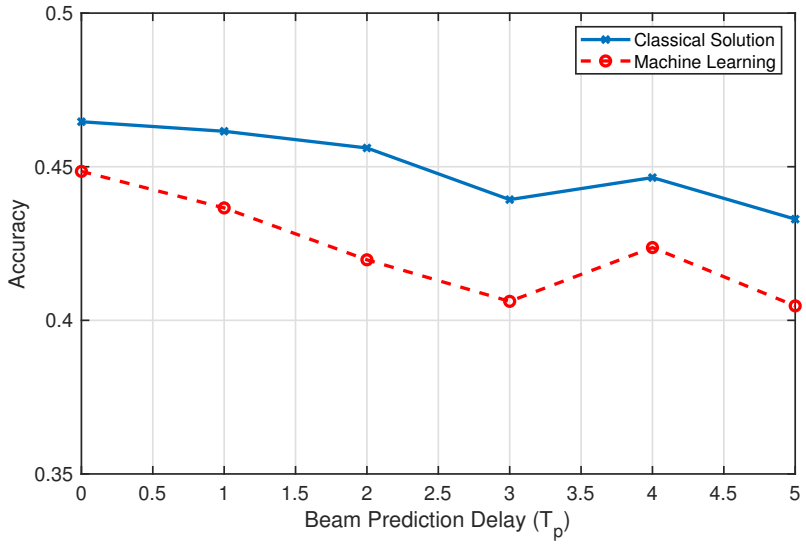


Figure 3.10: The Top-1 Beam Prediction Accuracy of the Classical and Machine Learning Solutions Proposed in Section 3.6.2. The Classical Solution Slightly Outperforms the ML Solution, While It Requires a Significant Effort in the Model Design and Its Tuning.

3.8.2 Beam Tracking

Next, we evaluate the performance of the beam-tracking solutions. In particular, the EKF-based tracking solution and the LSTM-based machine learning solutions are compared in terms of performance. For the EKF solution, the target vehicle is needed to be tracked within time samples, however, it is not necessarily limited to the T_o samples defined for the problem. To this end, for the classical solution, we consider $T_o = \infty$, utilizing all the relevant previous samples. Differently, the machine learning solution requires a fixed input size, possibly limited due to the complexity. For that, we select $T_o = 8$ radar maps to design the input to the machine learning model. As the beam prediction results have shown the higher accuracy of the range-angle maps, we only adopt them as the input of the LSTM model.

Beam Tracking Accuracy: The accuracy of the solutions are shown in Figure 3.10 for different values of the prediction delay. Interestingly, the classical solution offers $2-3\%$ better performance over different beam prediction delay values. This may be caused by the

comparably low number of samples in the dataset for the LSTM model. Meanwhile, the better performance of the classical solution indicates the well-fit of the constant velocity model. However, note that the classical beam tracking approach also utilizes a comparably simple machine learning classifier (SVM), which provides robustness to the final predictions. This may possibly eliminate some of the errors caused by the model by relying on the data samples. Both approaches for $T_p = 0$ present improvements over the beam prediction solutions, indicating the advantages of utilizing radar samples from a longer duration of time, although with the complexity disadvantage in the case of the LSTM solution ($\sim 640k$ parameters). Moreover, the LSTM solution is required to be run in series instead of parallel since the information from the previous samples is carried over to the following LSTM cells. Based on the results, it can be stated that if a movement model can fit well for the scenario, a classical solution may be preferable. However, as it requires fine-tuning of the parameters, deep learning models can be preferable thanks to their simplicity.

3.9 Summary

In this chapter, we developed and demonstrated, for the first time, the feasibility of radar-aided mmWave beam prediction approaches with real-world datasets. Later, we extended the initial beam prediction formulation and solutions to track the beams over time with a beam-tracking framework. The developed solutions leveraged deep neural networks and domain-knowledge radar processing to increase the beam prediction/tracking accuracy and reduce the inference/complexity overhead. We then evaluated the performance of the proposed solutions based on the real-world dataset DeepSense 6G, which comprises co-existing modalities including radar and mmWave beam training measurements. This real-world evaluation demonstrates the feasibility of radar-aided mmWave beam prediction and highlights its promising gains in enabling highly mobile mmWave/sub-THz communication applications.

Chapter 4

RADAR-AIDED BLOCKAGE PREDICTION IN REAL-WORLD

4.1 Overview

Millimeter wave (mmWave) and sub-terahertz communication systems rely mainly on line-of-sight (LOS) links between the transmitters and receivers. The sensitivity of these high-frequency LOS links to blockages, however, challenges the reliability and latency requirements of these communication networks. In this paper, we propose to utilize radar sensors to provide sensing information about the surrounding environment and moving objects, and leverage this information to proactively predict future link blockages before they happen. This is motivated by the low cost of the radar sensors, their ability to efficiently obtain important features such as the range, angle, velocity of the moving scatterers (candidate blockages), and their capability to capture radar frames at relatively high speed. We formulate the radar-aided proactive blockage prediction problem and develop two solutions for this problem based on classical radar object tracking and deep neural networks. The two solutions are designed to leverage domain knowledge and the understanding of the blockage prediction problem. To accurately evaluate the proposed solutions, we build a large-scale real-world dataset, based on the DeepSense framework, gathering co-existing radar and mmWave communication measurements of more than 10 thousand data points and various blockage objects (vehicles, bikes, humans, etc.). The evaluation results, based on this dataset, show that the proposed approaches can predict future blockages 1 second before they happen with more than 90% F_1 score (and more than 90% accuracy). These results, among others, highlight a promising solution for blockage prediction and reliability enhancement in future wireless mmWave and terahertz communication systems.

4.2 Introduction

Future wireless networks attempt to meet the increasing demand on high data rates, low latency, and high reliability. More extensive usage of the higher frequency bands, millimeter-wave (mmWave) and sub-terahertz (sub-THz), is one prominent direction [32, 66] for satisfying the high data rate demands. However, the propagation characteristics at these frequencies result in two important features for mmWave/sub-THz communication systems: (i) These systems rely mainly on line-of-sight (LOS) links to guarantee sufficient receive signal power, and (ii) this dependency on LOS links coupled with the high penetration loss at mmWave/THz bands make these communication systems very sensitive to blockages. In particular, if these links are blocked, for example by moving objects, this could cause sudden performance degradation or even a link disconnection, which highly challenges the reliability and latency of these networks. This motivates the research for approaches that overcome the blockage challenges in high-frequency (mmWave/sub-THz) wireless networks.

Given that the LOS link blockages depend heavily on the positions of the communication terminals and the blockages in addition to their geometric properties (size, height, etc.), sensing the environment could potentially provide useful information for proactively predicting link blockages before they happen [67]. Predicting future blockages enables the wireless network to make proactive decisions such as proactively handing off the user to another basestation or proactively switching to another beam. In [67], beam sequences and recurrent neural networks were leveraged to predict future stationary blockages, and in [68], the sub-6GHz channels were used to infer mmWave blockages. To enable dynamic (moving) blockage prediction, [69] showed that the moving blockages produce *pre-blockage wireless signatures* that can potentially be leveraged for proactive blockage prediction. This approach, though, is mainly useful for scenarios with a few moving objects. To overcome

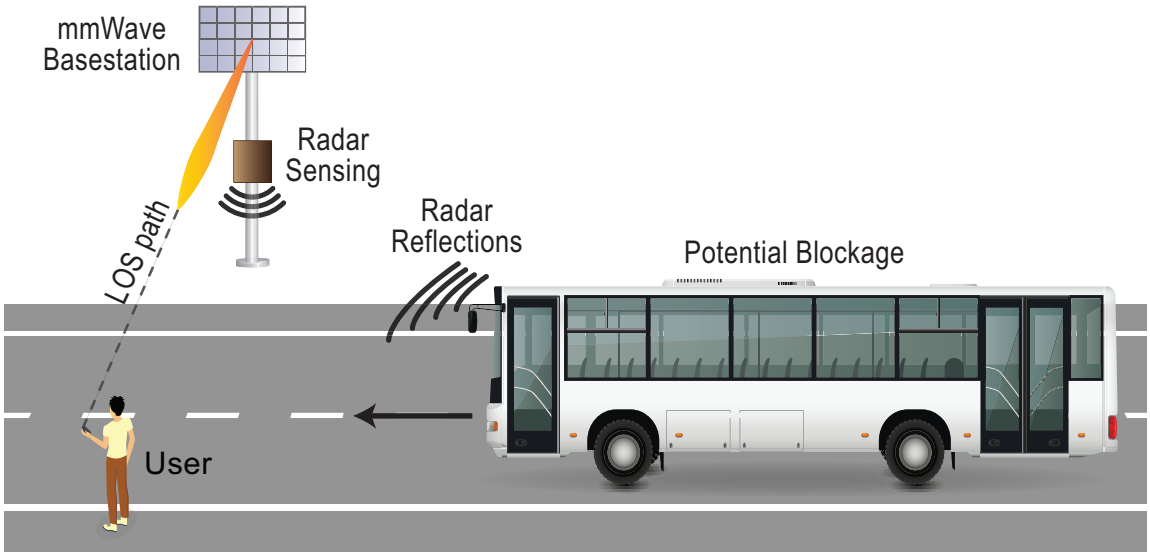


Figure 4.1: An Illustration of the Adopted System Model. The mmWave LOS Communication Between the User and Basestation Is About to Be Interrupted by the Moving Bus Due to the Potential Blockage of the LOS Path.

these limitations and enable a more scalable solution, [70] proposed to leverage red-green-blue (RGB) camera data which provide rich information about the moving objects in the surrounding environment. While cameras are relatively simple to deploy [48], their usage is sometimes associated with privacy concerns and its operation may be limited scenarios with low light or bad weather conditions. Another promising approach is to use LiDAR sensory data, as proposed in [71]. Using LiDAR, however, is mainly suitable for low-range scenarios and is associated with relatively high cost.

In this paper, we propose to leverage radar sensors to obtain useful information about the moving objects (candidate blockages) in the surrounding environment and use this information for proactive mmWave blockage prediction. The use of radar sensors (such as frequency-modulated continuous-wave (FMCW) radars) is motivated by (i) their off-the-shelf availability at relatively low-cost, (ii) their capability to measure velocity (Doppler) in addition to range and angle, (iii) their potential high-frequency low-latency measurements and (iv) the lack of privacy concerns with radar sensory data, and (v) the possibility to

be integrated with communication systems [44, 45]. With this motivation, we formalize the radar-aided proactive blockage prediction problem and present two alternative solutions: One is based on classical signal processing for radar object tracking and one that leverages deep neural networks. In the first approach, we develop a radar object tracking algorithm based on Unscented Kalman filter [72] whose states are later utilized with a k-nearest neighbors classification algorithm for blockage prediction. In the second approach, we utilize the range-angle maps obtained from the radar measurements and develop a deep neural network based solution to predict future blockages.

To evaluate the performance of the proposed solutions, we build a large-scale real-world dataset for an outdoor scenario, based on the DeepSense framework [52], with coexisting radar and mmWave communication data. The constructed dataset comprises around 10 thousand data points from more than 300 unique blockage trajectories including vehicles, bikes, and humans. With this dataset, our evaluation results show that the deep learning approach gives promising blockage prediction accuracy gains with the advantage of design simplicity. For example, the results indicate that the developed solutions could predicted future blockages 1 second before they happen with more than 90% F_1 score (and more than 90% accuracy). These results, among others, highlight a promising solution for blockage prediction and reliability enhancement in future wireless communication systems.

The rest of the paper is organized as follows. In Section 4.3, the system model is presented. Building upon the system model, Section 4.4 formulates the radar-aided blockage prediction problem. Section 4.5 and Section 4.6 propose the object tracking and the deep learning solutions, respectively, for the radar-aided blockage prediction problem. For the evaluation, Section 4.7 details the experimental setup and real-world dataset, and Section 4.8 presents the numerical results utilizing this dataset. Finally, the paper is concluded in Section 4.9.

4.3 System Model

The considered system comprises a basestation communicating with a stationary user. The basestation adopts two main components: (i) A mmWave communication transceiver equipped with a phased array communicating with the stationary user and (ii) an FMCW radar. The radar is utilized to sense the environment and predict the LOS link blockages that are caused by the objects (e.g., vehicles) moving between the base station and user. An illustration of the system model is shown in Figure 4.1. In the following two subsections, we briefly describe the system and signal models of the adopted communication and radar components.

4.3.1 Radar Model

In our system, the basestation is equipped with an FMCW radar. The radar device provides measurements for the communication environment around the base station, which could be leveraged for predicting future blockages. The radar captures one measurement every τ_f seconds. In each measurement, the FMCW radar transmits a frame of L chirps. Each chirp has a linearly increasing frequency starting at an initial frequency f_c and ending at a stop frequency $f_c + \mu t$, given by

$$s^{\text{tx}}_{\text{chirp}}(t) = \begin{cases} \sin(2\pi[f_c\, t + \frac{\mu}{2}\, t^2]) & \text{if } 0 \leq t \leq \tau_c, \\ 0 & \text{otherwise}, \end{cases} \tag{4.1}$$

where $\mu = B/\tau_c$ is the slope of the linear chirp signal with B and τ_c representing the bandwidth and duration of the chirp, respectively. As mentioned, each radar measurement is collected from a frame of L chirps, and the chirps are transmitted with a waiting time τ_s between them. After the transmission of L chirps, no other signals are transmitted until the

next frame. The transmit signal of the radar frame can be written as

$$s^{\text{tx}}_{\text{frame}}(t) = \sqrt{\mathcal{E}_t} \sum_{l=0}^{L-1} s_{\text{chirp}}(t - (\tau_c + \tau_s) \cdot l), \quad 0 \leq t \leq \tau_f \tag{4.2}$$

where $\sqrt{\mathcal{E}_t}$ is the transmitter gain.

The radar transmit signal is reflected on the different objects in the environment, and is received back at the radar. At the receiver, the signal obtained from an antenna is passed through a quadrature mixer that combines the transmit with receive signals, producing the in-phase and quadrature components. After that, a low-pass filter is applied to the mixed signals. The resulting signal, referred to as intermediate frequency (IF) signal, reflects the frequency and phase difference between the transmit and receive signals. If a single object exists in the environment, then the receive IF signal of a single chirp can be written as [57]

$$s^{\text{rx}}_{\text{chirp}}(t) = \sqrt{\mathcal{E}_t \mathcal{E}_r} \exp\left(j2\pi \left[\mu \tau_{rt} t + f_c \tau_{rt} - \frac{\mu}{2} \tau_{rt}^2\right]\right), \tag{4.3}$$

where $\sqrt{\mathcal{E}_r}$ is the channel gain of the object which depends on the radar cross section (RCS) and path-loss, $\tau_{rt} = 2d/c$ is the round-trip delay of the signal reflected from the object. The symbol d denotes the distance between the object and the radar, and c represents the speed of light.

The receive IF signal, $s^{\text{rx}}_{\text{chirp}}(t)$ is then sampled at the sampling rate of the ADC, f_s, producing S samples for each chirp. Given the L chirps per frame, and assuming an FMCW radar with M_r receive antennas (with an RF chain per antennas), each radar measurement produces $M_r \cdot S \cdot L$ ADC samples. We use $\mathbf{R} \in \mathbb{C}^{M_r \times S \times L}$ to denote the receive radar ADC samples (raw data) of each measurement. Please refer to [57, 58] for more information about the adopted FMCW radar and its hardware architecture. Next, we describe the communication and blockage models.

4.3.2 Communication Model

The considered base station employs a mmWave transceiver with M_{A} antennas and use it to communicate with a single-antenna mobile user. We adopt a narrowband channel model and write the channel between the base station and user as

$$\widetilde{\mathbf{h}} = \widetilde{\mathbf{h}}^{\mathrm{LOS}} + \widetilde{\mathbf{h}}^{\mathrm{NLOS}}, \tag{4.4}$$

where $\mathbf{h}^{\mathrm{LOS}}$ and $\mathbf{h}^{\mathrm{NLOS}}$ are the channel coefficients due to the LOS and NLOS paths. At the downlink, the basestation utilizes the beamforming vector $\mathbf{f} \in \mathbb{C}^{M_A}$ to transmit the symbol s_d to the user. With this model, the receive signal at the user can be expressed as

$$y = \sqrt{\mathcal{E}_c}\, \widetilde{\mathbf{h}}^H \mathbf{f} s_d + n, \tag{4.5}$$

where $\sqrt{\mathcal{E}_c}$ is the transmit gain of the basestation, and $n \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise with σ^2 being the variance. The beamforming vector $\mathbf{f}$ is assumed to be selected from a pre-defined codebook $\boldsymbol{\mathcal{F}}$, i.e., $\mathbf{f} \in \boldsymbol{\mathcal{F}}$ [42]. In particular, the basestation selects the optimal beamforming vector $\mathbf{f}^\star$ that maximizes the receive beamforming gain $|\mathbf{h}^H \mathbf{f}|^2$. In this work, Assuming that $f^\star$ is selected, we write the effective channel as

$$h = \widetilde{\mathbf{h}}^H \mathbf{f}^\star = h^{\mathrm{LOS}} + h^{\mathrm{NLOS}}, \tag{4.6}$$

with h^{LOS} and h^{NLOS} are the effective channel gains of the LOS and NLOS components.

4.3.3 Blockage Model

Adopting a block fading channel model, we define $h[t]$ and $\mathbf{R}[t]$ as the channel gain and radar measurements at time instance $t \in \mathbb{Z}^+$. Now, we can define the blockage indicator at time instance t by $\widetilde{b}[t] \in \{0, 1\}$, which indicates the LOS path being blocked ($\widetilde{b}[t] = 1$) or not ($\widetilde{b}[t] = 0$). With the blockage indicator, we can write the channel gain at time instance t as

$$h[t] = (1 - \widetilde{b}[t]) \cdot h^{\mathrm{LOS}}[t] + h^{\mathrm{NLOS}}[t]. \tag{4.7}$$

We note that for the mmWave and THz frequency communication bands considered in this paper, there are usually a limited number of NLOS paths [73], and the channel gains with the blockage are comparably smaller, i.e., $|h^{\mathrm{LOS}}| \gg |h^{\mathrm{NLOS}}|$. Next, we formulate the blockage prediction problem.

4.4 Radar Aided Blockage Prediction: Problem Formulation

In this section, building upon the system model described in Section 4.3, we define the radar-based blockage prediction problem. This paper aims to predict future blockages utilizing the current and previous radar measurements. Formally, we consider the latest (past and present) T_o radar observations to predict a blockage within the next T_p time-slots. Let us denote the set of the T_o latest radar measurements by

$$\mathbf{X}[t] = \{\mathbf{R}[t - T_o + 1],\ \ldots,\ \mathbf{R}[t]\}\,. \tag{4.8}$$

With this information, our purpose is to predict the blockage status in the following T_p time-slots, i.e., $\{t+1, \ldots, t+T_p\}$. If there is any blockage during these slots, the blockage status for the T_p slots, $b[t]$, is considered as blocked. Mathematically, we can write

$$b_t = \bigvee_{t_p=1}^{T_p} \widetilde{b}[t + t_p], \tag{4.9}$$

where $\vee$ is the logical *OR* operation. With this notation, we define a function $\Psi_{\mathbf{\Theta}}$ that maps the stack of the radar measurements $\mathbf{X}[t]$ to the blockage status $b[t]$. Mathematically, we can write

$$\Psi_{\mathbf{\Theta}} : \mathbf{X}[t] \rightarrow b[t]. \tag{4.10}$$

The function $\Psi_{\mathbf{\Theta}}$ (with its parameters $\mathbf{\Theta}$) returns the blockage status given the radar measurements. Hence, our purpose in this paper is to design a function $\Psi_{\mathbf{\Theta}}$ that approximates the function defined in (4.10) and the optimization of its parameters, $\mathbf{\Theta}$. With the optimal

function and parameters being denoted by $\Psi^\star$ and $\mathbf{\Theta}^\star$, we can formalize the problem by

$$\Psi^\star_{\mathbf{\Theta}^\star} = \arg\min_{\Psi_{\mathbf{\Theta}}} \frac{1}{T} \sum_{t=1}^{T} \mathcal{L}(\Psi_{\mathbf{\Theta}}(\mathbf{X}[t]), b[t]), \tag{4.11}$$

where T is the total number of time-slot samples and $\mathcal{L}(.,.)$ is the loss function of the predictions. In the next two sections, we present two proposed solutions for radar-aided blockage prediction: One is based on radar object tracking and the other one is based on deep learning.

4.5 Radar Aided Blockage Prediction: An Object Tracking Solution

In this section, we propose a radar-aided blockage prediction approach based on target/object tracking. The proposed solution applies the following processing pipeline on the radar measurements: (i) Obtaining range, velocity, and angle Fast Fourier Transformations (FFTs), (ii) detecting the targets by applying a constant rate false alarm rate (CFAR) algorithm [74], (iii) applying a density-based spatial clustering of applications with noise (DBSCAN) algorithm [75, 76] for clustering the detected points, (iv) determining the range/angle/velocity measurements of the detected/clustered objects, (v) associating the measurements of the objects detected in different time samples to form tracks, (vi) tracking the object states (position/velocity) with a Kalman filter, and (vii) predicting the future blockages based on the states (predicted position/velocity) of the objects. Next, we detail the various steps of the proposed approach.

Pre-processing: The radar measurements $\mathbf{R}[t]$ are first processed to obtain range, angle and velocity information. In this method, we apply the pre-processing to initially obtain the radar cube of the range, angle and velocity which is used for the detection of the moving blockage object. To extract the range, angle and velocity information from a radar measurement $\mathbf{R}[t]$, three FFTs are applied. In summary, (i) an FFT in the direction of the time samples, referred to as the range FFT, is applied to obtain the range information, (ii) an

FFT through the chirp samples, called as the Doppler FFT, is applied for the velocity information, and (iii) for the angle information, an FFT through the direction of the antenna samples, referred to as the angle FFT, is applied. We consider the FFTs of size N_S, N_L, and N_M, and denote the resulting processed radar information by $\mathbf{R}_{\mathrm{RC}}[t] \in \mathbb{C}^{N_M \times N_S \times N_L}$, referred to as the radar cube. If we denote the 3D FFT by $\mathcal{F}_{\mathrm{3D}}$, the radar cube can be mathematically expressed as $\mathbf{R}_{\mathrm{RC}}[t] = \mathcal{F}_{\mathrm{3D}}(\mathbf{R}[t])$. In the radar cube, the 2D matrices for each angular sample, transformed from the antenna samples by the angle FFT, contain the range-velocity maps. This information can be further reduced by summing over different angular samples. To clarify, we can write the transformation for the range-velocity maps, denoted by $\mathbf{R}_{\mathrm{RV}} \in \mathbb{R}^{N_M \times N_S}$, as

$$\mathbf{R}_{\mathrm{RV}}[t] = \sum_{n=1}^{N_M} |\mathbf{R}_{\mathrm{RC}}[t]_{(n,:,:)})|, \tag{4.12}$$

where the sub-indices of the radar cube denotes the corresponding elements of the cube. Now, the resulting range-velocity maps are ready for the detection of the objects.

Detection: For the detection of the points in the radar measurements, the constant rate false alarm rate (CFAR) algorithm [74] is applied. In the CFAR algorithm, a window and guard region is determined. For the cell-averaging, the average of the points within the window around the cell under test, excluding the guard-cells, are averaged. This average is multiplied by a constant to construct the detection threshold for the cell under test. In our approach, we adopt a comparably simple 2D cell-averaging CFAR on the range-velocity map to detect the points with reflections. We select the constant for the threshold by inspection. Then, we apply the algorithm for each point in the range-velocity map, and collect the detected points in a binary range-velocity map.

Clustering: After applying the CFAR detection algorithm, the detected points for the different objects need to be determined. To detail, the detected point do not necessarily present a single point for each single object. In addition, due to the noise and clutter, the

points are not necessarily completely grouped around the objects. Hence, the detected points of the map need to be grouped by a clustering algorithm with unknown number of clusters (objects). For this purpose, we apply the DBSCAN [75, 76] clustering algorithm to group the detected points into candidate link blockage objects.

Measurements of the objects: For each object returned by the DBSCAN algorithm, a measurement vector needs to be obtained for tracking. As the measurements, range, velocity and angle information can be extracted from the available radar cube. For the range and velocity, each detected point of an object in the binary range-velocity map correspond to a range and velocity pair. As the range and velocity of this object, we take the average of these pairs. For the angle of an object, first, the angle slices of the 3D radar cube, that correspond to the detection points of the object, are summed. Then, in this sum, the angle corresponding to the bin with the maximum value is taken as the angle measurement of the object.

Data association: With the range, angle and velocity measurements of each object extracted from the radar measurements, the objects in consecutive time samples are associated to construct the tracks of the objects. In particular, we resort to a weighted metric as the distance, which is the weighted sum of the Euclidean position (obtained by using the range and angle), and velocity. Based on this metric, the closest objects in the consecutive time samples are associated with each other, unless the distance between the objects does not exceed a pre-determined threshold value. To clarify, the measurements of the object may not be available in the following frame, and it should not be associated with any other detected object based on the distance. The threshold aims to ensure this is the case. Furthermore, if a track of data is not associated with any object in the last t_e measurements, it is assumed to be ended and removed from the list of active tracks. Next, we apply a Kalman filter to the constructed object tracks.

Kalman filter: In this step, the noisy measurements of the tracks are passed through

a Kalman filter to estimate the states of the objects. Specifically, the Kalman filter adopts the state transition and observation/measurement models to estimate the current state by utilizing the previous state and the current measurements. To detail, the state update model is utilized for the prediction of the current state from the previous state. Then, via the measurement model, the measurements corresponding to the predicted next state are estimated. After obtaining the actual measurement, the difference between the prediction and actual value is utilized for the update step of the Kalman filter, providing a robust estimate for the current state.

For the state transition model of this filter, we follow the constant velocity model [61], and define the states of the objects as the position and velocity in the Cartesian coordinates with the notation $\mathbf{z}[t] = [x[t], y[t], v_x[t], v_y[t]]$. In the constant velocity model, the state is updated as

$$\mathbf{z}[t+1] = \big[x[t] + \tau v_x[t],\ v_x[t],\ y[t] + \tau v_y[t],\ v_y[t]\big], \tag{4.13}$$

For the measurement model, we adopt the augmented measurement model proposed in [77]. In this model, the range, angle and velocity are the measurements, and the Cartesian position and velocity are the states. Mathematically, it is written by

$$\rho = \sqrt{x^2 + y^2}, \tag{4.14}$$

$$v = \frac{x v_x + y v_y}{\sqrt{x^2 + y^2}}, \tag{4.15}$$

$$\theta = \arctan \frac{y}{x}, \tag{4.16}$$

where ρ, v and θ denote the range, velocity and angle measurements, respectively.

In this work, due to the non-linearity of the observation model, an Unscented Kalman filter [72] is adopted for the tracking of the objects. For the multiple active objects, different Kalman filters initiated. To distinguish, we denote the number of objects tracked with separate Kalman filters at time instance t by $O[t]$, and the estimated state of the object $o \in \{1, \ldots, O[t]\}$ by $\mathbf{z}^o[t]$.

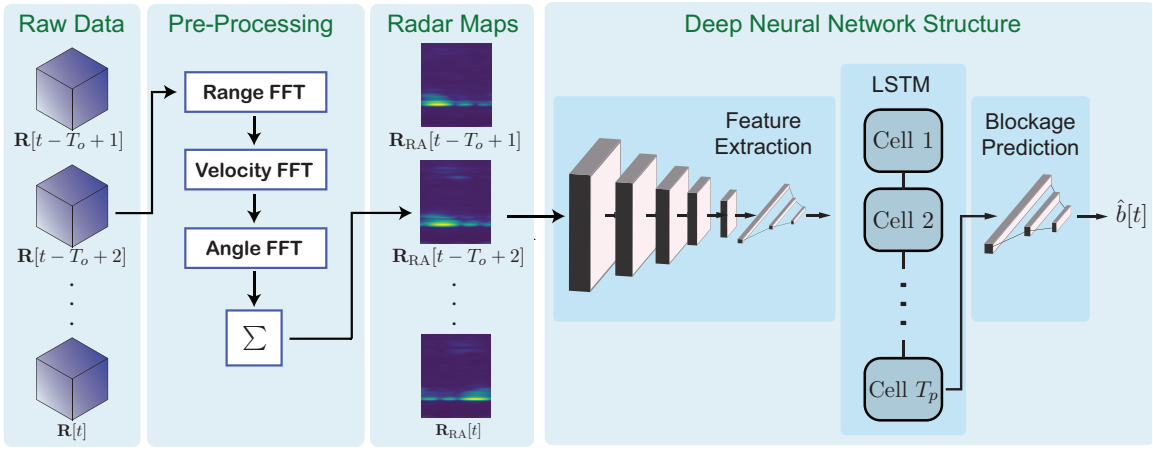


Figure 4.2: The Schema of the Proposed Deep Learning Solution. The Raw Data Is First Processed to Obtain the Range-Angle Maps by Applying the Signal Processing Techniques. The Series of the Range-Angle Maps Are Then Jointly Fed to a Neural Network That Comprises Feature Extraction, LSTM and Blockage Prediction Sub-networks.

Blockage Prediction: For the prediction of the blockages, different types of methods can be considered. For instance, in a more interpretable manner, the current states of the objects can be projected to the future blockage instances via the constant velocity model of the Kalman filter (4.13), and a region to determine the blockage status for each of these future instances can be selected. Nevertheless, from a data-centric and potentially robust manner, we aim to classify the blockage status based on the current object states. Specifically, the states of the objects are stacked with the addition of zeros padding in a vector given by

$$\mathbf{X}_{track}[t] = \left[\mathbf{z}^1[t], \ldots, \mathbf{z}^{O[t]}[t], \mathbf{0}, \ldots, \mathbf{0}\right], \tag{4.17}$$

where $\mathbf{X}_{track}[t] \in \mathbb{R}^{4 \cdot \max_t O[t]}$. With this definition, we can finally apply the classification method that aims to return $b[t]$ from the given states $\mathbf{X}_{track}[t]$. For the classification, we empirically test several classical machine learning classifiers, and adopt k-nearest neighbor (k-NN) algorithm that performs the best over the dataset.

4.6 Radar Aided Blockage Prediction: A Deep Learning Solution

In this section, we propose a deep learning solution for the radar-aided blockage prediction problem defined in Section 4.4. The solution learns a mapping function Ψ, in (4.11), and its variable parameters to solve the considered problem. Our solution in this section does not just adopt a deep neural network model, but also incorporate a domain-knowledge pre-processing approach to present interpretable inputs to the learning model. Specifically, the considered approach starts by pre-processing the radar measurements to extract the range-angle maps. Then, leveraging the understanding of the object tracking problem, we construct a deep neural network architecture, consisting of three stages that aim to extract the relevant features, exploit the sequential correlation, and make classification decisions. This is done via a combination of convolutional neural networks (CNNs), LSTM networks, and fully-connected layers, which complement each other and present a promising solution. In particular, the range-angle maps are fed to the developed model that comprises (i) a CNN-based feature extraction component to extract the essential information from the maps, (ii) an LSTM component to take advantage of the time correlations, and (iii) a linear prediction layer to return the blockage status. The proposed solution, along with its components, are depicted in Figure 4.2. Next, we detail the components of the developed solution, namely, pre-processing; feature extraction, LSTM, and prediction layers.

Pre-processing: In this step, we follow a similar pre-processing approach to that proposed in Section 4.5. In summary, we apply the range, velocity and angle FFTs of size N_S, N_L, and N_M, respectively, and obtain the radar cube $\mathbf{R}_{\mathrm{RC}}[t] \in \mathbb{C}^{N_M \times N_S \times N_L}$ from the measurement $\mathbf{R}[t]$. In the radar cube, the 2D matrices for each chirp sample contains the range-angle maps, which can be further reduced by summing over different chirp samples. Thus, we can write the transformation for the range-angle maps, denoted by $\mathbf{R}_{\mathrm{RA}} \in \mathbb{R}^{N_M \times N_S}$, as

$$\mathbf{R}_{\mathrm{RA}}[t] = \sum_{n=1}^{N_L} |\mathbf{R}_{\mathrm{RC}}[t]_{(:,:,n)})|. \tag{4.18}$$

The resulting range-angle maps of different time samples are ready for the deep learning processing.

Feature Extraction: The range-angle maps contain information about the receive power levels for each point in the maps. However, not all of this information is useful for the prediction of the blockages since only the moving objects are relevant and may cause a blockage. Hence, the irrelevant information can be minimized by extracting a smaller number of features from the range-angle maps. This lower-dimensional representation helps to ease the complexity of the following LSTM architecture. To exemplify a similar approach from the video processing, an object detector network can be adopted to find the objects before tracking them through the LSTM layers. In addition, this dimensionality reduction does not necessarily degrade the performance [78].

Therefore, as the first part of the neural network, we adopt a CNN architecture to reduce the dimensionality and extract the essential local features, relying on the powerful capabilities of CNN networks in similar complex tasks [79]. For this network architecture, we adopt a sequence of the convolutional, average pooling and fully-connected layers. In our design, this part of the network does not aim to extract time-dependent information, and hence, the range-angle maps from different time samples can be fed to the same network separately. Similarly, the network can be trained with the gradient due to the each output separately, providing more samples and faster training opportunity. Formally, the network takes a single range-angle map $\mathbf{R}_{\mathrm{RA}}[t]$ as the input, passes the input through its layers, and returns the extracted features of this map, $\mathbf{r}_{\mathrm{RA}}[t]$, as the output. The architecture of the feature extraction network is summarized in Table 4.1, which is adopted from the deep learning model that has shown successful results [80] for the range-angle map beam prediction task.

Long-Term Short-Memory Networks: After the feature extraction, the dependency of the features across the different time samples can be captured. For this purpose, the

Table 4.1: Architecture of the Feature Extraction Network.

NN Layers	Range-Angle ($\mathbf{R}_{\mathrm{RA}}$)
Input	$1 \times 256 \times 64$
CNN-1	Output Channels: 4, Kernel: (3, 3), Activation: ReLU
CNN-2	Output Channels: 8, Kernel: (3, 3), Activation: ReLU
CNN-3	Output Channels: 16, Kernel: (3, 3), Activation: ReLU
AvgPool-1	Kernel: (2, 2)
CNN-4	Output Channels: 4, Kernel: (3, 3), Activation: ReLU
AvgPool-2	Kernel: (2, 2)
CNN-5	Output Channels: 2, Kernel: (3, 3), Activation: ReLU
AvgPool-3	Kernel: (2, 2)
FC-1	Input Size: 512, Output Size: 256, Activation: ReLU
FC-2	Input Size: 256, Output Size: 64, Activation: ReLU

family of recurrent neural networks, which contains sequential connections between the cells of different inputs, can be utilized. In this work, we adopt the LSTM [64] networks due to their successful radar applications such as object tracking and classification [62] and hand gesture recognition [63].

To detail, an LSTM network consists of multiple LSTM cells, each taking a single entry of the time-sequence data. These LSTM cells are connected to each other in a sequential manner, and each can return an output vector, resulting in an output sequence of these vectors. For the blockage prediction, we only adopt the vector returned from the latest cell. Formally, the network takes $\{\mathbf{r}_{RA}[t-v+1], \ldots, \mathbf{r}_{RA}[t]\}$ as the input, and the last cell returns the intermediate output, $\widetilde{\mathbf{r}}[t]$ of the size of $\mathbf{r}_{\mathrm{RA}}[t]$, to be utilized by the classification network.

Blockage Prediction: Finally, the output of the LSTM is fed to another set of fully-connected neural network layers to obtain the prediction of the blockage. The input of the blockage prediction layers is of the size of a feature vector extracted from a single range-

angle map. For the prediction of the blockage, a final set of layers returning the blockage prediction output is required. For this purpose, we utilize a simple neural network of a single fully-connected layer. This final layer only returns a single prediction as a soft information, i.e., $\hat{b}'[t] \in [0, 1]$, which is later converted to the binary value of the blockage prediction, $\hat{b}[t] = \mathbf{1}\{\hat{b}'[t] > 0.5\}$.

Neural Network Training and Loss Function: The neural network is trained by adopting the formulation of (4.11), only over the parameters $\mathbf{\Theta}$. To clarify, as the design of the function Ψ is fixed by the proposed process, we only aim to find the parameters that minimize the loss over the data samples by

$$\mathbf{\Theta}^{\star} = \arg\min_{\mathbf{\Theta}} \frac{1}{L} \sum_{t=1}^{T} \mathcal{L}\left(\hat{b}'[t], b[t]\right), \tag{4.19}$$

where the loss function is defined as the binary cross-entropy as the problem is a binary classification problem. The binary cross-entropy function can be written by

$$\mathcal{L}(b[t], \hat{b}'[t]) = b[t] \log(\hat{b}'[t]) + (1 - b[t]) \log(1 - \hat{b}'[t]).$$

For the neural network objective defined in (4.19), the error can be computed at the output of the network, and it can be back-propagated through the layers with gradient descent (or alternative) methods, optimizing the parameters $\mathbf{\Theta}$.

4.7 Experimental Setup and Real-World Dataset

For a realistic evaluation of the proposed radar and machine learning aided blockage prediction solution, we collected a large-scale real-world dataset using a hardware testbed with co-existing radar and wireless mmWave equipment, following the DeepSense dataset structure [52]. Using the collected measurements/raw dataset, we built our development dataset for the radar-aided blockage prediction task. In this section, we describe our testbed, raw measurement database, and development dataset.

(a) The basestation antenna array and radar device on the setup

(b) The system setup with a car on sight

(c) Satellite image of the scenario

Figure 4.3: The Testbed and Scenario Details Are Shown From (a) the Front-View, (b) the Back-View With a Potential Blockage Car on Sight, (c) the Satellite. In (a), a Closer Look on the Testbed With mmWave Array and FMCW Radar Is Provided. As Shown in (b) and (c), the Transmitter Is Located on the East Side of the Road While the Receiver on the West.

4.7.1 DeepSense Tesbed-3

We adopt Testbed 3 of the DeepSense 6G dataset [52] for the data collection. Testbed 3 comprises two units: (i) Unit 1, a fixed receiver acting as a basestation and (ii) Unit 2, a static transmitter. Unit 1 includes a 60 GHz uniform linear array (ULA) with $M_{\mathrm{A}} = 16$ elements, and an FMCW radar board (TI AWR2243BOOST) equipped with 3 transmitter and 4 receiver antennas. Meanwhile, Unit 2 comprises an omni-directional static transmitter. The phased array of Unit 1 utilizes an over-sampled beamforming codebook of 64 receive beams. In the radar, only a single transmit antenna along with the 4 receive antennas are activated. The radar (chirp) parameters are selected based on the short-range radar example of TI [58], providing a maximum range of 45m and velocity of 56 km/s. The bandwidth of the utilized chirp frame covers $B = 750$ MHz bandwidth with a chirp slope of $\mu = 15$ MHZ/us over $L = 128$ chirps/frame and $S = 256$ samples/chirp. Next, we detail the collection scenario and development dataset. In Figure 4.3(a), a picture of the testbed is presented.

Table 4.2: Scenario 30: McAlister Stationary Blockage.

Testbed	3
Number of Instances	76000
Number of Units	2
Total Data Modalities	RGB images, LiDAR and radar 64-dimensional received power vector GPS locations
Unit 1	
Type	Stationary
Hardware elements	RGB camera, LiDAR, radar mmWave phased array receiver GPS receiver
Data Modalities	RGB images, LiDAR and radar measurements 64-dimensional received power vector GPS location
Unit 2	
Type	Stationary
Hardware elements	mmWave omni-directional transmitter, GPS receiver
Data Modalities	GPS location

4.7.2 DeepSense Scenario 30

To evaluate our solution, we construct Scenario 30 of the DeepSense 6G dataset [52]. In this scenario, a base station is placed on the sidewalk of a road, directed towards the transmitter, which is placed on the other side of the road. The transmission is blocked when the buses, cars, bicycles and pedestrians are passing through the LOS path. The received power via each beamforming vector and radar measurements are saved continuously to be processed later. In the construction of the dataset, the beam providing the most power and the corresponding power level are saved as the optimal beamforming vector and the max-

imum power level. For labeling the blockage status of the samples, first, a threshold level for the maximum receive power level is determined. The samples providing power level below this threshold are considered as blockages, which are further confirmed manually through the inspection of the RGB images that are captured from the camera of Unit-1. The sampling periodicity is determined as 9 samples/s. The other details of the testbed and raw measurement database is summarized in Table 4.2. We also illustrate the scenario details with the pictures in Figure 4.3.

4.7.3 Development Dataset

To build the development dataset for the considered blockage prediction task, we reduce the number of measurements by keeping only the data points relevant to the blockage, and generate the data samples of sub-sequences for the blockage task. In the following, we refer to each measurement as the data point, and each set of measurements for blockage prediction task as a sample. For the development dataset, we first filtered the raw dataset to keep only 36 data points before a blockage and 10 data points after a blockage, including the first blockage instance. The filtered dataset comprises 14624 data points. These data points consist of 307 unique blockage sequences, which are later utilized to construct the input time-series samples and blockage/status labels of different lengths. For the training, validation and test samples, the sequences are split via $70/20/10\%$ ratio to provide unseen blockage sequences in the test and validation sets. Further, an observation of $T_o = 8$ and a prediction window of $T_p = 10$, we generate the sub-sequences of data. These sub-sequences are later utilized to be fed into the machine learning model and to generate the labels, b, from the individual blockage status, $\widetilde{b}$. To emphasize, $T_p = 10$ is a soft selection and used for the generation of the samples (sub-sequences), i.e., $\mathbf{X}[t]$ and $\widetilde{b}[t+t_p]$ $\forall t_p \in \{1, \ldots, T_p\}$. The final development dataset comprises 6965 training, 1808 validation and 907 test samples. The final labels, $b[t]$ are later generated for each sample based on the

selected T_p value, as presented in the section Section 4.8.

4.8 Evaluation Results

In this section, we evaluate the proposed blockage prediction solutions, namely the object tracking and the LSTM/deep learning based solutions. In the object tracking method, we do not utilize the observation period T_o since the tracking updates are continuously performed after the initial detection of an object. To clarify, there is no input sequence size, which can be interpreted as $T_o = \infty$. The current states of the objects, that are updated through the data points, are adopted in the prediction of future blockages, as described in Section 4.5. In the deep learning solution, we adopt the prepared sub-sequences with $T_o = 8$, as described in Section 4.7. For different values of simulation parameters, a different deep learning model is trained. The training is carried out for up to 30 epochs with the Adam algorithm [65] using a learning rate of 10^{-3}. An early stopping criterion of 5 epochs is adopted to stop the training and save the model with minimum validation loss. As the sampling periodicity of the scenario is designed as 9 data points per second, the frame duration is taken as $\tau_f \approx 110$ ms. We also note that, in the following results, only the results related to the test set are illustrated.

Balance of the labels: The balance of the labels is important for the machine learning tasks. It is usually preferable to have a balanced dataset. In the case it is not provided, different metrics and methods may be applied for a better performance and its evaluation. In the blockage prediction task, it is expected to have an imbalanced dataset with mostly unblocked samples. Although only the relevant samples are kept in our development dataset, the labels are generated based on the blockage interval, and the balance of the dataset changes based on this value. To evaluate the balance, we first investigate the balance for different blockage interval values (T_p). There are 951 sub-sequences (samples) in test set. The percentage of the blocked (true) and unblocked (false) labels are illustrated

in Figure 4.4. *As seen in the figure, the dataset is imbalanced, especially for small blockage intervals, and a careful evaluation is required.* For this purpose, we adopt F_1 score as an evaluation metric. Specifically, F_1 score is defined as the harmonic mean of the precision and recall given by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4.20}$$

where TP, FP and FN represent true positives, false positives and false negatives, respectively. The F_1 score provides a better metric for the evaluation of the imbalanced classification problems by penalizing the extreme values of precision and recall, which indicates the accuracy of true predictions and accuracy for predicting true labels. *Further, in our simulations, we observe that the predictions from both the object detection and deep learning methods present a proportional set of prediction labels. This indicates a well-fit set of methods without requiring any extra attention.*

Performance versus blockage interval: In Figure 4.5, we show the accuracy and F_1 score of the methodologies for different time blockage interval values (T_p). In the figure, the accuracy of the predictions present high-values with around $92 - 97\%$. However, the F_1 scores does not reflect similar results due the imbalance of the dataset. The F_1 score of the radar aided blockage prediction solutions increase with the larger blockage interval. Especially for $T_p\tau = 110$ ms, both solutions perform poorly, potentially due to the low-angular resolution with 4 radar receive antennas. It also shows more successful predictions into the future, while having more problems with the prediction of the closer blockages. In addition, the gap between the object tracking and deep learning solutions stays stable over different time intervals, potentially indicating a similar error difference between the solutions. *The results highlights the delicate design requirements of the classical radar pipeline, while the deep learning solution easily attains significantly better results with the available data.*

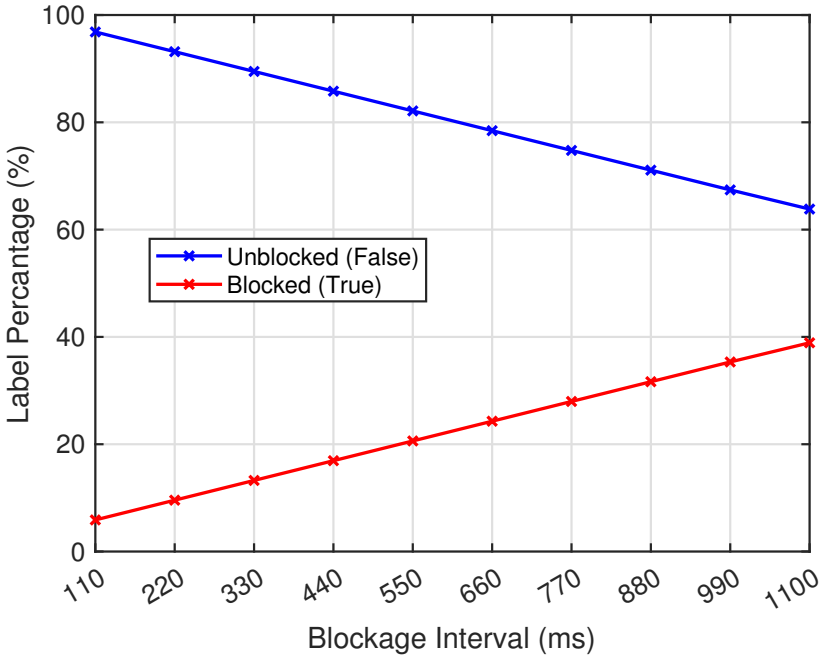


Figure 4.4: Distribution of the Blockage Status in the Ground Truth Values and the Predictions by the Two Proposed Methods. The Dataset Is Separated in a Comparably Balanced Manner.

Complexity: The direct inference duration comparison of the methods may not be fair due to the efficient implementation of the neural networks on the graphical processing units. However, we would like to comment on the complexity of the proposed solutions. In the object tracking method, the CFAR detection requires a 2D convolution (similar to a convolutional layer) with highly complex DBSCAN and Kalman filter applications of small number of variables. To clarify, the DBSCAN algorithm scales with the number of detected points and the Kalman filter scales with the number of tracked objects and states. *In the deep learning solution, the neural network adopted in the evaluations only consists of* $184,015$ *parameters and presents a very small overhead to the system, especially with the specialized devices for the deep learning architectures.*

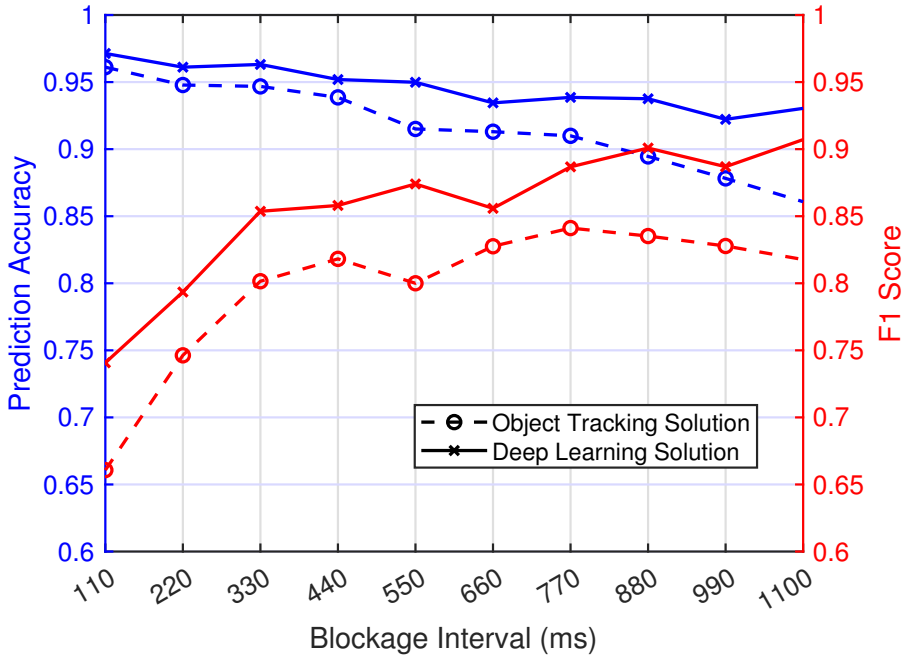


Figure 4.5: The Performance of the Proposed Approaches for Different Blockage Interval. Specifically, the Frame Duration Is Taken As $\tau_f = 110$ Ms and Different Number of Blockage Intervals, $T_p \in \{1, \ldots, 10\}$ Are Shown.

4.9 Summary

In this paper, we proposed radar aided blockage prediction approaches for mmWave and terahertz wireless networks. In particular, we developed two solutions based on (i) a standard radar object tracking pipeline along with a classifier based on the tracked object states, and (ii) a direct LSTM based deep learning solution using the range-angle maps. We evaluated our solutions based on a real-world large-scale dataset comprising co-existing radar and mmWave communication measurements. The results showed that the deep learning solution can achieve better blockage prediction results with a lower design complexity, while the object tracking solution comprises several layers of design and decision complexity. For example, the results indicated that the proposed deep learning approach can predict future blockages 1 second before they happen with an $F1$ score of more than 90% (and $> 90\%$ accuracy) while the object tracking method presented $\sim 80\%$ accuracy. These re-

sults, among others, demonstrate the promising gains of leveraging low-cost radar sensors to proactively predict blockages and enhance the reliability of mmWave networks.

Chapter 5

CELL-FREE ISAC MIMO SYSTEMS: JOINT SENSING AND COMMUNICATION BEAMFORMING

5.1 Overview

This chapter considers a cell-free integrated sensing and communication (ISAC) MIMO system, where distributed MIMO access points are jointly serving the communication users and sensing the targets. For this setup, we first develop two baseline approaches that separately design the sensing and communication beamforming vectors, namely communication-prioritized sensing beamforming and sensing-prioritized communication beamforming. Then, we consider the joint sensing and communication (JSC) beamforming design and derive the optimal structure of these JSC beamforming vectors based on a max-min fairness formulation. Further, based on the beamforming formulation, we propose a power allocation approach for fixed sensing and communication beams. The results show that the developed JSC beamforming is capable of achieving nearly the same communication signal-to-interference-plus-noise ratio (SINR) that of the communication-prioritized sensing beamforming solutions with almost the same sensing SNR of the sensing-prioritized communication beamforming approaches, and provide significant sensing gains over the power allocation, yielding a promising strategy for cell-free ISAC MIMO systems.

5.2 Introduction

The integration of sensing functions into the communication systems is envisioned to be an integral part of the 6G and future communication systems [81–83]. If the hardware and wireless resources are efficiently shared, this will enable the communication infrastruc-

ture to have sensing capabilities at minimal cost and open the sensing frequency bands for wireless communication operation. The sensing capabilities may be utilized to aid the communications for a better performance [6, 84, 85], and also enables interesting applications in security, healthcare, traffic management, etc. Achieving the joint sensing and communication operation, however, requires the careful design of the various aspects of the integrated sensing and communication (ISAC) systems, including the transmission waveform, the post-processing of the received signals, and the MIMO beamforming. While these problems have recently attracted increasing research interest, the prior work has mainly focused on the single ISAC basestation case. In practice, however, multiple ISAC basestations will operate in the same geographical region, frequency band, and time, causing interference on each other for both the sensing and communication functions. This motivates the coordination between these distributed nodes to improve both communication and sensing performance. This ultimately leads to cell-free ISAC MIMO systems, where distributed ISAC basestations jointly serve the same set of communication users and sense the same targets. With this motivation, this chapter investigates the joint sensing and communication beamforming design of these cell-free ISAC MIMO systems.

5.2.1 Prior Work

Distributed antenna systems and interference management in the multi-cell MIMO networks have been extensively studied in the literature [86–88]. With the possibility of more extensive coordination among the basestations, coordinated multi-point transmissions [89], and, more recently, with the densification of the networks, cell-free massive MIMO [90] have attracted significant interest. Cell-free massive MIMO is a concept where multiple access points (APs) jointly serve the user equipments (UEs) by transmitting messages to every user. Note that it is a distributed multi-user MIMO approach, and there are no limitations of cell boundaries. Due to its potential, various aspects of cell-free massive MIMO

have been extensively investigated for further improvements [2, 14, 15, 29, 91, 92]. For example, precoding techniques for cell-free massive MIMO are studied in [15], energy minimization in [14], fronthaul limitations in [2], scalability aspects in [29], and wireless fronthaul in [92]. Most of these studies, however, did not include the unification of the sensing and communication functions in cell-free massive MIMO networks.

The literature for joint sensing and communication (JSC), also called dual-functional radar-communication (DFRC), has mainly focused on the single node (basestation) scenarios [93–97]. For example, the design of the JSC waveform is studied in [93]. Specifically, the author investigated the JSC waveform design for correlated and uncorrelated waveforms and the trade-offs between communication and sensing. The authors in [94] proposed a sensing post-processing for JSC systems. For beamforming, the work in [95] investigated the JSC beamforming design of co-located MIMO system with monostatic radar that serves multiple users. The hybrid beamforming design for OFDM DFRC system is studied in [96]. The optimal beamforming solution for JSC with and without sensing signal's successive interference cancellation is provided in [98]. Along a similar direction, [97] formulated an outage-based beamforming problem and provided the optimal solution.

More relevantly, JSC with distributed nodes (basestations) has been investigated in a few papers [99–102] for power allocation and beamforming. In most of them, however, each user is served by a single AP [99–101]; hence, these studies focused on the interference and not considered a fully cell-free MIMO setup. For example, [100] proposes a transmit and receive beamforming optimization for JSC in a single basestation multi-user scenario, where the signals from a different cell are used to improve the sensing performance without communication interference. In [99], a power allocation problem for JSC is formulated. The problem, however, assumes a single antenna system, and every UE is served by a single AP. In [101], the authors proposed a JSC beamforming optimization for maximizing the detection probability, however, this problem only relied on each AP serving

a single UE. As the most relevant work, the optimization of the JSC power allocation for cell-free massive MIMO has been investigated in [102]. The authors in this work adopted fixed beam designs, i.e., regularized zero beamforming for the communication with the sensing beamforming in the nullspace of the communication channels without further optimization, and focused on optimizing the power allocated to these beams. For the solution, they proposed a convex-concave procedure. Since these cell-free ISAC MIMO systems rely mainly on beamforming in their dual-function operation, it is very important to optimize the design of these JSC beams, which to the best of our knowledge, has not been previously investigated. With this motivation, we propose and compare various beamforming strategies for the cell-free ISAC MIMO systems.

5.2.2 Contributions

To investigate the JSC transmit beamforming in cell-free massive MIMO systems, in this chapter, we consider a system model with many APs and UEs, where the APs jointly serve the UEs and sense the targets in the environment. With this model, we formulate beamforming and power allocation problems and develop various solutions. Our contributions in this chapter can be summarized as follows:

- For beamforming, we first present two baseline strategies that we call *communication-prioritized sensing* and *sensing-prioritized communication beamforming*. In these strategies, either the sensing or the communication beamforming is given the priority to be designed first without accounting for the other function, and then the beamforming of the other function is designed in a way that does not affect the performance of the higher-priority function. For the communication-prioritized sensing, we develop a communication beamforming optimization approach for the given sensing beamforming vector. In this approach, we cast the problem as a convex second-order cone program and provide the optimal solution.

- We consider the case when the sensing and communication beamforming is jointly designed. For this, we formulate a JSC beamforming problem that aims to maximize the sensing SNR while satisfying the communication SINR constraints. We then re-formulate this problem as a non-convex semidefinite problem (SDP) and apply semidefinite relaxation (SDR) to find the optimal beamforming structure for a large set of classes.

- As an alternative to the joint beamforming design problem, we develop a power allocation problem for the given beamforming vectors. Specifically, with fixed beams, we reformulate our beamforming problem as a power allocation problem, which is also an SDP. For this, we develop an approximate SDR-based solution.

We have extensively evaluated the proposed approaches and show that the JSC beamforming design provides near-optimal performance for both sensing and communication thanks to the co-design for the two functions.

Organization: In Section 5.3, we present our system model with the communication and sensing objectives. In Section 5.4 and 5.5, we respectively present the communication- and sensing-prioritized beamforming approaches. Then, we develop the joint sensing and communication beamforming optimization in Section 5.6, and the power allocation formulation and solution in Section 5.7. Finally, in Section 5.8, we provide the numerical results evaluating the developed solutions and present our conclusions of the chapter in Section 5.9.

Notation: We use the following notation throughout this chapter: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, $\mathcal{A}$ is a set. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ are transpose, Hermitian (conjugate transpose), conjugate, inverse, and pseudo-inverse of $\mathbf{A}$, respectively. $\|\mathbf{a}\|$ is the l_2-norm of $\mathbf{a}$ and $\|\mathbf{A}\|_{\mathrm{F}}$ is the Frobenius norms of $\mathbf{A}$. $\mathbf{I}$. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathbb{E}[\cdot]$ and $\otimes$ denote expectation and Kronecker

product, respectively. $\mathbb{S}^+$ is the set of hermitian positive semidefinite matrices.

5.3 System Model

We consider a cell-free massive MIMO ISAC system with M access points (APs) and U communication users, as illustrated in Figure 5.1. In the downlink, and without loss of generality, we assume that a subset $\mathcal{M}_t$ (out of the M APs) are transmitting communication and sensing waveforms to jointly serve the U users, where $|\mathcal{M}_t| = M_t$. Simultaneously, a subset $\mathcal{M}_r$ (out of the M APs) is receiving the possible reflections/scattering of the transmitted waveforms on the various targets/objects in the environment, with $|\mathcal{M}_r| = M_r$. It is important to note here that the subsets $\mathcal{M}_t$ and $\mathcal{M}_r$ may generally have no, partial, or full overlap, which means that none, some, or all the APs could be part of $\mathcal{M}_t$ and $\mathcal{M}_r$ and are simultaneously transmitting and receiving signals. The transmitting and receiving APs are equipped with N_t and N_r antennas. Further, for simplicity, all the APs are assumed to have digital beamforming capabilities, i.e., each antenna element has a dedicated radio frequency (RF) chain. The UEs are equipped with single antennas. The APs are connected to a central unit that allows joint design and processing, and they are assumed to be fully synchronized for both sensing and communication purposes.

5.3.1 Signal Model

In this subsection, we define the joint sensing and communication signal model for the downlink transmissions. The APs jointly transmit U communication streams, $\{x_u[\ell]\}_{u\in\mathcal{U}}$, and Q sensing streams, $\{x_q[\ell]\}_{q\in\mathcal{Q}}$, where $\mathcal{Q} = \{U+1, \ldots, U+Q\}$ and with ℓ denoting the ℓ's symbol in these communication/sensing streams. For ease of exposition, we also define the overall set of streams as $\mathcal{S} = \mathcal{U} \cup \mathcal{Q} = \{1, \ldots, S\}$ with $S = U + Q$. If $\mathbf{x}_m[\ell] \in \mathbb{C}^{N_t\times 1}$ denotes the transmit signal from the transmitting AP m due to the ℓ-th symbol, we can then

write

$$\mathbf{x}_m[\ell] = \underbrace{\sum_{u\in\mathcal{U}} \mathbf{f}_{mu} x_u[\ell]}_{\text{Communication}} + \underbrace{\sum_{q\in\mathcal{Q}} \mathbf{f}_{mq} x_q[\ell]}_{\text{Sensing}} = \sum_{s\in\mathcal{S}} \mathbf{f}_{ms} x_s[\ell], \tag{5.1}$$

where $x_s[\ell] \in \mathbb{C}$ is the ℓ-th symbol of the s-th stream, $\mathbf{f}_{ms} \in \mathbb{C}^{N_t \times 1}$ is the beamforming vector for this stream applied by AP m. The symbols are assumed to be of unit average energy, $\mathbf{E}[|x_s|^2] = 1$. The beamforming vectors are subject to the total power constraint, P_m, given as

$$\mathbf{E}[\|\mathbf{x}_m[\ell]\|^2] = \sum_{s\in\mathcal{S}} \|\mathbf{f}_{ms}\|^2 \leq P_m. \tag{5.2}$$

Further, by stacking the beamforming vectors of stream s of all the APs, we define the beamforming vector $\mathbf{f}_s$

$$\mathbf{f}_s = \begin{bmatrix} \mathbf{f}_{1s}^T & \dots & \mathbf{f}_{M_t s}^T \end{bmatrix}^T \in \mathbb{C}^{M_t N_t}. \tag{5.3}$$

For each stream s, we denote the sequence of L transmit symbols as $\mathbf{x}_s = \begin{bmatrix} x_s[1], \dots, x_s[L] \end{bmatrix}^T$. Given this notation, we make the following assumption, which is commonly adopted in the literature [95]: The messages of the radar and communication signals are statistically independent, i.e., $\mathbf{E}[\mathbf{x}_s \mathbf{x}_s^H] = \mathbf{I}$ and $\mathbf{E}[\mathbf{x}_s \mathbf{x}_{s'}^H] = \mathbf{0}$ for $s, s' \in \mathcal{S}$ with $s \neq s'$. Note that the radar signal generation with these properties may be achieved through pseudo-random coding [95].

5.3.2 Communication Model

We denote the communication channel between UE u and AP m as $\mathbf{h}_{mu} \in \mathbb{C}^{N_t \times 1}$. Further, by stacking the channels between user u and all the APs, we construct $\mathbf{h}_u \in \mathbb{C}^{M_t N_t \times 1}$. Next, considering a block fading channel model, where the channel remains constant over the transmission of the L symbols, we can write the received signal at UE u

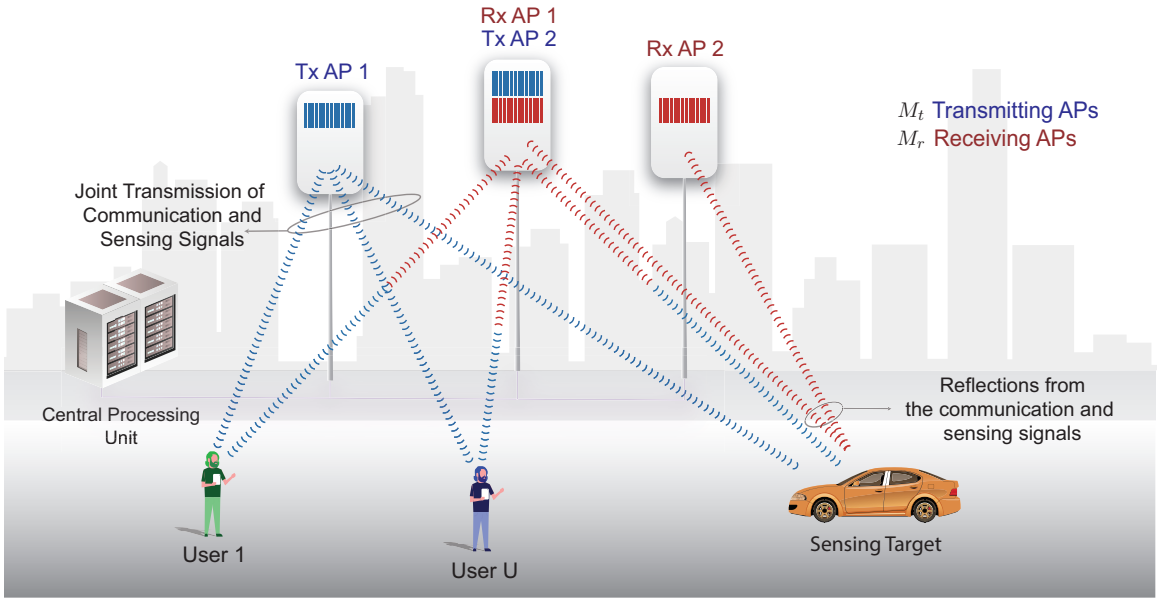


Figure 5.1: The System Model With the Joint Sensing and Communication Transmissions Is Illustrated. The APs Serve Multiple Users While Aiming to Sense the Target.

as

$$\begin{aligned} y_u^{(\mathrm{c})}[\ell] &= \sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{x}_m[\ell] + n_u \\ &= \underbrace{\sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mu} x_u[\ell]}_{\text{Desired Signal (DS)}} + \underbrace{\sum_{u'\in\mathcal{U}\setminus\{u\}} \sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mu'} x_{u'}[\ell]}_{\text{Multi-user Interference (MUI)}} \\ &\quad + \underbrace{\sum_{q\in\mathcal{Q}} \sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mq} x_q[\ell]}_{\text{Sensing Interference (SI)}} + \underbrace{n_u[\ell]}_{\text{Noise}}, \end{aligned} \tag{5.4}$$

where $n_u[\ell] \sim \mathcal{CN}(0, \sigma_u^2)$ is the receiver noise of UE u. Then, the communication SINR of UE u can be obtained as

$$\mathrm{SINR}_u^{(\mathrm{c})} = \frac{\mathbb{E}[|\mathrm{DS}|^2]}{\mathbb{E}[|\mathrm{MUI}|^2] + \mathbb{E}[|\mathrm{SI}|^2] + \mathbb{E}[|\mathrm{Noise}|^2]}. \tag{5.5}$$

In terms of the individual beamforming variables, $\{\mathbf{f}_{ms}\}$, this SINR can be written as

$$\mathrm{SINR}_u^{(\mathrm{c})} = \frac{\left|\sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mu}\right|^2}{\sum_{u'\in\mathcal{U}\setminus\{u\}} \left|\sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mu'}\right|^2 + \sum_{q\in\mathcal{Q}} \left|\sum_{m\in\mathcal{M}_t} \mathbf{h}_{mu}^H \mathbf{f}_{mq}\right|^2 + \sigma_u^2}, \tag{5.6}$$

which is to be utilized in Section 5.5. Further, we also write this expression in terms of the stacked vector variables of each UE u as in (5.3), as

$$\text{SINR}_u^{(\text{c})} = \frac{\left|\mathbf{h}_u^H \mathbf{f}_u\right|^2}{\sum_{u' \in \mathcal{U} \setminus \{u\}} \left|\mathbf{h}_u^H \mathbf{f}_{u'}\right|^2 + \sum_{q \in \mathcal{Q}} \left|\mathbf{h}_u^H \mathbf{f}_q\right|^2 + \sigma_u^2}. \tag{5.7}$$

5.3.3 Sensing Model

For the sensing channel model, we consider a single-point reflector, as commonly adopted in the literature [102, 103]. Specifically, the transmit signal is scattered from the single-point reflector and received by the receiving APs in $\mathcal{M}_r$. With a single path model, the channel between the transmitting AP m_t and the receiving AP m_r through the reflector is defined as

$$\mathbf{G}_{m_t m_r} = \alpha_{m_t m_r} \mathbf{a}(\theta_{m_r}) \mathbf{a}^H(\theta_{m_t}), \tag{5.8}$$

where $\alpha_{m_t m_r} \sim \mathcal{CN}(0, \zeta_{m_t m_r}^2)$ is the combined sensing channel gain, which includes the effects due to the path-loss and radar cross section (RCS) of the target. $\mathbf{a}(\theta)$ is the array response vector. The angles of departure/arrival of the transmitting AP m_t and receiving AP m_r from the point reflector are respectively denoted by θ_{m_t} and θ_{m_r}. We consider the Swerling-I model for the sensing channel [104], which assumes that the fluctuations of RCS are slow and the sensing channel does not change within the transmission of the L sensing and communication symbols in $\mathbf{x}_s$. With this model, the signal received at AP m_r at instant ℓ can be written as

$$\begin{aligned} \mathbf{y}_{m_r}^{(\text{s})}[\ell] &= \sum_{m_t \in \mathcal{M}_t} \mathbf{G}_{m_t m_r} \, \mathbf{x}_{m_t}[\ell] + \mathbf{n}_{m_r}[\ell] \\ &= \sum_{m_t \in \mathcal{M}_t} \alpha_{m_t m_r} \mathbf{a}(\theta_{m_r}) \mathbf{a}^H(\theta_{m_t}) \, \mathbf{x}_{m_t}[\ell] + \mathbf{n}_{m_r}[\ell], \end{aligned} \tag{5.9}$$

where $\mathbf{n}_{m_r}[\ell] \in \mathbb{C}^{N_r}$ is the receiver noise at AP m_r and has the distribution $\mathcal{CN}(0, \varsigma_{m_r}^2 \mathbf{I})$. To write the received radar signal due to the L symbols in a compact form, we introduce

$$\overline{\mathbf{F}}_m = [\mathbf{f}_{m1}, \ldots, \mathbf{f}_{mS}] \in \mathbb{C}^{N_t \times S}, \tag{5.10}$$

$$\overline{\mathbf{X}} = [\mathbf{x}_1, \ldots, \mathbf{x}_S]^T \in \mathbb{C}^{S \times L}. \tag{5.11}$$

Then, we can write the transmit signal from each AP m_t, in (5.1), due to the L symbols as

$$\mathbf{X}_{m_t} = \overline{\mathbf{F}}_{m_t} \overline{\mathbf{X}} \in \mathbb{C}^{N_t \times L}. \tag{5.12}$$

With that, we can re-write the sensing signal in (5.9) at each receiving AP m_r, due to the L symbols, in a compact form as

$$\mathbf{Y}_{m_r}^{(\mathrm{s})} = \underbrace{\sum_{m_t \in \mathcal{M}} \underbrace{\alpha_{m_t m_r} \mathbf{a}(\theta_{m_r}) \mathbf{a}^H(\theta_{m_t}) \overline{\mathbf{F}}_{m_t}}_{\triangleq \overline{\mathbf{G}}_{m_t m_r}}}_{\triangleq \overline{\mathbf{G}}_{m_r}} \overline{\mathbf{X}} + \mathbf{N}_{m_r}, \tag{5.13}$$

with $\overline{\mathbf{G}}_{m_r}$ denoting the beam-space sensing channel of the receiving AP m_r and the receive noise matrix $\mathbf{N}_{m_r} = [\mathbf{n}_{m_r}[1], \ldots, \mathbf{n}_{m_r}[L]]$.

To define a general sensing objective that is correlated with the performance of various sensing tasks (e.g., detection, range/Doppler/angle estimation and tracking), we adopt the joint SNR of the received signals as the sensing objective. Note that the utilization of the joint SNR requires a joint processing of the radar signal at the $\mathcal{M}_r$ sensing receivers. The sensing SNR can be written as

$$\mathrm{SNR}^{(\mathrm{s})} = \frac{\mathbb{E}\left[\sum_{m_r \in \mathcal{M}_r} \left\| \overline{\mathbf{G}}_{m_r} \overline{\mathbf{X}} \right\|_{\mathrm{F}}^2\right]}{\mathbb{E}\left[\sum_{m_r \in \mathcal{M}_r} \left\| \mathbf{N}_{m_r} \right\|_{\mathrm{F}}^2\right]} = \frac{\sum_{m_r \in \mathcal{M}_r} \sum_{m_t \in \mathcal{M}_t} \zeta_{m_t m_r}^2 \left\| \mathbf{a}^H(\theta_{m_t}) \overline{\mathbf{F}}_{m_t} \right\|^2}{\sum_{m_r \in \mathcal{M}_r} \varsigma_{m_r}^2}, \tag{5.14}$$

where the derivation is provided in Appendix A. Recall that $\zeta_{m_t m_r}^2$ denotes the variance of the combined sensing channel gain and $\varsigma_{m_r}^2$ is the variance of the radar receiver noise. The sensing SNR is scaled with the contribution of all the *communication and sensing* streams.

Our objective is then to design the cell-free communication beamforming $\{\mathbf{f}_u\}_{u \in \mathcal{U}}$ and the sensing beamforming $\{\mathbf{f}_q\}_{q \in \mathcal{Q}}$ to optimize the communication SINR and the sensing

SNR defined in (5.7) and (5.14). It is important to note here that, in this chapter, we focus on the beamforming design problem assuming that the communication channel and the sensing target angles are known to the transmit APs. Extending this work to include the channels/angles problem is an interesting future research direction. In the next three sections, we present the proposed beamforming strategies for communication-prioritized sensing, sensing-prioritized communication, and joint sensing and communication.

5.4 Communication-Prioritized Sensing Beamforming Design

In this section, we investigate the scenario where the communication has a higher priority, and where the communication beams are already designed a priori. In this case, the objective is to design the sensing beams to optimize the sensing performance while not affecting the communication performance (i.e., not causing any interference to the U communication users). Note that in this section and the next section, Section 5.5, we assume that $Q = 1$ since we have one sensing target and that the total power is divided with a fixed ratio ρ, leading to $P_m^{\mathrm{c}} = \rho P_m$ for the communication power and $P_m^{\mathrm{s}} = P_m - P_m^{\mathrm{c}}$ for the sensing power. This makes it interesting to explore the joint optimization of the beamforming and power allocation in cell-free ISAC MIMO systems, which is presented in Section 5.7. Next, we present two sensing beamforming design solutions for the cases (i) when the communication users are not present and when (ii) they are present.

Conjugate Sensing Beamforming: When the communication users are not present (i.e., $U = 0$), for example, during downtimes, the system can completely focus on the sensing function. In this case, and given the single target sensing model, the conjugate sensing beamforming solution becomes optimal, as it directly maximizes the sensing SNR. With this solution, the sensing beamforming vectors can be written as

$$\mathbf{f}_{mq}^{\mathrm{CB}} = \sqrt{\frac{p_{mq}}{N_t}}\, \mathbf{a}(\theta_m), \tag{5.15}$$

where $p_{mq} = P_m^s$ is the power allocated for the sensing beam.

Communication-Prioritized Optimal Sensing Solution: When the communication users exist (i.e., $U \geq 1$), and since the communication has a higher priority, a straightforward optimal sensing beamforming approach is to project the optimal sensing beams (constructed through conjugate beamforming) to the null-space of the communication channels. This way, the interference contribution of the sensing beam to the communication channels is eliminated while the sensing SNR is maximized within the communication null space. Let $\mathbf{H}_m = [\mathbf{h}_{m1}, \ldots, \mathbf{h}_{mU}] \in \mathbb{C}^{N_t \times U}$ denote the full channel matrix from the transmit AP m to all the UEs, then the NS sensing beamforming can be constructed as

$$\mathbf{f}_{mq}^{\mathrm{NS}} = \sqrt{p_{mq}} \frac{\left(\mathbf{I} - \mathbf{H}_m \left(\mathbf{H}_m^H \mathbf{H}_m\right)^\dagger \mathbf{H}_m^H\right) \mathbf{a}(\theta_m)}{\left\| \left(\mathbf{I} - \mathbf{H}_m \left(\mathbf{H}_m^H \mathbf{H}_m\right)^\dagger \mathbf{H}_m^H\right) \mathbf{a}(\theta_m) \right\|}, \tag{5.16}$$

where we again set the allocated power $p_{mq} = P_m^s$ as we consider a single sensing beam.

5.5 Sensing-Prioritized Communication Beamforming Design

In this section, we consider the scenario where the sensing has a higher priority, and where the sensing beams are already designed a priori. In this case, the objective is to design the communication beams to optimize the communication performance while *minimizing* the impact of the sensing interference. It is important to note here that an interesting difference between the communication and sensing optimization problems is that while the sensing signals could cause interference that degrades the communication performance, the communication signals could generally be leveraged to further enhance the sensing performance. Next, we present two communication beamforming design solutions for the cases when (i) the sensing target is not present and when (ii) it is present.

Regularized Zero-forcing Beamforming: When the sensing target is not present, i.e., $Q = 0$, a near-optimal communication beamforming design is the regularized zero-forcing (RZF) [105]. This solution allows a trade-off between the multi-user interference and noise

terms of the SINR through a regularization parameter λ, that is added to the ZF beamforming:

$$\tilde{\mathbf{f}}_u^{\mathrm{RZF}} = \left(\lambda \mathbf{I} + \sum_{u' \in \mathcal{U}} \mathbf{h}_{u'} \mathbf{h}_{u'}^H \right)^{-1} \mathbf{h}_u, \tag{5.17}$$

which then can be normalized to satisfy the power constraints, i.e.,

$$\mathbf{f}_{mu}^{\mathrm{RZF}} = \sqrt{p_{mu}} (\tilde{\mathbf{f}}_{mu}^{\mathrm{RZF}} / |\tilde{\mathbf{f}}_{mu}^{\mathrm{RZF}}|). \tag{5.18}$$

We here again adopt $p_{mu} = P_m / U$ with an equal power between the beams. For the RZF, it is preferable to have a higher regularization parameter in the scenarios with a higher noise and smaller regularization parameter in scenarios with more interference. For further details, we refer to [105].

Sensing-Prioritized Optimal Communication Solution: For the case when the sensing beam is designed a priori, we derive a max-min fair rate optimal communication beamforming solution. First, this max-min problem can be written as

$$\text{(P1.1):} \quad \max_{\{\mathbf{f}_{mu}\}} \min_{u} \quad \mathrm{SINR}_u^{(\mathrm{c})} \tag{5.19a}$$

$$\text{s.t.} \quad \sum_{u \in \mathcal{U}} \|\mathbf{f}_{mu}\|^2 \leq P_m^c, \quad \forall m \in \mathcal{M}_t, \tag{5.19b}$$

where the objective is quasiconvex [36] and shows a similar structure to the optimal beamforming formulation for the cell-free massive MIMO networks with only the communication objective [91]. For a given minimum SINR constraint γ, the problem can be written as the following feasibility problem

$$\text{(P1.2):} \quad \text{find} \quad \{\mathbf{f}_{mu}\} \tag{5.20a}$$

$$\text{s.t.} \quad \mathrm{SINR}_u^{(\mathrm{c})} \geq \gamma, \quad \forall u \in \mathcal{U}, \tag{5.20b}$$

$$\sum_{u \in \mathcal{U}} \|\mathbf{f}_{mu}\|^2 \leq P_m^c, \quad \forall m \in \mathcal{M}_t. \tag{5.20c}$$

Here, we note that the SINR constraint (5.20b) is in a fractional form. This, however, can be converted to a second-order cone constraint. For this purpose, we can re-write the

constraint as

$$\left(1+\frac{1}{\gamma}\right)\left|\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu}\right|^2 \geq \sum_{u'\in\mathcal{U}}\left|\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu'}\right|^2+\sum_{q\in\mathcal{Q}}\left|\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mq}\right|^2+\sigma_u^2. \quad (5.21)$$

Now, taking the square root of both sides, we can convert the given form to a second-order cone constraint. The square root, however, leaves an absolute on the left-hand side, which is a non-linear function. This can be simplified as the real part of the variable [105], since any angular rotation ($e^{-j\psi}$) to the expression inside the absolute does not change the value, i.e.,

$$|\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu}| = |\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu}e^{-j\psi}| = \text{Re}\left\{\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu}\right\}. \quad (5.22)$$

This can be seen as selecting the optimal solution with a specific angular rotation from the set of infinite rotations $\psi \in [0, 2\pi)$. Finally, we can write the constraint (5.20b) as a second-order cone as follows

$$\left(1+\frac{1}{\gamma}\right)^{\frac{1}{2}}\text{Re}\left\{\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mu}\right\} \geq \left\|\begin{array}{c}\sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{m1}\\ \vdots \\ \sum_{m\in\mathcal{M}_t}\mathbf{h}_{mu}^H\mathbf{f}_{mS}\\ \sigma_u\end{array}\right\|. \quad (5.23)$$

When (5.20b) is replaced with (5.23), it results in a second-order cone problem and can be solved by the convex solvers [106]. Using the bisection algorithm, the maximum SINR value, $\gamma^\star$, can be obtained by solving the convex feasibility problem (5.20) for different values of γ within a predetermined range $[\gamma_{\text{min}}, \gamma_{\text{max}}]$. This computes the optimal solution to (5.19).

5.6 Joint Sensing and Communication: Beamforming Optimization

A more desirable approach for cell-free joint sensing and communication MIMO systems is to jointly optimize the beamforming vectors for the sensing and communication

functions. Specifically, our objective is to maximize the sensing SNR together with the communication SINR of the UEs. Towards this objective, we reformulate (5.20) as a sensing SNR maximization problem by (i) adding the sensing SNR maximization as an objective to the feasibility problem, and (ii) generalizing the minimum communication SINR limit, γ, individually for each UE with γ_u. Then, the JSC beamforming optimization problem can be written as

$$\text{(P2.1):} \quad \max_{\{\mathbf{f}_{ms}\}} \quad \text{SNR}^{(\text{s})} \tag{5.24a}$$

$$\text{s.t.} \quad \text{SINR}_u^{(\text{c})} \geq \gamma_u, \quad \forall u \in \mathcal{U}, \tag{5.24b}$$

$$\sum_{s \in \mathcal{S}} \|\mathbf{f}_{ms}\|^2 \leq P_m, \quad \forall m \in \mathcal{M}_t, \tag{5.24c}$$

where the objective, i.e., the maximization of the convex SNR expression, $\text{SNR}^{(\text{s})}$, is non-convex and the problem is a non-convex quadratically constrained quadratic program (QCQP). Hence, a similar approach to the beamforming optimization in the previous section can not be adopted. The problem in (5.24), however, can be cast as a semidefinite program, which allows applying a semidefinite relaxation for the non-convex objective [107]. With the relaxation, the problem becomes convex, and the optimal solution can be obtained with the convex solvers. After that, the result obtained from the relaxed problem can be cast to the original problem's space with a method designed specifically for the problem. In the following, we present the details of our approach.

To reformulate (5.24) as an SDP, we first re-define the beamforming optimization variables as matrices: $\mathbf{F}_s = \mathbf{f}_s \mathbf{f}_s^H, \forall s \in \mathcal{S}$. Writing (5.24) in terms of $\mathbf{F}_s$ instead of $\mathbf{f}_s$ eliminates the quadratic terms in the sensing SNR and communication SINR expressions. This SDP formulation, however, by construction introduces two new constraints: (i) The convex hermitian positive semi-definiteness constraint $\mathbf{F}_s \in \mathbb{S}^+$, where $\mathbb{S}^+$ is the set of hermitian positive semidefinite matrices, and (ii) the non-convex rank-1 constraint $\text{rank}(\mathbf{F}_s) = 1$. Further, we need to write the problem (P2.1) in terms of these newly introduced variables,

$\{\mathbf{F}_s\}$. For this purpose, we define the AP selection matrix, $\mathbf{D}_m \in \mathbb{R}^{MN_t \times MN_t}$, where each element of this matrix is given by

$$[\mathbf{D}_m]_{ij} = \begin{cases} 1 & \text{if } (m-1)N_t + 1 \leq i \leq mN_t \text{ with } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{5.25}$$

where the only non-zero elements of the $\mathbf{D}_m$ is the identity matrix placed at the m-th cross diagonal $N_t \times N_t$ block matrix. To write the sensing SNR in a compact form, we define $\mathbf{A} = \sum_{m_t \in \mathcal{M}_t} \bar{\zeta}_{m_t} \mathbf{D}_{m_t}, \overline{\mathbf{A}} \mathbf{D}_{m_t}$, where $\overline{\mathbf{A}} = \bar{\mathbf{a}}\,\bar{\mathbf{a}}^H$ with $\bar{\mathbf{a}} = [\mathbf{a}(\theta_1)^T, \ldots, \mathbf{a}(\theta_{M_t})^T]^T$, and $\bar{\zeta}_{m_t} = \sum_{m_r \in \mathcal{M}_r} \zeta^2_{m_t, m_r}$. Now, we can write the objective of (5.24) (sensing SNR) in terms of $\mathbf{A}$ as

$$\text{SNR}^{(\text{s})} = \frac{\text{Tr}\left(\mathbf{A} \sum_{s \in \mathcal{S}} \mathbf{F}_s\right)}{\sum\limits_{m_r \in \mathcal{M}_r} \varsigma^2_{m_r}}. \tag{5.26}$$

where the derivation is provided in Appendix B.

For the constraints of the problem in (5.24), we define $\mathbf{Q}_u = \mathbf{h}_u \mathbf{h}_u^H$ and re-write the SINR in (5.7) in terms of the new variables as

$$\text{SINR}_u^{(\text{c})} = \frac{\text{Tr}\left(\mathbf{Q}_u \mathbf{F}_u\right)}{\sum\limits_{u' \in \mathcal{U} \setminus \{u\}} \text{Tr}\left(\mathbf{Q}_u \mathbf{F}_{u'}\right) + \sum\limits_{q \in \mathcal{Q}} \text{Tr}\left(\mathbf{Q}_u \mathbf{F}_q\right) + \sigma_u^2}. \tag{5.27}$$

With this, we can write the constraint in (5.24b) and the power constraint in (5.24c) as

$$\left(1 + \gamma_u^{-1}\right) \text{Tr}\left(\mathbf{Q}_u \mathbf{F}_u\right) - \text{Tr}\left(\mathbf{Q}_u \sum_{s \in \mathcal{S}} \mathbf{F}_s\right) \geq \sigma_u^2, \tag{5.28}$$

$$\sum_{s \in \mathcal{S}} \text{Tr}\left(\mathbf{D}_m \mathbf{F}_s\right) \leq P_m, \quad \forall m \in \mathcal{M}_t. \tag{5.29}$$

Here, we notice that for the objective and constraints, we can simplify the sensing variables by defining $\mathbf{F}_{\mathcal{Q}} = \sum_{q \in \mathcal{Q}} \mathbf{F}_q$, since the sensing variables $\mathbf{F}_q$ only appear in the defined summation form. For the optimality of the problem, however, this variable needs to have at most rank Q, so that we can construct Q beamforming vectors [1] . Then, by

[1] If the rank of this variable is less than Q, some of the beamforming vectors are not needed, and can be selected as zero.

collecting the expressions together, we can write the SDP form of our problem (P2-QCQP) as

$$\begin{aligned} \text{(P2.1-SDP):} \qquad \max_{\{\mathbf{F}_u\},\mathbf{F}_{\mathcal{Q}}} \quad & \mathrm{Tr}\left(\mathbf{A}\sum_{s\in\mathcal{S}}\mathbf{F}_s\right) \\ \text{s.t.} \quad & \text{(5.28) and (5.29)}, \\ & \mathrm{rank}(\mathbf{F}_u)=1, \quad \forall u\in\mathcal{U}, \\ & \mathbf{F}_u\in\mathbb{S}^+, \quad \forall u\in\mathcal{U}. \\ & \mathrm{rank}(\mathbf{F}_{\mathcal{Q}})\leq Q, \\ & \mathbf{F}_{\mathcal{Q}}\in\mathbb{S}^+ \end{aligned} \tag{5.30}$$

which can be relaxed by removing the rank constraints as follows

$$\begin{aligned} \text{(P2.1-SDR):} \qquad \max_{\{\mathbf{F}_u\},\mathbf{F}_{\mathcal{Q}}} \quad & \mathrm{Tr}\left(\mathbf{A}\sum_{s\in\mathcal{S}}\mathbf{F}_s\right) \\ \text{s.t.} \quad & \text{(5.28) and (5.29)}, \\ & \mathbf{F}_u\in\mathbb{S}^+, \quad \forall u\in\mathcal{U} \\ & \mathbf{F}_{\mathcal{Q}}\in\mathbb{S}^+. \end{aligned} \tag{5.31}$$

This relaxed problem (P2.1-SDR) can be solved via CVX and convex SDP solvers [106, 108]. Then, if the matrices obtained by this solution, denoted by $\{\mathbf{F}'_u\}$, are rank-1, and $\mathbf{F}'_{\mathcal{Q}}$ is at most rank-Q, then they are optimal for (5.30). The optimal user beamforming vectors, $\mathbf{f}_u$, in this case, can be obtained as the eigenvector of $\mathbf{F}'_u$. Similarly, $\{\mathbf{f}_q\}$ can be constructed as the Q eigenvectors of $\mathbf{F}'_{\mathcal{Q}}$. For the case the user matrices are not rank-1, we make the following proposition.

Proposition 1. *There exists a solution to the problem* (5.30)*, denoted by* $\{\mathbf{F}''_u\}$ *and* $\mathbf{F}''_{\mathcal{Q}}$*, that satisfies* $\mathrm{rank}(\mathbf{F}''_u)=1$, $\forall u\in\mathcal{U}$ *and*

$$\mathbf{F}''_{\mathcal{Q}}=\mathbf{F}'_{\mathcal{Q}}+\sum_{u\in\mathcal{U}}\mathbf{F}'_u-\sum_{u\in\mathcal{U}}\mathbf{F}''_u. \tag{5.32}$$

where $\mathbf{F}'_{\mathcal{Q}}$ *and* $\{\mathbf{F}'_u\}$ *are the solutions of the SDP problem in (P2.1-SDR). The communication beamforming vectors of this solution can be given as*

$$\mathbf{f}''_u = (\mathbf{h}_u^H \mathbf{F}'_u \mathbf{h}_u)^{-\frac{1}{2}} \mathbf{F}'_u \mathbf{h}_u. \tag{5.33}$$

Further, if $\mathrm{rank}(\mathbf{F}''_{\mathcal{Q}}) \leq Q$*, the optimal sensing beamforming vectors of this solution can be constructed by*

$$\mathbf{f}''_q = \sqrt{\lambda_{q-U}}\, \mathbf{u}_{q-U}, \tag{5.34}$$

with λ_i *and* $\mathbf{u}_i$ *being the* i*-th largest eigenvalue of* $\mathbf{F}''_{\mathcal{Q}}$ *and the corresponding eigenvector.*

The proof extends the solution in [95, Theorem 1], which we provide in Appendix C. For $\mathrm{rank}(\mathbf{F}''_{\mathcal{Q}}) \leq Q$, the solution obtained is from Proposition 1 optimal. In the case $\mathrm{rank}(\mathbf{F}''_{\mathcal{Q}}) > Q$, however, (5.34) will not lead to the optimal solution. We will examine the performance of this approximation in Figure 5.6. Next, we investigate the value of Q required to satisfy the optimality.

5.6.1 How Many Sensing Streams Do We Need?

In the formulations of (P2.1-SDP) and (P2.1-SDR) in Section 5.6, the number of sensing streams is kept generic with the variable Q. In reality, however, it would be preferable to have as few as possible sensing streams. To that end, it is interesting to investigate how many sensing beams are needed to achieve optimal sensing performance. For this objective, we can further investigate (P2.1-SDR) to find the constraints on the optimal sensing solution. Specifically, we attempt to solve the problem (P2.1-SDR). Since this problem is convex, it satisfies the strong duality [36, 109]. Then, we can derive the dual problem as

$$\begin{aligned} \text{(D2.1-SDR):} \min_{\{\lambda_u\},\{\nu_m\}} \quad & \sum_m \nu_m P_m - \sum_u \lambda_u \sigma_u^2 \\ \text{s.t.} \quad & \mathbf{B}_u \preceq 0, \quad \forall u \in \mathcal{U}, \\ & \mathbf{B}_{\mathcal{Q}} \preceq 0, \end{aligned} \tag{5.35}$$

where $\{\lambda_u \geq 0\}$, $\{\nu_m \geq 0\}$ are the Lagrangian coefficients corresponding to the SINR and power constraints, respectively, and

$$\mathbf{B}_u = \mathbf{A} + \lambda_u \gamma_u^{-1} \mathbf{Q}_u - \sum_{u' \in \mathcal{U} \setminus \{u\}} \lambda_{u'} \mathbf{Q}_{u'} - \sum_m \nu_m \mathbf{D}_m, \tag{5.36}$$

$$\mathbf{B}_{\mathcal{Q}} = \mathbf{A} - \sum_{u' \in \mathcal{U}} \lambda_{u'} \mathbf{Q}_{u'} - \sum_m \nu_m \mathbf{D}_m. \tag{5.37}$$

The derivation of the dual function is provided in Appendix D. Further, we make the following remark on the definition of the new variables in the dual problem.

Remark 1. From the definition of the new variables, $\mathbf{B}_u$ and $\mathbf{B}_{\mathcal{Q}}$, we also have the relation

$$\mathbf{B}_u = \mathbf{B}_{\mathcal{Q}} + \lambda_u (1 + \gamma_u^{-1}) \mathbf{Q}_u. \tag{5.38}$$

Let us assume that there exists a feasible set of primal-dual optimal variables, i.e., $\{\mathbf{F}_u^\star\}$, $\mathbf{F}_{\mathcal{Q}}^\star$, $\{\lambda_u^\star\}, \{\eta_m^\star\}$, and the corresponding variables $\mathbf{B}_{\mathcal{Q}}^\star$ and $\{\mathbf{B}_u^\star\}$. With the Karush-Kuhn-Tucker (KKT) conditions [36], we have the complementary slackness for the semidefinite constraints given as

$$\begin{aligned} \mathbf{B}_u^\star \mathbf{F}_u^\star &= \mathbf{0} \\ \mathbf{B}_{\mathcal{Q}}^\star \mathbf{F}_{\mathcal{Q}}^\star &= \mathbf{0} \end{aligned} \tag{5.39}$$

which shows that $\mathbf{F}_{\mathcal{Q}}^\star$ is in the nullspace of $\mathbf{B}_{\mathcal{Q}}^\star$. We can further refine this condition with the following proposition.

Proposition 2. *The sensing beamforming matrix is in the nullspace of* $\mathbf{A} - \sum \nu_m^\star \mathbf{D}_m$. *In addition, it is in the nullspace of* $\mathbf{Q}_u$ *for any user with* $\lambda_u^\star > 0$.

Proof: We have

$$\begin{aligned} \lambda_u^\star (1 + \gamma_u^{-1}) \mathrm{Tr}\left(\mathbf{Q}_u \bar{\mathbf{F}}_{\mathcal{Q}}^\star\right) &= \mathrm{Tr}\left((\mathbf{B}_u^\star - \mathbf{B}_{\mathcal{Q}}^\star) \bar{\mathbf{F}}_{\mathcal{Q}}^\star\right) \\ &= \mathrm{Tr}\left(\mathbf{B}_u^\star \bar{\mathbf{F}}_{\mathcal{Q}}^\star\right) \\ &\leq \max_{\bar{\mathbf{F}}_{\mathcal{Q}} \succeq 0} \mathrm{Tr}\left(\mathbf{B}_u^\star \bar{\mathbf{F}}_{\mathcal{Q}}\right) = 0, \end{aligned} \tag{5.40}$$

where the first equality is due to definition (Remark 1), the second equality by the complementary slackness condition given in (5.39), and the latter inequality due to the multiplication of the negative and positive semidefinite matrices.

This proposition shows that a sensing matrix will be in the nullspace of stacked UE channels, $\mathbf{h}_u$, for every UE u that satisfies the SINR constraint at the equality. On the other hand, when the equality of the SINR constraint is not satisfied, we have $\lambda_u^\star = 0$ due to the complementary slackness condition of the SINR constraint. This leads to $\mathbf{B}_u^\star = \mathbf{B}_{\mathcal{Q}}^\star$ as shown in Remark 1. Combined with the complementary slackness conditions in (5.39), it results in $\mathbf{F}_u^\star$ and $\mathbf{F}_{\mathcal{Q}}^\star$ being in the same space defined as the nullspace of the channels of the UEs with $\lambda_u^\star > 0$ and the sensing direction via $\mathbf{A} - \sum \nu_m^\star \mathbf{D}_m$. There is, however, no enforcement towards the direction of the user itself because $\lambda_u^\star(1 + \gamma_u^{-1})\mathbf{Q}_u = 0$. Then, this case is likely to appear only if there is sufficient SINR with the transmission towards the sensing direction from all the APs (e.g., the UE and target are at the same location). Hence, it is trivial and not of significant interest. With this observation, we focus on the case with $\lambda_u^\star > 0$ for every UE.

In Proposition 2, we also have the nullspace of $\mathbf{A} - \sum \nu_m^\star \mathbf{D}_m$ to define the sensing matrix. We note that $\sum \nu_m^\star \mathbf{D}_m$ is a diagonal matrix with the diagonal of each block having the same value $\nu_m^\star$.

Remark 2. The sensing SNR matrix, $\mathbf{A}$, is a block diagonal matrix of rank-1 blocks, i.e.,

$$\mathbf{A} = \mathrm{diag}(\bar{\zeta}_1 \mathbf{a}(\theta_1)\mathbf{a}^H(\theta_1), \ldots, \bar{\zeta}_{M_t} \mathbf{a}(\theta_{M_t})\mathbf{a}^H(\theta_{M_t})). \tag{5.41}$$

This fact can be seen by the definition of $\mathbf{A} = \sum_{m_t \in \mathcal{M}_t} \bar{\zeta}_{m_t} \mathbf{D}_{m_t} \overline{\mathbf{A}} \mathbf{D}_{m_t}$, where the multiplication of a matrix from both sides with the selection matrix, $\mathbf{D}_{m_t}$, results in the block diagonal of the selected entries. Further, with a slight abuse of notation, each block of the diagonal $\mathbf{A}_{m_t} \triangleq \bar{\zeta}_{m_t} \mathbf{a}(\theta_{m_t})\mathbf{a}^H(\theta_{m_t}) = \mathbf{D}_{m_t} \mathbf{A} \mathbf{D}_{m_t}$ is a weighted outer product of an array response vector. Therefore, we have $\mathrm{rank}(\mathbf{A}_{m_t}) = 1, \forall m_t \in \mathcal{M}_t$ and $\mathrm{rank}(\mathbf{A}) = M_t$.

With Remark 2, we can further refine our space as $\mathbf{A} - \sum \nu_m^\star \mathbf{D}_m = \sum(\mathbf{A}_m - \nu_m^\star \mathbf{I}_m)$, where each component, $\mathbf{A}_m - \nu_m^\star \mathbf{I}_m$, is a diagonal block. For $\nu_m^\star > 0$, each of these components can have a single nullspace vector $\mathbf{f}_{mq} = \mathbf{a}(\theta_m)$ if $\nu_m^\star = \bar{\zeta}_m$. Note that this only constructs a part of the sensing matrix, not the full domain $\mathbf{f}_q$. If $\nu_m^\star \neq \bar{\zeta}_m$, the component is full rank, and the nullspace is empty. In other words, the existence of sensing beams is independently determined at each AP based on $\nu_m^\star$ and only available if $\nu_m^\star = \bar{\zeta}_m$. For any other $\nu_m^\star > 0$, there is no sensing stream. Then, the sensing beams can be written in the form of $\mathbf{f}_q = [\eta_{1q}\mathbf{a}(\theta_1), \ldots, \eta_{Mq}\mathbf{a}(\theta_M)]$ for some $\eta_{mq} \in \mathbb{C}$, and this vector is in the nullspace of $\mathbf{h}_u$ for every UE with $\lambda_u^\star > 0$. To that end, further investigating the performance of the nullspace sensing beam with the suboptimal beamforming solutions is interesting. As an alternative to beamforming optimization, in the next section, we develop a power allocation approach for the pre-determined beamforming vectors. Before moving on, we further conclude the limitations on the sensing streams.

Proposition 3. *For $\nu_m^\star > 0 \; \forall m \in \mathcal{M}_t$, the maximum number of sensing streams is limited by*

$$\operatorname{rank}(\mathbf{F}_{\mathcal{Q}}^\star) \leq M_t. \tag{5.42}$$

The proof follows Remark 2, where the minimum rank of $\mathbf{A} - \sum_m \nu_m^\star \mathbf{D}_m = (N_t - 1)M_t$, and its nullspace can have at most M_t dimensions. Further, we can refine the limit in the case of random Rayleigh channels as follows.

Proposition 4. *For $\nu_m^\star > 0 \; \forall m \in \mathcal{M}_t$ and $\mathbf{h}_{mu} \sim \mathcal{CN}(0, \mathbf{I})$, the maximum number of sensing streams is limited by*

$$\operatorname{rank}(\mathbf{F}_{\mathcal{Q}}^\star) \leq \max\{M_t - U, 0\}. \tag{5.43}$$

with probability 1.

The proof follows the fact that the probability of any $\mathbf{h}_u$ drawn from random Gaussian distribution can be spanned by the space constructed by $\mathbf{A} - \sum_m \nu_m^\star \mathbf{D}_m$ with probability

0. Hence, each UE reduces the available dimensions for the sensing stream with probability 1.

As shown in the Proposition 3, the number of sensing streams upper bounded by the number of APs. More interestingly, in the case of Rayleigh channels, the number of sensing streams is limited to $M_t - U$, i.e., the difference between the number of transmitting APs and UEs. From this result, if the number of transmitting APs is smaller than the number of UEs, no sensing streams is required. However, in the cell-free massive MIMO regime where the number of APs are much larger than the number of UEs, approximately a single stream for each AP may be required. We will further investigate the suboptimality of the solutions with lower number of sensing streams in Section 5.8.

5.7 Joint Sensing and Communication: Power Allocation with Fixed Beams

In this section, we develop a power allocation formulation. First, we note that the formulated beamforming optimization problem jointly optimizes the power along with the beams since the power constraints of the beams are set to satisfy the power constraints at each AP. Another interesting case with the cell-free massive MIMO, however, is to allocate the power for pre-determined suboptimal beams. For this purpose, in this section, we develop a power allocation formulation for given beams. Differently from the approach developed for power optimization in [102], which adopts an iterative convex-concave programming approach and does not guarantee the optimality, we maintain the SDP framework of our chapter and develop a power allocation approach with the SDR relaxation. This approach can also provide an upper-bound with the relaxation.

Mathematically, let us denote the pre-determined unit-power beamforming vectors and power coefficients by $\{\bar{\mathbf{f}}_{mq}\}$ and $\{p_{mq}\}$. With this notation, the beamforming vectors in the previous formulation in (5.30) can be written as $\mathbf{f}_{mq} = \sqrt{p_{mq}}\bar{\mathbf{f}}_{mq}$. In this model, the fixed beamforming vectors can be selected by the approaches given in Section 5.4 and

Section 5.5. With this definition, we can rewrite our JSC objective in terms of the power variables of the beams, p_{mq}. First, we re-write the JSC optimization problem in terms of these variables as

$$\text{(P3.1a):} \max_{\{p_{ms}\}} \quad \sum_{m\in\mathcal{M}_t}\sum_{m_r\in\mathcal{M}_r}\zeta^2_{m,m_r}\sum_{s\in\mathcal{S}} p_{ms}\left|\mathbf{a}^H(\theta_m)\bar{\mathbf{f}}_{ms}\right|^2 \tag{5.44a}$$

$$\text{s.t.} \quad \frac{\left|\sum_{m\in\mathcal{M}_t}\sqrt{p_{mu}}\mathbf{h}^H_{mu}\bar{\mathbf{f}}_{mu}\right|^2}{\sum_{u'\in\mathcal{U}\setminus\{u\}}\left|\sum_{m\in\mathcal{M}_t}\sqrt{p_{mu'}}\mathbf{h}^H_{mu}\bar{\mathbf{f}}_{mu'}\right|^2+\sum_{q\in\mathcal{Q}}\left|\sum_{m\in\mathcal{M}_t}\sqrt{p_{mq}}\mathbf{h}^H_{mu}\bar{\mathbf{f}}_{mq}\right|^2+\sigma_u^2}\geq\gamma_u, \tag{5.44b}$$

$$\sum_{s\in\mathcal{S}} p_{ms}\leq P_m, \quad \forall m\in\mathcal{M}_t, \tag{5.44c}$$

Next, to simplify the expression, we define the effective channel of UE u and AP m due to the stream u' as $\rho_{muu'}=\mathbf{h}^H_{mu}\bar{\mathbf{f}}_{mu'}$ and the sensing channel gain due to the stream s of AP m as $\varrho_{ms}=\left|\mathbf{a}^H(\theta_m)\bar{\mathbf{f}}_{ms}\right|^2\sum_{m_r\in\mathcal{M}_r}\zeta^2_{mm_r}$. In terms of these variables, we can rewrite the problem in (5.44) as

$$\text{(P3.1b):} \max_{\{p_{ms}\}} \quad \sum_{m\in\mathcal{M}_t}\sum_{s\in\mathcal{S}} p_{ms}\varrho_{ms} \tag{5.45a}$$

$$\text{s.t.} \quad {\gamma_u}^{-1}\left|\sum_{m\in\mathcal{M}_t}\sqrt{p_{mu}}\rho_{muu}\right|^2\geq\sum_{s\in\mathcal{S}\setminus\{u\}}\left|\sum_{m\in\mathcal{M}_t}\sqrt{p_{ms}}\rho_{mus}\right|^2+\sigma_u^2, \quad \forall u\in\mathcal{U}, \tag{5.45b}$$

$$\sum_{s\in\mathcal{S}} p_{ms}\leq P_m, \quad \forall m\in\mathcal{M}_t, \tag{5.45c}$$

This problem, however, is difficult to solve since (i) includes the square root of the power terms, i.e., $\{\sqrt{p_{ms}}\}$, and (ii) contains the summation inside the absolute terms. For (i), we can write the problem in terms of the square root power terms, $\{\sqrt{p_{ms}}\}$. For (ii), we define the per-stream vector form of the power coefficients by stacking the power coefficients of every AP for a given stream, similar to the one applied in (5.7), given as $\mathbf{p}_s=[\sqrt{p_{1s}},\ldots,\sqrt{p_{Ms}}]^T$. To complement this variable in our new formulation, we also define the vectors $\boldsymbol{\rho}_{us}=[\rho_{1us},\ldots,\rho_{Mus}]$, and $\boldsymbol{\varrho}_s=[\varrho_{1s},\ldots,\varrho_{Ms}]$. In addition, we define

the AP selection matrix for the power allocation formulation, $\widetilde{\mathbf{D}}_m \in \mathbb{R}^{M\times M}$, where each element of this matrix is given as

$$[\widetilde{\mathbf{D}}_m]_{ij} = \begin{cases} 1 & \text{if } i=j=m, \\ 0 & \text{otherwise.} \end{cases} \tag{5.46}$$

This variable allows rewriting the power constraint in terms of the stacked variable. Then, we can re-write the problem in terms of the newly defined variables as

$$\text{(P3.1c):} \max_{\{\mathbf{p}_s\}} \quad \sum_{s\in\mathcal{S}} (\mathbf{p}_s \odot \mathbf{p}_s)^T \boldsymbol{\varrho}_s \tag{5.47a}$$

$$\text{s.t.} \quad \gamma_u^{-1} \left|\mathbf{p}_u^T \boldsymbol{\rho}_{uu}\right|^2 \geq \sum_{s\in\mathcal{S}\setminus\{u\}} \left|\mathbf{p}_s^T \boldsymbol{\rho}_{us}\right|^2 + \sigma_u^2, \quad \forall u \in \mathcal{U}, \tag{5.47b}$$

$$\sum_{s\in\mathcal{S}} \left\|\widetilde{\mathbf{D}}_m \mathbf{p}_s\right\|^2 \leq P_m, \quad \forall m \in \mathcal{M}_t, \tag{5.47c}$$

where the problem is a non-convex QCQP due to the maximization of the sensing SNR, which is a quadratic function of the variable $\mathbf{p}$. Similar to the Section 5.6, we can transform it into an SDP and apply SDR. For this purpose, we define the new optimization variables for the SDP, i.e., $\mathbf{P}_s = \mathbf{p}_s\mathbf{p}_s^T$, which, by definition, introduces two constraints on the problem (i) The convex symmetric positive semi-definiteness constraint $\mathbf{P}_s \in \mathbb{S}^+$, where $\mathbb{S}^+$ is the set of symmetric positive semidefinite matrices, and (ii) the non-convex rank-1 constraint $\mathrm{rank}(\mathbf{P}_s) = 1$. To complement this variable in our formulation, we also define $\boldsymbol{\Gamma}_{us} = \boldsymbol{\rho}_{us}\boldsymbol{\rho}_{us}^H$, and $\boldsymbol{\Gamma}'_s = \mathrm{diag}(\boldsymbol{\varrho}_s)$. Then, (5.47) can be written as an SDP in terms of these

variables, given by

$$\text{(P3.1-SDP):} \max_{\{\mathbf{P}_s\}} \quad \sum_{s\in\mathcal{S}} \mathrm{Tr}\left(\mathbf{P}_s\mathbf{\Gamma}'_s\right) \tag{5.48a}$$

$$\text{s.t.} \quad \gamma_u^{-1}\mathrm{Tr}\left(\mathbf{P}_u\mathbf{\Gamma}_{uu}\right) \geq \sum_{s\in\mathcal{S}\setminus\{u\}} \mathrm{Tr}\left(\mathbf{P}_s\mathbf{\Gamma}_{us}\right) + \sigma_u^2, \quad \forall u \in \mathcal{U}, \tag{5.48b}$$

$$\sum_{s\in\mathcal{S}} \mathrm{Tr}\left(\mathbf{P}_s\widetilde{\mathbf{D}}_m\right) \leq P_m, \quad \forall m \in \mathcal{M}_t, \tag{5.48c}$$

$$\mathbf{P}_s \in \mathbb{S}^+ \tag{5.48d}$$

$$\mathrm{rank}(\mathbf{P}_s) = 1 \tag{5.48e}$$

The given problem is non-convex due to the rank-1 constraint. To obtain a convex problem, we apply SDR by removing this constraint. Then, the relaxed formulation for the power allocation can be given as

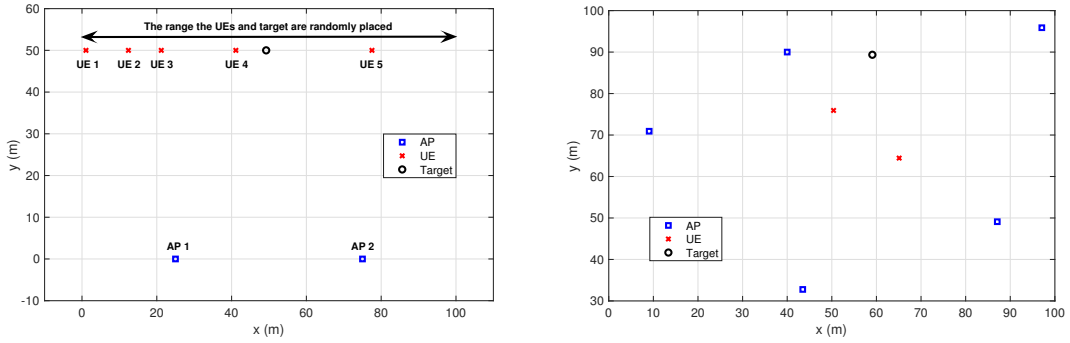
$$\text{(P3.1-SDR):} \max_{\{\mathbf{P}_s\}} \quad \sum_{s\in\mathcal{S}} \mathrm{Tr}\left(\mathbf{P}_s\mathbf{\Gamma}'_s\right) \tag{5.49a}$$

$$\text{s.t.} \quad \gamma_u^{-1}\mathrm{Tr}\left(\mathbf{P}_u\mathbf{\Gamma}_{uu}\right) \geq \sum_{s\in\mathcal{S}\setminus\{u\}} \mathrm{Tr}\left(\mathbf{P}_s\mathbf{\Gamma}_{us}\right) + \sigma_u^2, \quad \forall u \in \mathcal{U}, \tag{5.49b}$$

$$\sum_{s\in\mathcal{S}} \mathrm{Tr}\left(\mathbf{P}_s\widetilde{\mathbf{D}}_m\right) \leq P_m, \quad \forall m \in \mathcal{M}_t, \tag{5.49c}$$

$$\mathbf{P}_s \in \mathbb{S}^+ \tag{5.49d}$$

where the solution can be obtained by the convex solvers. This, however, only results in the matrices $\{\mathbf{P}_s^\star\}$, which is not necessarily rank-1, and the reconstruction of the individual power variables, $\{\mathbf{p}_s^\star\}$, is required. To reconstruct the solution, one can apply a heuristic approach inspired by the solution in Section 5.6 or develop different approaches with the randomization techniques [107]. For our purposes of evaluating the beamforming against the power optimization, we adopt a heuristic method by utilizing the most significant eigenvector for each beam, and also rely on the solution obtained by the (P3.1-SDR), which provides an upper bound on the power optimization. With the solution completed, we next evaluate our results.



(a) Line setup

(b) Square setup

Figure 5.2: The Simulation Placement Is Illustrated. For Different Realizations, the AP Positions Are Fixed. In (a), the UEs and Target Are Randomly Placed Over the y-Axis, While in (B), the UEs and Target Are Randomly Placed Over the Square Area of $100\text{m}\times100\text{m}$.

5.8 Results

In this section, we evaluate the performance of the proposed beamforming solutions for cell-free ISAC MIMO systems.

5.8.1 Evaluated Solutions

For this setup, we compare the following solutions:

(i) **NS Sensing - RZF Comm** which designs the sensing beam as conjugate beamforming projected on the null space of the communication channels as in (5.16) and implements the communications beams according to the RZF design in (5.17).

(ii) **NS Sensing - OPT Comm** which has the same sensing beam design as in (i) but designs the communication beam based on the max-min optimization in (5.20).

(iii) **CB Sensing - OPT Comm** which first designs the sensing beam as the conjugate beamforming in (5.15) and then designs the communication beams to solve the max-min optimization in (5.20).

(iv) **JSC Beam Optimization** which implements the communication and sensing beams

based on the SDR problem in (5.31) and Proposition 1, which jointly optimizes the beamforming vectors based on the communication and sensing functions.

(v) **JSC Power Optimization** which implements the communication and sensing beam powers based on the SDR problem in (5.49), that jointly optimizes the power coefficients for the given beams based on the communication and sensing functions. The beamforming vectors of this solution are taken as in the (i) *NS Sensing - RZF Comm* approach.

(vi) **JSC Beam SDR UB** which applies the matrix solution for the communication and sensing beams based on the SDR problem in (5.31). There is no rank constraint on the beams; hence, it is an upper bound.

(vii) **JSC Power SDR UB** which applies the matrix solution for the communication and sensing beam powers based on the SDR problem in (5.49), while the communication beams are set according to the RZF design in (5.17) and a single sensing beam as the NS sensing in (5.16). There is no rank constraint on the power variables; hence, it is an upper bound for the power allocation methods.

5.8.2 LoS Channels

In particular, we consider a scenario where $\mathcal{M}_t = \mathcal{M}_r$ with two APs placed at $(25, 0)$ and $(75, 0)$ in the Cartesian coordinates, as shown in Figure 5.2(a). Each AP is equipped with a uniform linear array (ULA) along the x axis of $N_t = N_r = 16$ antennas. At $y = 50m$, we randomly place one sensing target and the $U = 5$ communications users along the x-axis. Specifically, the x coordinates of these locations are drawn from a uniform distribution in $[0, 100]$. For the communication channels, we adopt a LOS channel model and take $\sigma_u^2 = 1$. For the sensing channels, we adopt the parameters $\varsigma_{m_r}^2 = 1$ and $\zeta_{m_t m_r} = 0.1$. The transmit power of the APs is $P_m = 0$dBW and the number of sensing streams

$Q = 1$. In the following, we average the results over 1000 realizations.

Providing NS Sensing - OPT Comm SINR for All UEs

With the defined setup, we first focus on an equal rate case, i.e., γ_u is the same for every UE. For the selection of this value, we adopt the minimum UE SINR obtained from solution (ii). Further, we do not include the power optimization solution as it is not able to satisfy the SINR constraints for the most cases with $\gamma > 0.2$.

Sensing and Communication Power Allocation: We first investigate the sensing and communication performance for different power allocation ratios. Specifically, in Figure 5.3, we show the sensing SNR and minimum communication SINR of UEs achieved by the different beamforming solutions for different values of $\rho \in (0, 1)$. It is important to note here that for the beamforming solutions (i)-(iii), the communication and sensing beams are separately designed, and we directly allocate the communication and sensing powers based on the ratio ρ. For the JSC beam optimization solution (iv), it implements the beamforming design in Proposition 1, which optimizes both the structure of the beams and the power allocation. Therefore, and for the sake of comparing with the other approaches, we plot the JSC optimization curve in Figure 5.3 by setting the communication SINR threshold to be equal to the achieved SINR by solution (ii). This still respects the total power constraint, which is taken care of by (5.29). As seen in the figure, the first two solutions, (i) and (ii), achieve better communication SINR and less sensing SNR compared to solution (iii). This is expected as solution (iii) aims to maximize the sensing performance, irrespective of the communication, and hence, it causes some interference to the communication users. *Interestingly, while achieving the best communication performance of the separate solutions, the joint solution provides very similar sensing performance to the MF sensing.* This highlights the gain of the developed JSC beamforming design.

Target distance to closest UE: To further investigate how the different beamforming

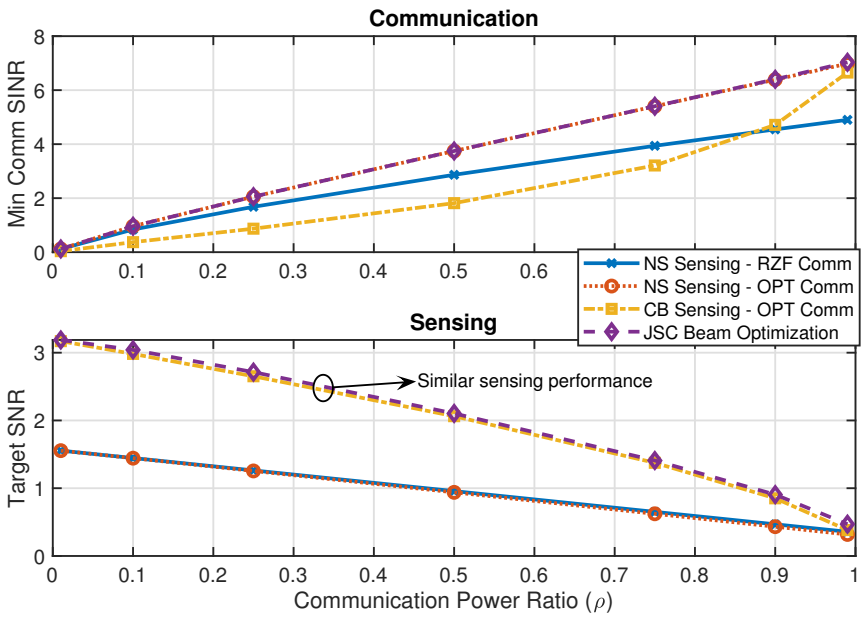


Figure 5.3: Performance of the Solutions for Different Power Allocation Ratios for the Communications and Sensing. The Proposed JSC Optimization Provides a Significant Gain for Sensing While Satisfying the Best Communication SINR.

approaches impact the trade-off between the sensing and communication performance, we evaluate this performance versus the distance between the sensing target and closest communication UE in Figure 5.4. Note that, intuitively, as the sensing target gets closer to the communication users, the overlap between the communication and sensing channels' subspaces increases, which can benefit or penalize the communication and sensing performance depending on the beamforming design. In Figure 5.4, we set the power ratio as 0.5 for the communication and sensing operation. This figure shows that for the smaller distances/separation between the sensing target and communication users, the conjugate beamforming sensing solution (solution (iii)) optimizes the sensing performance but causes non-negligible interference to the communication, which significantly degrades its performance. On the other side, solutions (i) and (ii), which prioritize the communication and keep the sensing beamforming in the null-space of the communication channels, optimize the communication SINR and degrade the sensing SNR. For the SINR constraint of the JSC

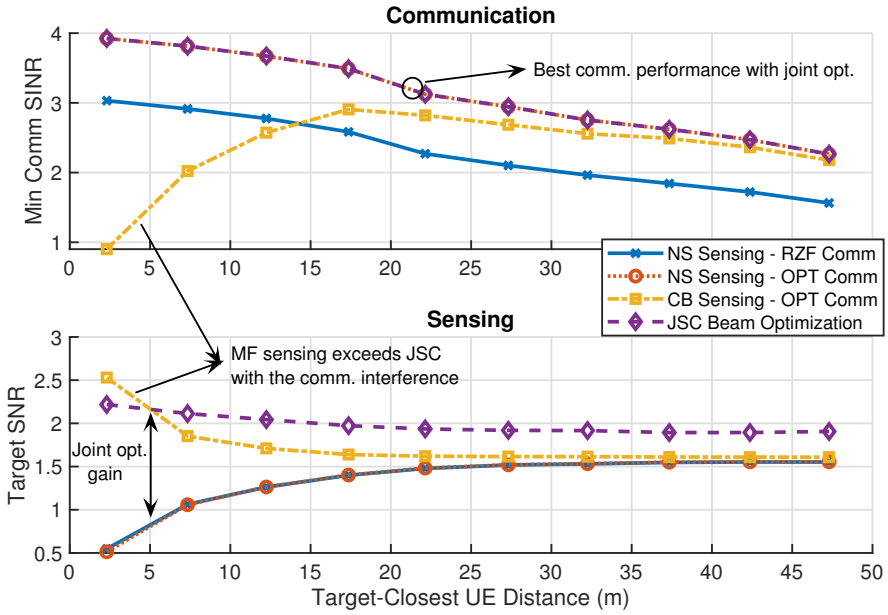


Figure 5.4: Performance of the Solutions Versus the Distance Between the Target and Closest AP. The Proposed JSC Optimization Provides Almost a Constant Sensing SNR for Different Distances, With a Significant Gain Over the NS Solutions.

optimization, we again adopt the SINR obtained by solution (ii), which achieves the best communication performance. Hence, the achieved communication SINR of this solution and JSC beam optimization are the same. *The sensing SNR, however, enjoys the advantage of the joint beam optimization. Specifically, it provides almost a constant sensing performance for different target-closest UE distances: Achieving a close sensing performance to solution (i) when the separation between the sensing target and communication users is small and exceeds the performance of all the other three solutions when this separation is large, which highlights the potential of the joint beamforming design.*

Providing NS Sensing - RZF Comm SINR for each UE

Now, to further investigate the performance of the joint optimization, we select each γ_u individually as their SINR obtained from solution (i). Differently from the previous approach (minimum communication rate), we test the imbalanced rates obtained from (i), and eval-

uate the the mean communication rate, instead of the minumum. In Figure 5.5, we provide the closest distance figure with these SINR constraints. In the figure, the beamforming (with $Q = 0$) and power solutions with corresponding upper bounds, (iv) with (vi) and (v) with (vii), achieve the same results, hence only (iv) and (v) are illustrated. Compared to the previous figure, the mean SINR of solution (i) shows less degradation with larger distances, thanks to being able to exploit the imbalance for the mean rates. On the other hand, the beamforming optimization (iv) achieves the same SINR and provides a significant sensing SNR gain over all other solutions without any sensing beams, as indicated by Proposition 4. This shows a similar advantage to the previous case. For very small target-closest UE distances, (iv) the beamforming optimization is worse for sensing than (iii) CB Sensing - OPT Comm, which cannot achieve similar communication SINRs due to the high interference. At the communication part, (v) the power optimization solution achieves higher average communication SINR with very close distances since allocating the power onto the UE with the closest distance provides more gain for sensing than the NS sensing beam. In the general case, however, (iv) the beamforming optimization provides significant sensing gain over all the solutions, showing a similar pattern to the previous case.

5.8.3 Rayleigh Channels

As we have only investigated a simplified setup so far to examine the effects, we now provide a more realistic setup. In this setup, the AP and UEs are placed over a square area of 100m$\times 100$m. We utilize the $f_c = 28$GHz band and place $M_t = M_r = 5$ APs, each equipped with a ULA of $N_t = 8$ antennas. To show the need for additional sensing streams, based on our observations in Section 5.6, we take the number of UEs as 2. The setup is illustrated in Figure 5.2(b). For the path-loss, we adopt the 3GPP UMi path loss [38] given as PL $= -32.4 - 21\log_{10}(\text{distance}) - 20\log_{10}(f_c)$. Further, we assume Rayleigh fading for the AP-UE channels. The receiver noise at the UEs is taken as -135dBm.

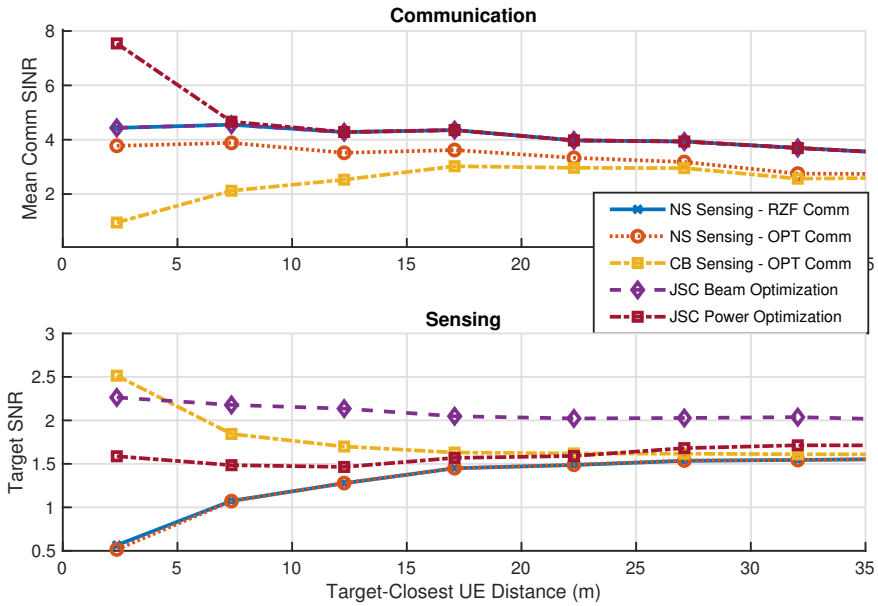


Figure 5.5: Performance of the Mean SINR Versus the Distance Between the Target and Closest AP. The Optimizations Are Carried Out to Satisfy Individual Rates Achieved by RZF. A Similar Pattern to the Previous Case Is Observed.

To evaluate the performance in this setup, we set the minimum communication SINR threshold as 10dB for all UEs. Within the evaluations, we only use solutions beam and power optimization solutions, as they are able to conform to the SINR constraints while maximizing the sensing SNR. In Figure 5.6, we show the achieved sensing SNR and communication SINR values with different number of UEs. As expected from Proposition 4, we need $Q = M - U$ streams to achieve the beamforming upperbound. To that end, the sensing SNR provided by any beamforming optimization (iv) curve with $Q \geq M - U$ achieves the same value, while satisfying the communication constraints. The power optimization can exceed the solution obtained by a single sensing beam with $U \leq 2$, which shows the case it may be preferable if the sensing streams are limited, and there are more APs than UEs. This result also highlight the take advantage of the beamforming solution. In addition, we note that the provided heuristic rank-1 optimization does not provide sufficient SNR, and better heuristic methods for the rank-1 constructions are required. We, however, leave this

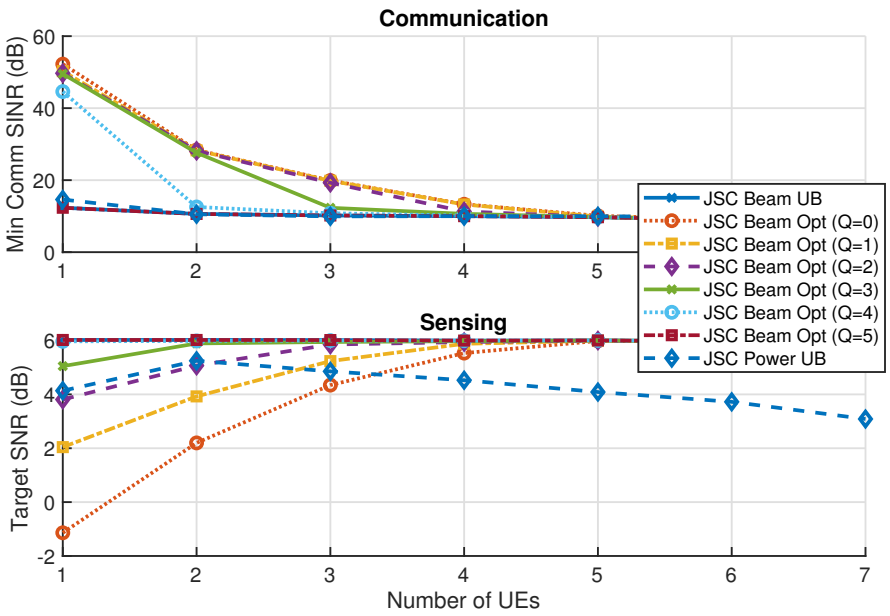


Figure 5.6: Performance of the SDR Based Optimization Solutions With Varying Number of Sensing Streams and Number of UEs.

as a future work for researchers as our focus of the work is to investigate JSC beamforming, and also compare it with the potential of power allocation. As shown in the figure, adding sensing streams can provide advantages for the cell-free massive MIMO systems, and this can allow further gain over the suboptimal beams with power allocation.

5.9 Summary

In this chapter, we investigated downlink beamforming for the joint sensing and communication in cell-free massive MIMO systems. Specifically, we designed communication-prioritized sensing beamforming and sensing-prioritized communication beamforming solutions as the baseline. Further, we have developed an optimal solution for the JSC beamforming. The results showed the advantage of the joint optimization, where the developed JSC beamforming is capable of achieving nearly the SINR that of the communication-prioritized sensing beamforming solutions with almost the same sensing SNR of the sensing-prioritized communication beamforming approaches.

Chapter 6

SUMMARY AND FUTURE WORK

6.1 Summary

Chapter 2 offers innovative solutions to address the challenges posed by fiber-based fronthaul in cell-free massive MIMO systems, enabling its scalability. Specifically, wireless fronthaul and wired/wireless mixed fronthaul architectures are proposed for the cell-free massive MIMO. With the formulation of an end-to-end data rate optimization problem and proposed low-complexity joint beamforming and resource allocation solution, the new architectures demonstrated the potential to achieve data rates comparable to optical fiber-based fronthaul under realistic conditions. This presents a promising pathway for realizing the benefits of cell-free massive MIMO in practical deployments, effectively mitigating infrastructure and deployment challenges. The chapter showed the scalability of these architectures, highlighting their capability to support a large number of access points while meeting practical fronthaul bandwidth requirements. Overall, the introduced wireless-fronthaul based architectures offer a way to achieve the theoretical gains in scalable practical implementation of cell-free massive MIMO systems.

Chapter 3 unveils a pioneering real-world demonstration employing machine learning for radar-aided beam prediction in a practical vehicular communication scenario. By harnessing radar sensory data at communication terminals, crucial awareness of transmitter/receiver locations and the surrounding environment is obtained. This awareness proves instrumental in mitigating or eliminating beam training overhead in millimeter wave (mmWave) and sub-terahertz (THz) MIMO communication systems, thereby unlocking possibilities for highly-mobile, low-latency applications. The deep learning-based radar-

aided beam prediction approaches developed in this chapter leverage domain knowledge in radar signal processing to optimize performance, complexity, and inference time. Utilizing the DeepSense 6G real-world dataset, which integrates mmWave beam training and radar measurements, the proposed algorithms not only eliminate radar/communication calibration overhead but also achieve an impressive top-5 beam prediction accuracy of approximately 90%, while saving a remarkable 93% of beam training overhead. This chapter marks a promising leap forward in addressing beam management challenges in mmWave/THz communication systems, showcasing the feasibility and efficacy of radar-aided mmWave beam prediction in real-world applications. Furthermore, the extension of the beam prediction framework to include beam tracking reinforces the potential of this approach for enabling highly mobile mmWave/sub-THz communication applications.

Chapter 4 introduced radar-aided blockage prediction approaches for millimeter wave (mmWave) and terahertz communication systems, addressing the challenges posed by line-of-sight (LOS) links' sensitivity to blockages. Leveraging radar sensors to gather crucial information about the surrounding environment and moving objects, two solutions are proposed proactively predict future link blockages: classical radar object tracking and deep neural networks. The evaluation, conducted on a large-scale real-world dataset from the DeepSense dataset, demonstrated that both approaches can predict future blockages one second before they occur with over 90% F1 score and accuracy. In addition, the deep learning solution showed superior blockage prediction results with lower design complexity compared to the object tracking method. These results highlight the promising potential of utilizing low-cost radar sensors to enhance the reliability of mmWave and terahertz communication systems by proactively predicting and mitigating blockages.

Chapter 5 explores the design to achieve joint sensing and communication (JSC) in cell-free ISAC MIMO systems, where distributed APs simultaneously serve communication users and sense targets. Specifically, the chapter aims to develop solutions for the beam-

forming and power allocation. Initially, two baseline approaches are developed, focusing on separate design considerations for sensing and communication beamforming vectors: communication-prioritized sensing beamforming and sensing-prioritized communication beamforming. Then, the chapter proposes a joint sensing and communication beamforming design problem based on a max-min fairness formulation, and derives the optimal solution. Additionally, leveraging the derived beamforming formulation, a power allocation approach for pre-determined sensing and communication beams is proposed. The evaluation demonstrated that the developed JSC beamforming achieves nearly the same communication signal-to-interference-plus-noise ratio (SINR) as communication-prioritized sensing beamforming, maintaining a comparable sensing SNR to sensing prioritized communication beamforming. The JSC beamforming optimization also provided remarkable sensing gains over power allocation, presenting a promising strategy for cell-free ISAC MIMO systems and further enabling ISAC for future communication systems.

6.2 Future Work

For future work, it is interesting to deepen the investigation on cell-free ISAC MIMO systems. Specifically, the results of this dissertation has revealed following directions for future research.

- Bringing the sensing for communication further into real-world, joint design of user identification, beam tracking, blockage prediction and handover would be interesting to develop. Specifically, support for tracking multiple users, their beamforming vectors and potential blockages can be provided via identification of the communication user. To that end, all these problems can be resolved with a joint design, which can enhance the performance and minimize the latency.

- It is interesting to utilize machine learning in the design of cell-free integrated sensing

and communication systems since the computation and communication complexities of these systems increase significantly with a larger number of arrays, which is a necessity for the cell-free massive MIMO systems. With machine learning, it can be possible to efficiently resolve the design problems with low computation and communication complexities.

- The joint processing of the radar signals causes significant overhead to the cell-free MIMO systems and suffers from significant fronthaul requirements for the dense and distributed placement. Therefore, it is interesting to extend the wireless fronthaul for cell-free ISAC MIMO systems, and develop communication-limited joint/distributed signal processing approaches for this architecture.

- The sensing aided communication approaches can also be realized in distributed MIMO system. Achieving this, it would be interesting to develop sensing-aided communication solutions for the cell-free ISAC MIMO systems. Development of such system would entail sensing through bi-static or multi-static radars, and then tracking the objects/vehicles within the measurements of this sensing and designing both communication and sensing based on these observations. With that, it would be possible to realize the gains of sensing-aided communication for next-generation communication and sensing systems.

REFERENCES

[1] Hamed Masoumi and Mohammad Javad Emadi. Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments. *IEEE Transactions on Wireless Communications*, 19(2):1038–1053, 2019.

[2] Guillem Femenias and Felip Riera-Palou. Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity. *IEEE Access*, 7:44596–44612, 2019.

[3] Guillem Femenias and Felip Riera-Palou. Reduced-complexity downlink cell-free mmWave massive MIMO systems with fronthaul constraints. In *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, pages 1–5, A Coruna, Spain, September 2019.

[4] Priyabrata Parida, Harpreet S Dhillon, and Andreas F Molisch. Downlink performance analysis of cell-free massive MIMO with finite fronthaul capacity. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pages 1–6. IEEE, 2018.

[5] Ahmed Alkhateeb, Sam Alex, Paul Varkey, Ying Li, Qi Qu, and Djordje Tujkovic. Deep learning coordinated beamforming for highly-mobile millimeter wave systems. *IEEE Access*, 6:37328–37348, 2018.

[6] Fan Liu, Weijie Yuan, Christos Masouros, and Jinhong Yuan. Radar-assisted predictive beamforming for vehicular links: Communication served by sensing. *IEEE Transactions on Wireless Communications*, 19(11):7704–7719, 2020.

[7] Anum Ali, Nuria González-Prelcic, and Amitava Ghosh. Millimeter wave V2I beam-training using base-station mounted radar. In *2019 IEEE Radar Conference (RadarConf)*, pages 1–5. IEEE, 2019.

[8] Hien Quoc Ngo, Alexei Ashikhmin, Hong Yang, Erik G Larsson, and Thomas L Marzetta. Cell-free massive MIMO versus small cells. *IEEE Transactions on Wireless Communications*, 16(3):1834–1850, March 2017.

[9] Giovanni Interdonato, Emil Björnson, Hien Quoc Ngo, Pål Frenger, and Erik G Larsson. Ubiquitous cell-free massive MIMO communications. *EURASIP J. Wireless Commun. and Netw.*, 2019(1):197, August 2019.

[10] Sivarama Venkatesan, Angel Lozano, and Reinaldo Valenzuela. Network MIMO: Overcoming intercell interference in indoor wireless systems. In *Conf. Rec. 41st Asilomar Conf. Signals, Syst., Comput.*, pages 83–87, Pacific Grove, CA, USA, November 2007.

[11] Dongming Wang, Jiangzhou Wang, Xiaohu You, Yan Wang, Ming Chen, and Xiaoyun Hou. Spectral efficiency of distributed MIMO systems. *IEEE Journal on Selected Areas in Communications*, 31(10):2112–2127, October 2013.

[12] Ralf Irmer, Heinz Droste, Patrick Marsch, Michael Grieger, Gerhard Fettweis, Stefan Brueck, Hans-Peter Mayer, Lars Thiele, and Volker Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *IEEE Communications Magazine*, 49(2):102–111, February 2011.

[13] Patrick Marsch and Gerhard P Fettweis. *Coordinated Multi-Point in Mobile Communications: from theory to practice*. Cambridge University Press, 2011.

[14] Hien Quoc Ngo, Le-Nam Tran, Trung Q Duong, Michail Matthaiou, and Erik G Larsson. On the total energy efficiency of cell-free massive MIMO. *IEEE Trans. Green Commun. Netw.*, 2(1):25–39, March 2017.

[15] Elina Nayebi, Alexei Ashikhmin, Thomas L Marzetta, Hong Yang, and Bhaskar D Rao. Precoding and power optimization in cell-free massive MIMO systems. *IEEE Transactions on Wireless Communications*, 16(7):4445–4459, July 2017.

[16] Stefano Buzzi and Carmen D'Andrea. Cell-free massive MIMO: User-centric approach. *IEEE Wireless Commun. Lett.*, 6(6):706–709, December 2017. doi: 10.1109/LWC.2017.2734893.

[17] Muhammad Alrabeiah and Ahmed Alkhateeb. Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency. In *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, pages 1465–1470, Pacific Grove, CA, USA, November 2019.

[18] Manijeh Bashar, Kanapathippillai Cumanan, Alister G Burr, Hien Quoc Ngo, Mérouane Debbah, and Pei Xiao. Max–min rate of cell-free massive MIMO uplink with optimal uniform quantization. *IEEE Transactions on Communications*, 67 (10):6796–6815, October 2019.

[19] Manijeh Bashar, Kanapathippillai Cumanan, Alister G Burr, Hien Quoc Ngo, and Mérouane Debbah. Cell-free massive MIMO with limited backhaul. In *IEEE Int. Conf. on Commun. (ICC)*, pages 1–7, Kansas City, MO, USA, May 2018.

[20] Yun Zhu, Yong Niu, Jiade Li, Dapeng Oliver Wu, Yong Li, and Depeng Jin. QoS-aware scheduling for small cell millimeter wave mesh backhaul. In *IEEE Int. Conf. on Commun. (ICC)*, pages 1–6, Kuala Lumpur, Malaysia, May 2016.

[21] Wanming Hao and Shouyi Yang. Small cell cluster-based resource allocation for wireless backhaul in two-tier heterogeneous networks with massive MIMO. *IEEE Transactions on Vehicular Technology*, 67(1):509–523, January 2018.

[22] Zhen Gao, Linglong Dai, De Mi, Zhaocheng Wang, Muhammad Ali Imran, and Muhammad Zeeshan Shakir. MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network. *IEEE Wireless Communications*, 22(5):13–21, October 2015.

[23] Tadilo Endeshaw Bogale and Long Bao Le. Massive MIMO and mmWave for 5G wireless HetNet: Potential benefits and challenges. *IEEE Veh. Technol. Mag.*, 11(1): 64–75, March 2016.

[24] Reuben George Stephen and Rui Zhang. Joint millimeter-wave fronthaul and OFDMA resource allocation in ultra-dense CRAN. *IEEE Transactions on Communications*, 65(3):1411–1423, March 2017.

[25] Bin Hu, Cunqing Hua, Jun Zhang, Cailian Chen, and Xinping Guan. Joint fronthaul multicast beamforming and user-centric clustering in downlink C-RANs. *IEEE Transactions on Wireless Communications*, 16(8):5395–5409, August 2017.

[26] Bin Hu, Cunqing Hua, Cailian Chen, and Xinping Guan. Joint beamformer design for wireless fronthaul and access links in C-RANs. *IEEE Transactions on Wireless Communications*, 17(5):2869–2881, May 2018.

[27] Ahmed Alkhateeb, Geert Leus, and Robert W. Heath. Multi-layer precoding for full-dimensional massive MIMO systems. In *Proc. 48th Asilomar Conf. Signals, Syst., Comput.*, pages 815–819, November 2014. doi: 10.1109/ACSSC.2014.7094563.

[28] Thomas Pfeiffer. Next generation mobile fronthaul and midhaul architectures. *J. Opt. Commun. Netw.*, 7(11):B38–B45, November 2015.

[29] Emil Björnson and Luca Sanguinetti. Scalable cell-free massive MIMO systems. *IEEE Transactions on Communications*, 68(7):4247–4261, July 2020.

[30] Deep Shrestha, Zhibo Pang, and Dacfey Dzung. Precise clock synchronization in high performance wireless communication for time sensitive networking. *IEEE Access*, 6:8944–8953, 2018.

[31] Ryan Rogalin, Ozgun Y Bursalioglu, Haralabos Papadopoulos, Giuseppe Caire, Andreas F Molisch, Antonios Michaloliakos, Vlad Balan, and Konstantinos Psounis. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. *IEEE Transactions on Wireless Communications*, 13(4):1815–1831, April 2014.

[32] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):436–453, April 2016.

[33] Aissa Ikhlef, Diomidis S Michalopoulos, and Robert Schober. Max-max relay selection for relays with buffers. *IEEE Transactions on Wireless Communications*, 11 (3):1124–1135, 2012.

[34] Zihuan Wang, Qian Liu, Ming Li, and Wolfgang Kellerer. Energy efficient analog beamformer design for mmWave multicast transmission. *IEEE Trans. Green Commun. Netw.*, 3(2):552–564, June 2019.

[35] Jung-Chieh Chen and Jyh-Horng Wen. Improved analog beamformer design for millimeter-wave multicast transmission with large antenna arrays. *IEEE Communications Letters*, August 2020.

[36] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[37] Giovanni Interdonato, Pal Frenger, and Erik G Larsson. Scalability aspects of cell-free massive MIMO. In *IEEE Int. Conf. on Commun. (ICC)*, pages 1–6, Shanghai, China, May 2019. IEEE.

[38] 3GPP TR 38.901 version 14.1.1 Release 14. Study on channel model for frequencies from 0.5 to 100 GHz. Technical report, , 2017.

[39] Xiaofeng Li and Ahmed Alkhateeb. Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 800–805, Pacific Grove, CA, USA, November 2019. doi: 10.1109/IEEECONF44664.2019.9048966.

[40] Muhammad Alrabeiah, Yu Zhang, and Ahmed Alkhateeb. Neural networks based beam codebooks: Learning mmWave massive MIMO beams that adapt to deployment and hardware. *arXiv preprint arXiv:2006.14501*, 2020.

[41] Yu Zhang and Ahmed Alkhateeb. Learning reflection beamforming codebooks for arbitrary RIS and non-stationary channels. *arXiv preprint arXiv:2109.14909*, 2021.

[42] Yu Zhang, Muhammad Alrabeiah, and Ahmed Alkhateeb. Reinforcement learning of beam codebooks in millimeter wave and terahertz MIMO systems. *IEEE Transactions on Communications*, pages 1–1, 2021. doi: 10.1109/TCOMM.2021.3126856.

[43] Brian P Ginsburg, Karthik Subburaj, Sreekiran Samala, Karthik Ramasubramanian, Jasbir Singh, Sumeer Bhatara, Sriram Murali, Dan Breen, Meysam Moallem, Krishnanshu Dandu, et al. A multimode 76-to-81GHz automotive radar transceiver with autonomous monitoring. In *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, pages 158–160. IEEE, 2018.

[44] Preeti Kumari, Junil Choi, Nuria González-Prelcic, and Robert W. Heath. Ieee 802.11ad-based radar: An approach to joint vehicular communication-radar system. *IEEE Transactions on Vehicular Technology*, 67(4):3012–3027, 2018. doi: 10.1109/TVT.2017.2774762.

[45] Abdelrahman Taha, Qi Qu, Sam Alex, Ping Wang, William L. Abbott, and Ahmed Alkhateeb. Millimeter wave MIMO-based depth maps for wireless virtual and augmented reality. *IEEE Access*, 9:48341–48363, 2021. doi: 10.1109/ACCESS.2021.3067839.

[46] Muhammad Alrabeiah and Ahmed Alkhateeb. Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels. *IEEE Transactions on Communications*, 68(9):5504–5518, 2020.

[47] Anum Ali, Nuria González-Prelcic, and Robert W. Heath. Millimeter wave beam-selection using out-of-band spatial information. *IEEE Transactions on Wireless Communications*, 17(2):1038–1052, 2018. doi: 10.1109/TWC.2017.2773532.

[48] Gouranga Charan, Tawfik Osman, Andrew Hredzak, Ngwe Thawdar, and Ahmed Alkhateeb. Vision-position multi-modal beam prediction using real millimeter wave datasets. *arXiv preprint arXiv:2111.07574*, 2021.

[49] Vutha Va, Takayuki Shimizu, Gaurav Bansal, and Robert W. Heath. Position-aided millimeter wave V2I beam alignment: A learning-to-rank approach. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–5, 2017. doi: 10.1109/PIMRC.2017.8292679.

[50] Muhammad Alrabeiah, Andrew Hredzak, and Ahmed Alkhateeb. Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5. IEEE, 2020.

[51] Weijie Yuan, Fan Liu, Christos Masouros, Jinhong Yuan, Derrick Wing Kwan Ng, and Nuria González-Prelcic. Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach. *IEEE Transactions on Wireless Communications*, 20(3):1442–1456, 2021. doi: 10.1109/TWC.2020.3033776.

[52] Ahmed Alkhateeb, Gouranga Charan, Tawfik Osman, Andrew Hredzak, João Morais, Umut Demirhan, and Nikhil Srinivas. Deepsense 6G: A large-scale real-world multi-modal sensing and communication dataset. *arXiv preprint arXiv:2211.09769*, 2022.

[53] Jau-Jr Lin, Yuan-Ping Li, Wei-Chiang Hsu, and Ta-Sung Lee. Design of an FMCW radar baseband signal processing system for automotive application. *SpringerPlus*, 5(1):1–16, 2016.

[54] Guoqiang Zhang, Haopeng Li, and Fabian Wenger. Object detection and 3D estimation via an FMCW radar using a fully convolutional network. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4487–4491. IEEE, 2020.

[55] B Dekker, S Jacobs, AS Kossen, MC Kruithof, AG Huizing, and M Geurts. Gesture recognition with a low power FMCW radar and a deep convolutional neural network. In *2017 European Radar Conference (EURAD)*, pages 163–166. IEEE, 2017.

[56] Xinrong Li, Xiaodong Wang, Qing Yang, and Song Fu. Signal processing for TDM MIMO FMCW millimeter-wave radar sensors. *IEEE Access*, 9:167959–167971, 2021.

[57] Cesar Iovescu and Sandeep Rao. The fundamentals of millimeter wave sensors. *Texas Instruments*, pages 1–8, 2017.

[58] Vivek Dham. Programming chirp parameters in ti radar devices. *Application Report SWRA553, Texas Instruments*, 2017.

[59] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, 2014.

[60] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12): 1565–1567, 2006.

[61] Robin Schubert, Eric Richter, and Gerd Wanielik. Comparison and evaluation of advanced motion models for vehicle tracking. In *2008 11th International Conference on Information Fusion*, pages 1–6. IEEE, 2008.

[62] Tokihiko Akita and Seiichi Mita. Object tracking and classification using millimeter-wave radar based on LSTM. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1110–1115. IEEE, 2019.

[63] Jae-Woo Choi, Si-Jung Ryu, and Jong-Hwan Kim. Short-range radar based real-time hand gesture recognition using LSTM encoder. *IEEE Access*, 7:33610–33618, 2019.

[64] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. *Neural computation*, 12(10):2451–2471, 2000.

[65] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[66] Theodore S Rappaport, Yunchou Xing, Ojas Kanhere, Shihao Ju, Arjuna Madanayake, Soumyajit Mandal, Ahmed Alkhateeb, and Georgios C Trichopoulos. Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond. *IEEE access*, 7:78729–78757, 2019.

[67] Ahmed Alkhateeb, Iz Beltagy, and Sam Alex. Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1055–1059, 2018. doi: 10.1109/GlobalSIP.2018.8646438.

[68] Muhammad Alrabeiah and Ahmed Alkhateeb. Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels. *IEEE Transactions on Communications*, 68(9):5504–5518, 2020.

[69] Shunyao Wu, Muhammad Alrabeiah, Chaitali Chakrabarti, and Ahmed Alkhateeb. Blockage prediction using wireless signatures: Deep learning enables real-world demonstration. *arXiv preprint arXiv:2111.08242*, 2021.

[70] Gouranga Charan, Muhammad Alrabeiah, and Ahmed Alkhateeb. Vision-aided 6G wireless communications: Blockage prediction and proactive handoff. *IEEE Transactions on Vehicular Technology*, pages 1–1, 2021. doi: 10.1109/TVT.2021.3104219.

[71] Shunyao Wu, Chaitali Chakrabarti, and Ahmed Alkhateeb. Lidar-aided mobile blockage prediction in real-world millimeter wave systems. *arXiv preprint arXiv:2111.09581*, 2021.

[72] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000.

[73] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. Millimeter wave mobile communications for 5G cellular: It will work! *IEEE access*, 1:335–349, 2013.

[74] Hermann Rohling. Radar cfar thresholding in clutter and multiple target situations. *IEEE transactions on aerospace and electronic systems*, (4):608–621, 1983.

[75] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[76] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[77] Prachi Vaishnav and Avik Santra. Continuous human activity classification with unscented kalman filter tracking using FMCW radar. *IEEE Sensors Letters*, 4(5): 1–4, 2020.

[78] Tara N Sainath, Oriol Vinyals, Andrew Senior, and Haşim Sak. Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4580–4584. IEEE, 2015.

[79] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. IEEE, 2017.

[80] Umut Demirhan and Ahmed Alkhateeb. Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration. *arXiv preprint arXiv:2111.09676*, 2021.

[81] Fan Liu, Christos Masouros, Athina P Petropulu, Hugh Griffiths, and Lajos Hanzo. Joint radar and communication design: Applications, state-of-the-art, and the road ahead. *IEEE Transactions on Communications*, 68(6):3834–3862, 2020.

[82] Fan Liu, Yuanhao Cui, Christos Masouros, Jie Xu, Tony Xiao Han, Yonina C Eldar, and Stefano Buzzi. Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond. *IEEE Journal on Selected Areas in Communications*, 2022.

[83] Umut Demirhan and Ahmed Alkhateeb. Integrated sensing and communication for 6G: Ten key machine learning roles. *IEEE Communications Magazine*, 2023.

[84] Umut Demirhan and Ahmed Alkhateeb. Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2655–2660. IEEE, 2022.

[85] Umut Demirhan and Ahmed Alkhateeb. Radar aided proactive blockage prediction in real-world millimeter wave systems. In *ICC 2022-IEEE International Conference on Communications*, pages 4547–4552. IEEE, 2022.

[86] Robert W Heath Jr, Tao Wu, Young Hoon Kwon, and Anthony CK Soong. Multiuser MIMO in distributed antenna systems with out-of-cell interference. *IEEE Transactions on Signal Processing*, 59(10):4885–4899, 2011.

[87] David A Schmidt, Changxin Shi, Randall A Berry, Michael L Honig, and Wolfgang Utschick. Comparison of distributed beamforming algorithms for MIMO interference networks. *IEEE Transactions on Signal Processing*, 61(13):3476–3489, 2013.

[88] Kan Zheng, Long Zhao, Jie Mei, Bin Shao, Wei Xiang, and Lajos Hanzo. Survey of large-scale MIMO systems. *IEEE Communications Surveys & Tutorials*, 17(3): 1738–1760, 2015.

[89] Gaurav Nigam, Paolo Minero, and Martin Haenggi. Coordinated multipoint joint transmission in heterogeneous networks. *IEEE Transactions on Communications*, 62(11):4134–4146, 2014.

[90] Hien Quoc Ngo, Alexei Ashikhmin, Hong Yang, Erik G Larsson, and Thomas L Marzetta. Cell-free massive MIMO versus small cells. *IEEE Transactions on Wireless Communications*, 16(3):1834–1850, 2017.

[91] Andong Zhou, Jingxian Wu, Erik G Larsson, and Pingzhi Fan. Max-min optimal beamforming for cell-free massive MIMO. *IEEE Communications Letters*, 24(10): 2344–2348, 2020.

[92] Umut Demirhan and Ahmed Alkhateeb. Enabling cell-free massive MIMO systems with wireless millimeter wave fronthaul. *IEEE Transactions on Wireless Communications*, 21(11):9482–9496, 2022.

[93] Fan Liu, Longfei Zhou, Christos Masouros, Ang Li, Wu Luo, and Athina Petropulu. Toward dual-functional radar-communication systems: Optimal waveform design. *IEEE Transactions on Signal Processing*, 66(16):4264–4279, 2018.

[94] Stefano Fortunati, Luca Sanguinetti, Fulvio Gini, Maria Sabrina Greco, and Braham Himed. Massive MIMO radar for target detection. *IEEE Transactions on Signal Processing*, 68:859–871, 2020.

[95] Xiang Liu, Tianyao Huang, Nir Shlezinger, Yimin Liu, Jie Zhou, and Yonina C Eldar. Joint transmit beamforming for multiuser MIMO communications and MIMO radar. *IEEE Transactions on Signal Processing*, 68:3929–3944, 2020.

[96] Ziyang Cheng, Zishu He, and Bin Liao. Hybrid beamforming design for OFDM dual-function radar-communication system. *IEEE Journal of Selected Topics in Signal Processing*, 15(6):1455–1467, 2021.

[97] Ahmad Bazzi and Marwa Chafii. On outage-based beamforming design for dual-functional radar-communication 6g systems. *IEEE Transactions on Wireless Communications*, 2023.

[98] Haocheng Hua, Jie Xu, and Tony Xiao Han. Optimal transmit beamforming for integrated sensing and communication. *arXiv preprint arXiv:2104.11871*, 2021.

[99] Yi Huang, Yuan Fang, Xinmin Li, and Jie Xu. Coordinated power control for network integrated sensing and communication. *arXiv preprint arXiv:2203.09032*, 2022.

[100] Na Zhao, Yunlong Wang, Zhibo Zhang, Qing Chang, and Yuan Shen. Joint transmit and receive beamforming design for integrated sensing and communication. *IEEE Communications Letters*, 26(3):662–666, 2022.

[101] Gaoyuan Cheng and Jie Xu. Coordinated transmit beamforming for multi-antenna network integrated sensing and communication. *arXiv preprint arXiv:2211.01085*, 2022.

[102] Zinat Behdad, Özlem Tuğfe Demir, Ki Won Sung, Emil Björnson, and Cicek Cavdar. Power allocation for joint communication and sensing in cell-free massive MIMO. *arXiv preprint arXiv:2209.01864*, 2022.

[103] Fan Liu, Ya-Feng Liu, Ang Li, Christos Masouros, and Yonina C. Eldar. Cramér-rao bound optimization for joint radar-communication beamforming. *IEEE Transactions on Signal Processing*, 70:240–253, 2022.

[104] Mark A Richards, Jim Scheer, William A Holm, and William L Melvin. *Principles of modern radar*, volume 1. Citeseer, 2010.

[105] Emil Björnson, Mats Bengtsson, and Björn Ottersten. Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]. *IEEE Signal Processing Magazine*, 31(4):142–148, 2014.

[106] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[107] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[108] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. SDPT3—a MATLAB software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

[109] Jie Xu, Liang Liu, and Rui Zhang. Multiuser MISO beamforming for simultaneous wireless information and power transfer. *IEEE Transactions on Signal Processing*, 62(18):4798–4810, 2014.

APPENDIX A

DERIVATION OF THE SENSING SNR

The expectation of the nominator can be simplified as

$$\mathbb{E}\left[\sum_{m_r\in\mathcal{M}_r}\left\|\overline{\mathbf{G}}_{m_r}\overline{\mathbf{X}}\right\|_{\mathrm{F}}^2\right] = \sum_{m_r\in\mathcal{M}_r}\mathrm{Tr}\left\{\mathbb{E}\left[\overline{\mathbf{X}}\,\overline{\mathbf{X}}^H\overline{\mathbf{G}}_{m_r}^H\overline{\mathbf{G}}_{m_r}\right\}\right] \tag{A.1}$$

$$= \sum_{m_r\in\mathcal{M}_r}\mathrm{Tr}\Bigg\{\underbrace{\mathbb{E}\left[\overline{\mathbf{X}}\,\overline{\mathbf{X}}^H\right]}_{=L\cdot\mathbf{I}}\mathbb{E}\left[\overline{\mathbf{G}}_{m_r}^H\overline{\mathbf{G}}_{m_r}\right]\Bigg\} \tag{A.2}$$

$$\begin{aligned}= L\sum_{m_r\in\mathcal{M}_r}\sum_{m_t\in\mathcal{M}_t}\Bigg(&\mathrm{Tr}\,\mathbb{E}\left[\overline{\mathbf{G}}_{m_tm_r}\overline{\mathbf{G}}_{m_tm_r}^H\right]\\ &+L\sum_{m_t'\in\mathcal{M}_t\setminus\{m_t\}}\mathrm{Tr}\,\mathbb{E}\left[\overline{\mathbf{G}}_{m_tm_r}\overline{\mathbf{G}}_{m_t'm_r}^H\right]\Bigg)\end{aligned} \tag{A.3}$$

where (A.1) and (A.2) are obtained by applying the expansion of the Frobenius norm, interchanging expectation and trace, and permutating the inner terms of the trace operation several times. To obtain (A.3), we apply the definition $\overline{\mathbf{G}}_{m_r}=\sum_{m_t}\overline{\mathbf{G}}_{m_tm_r}$ given in (5.13), and re-organize the multiplication terms. Further, for (A.3), due to the expectation over the random variables $\{\alpha_{m_tm_r}\}$ and independence of them, we have $\mathbb{E}\left[\overline{\mathbf{G}}_{m_tm_r}\overline{\mathbf{G}}_{m_t'm_r}^H\right]=0$, which makes the latter line of (A.3) zero. For the former, we have

$$\mathrm{Tr}\,\mathbb{E}\left[\overline{\mathbf{G}}_{m_tm_r}\overline{\mathbf{G}}_{m_tm_r}^H\right] = \mathrm{Tr}\,\mathbb{E}\left[|\alpha_{m_tm_r}|^2\,\mathbf{a}(\theta_{m_r})\mathbf{a}^H(\theta_{m_t})\overline{\mathbf{F}}_{m_t}\overline{\mathbf{F}}_{m_t}^H\mathbf{a}(\theta_{m_t})\mathbf{a}^H(\theta_{m_r})\right] \tag{A.4}$$

$$= \zeta_{m_tm_r}^2\mathrm{Tr}\,\mathbb{E}\left[\mathbf{a}^H(\theta_{m_t})\overline{\mathbf{F}}_{m_t}\overline{\mathbf{F}}_{m_t}^H\mathbf{a}(\theta_{m_t})\underbrace{\mathbf{a}^H(\theta_{m_r})\mathbf{a}(\theta_{m_r})}_{=N_r}\right] \tag{A.5}$$

$$= \zeta_{m_tm_r}^2N_r\left\|\mathbf{a}^H(\theta_{m_t})\overline{\mathbf{F}}_{m_t}\right\|^2. \tag{A.6}$$

For the denominator, we can write

$$\mathbb{E}\left[\sum_{m_r\in\mathcal{M}_r}\|\mathbf{N}_{m_r}\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[\sum_{m_r\in\mathcal{M}_r}\sum_{\ell=1}^{L}\|\mathbf{n}_{m_r}[\ell]\|^2\right] = LN_r\sum_{m_r\in\mathcal{M}_r}\varsigma_{m_r}^2. \tag{A.7}$$

Finally, combining (A.3), (A.6), and (A.7) in (5.14), we obtain the result.

APPENDIX B

DERIVATION OF THE SENSING SNR IN SDP FORM

To simplify the sensing SNR expression given in (5.14) in the SDP form, we can write the nominator as

$$\mathbb{E}\left[\sum_{m_r\in\mathcal{M}_r}\left\|\overline{\mathbf{G}}_{m_r}\overline{\mathbf{X}}\right\|^2\right] = N_rL\sum_{m_t\in\mathcal{M}_t}\sum_{m_r\in\mathcal{M}_r}\zeta_{m_t,m_r}^2\sum_{s\in\mathcal{S}}\left|\mathbf{a}^H(\theta_{m_t})\mathbf{f}_{m_ts}\right|^2 \tag{B.1}$$

$$= N_rL\sum_{m_t\in\mathcal{M}_t}\sum_{m_r\in\mathcal{M}_r}\sum_{s\in\mathcal{S}}\left|\zeta_{m_t,m_r}\overline{\mathbf{a}}^H\mathbf{D}_{m_t}\mathbf{f}_s\right|^2 \tag{B.2}$$

$$= N_rL\sum_{m_t\in\mathcal{M}_t}\sum_{m_r\in\mathcal{M}_r}\sum_{s\in\mathcal{S}}\mathrm{Tr}\left(\zeta_{m_t,m_r}^2\overline{\mathbf{a}}^H\mathbf{D}_{m_t}\mathbf{f}_s\mathbf{f}_s^H\mathbf{D}_{m_t}\overline{\mathbf{a}}\right) \tag{B.3}$$

$$= N_rL\sum_{m_t\in\mathcal{M}_t}\sum_{m_r\in\mathcal{M}_r}\sum_{s\in\mathcal{S}}\mathrm{Tr}\left(\zeta_{m_t,m_r}^2\mathbf{D}_{m_t}\overline{\mathbf{A}}\mathbf{D}_{m_t}\mathbf{F}_s\right) \tag{B.4}$$

$$= N_rL\mathrm{Tr}\left(\sum_{m_t\in\mathcal{M}_t}\mathbf{D}_{m_t}\overline{\mathbf{A}}\mathbf{D}_{m_t}\sum_{m_r\in\mathcal{M}_r}\zeta_{m_t,m_r}^2\sum_{s\in\mathcal{S}}\mathbf{F}_s\right) \tag{B.5}$$

$$= N_rL\mathrm{Tr}\left(\sum_{m_t\in\mathcal{M}_t}\mathbf{D}_{m_t}\overline{\mathbf{A}}\mathbf{D}_{m_t}\bar{\zeta}_{m_t}\sum_{s\in\mathcal{S}}\mathbf{F}_s\right) \tag{B.6}$$

$$= N_rL\mathrm{Tr}\left(\mathbf{A}\sum_{s\in\mathcal{S}}\mathbf{F}_s\right) \tag{B.7}$$

where we obtain (B.2) by the definitions of $\overline{\mathbf{a}}$, $\mathbf{D}_{m_t}$, and $\mathbf{f}_s$, (B.3) by $|\mathbf{X}|^2 = \mathrm{Tr}\left(\mathbf{X}\mathbf{X}^H\right)$, (B.4) by cyclic permutation property of the trace operation and the definitions of $\mathbf{F}_s$ and $\overline{\mathbf{A}}$, (B.5) by rearranging the summations, (B.6) by defining $\bar{\zeta}_{m_t} = \sum_{m_r\in\mathcal{M}_r}\zeta_{m_t,m_r}^2$, and (B.7) by defining $\mathbf{A} = \sum_{m_t\in\mathcal{M}_t}\bar{\zeta}_{m_t}\mathbf{D}_{m_t}\overline{\mathbf{A}}\mathbf{D}_{m_t}$.

APPENDIX C

OPTIMALITY OF JOINT SENSING AND COMMUNICATION BEAMFORMING

This extends the proof in [95]. For this purpose, we first note that in the problem formulation in (5.31), the sensing variable, $\mathbf{F}_{\mathcal{Q}}$, are utilized together as a summation of all of the streams, both in the objective and constraints. Hence, if we define $\bar{\mathbf{F}} = \sum_{s\in\mathcal{S}} \mathbf{F}_s$, we can eliminate the sensing term $\mathbf{F}_{\mathcal{Q}}$, and apply the optimization in terms of the user streams $\mathbf{F}_u$ and $\bar{\mathbf{F}}$. To that end, we re-formulate the problem (5.31) as

$$\max_{\{\mathbf{F}_u\},\bar{\mathbf{F}}} \quad \mathrm{Tr}\left(\mathbf{A}\bar{\mathbf{F}}\right) \tag{C.1a}$$

$$\left(1+\gamma_u^{-1}\right)\mathrm{Tr}\left(\mathbf{Q}_u\mathbf{F}_u\right) - \mathrm{Tr}\left(\mathbf{Q}_u\bar{\mathbf{F}}\right) \geq \sigma_u^2, \quad \forall u\in\mathcal{U} \tag{C.1b}$$

$$\mathrm{Tr}\left(\mathbf{D}_m\bar{\mathbf{F}}\right) = P_m, \quad \forall m\in\mathcal{M}_t, \tag{C.1c}$$

$$\mathbf{F}_u \in \mathbb{S}^+, \quad \forall u\in\mathcal{U}, \tag{C.1d}$$

$$\bar{\mathbf{F}} - \sum_{u\in\mathcal{U}} \mathbf{F}_u \in \mathbb{S}^+, \tag{C.1e}$$

$$\bar{\mathbf{F}} \in \mathbb{S}^+. \tag{C.1f}$$

Let us denote the variables obtained by the solution of this problem by $\{\mathbf{F}'_u\}$ and $\bar{\mathbf{F}}'$. Using this solution, we aim to construct an alternative optimal solution of rank-1. For this purpose, we construct the following rank-1 set of solutions

$$\bar{\mathbf{F}}'' = \bar{\mathbf{F}}', \quad \mathbf{F}''_u = \mathbf{f}''_u(\mathbf{f}''_u)^H, \quad \mathbf{f}''_u = (\mathbf{h}_u^H\mathbf{F}'_u\mathbf{h}_u)^{-\frac{1}{2}}\mathbf{F}'_u\mathbf{h}_u. \tag{C.2}$$

whose optimality needs to be proved. For this, we need to check if (i) the value of the objective is the same and (ii) the constraints are satisfied. First, the objective only contains the summation variable and provides the optimal value by definition. For (C.1b), we define $v_u = (\mathbf{h}_u^H\mathbf{F}'_u\mathbf{h}_u)^{-\frac{1}{2}}$, and write

$$\mathrm{Tr}\left(\mathbf{Q}_u\mathbf{F}''_u\right) = \mathrm{Tr}\left(\mathbf{h}_u v_u^2 \mathbf{h}_u^H\mathbf{F}'_u\mathbf{h}_u\mathbf{h}_u^H\mathbf{F}'^H_u\right) = \mathrm{Tr}\left(\mathbf{Q}_u\mathbf{F}'_u\right), \tag{C.3}$$

where we used the cyclic permutation property of the trace and $\mathbf{F}'^H_u = \mathbf{f}'_u\mathbf{f}'^H_u = \mathbf{F}'_u$. With the addition of $\bar{\mathbf{F}}'' = \bar{\mathbf{F}}'$, (C.1b) is satisfied. Similarly, the constraints (C.1c) and (C.1f) are already satisfied by $\bar{\mathbf{F}}'' = \bar{\mathbf{F}}'$. Further, (C.1d) and the solution being rank-1 are also satisfied by the definition of $\mathbf{F}''_u$ in (C.2). For (C.1e), we have

$$\mathbf{v}^H(\mathbf{F}'_u - \mathbf{F}''_u)\mathbf{v} = \mathbf{v}^H\mathbf{F}'_u\mathbf{v} - (\mathbf{h}_u^H\mathbf{F}'_u\mathbf{h}_u)^{-1}\left|\mathbf{v}^H\mathbf{F}'_u\mathbf{h}_u\right|^2. \tag{C.4}$$

From the Cauchy-Schwarz inequality, we have $(\mathbf{v}^H\mathbf{F}'_u\mathbf{v})(\mathbf{h}_u^H\mathbf{F}'_u\mathbf{h}_u) \geq \left|\mathbf{v}^H\mathbf{F}'_u\mathbf{h}_u\right|^2$. Inserting this into (C.4), we obtain $\mathbf{v}^H(\mathbf{F}'_u - \mathbf{F}''_u)\mathbf{v} \geq 0$, which leads to $\mathbf{v}^H\mathbf{F}''_u\mathbf{v} \geq 0$ since it is the summation of two semidefinite matrices, $\mathbf{F}'_u - \mathbf{F}''_u$ and $\mathbf{F}'_u$. Finally, (C.1e) can be shown via

$$\bar{\mathbf{F}}'' - \sum_{u\in\mathcal{U}} \mathbf{F}''_u = \bar{\mathbf{F}}' - \sum_{u\in\mathcal{U}} \mathbf{F}'_u + \sum_{u\in\mathcal{U}}(\mathbf{F}'_u - \mathbf{F}''_u) \tag{C.5}$$

which again leads to the summation of semidefinite matrices.

Finally, for constructing the sensing matrices of the solution, we want to find Q rank-1 matrices whose summation is $\sum_{q\in\mathcal{Q}} \mathbf{F}''_q$. For this purpose, we can utilize the eigendecomposition, i.e., $\sum_Q \mathbf{F}''_q = \mathbf{U}\Lambda\mathbf{U}^H = \sum_{q'=1}^{Q'} \lambda_{q'}\mathbf{u}_{q'}\mathbf{u}_{q'}^H$, and take the largest Q eigenvectors

as the beams via $\mathbf{f}_q'' = \sqrt{\lambda_{u'}}\,\mathbf{u}_{q'}$. Here, it is important to note that it is only possible if the rank of the summation, $Q' = \mathrm{rank}(\sum_{q\in\mathcal{Q}} \mathbf{F}_q'')$, is smaller than or equal to the number of the sensing streams, Q. Otherwise, the summation cannot be constructed by Q rank-1 matrices, and the solution is approximate.

APPENDIX D

DERIVATION OF DUAL SEMIDEFINITE JOINT SENSING AND COMMUNICATION PROBLEM

For the dual of (P2.1-SDR), we first write the Lagrangian function as follows.

$$\begin{aligned}
&\mathcal{L}(\{\mathbf{F}_u\}, \mathbf{F}_{\mathcal{Q}}, \{\mathbf{Z}_u\}, \mathbf{Z}_{\mathcal{Q}}, \{\lambda_u\}, \{\nu_m\}) \\
&= \sum_u \mathrm{Tr}\left(\mathbf{A}\mathbf{F}_u\right) + \mathrm{Tr}\left(\mathbf{A}\mathbf{F}_{\mathcal{Q}}\right) + \sum_u \lambda_u \left(1+\gamma_u^{-1}\right) \mathrm{Tr}\left(\mathbf{Q}_u\mathbf{F}_u\right) - \sum_u \sum_{u'} \lambda_{u'} \mathrm{Tr}\left(\mathbf{Q}_{u'}\mathbf{F}_u\right) \\
&\quad - \sum_{u'} \lambda_{u'} \mathrm{Tr}\left(\mathbf{Q}_{u'}\mathbf{F}_{\mathcal{Q}}\right) - \sum_u \lambda_u \sigma_u^2 - \sum_u \sum_m \nu_m \mathrm{Tr}\left(\mathbf{D}_m\mathbf{F}_u\right) - \sum_m \nu_m \mathrm{Tr}\left(\mathbf{D}_m\mathbf{F}_{\mathcal{Q}}\right) + \sum_m \nu_m P_m \\
&\quad + \sum_u \mathrm{Tr}\left(\mathbf{Z}_u\mathbf{F}_u\right) + \mathrm{Tr}\left(\mathbf{Z}_{\mathcal{Q}}\mathbf{F}_{\mathcal{Q}}\right)
\end{aligned} \tag{D.1}$$

where $\{\lambda_u\} \geq 0$, $\{\nu_m\} \geq 0$, $\{\mathbf{Z}_u\} \succeq 0$, and $\mathbf{Z}_{\mathcal{Q}} \succeq 0$ are the Lagrangian variables corresponding to the SINR constraints, AP power constraints, and the semidefiniteness constraints for the user matrices and the sensing matrix. Collecting all the multiplications with $\mathbf{F}_u$, and $\mathbf{F}_{\mathcal{Q}}$, we can rewrite the Lagrangian function in a compact form as

$$\begin{aligned}
&\mathcal{L}(\{\mathbf{F}_u\}, \mathbf{F}_{\mathcal{Q}}, \{\mathbf{Z}_u\}, \mathbf{Z}_{\mathcal{Q}}, \{\lambda_u\}, \{\nu_m\}) \\
&\quad = \sum_m \nu_m P_m - \sum_u \lambda_u \sigma_u^2 + \sum_u \mathrm{Tr}\left((\mathbf{B}_u + \mathbf{Z}_u)\mathbf{F}_u\right) + \mathrm{Tr}\left((\mathbf{B}_{\mathcal{Q}} + \mathbf{Z}_{\mathcal{Q}})\mathbf{F}_{\mathcal{Q}}\right),
\end{aligned} \tag{D.2}$$

Then, we note that supremum of Lagrangian for $\mathbf{F}_u$ and $\mathbf{F}_q$ is only bounded if $\mathbf{B}_u + \mathbf{Z}_u = \mathbf{0}$. Thus, replacing the variable $\mathbf{Z}_u \geq 0$ with $\mathbf{B}_u \leq 0$, and similarly, for the sensing matrix, we can derive the dual problem via

$$\begin{aligned}
&\min \sup_{\{\mathbf{f}_u\}, \mathbf{f}_q} \mathcal{L}(\{\mathbf{f}_u\}, \bar{\mathbf{f}}_q, \{\lambda_u\}, \{\nu_m\}) \\
&\quad = \begin{cases} \sum_m \nu_m P_m - \sum_u \lambda_u \sigma_u^2 & \text{if } \mathbf{B}_q \preceq 0 \text{ and } \mathbf{B}_u \preceq 0,\ \forall u \in \mathcal{U} \\ \infty & \text{otherwise.} \end{cases}
\end{aligned} \tag{D.3}$$

as given in (P2.1-SDR).