UNIVERSITY OF CALIFORNIA

Los Angeles

Millimeter-Wave Channel Estimation with True-Time-Delay Arrays and Its Use for Network Performance Optimization

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical and Computer Engineering

by

Veljko Boljanovic

2022

ABSTRACT OF THE DISSERTATION

Millimeter-Wave Channel Estimation with True-Time-Delay Arrays
and Its Use for Network Performance Optimization

by

Veljko Boljanovic
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2022
Professor Danijela Cabric, Chair

Millimeter-wave frequencies offer abundant spectrum and high data rates, but their use comes at the cost of a severe propagation loss. The loss can be alleviated by establishing directional communication links with high beamforming gains between the base stations and users and by using dense network deployments.

Beam training is a procedure that reveals the best beam steering directions in a wireless channel. When performed with conventional phased arrays that have many antenna elements, beam training imposes a large overhead. In this work, we introduce true-time-delay array architectures as promising alternatives for phased arrays to solve the overhead problem. Compared to phased arrays, true-time-delay arrays can synthesize frequency-dependent beams and thus probe all angular directions simultaneously using different signal frequencies. We leverage this property to develop and analyze low-complexity digital signal processing algorithms for fast and accurate millimeter-wave beam training.

Unlike beam training, channel estimation has the goal to estimate all parameters of a sparse millimeter-wave channel. We exploit the channel sparsity and frequency-dependent

beams of true-time-delay arrays to develop a frequency-domain compressive sensing based algorithm for channel estimation. We also analyze the performance of the developed algorithm in the presence of practical hardware impairments and we derive the lower bounds on the variances of channel parameter estimators.

In dense millimeter-wave networks with a small inter-site distance and a large number of users, directional beams can cause significant interference and prevent data-hungry users from satisfying their rate requirements. The user experience and the overall network performance can be optimized through coordinated user association and beam scheduling on a network level. Given the channel estimates between different pairs of base stations and users, we develop and analyze a new multi-step optimization framework for joint user association and beam scheduling. The main goal of the framework is to maximize the number of users with satisfied rate requirements while simultaneously suppressing the inter- and intra-cell interference. Since the framework includes NP-hard optimization problems, we propose an algorithm that attains a sub-optimal solution in polynomial-time.

The dissertation of Veljko Boljanovic is approved.

Gregory J. Pottie

Richard D. Wesel

Lara Dolecek

Danijela Cabric, Committee Chair

University of California, Los Angeles

2022

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Danijela Cabric for her continuous support and guidance over the last five years. I would also like to thank the members of my committee Prof. Gregory Pottie, Prof. Richard Wesel, and Prof. Lara Dolecek for their helpful comments.

During my Ph.D., I collaborated with many talented researchers and I would like to thank them for their help and useful discussions. I want to extend my gratitude to Dr. Han Yan and Dr. Shamik Sarkar from the CORES lab at the University of California, Los Angeles. I also want to thank my collaborators from Washington State University, including Mr. Chung-Ching Lin, Mr. Soumen Mohapatra, Dr. Erfan Ghaderi, Prof. Deukhyoun heo, and Prof. Subhanshu Gupta.

I would like to thank Mr. Benjamin Domae, Mr. Aditya Wadaskar, Mr. Chung-Ching Lin, Ms. Qiuyan Xu, and Dr. Han Yan for providing me with the opportunity to expand my knowledge while collaborating with them on their projects.

I was lucky to share the lab with some wonderful people during my graduate studies. I want to thank Mr. Enes Krijestorac, Mr. Benjamin Domae, Mr. Samurdhi Karunaratne, Mr. Aditya Wadaskar, Mr. Donar Li, Mr. Agon Memedi, Dr. Shailesh Chaudhari, Dr. Ghaith Hattab, Dr. Han Yan, Dr. Shamik Sarkar, and Dr. Samer Hanna for all the fun times we had both inside and outside the lab. I want to especially thank Samer for being a great friend and apartment mate.

I would like to thank the University of California, Los Angeles, National Science Foundation, and Semiconductor Research Corporation for funding my research.

Finally, I would like to thank my family. Thank you for constantly supporting me throughout my Ph.D. journey. Without your support, I would not have come this far.

VITA

2011–2015	B.S. in Electrical and Computer Engineering, University of Novi Sad, Novi Sad, Serbia.
2015–2016	M.S. in Electrical and Computer Engineering, University of Novi Sad, Novi Sad, Serbia.
2017-2022	Graduate Student Researcher and Teaching Assistant, UCLA
2019	Summer Intern, Qualcomm, San Diego, CA.
2020	Summer Intern, Qualcomm, San Diego, CA.
2021	Summer Intern, Qualcomm, San Diego, CA.

SELECTED PUBLICATIONS

[1] V. Boljanovic, H. Yan, C.-C. Lin, S. Mohapatra, D. Heo, S. Gupta, and D. Cabric, "Fast Beam Training With True-Time-Delay Arrays in Wideband Millimeter-Wave Systems," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 4, pp. 1727-1739, April 2021, doi: 10.1109/TCSI.2021.3054428

[2] V. Boljanovic, H. Yan, E. Ghaderi, D. Heo, S. Gupta and D. Cabric, "Design of Millimeter-Wave Single-Shot Beam Training for True-Time-Delay Array," in 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2020, pp. 1-5, doi: 10.1109/SPAWC48557.2020.9154233

[3] V. Boljanovic and D. Cabric, "Joint mmW AoD and AoA Estimation using One OFDM Symbol and Frequency-Dependent Beams," to be submitted.

[4] V. Boljanovic and D. Cabric, "Joint Single-Shot AoD/AoA Estimation in mmW Systems and Analysis under Hardware Impairments," to be submitted.

[5] V. Boljanovic and D. Cabric, "Millimeter-Wave Wideband Channel Estimation Using Analog True-Time-Delay Array Under Hardware Impairments," in *Journal of Signal Processing Systems*, 2022, doi: 10.1007/s11265-022-01771-6

[6] V. Boljanovic and D. Cabric, "Compressive Estimation of Wideband mmW Channel using Analog True-Time-Delay Array," in 2021 IEEE Workshop on Signal Processing Systems (SiPS), 2021, pp. 170-175, doi: 10.1109/SiPS52927.2021.00038

[7] V. Boljanovic, S. Sarkar, and D. Cabric, "Millimeter-Wave User Association and Low-Interference Beam Scheduling (Invited Paper)," in 6th ACM Workshop on Millimeter-Wave Networks and Sensing Systems, 2022, pp. 7-12, doi: 10.1145/3555077.3556471

[8] V. Boljanovic, S. Sarkar, and D. Cabric, "Joint User Association and Beam Scheduling with Interference Management in Dense mmW Networks," to be submitted.

[9] V. Boljanovic, H. Yan and D. Cabric, "Tracking Sparse mmWave Channel under Time Varying Multipath Scatterers: (Invited Paper)," in 2018 52nd Asilomar Conference on Signals, Systems, and Computers, 2018, pp. 1274-1279, doi: 10.1109/ACSSC.2018.8645308

CHAPTER 1

Introduction

1.1 Motivation

Due to the shortage of sub-6 GHz spectrum, millimeter-wave (mmW) communication networks gained an important role in the deployment of cellular systems. Abundant spectrum at mmW frequencies with the total of 3.85 GHz bandwidth [FCC16] at the lower mmW band (28, 37, and 39 GHz) offers peak rates in the order of several gigabits per second [GTC14]. Future generations of mmW networks will operate in the upper mmW frequency band where more than 10 GHz of bandwidth can be used to meet the ever increasing demands [YXX19, ZXM19].

Large available bandwidth, however, comes at the cost of less favorable propagation conditions. Recent measurement campaigns showed that mmW signals experience severe propagation loss in both the lower and upper mmW bands [ALS14]. For this reason, mmW channels are sparse, i.e., only a few multipath components reach the receiver. Due to a small wavelength at mmW frequencies, the base station (BS) and user equipment (UE) can be equipped with large antenna arrays and establish a directional link with high beamforming gain to compensate for the propagation loss. Unlike in sub-6 GHz acMIMO communication with fully digital arrays where each antenna element has a dedicated radio frequency (RF) chain, large mmW arrays have analog architecture with phase shifters (PS) and a single RF chain, or hybrid analog-digital architecture with a few RF chains to make chipsets power and cost efficient. Consequently, the majority of channel probing procedures required to

establish a directional link between the BS and UE have been developed for PS-based analog and hybrid arrays. The overhead associated with these procedures is directly proportional to the number of antenna elements in the array and inversely proportional to the number of beams that can be steered at once, i.e., the number of RF chains. For this reason, previously developed procedures for PS-based arrays with a few RF chains experience an overhead bottleneck when massive antenna arrays are used. Therefore, future mmW systems, especially those deployed in the upper mmW band, will require fundamental rethinking of array architectures and corresponding channel probing procedures to eliminate the problem of high overhead.

Another way to compensate for the propagation loss at mmW frequencies and improve the link budget is by using a much denser network deployment, where the inter-site distance is in the order of a hundred meters [BIB15]. Due to the use of highly directional beams, it is widely accepted that mmW communication is noise-limited rather than interference-limited. However, with further densification of mmW networks, the directional inter- and intra-cell interference cannot be neglected [ABK17]. In particular, the main-lobe of highly directional beams can cause significant interference and prevent data-hungry UEs from satisfying their data rate requirements. Based on this, two important questions arise in dense mmW networks: 1) Since a UE can be relatively close to multiple BSs, which BS should it be associated with? 2) How can the directional interference be avoided or suppressed? In the current mmW networks, these questions are answered individually by the UEs and BSs. For example, each UE can select the BS it wants to be associated with based on the measured quality of communication links. On the other hand, each BS can optimize its beamforming vectors to suppress the intra-cell interference. While this approach is simple, it can lead to a poor network performance overall. Specifically, user association is not balanced as many UEs can select the same BS. Beam scheduling at that BS can experience significant delays, while the beams at other BSs remain under-utilized. Additionally, simple optimization of BS beams does not suppress the inter-cell interference in the network. Actual network

performance improvements can be ensured through optimization frameworks that consider multiple BSs and UEs at the same time [LWC16]. If designed properly, optimization frameworks can enable both a balanced user association and a coordinated suppression of the inter- and intra-cell interference in the network.

1.2 Challenges and Objectives

1.2.1 Estimation of Best Beam Steering Directions

One of the most important physical-layer procedures in mmW networks is beam training, which enables the BS and UE to establish a directional link by iteratively searching for the best beam steering directions, i.e., the angle of departure (AoD) and angle of arrival (AoA) of the most dominant propagation path [HGR16]. As the number of antenna elements at the BSs and UEs increase, the beam widths get narrower and the number of candidate beam pairs that need to be checked increases significantly. Although beam training implementation is simple, the overhead created by the procedure becomes a bottleneck when BSs and UEs have large arrays. This problem cannot be solved with analog and hybrid PS-based arrays due to their low flexibility and inability to probe a large number of beams simultaneously. On the other hand, true-time-delay (TTD) arrays can leverage their frequency-dependent beams to probe all angular directions simultaneously using different signal frequencies. Our objective is to design and analyze digital signal processing (DSP) algorithms for fast mmW beam training using analog and hybrid TTD arrays.

1.2.2 Estimation of Sparse Millimeter-Wave Channel

Channel estimation is a procedure for acquiring the full knowledge of a wireless channel, and it includes the estimation of the AoDs, AoAs, and channel gains of all multipath components [HGR16]. Having the full channel knowledge provides a series of benefits, including the

ability to optimize the BS and UE beamforming vectors, boost the data rate through spatial multiplexing, perform optimal power allocation, determine the backup links to be used in the case of a link failure, and others. The majority of the existing algorithms were developed for analog and hybrid PS-based arrays. However, these arrays experience a similar problem like in beam training - their inability to probe many angular directions simultaneously leads to a high required overhead. Our objective is to exploit the sparsity of mmW channels and frequency-dependent beams of TTD arrays to design a high-accuracy channel estimation DSP algorithm that requires a low overhead. We also want to understand the impact of practical hardware impairments in TTD arrays on the designed algorithm.

1.2.3 Network Performance Optimization

The UE uses the acquired channel state information in the process of cell selection to decide which BS it wants to be associated with [3GP19a]. Although the cell selection is straightforward and already used in practice, it is an inefficient way of associating the UEs when the overall network performance needs to be optimized. An alternative way to perform user association and beam scheduling is through an optimization framework. A properly designed optimization framework aims to maximize a specific network utility function, while considering important service requirements, properties of directional communication, and features of mmW transceivers. However, the existing frameworks are often mathematically too complex and not comprehensive enough. For example, the utility function is commonly non-convex and focused on maximization of the network sum rate, while the set of constraints do not necessarily suppress the interference and capture important properties of mmW communication. In addition, with maximization of the network sum rate, available serving beams are scheduled to the UEs with good channels, while other UEs' rate requirements are often not satisfied. Our objective is to design and mathematically formulate a less complex, yet more comprehensive framework with a different utility function and a set of constraints that consider important properties of mmW communication.

1.3 Contributions

The contributions of the dissertation are summarized as follows.

1.3.1 Beam Training using True-Time-Delay Arrays

We introduce TTD array architectures for fast mmW beam training, including analog, hybrid sub-array based, and fully connected hybrid TTD arrays. The architectures are compared in terms of the beam training hardware requirements, dependency of beam training on the basic system parameters and TTD hardware constraints, and robustness to hardware impairments and quantization errors in analog-to-digital converters (ADC). In addition, we propose a benchmark emulation of frequency-dependent TTD-based beam training using a fully digital array to analyze the advantages and disadvantages of TTD array architectures for mmW beam training. We develop a high-accuracy DSP algorithm for UE beam training that requires a single wideband symbol. Using numerical simulations, we show that the algorithm achieves a sub-degree angle estimation accuracy and that it is highly robust to hardware impairments in TTD arrays. Further, we extend the idea of beam training with frequency-dependent beams to the joint beam training between the BS and UE. We analyze the beam pair misalignment probability in the fast joint beam training and we show that the probability is at least as low as in the conventional exhaustive beam sweeping (EBS) with PS-based arrays. The joint beam training is evaluated in realistic frequency-selective mmW channels and it is shown that it outperforms the EBS while requiring a lower overhead.

1.3.2 Channel Estimation using True-Time-Delay Arrays

We develop a compressive sensing (CS) based DSP algorithm to estimate sparse mmW channels using an analog TTD array. The algorithm is designed in the frequency domain and it is based on per-sub-band processing that exploits frequency-dependent beams of the TTD array and reduces the complexity compared to the state-of-the-art. Assuming antenna

arrays without hardware errors, the proposed DSP algorithm is compared with related state-of-the-art approaches designed for PS-based antenna arrays in terms of the required number of training frames, channel estimation accuracy, and computational complexity. The results indicate that the algorithm has a better accuracy, lower overhead, and lower complexity than the state-of-the-art. The proposed algorithm is also evaluated in the presence of impairments in the analog TTD array and the results reveal notable robustness to hardware errors. We extended our analysis of the impact of hardware impairments by linearizing the received signal model and deriving the Cramér-Rao lower bound (CRLB) for the parameters of line-of-sight (LoS) mmW channels, including the AoD, AoA, and phase of the complex channel gain. We propose a gradient descent based refinement of channel parameters to improve the estimation accuracy and close the gap between the algorithm performance and the derived CRLBs.

1.3.3 User Association and Low-Interference Beam Scheduling

We propose a new optimization framework where user association and low-interference beam scheduling are performed jointly in a centralized sub-network. The framework consists of three sequential steps: Step 1 - maximization of the number of UEs with fully satisfied rate requirements; Step 2 - maximization of the number of UEs with partially satisfied rate requirements using the remaining serving beams; Step 3 - design of hybrid precoders and combiners, followed by power allocation on a sub-network level to boost the rates of partially satisfied UEs. We mathematically formulate linear optimization problems in all three steps. This includes the design of multi-criterion objective functions using scalarization and the design of constraints that consider UEs' rate requirements and suppress the inter- and intra-cell interference. We explain that the optimization problems in Step 1 and Step 2 are NP-hard, and then we design a heuristic algorithm based on relaxation, rounding, and resource pruning to obtain sub-optimal solutions. We also analyze the complexity of the proposed algorithm and show that the solutions are obtained in polynomial time. Using

realistic mmW channels where co-located UEs can experience significant interference, we evaluate the impact of the interference management on the average signal-to-interference-plus-noise ratio (SINR) per associated link. We also compare the proposed optimization framework with existing baseline approaches, including the naive greedy association and the maximum sum rate association. The comparison is done in terms of average number of UEs with satisfied rate requirements, network sum rate, and transmit power usage per BS. The results reveal the advantages of the proposed framework over the existing baseline approaches.

1.4 Thesis Organization

The rest of the dissertation is organized as follows

- Chapter 2: We present TTD array architectures, DSP algorithms, and analysis for fast mmW beam training. This chapter is an extension of our previous publication [BYL21].
- Chapter 3: We present the CS-based DSP algorithm and analysis for TTD-based estimation of mmW channels. This chapter is based on our previous publication [BC22].
- Chapter 4: We present the three-step optimization framework for joint user association and low-interference beam scheduling in mmW networks. This chapter is an extension of our previous publication [BSC22a].
- Chapter 5: We summarize the research contributions and outline future research directions.

CHAPTER 2

Beam Training using True-Time-Delay Arrays

2.1 Introduction

Abundance of spectrum at mmW frequencies is seen as the key enabler of high data rates in the fifth generation (5G) of cellular systems [ABC14]. However, communication at mmW frequencies is challenged by severe propagation loss [RXM17]. To compensate for this loss, the BS and UE rely on directional communication with large antenna arrays that can provide high beamforming gains. A common way to establish a directional link between the BS and UE without explicit channel estimation is through *beam training* [HGR16], a procedure that identifies the dominant AoD and AoA, i.e., the best pair of steering directions, in the wireless multipath channel. Apart from aligning the beams for data communication, knowledge of the AoA and AoD is of utmost importance for other applications in practical mmW systems, including interference nulling and localization [WML16]. With large antenna arrays and narrow beams at the BS and UE, finding the optimal beam pair, but keeping the training overhead and computational complexity low, is a challenging task which has recently attracted significant attention from researchers.

2.2 Prior Work on mmW Beam Training

Early work on 5G mmW communications usually assumed that the BS and UE have power-efficient analog phased arrays with a single RF chain. Phased arrays have adjustable PSs in all antenna branches, which allows them to coherently steer or combine the signal from

a desired direction. However, since there is a single RF chain, only one steering/combining beam can be synthesized at the time, which limits the flexibility of phased arrays. For this reason, the existing beam training approaches for phased arrays include different variations of the EBS [HPR15, JPY15, LLZ20, CM21]. In the EBS, the BS and UE sequentially check all possible beam pairs to identify the one with the highest received signal power, which corresponds to the dominant AoD and AoA. Due to its simple implementation, the variations of the EBS have been proposed and used in practical communication systems [iee21]. The main problem of sweeping is a large training overhead since the required number of training symbols scales linearly with the number of antennas.

Previous work that address the problem of large beam training overhead can be roughly divided into two groups. The first group of works intends to leverage DSP techniques to reduce the required number of training symbols when the BS and UE are equipped with phased arrays. Specifically, several algorithms that exploit the sparsity of mmW channel and treat beam training as a compressive sensing problem have been proposed recently [ZLC20, YC19, Baj19, RMZ17]. It has been shown that the overhead of the compressive algorithms scales logarithmically with the number of antennas [RMZ17]. The second group of works aims to speed up channel probing by using different array architectures [DKS14, BHR14, BHM16, AEL14, DPW17, NZL17, VAG17, GYS20, JSR20, RJ20, YBC19, BYG20, BYL21]. Fully digital array architectures, where each antenna has a dedicated RF chain, offer the highest flexibility in channel probing. Having the signal samples in all antenna branches allows these arrays to probe numerous angular directions simultaneously through DSP for fast angle estimation [DKS14, BHR14, BHM16]. While digital architecture is appealing for BSs [Yan19], it may not be suitable for UEs due to high power consumption. Power-efficient hybrid array with multiple digitally connected phased arrays is another way to enable simultaneous probing of the channel [WFW21]. Further, DSP in hybrid arrays can be leveraged for the design of training codebooks with adaptive sector beams for hierarchical AoD and AoA estimation, which can reduce the beam training overhead [AEL14, DPW17, NZL17]. In

hierarchical approaches, information of the AoD/AoA is iteratively refined using narrower probing sectors in each iteration. However, the initial sector beams have more power in sidelobes and thus probability of early sector misdetection is high [DKS14]. Additionally, the required feedback about the chosen sector in each iteration negatively affects the overhead. Besides hierarchical approaches, various compressive algorithms were proposed for hybrid arrays [VAG17], but mainly in the context of channel estimation.

In an effort to further reduce the required overhead, recent work proposed the use of frequency-dependent beam steering in mmW beam training [GYS20, JSR20, RJ20, YBC19, BYG20, BYL21]. In [GYS20], the authors investigated fast beam training with a leaky wave antenna, which can probe all angular directions simultaneously by using different frequency components of the signal. Similarly, recently reemerged TTD arrays can facilitate beam training based on frequency-to-angle mapping [JSR20, RJ20, YBC19]. Compared to phased arrays, TTD arrays have delay elements along with PSs in all antenna branches, which allow them to synthesize frequency-dependent beams. Early implementations of TTD arrays relied on delay lines in all antenna branches [CH10], but this approach suffered from low scalability in terms of required area and power efficiency when the array size becomes large. Further, limited delay range at RF is insufficient to achieve frequency dispersive beam training as proposed in this work. Recent advancement in TTD arrays with baseband delay elements and large delay range-to-resolution ratios [GSR19, GPB20], improved the scalability and thus enabled the realization of fast beam training schemes with large arrays.

2.2.1 Contributions

In this chapter, we introduce and analyze TTD array architectures and DSP algorithms for fast beam training in wideband mmW channels. The contributions of this chapter can be summarized as follows:

- We introduce TTD architectures, including analog, hybrid sub-array based, and fully

connected hybrid architectures, for fast mmW beam training with frequency-dependent beams. We also propose a benchmark emulation of frequency-dependent TTD-based beam training using a fully digital array and time-domain DSP, to analyze the advantages and disadvantages of TTD array architectures for beam training.

- We develop a TTD beam training codebook and a DSP algorithm for a single-symbol UE beam training with a high angle estimation accuracy.
- We perform a thorough comparison of analog TTD, hybrid sub-array based TTD, and benchmark fully digital arrays in terms of the dependency of UE beam training on the basic system parameters and TTD hardware constraints, and robustness to hardware impairments and quantization errors in ADC.
- We develop codebooks and a DSP algorithm for joint beam training between the BS and UE using frequency-dependent beams.
- We analyze the beam pair misalignment probability in joint beam training in the presence of Gaussian noise. The misalignment probability in the EBS is included as the benchmark.
- The performance of the proposed joint beam training algorithm is evaluated in realistic wideband mmW channels generated in Quadriga [JRB19]. The results show that the proposed algorithm outperforms the EBS while requiring a lower overhead.

2.2.2 Organization and Notation

The rest of the chapter is organized as follows. In Sec. 2.3, we introduce the three TTD architectures and benchmark fully digital array. Sec. 2.4 introduces a wideband system model. Sec. 2.5 describes the design of a beam training codebook and a DSP algorithm design for the UE beam training. In Sec. 2.6, we compare the considered TTD array architectures when they are used for UE beam training. In Sec. 2.7, we present the design of a joint

beam training between the BS and UE using frequency-dependent beams. The comparison between the designed joint beam training and EBS is presented in Sec. 2.8. In Sec 2.9, joint beam training is evaluated in realistic wideband mmW channels. Sec. 2.10 summarizes the conclusions and ideas for the future work.

Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. The n-th element of $\mathbf{a}$ is denoted by $[\mathbf{a}]_n$. Hermitian transpose is denoted by $(.)^{\mathrm{H}}$.

2.3 TTD Array Architectures for Beam Training

The realization and performance of TTD beam training schemes heavily depends on the underlying TTD hardware. The design of a fast high performance beam training scheme imposes a challenging delay range requirement on TTD circuits, which raises the question of a *beam-training-efficient* TTD array architecture. Here, the efficiency depends the number of symbols used in beam training, angle estimation accuracy, and array power consumption. To address this question, we propose three uniform linear array architectures with baseband TTD elements, including analog, hybrid sub-array based, and fully connected hybrid arrays. Further, we extensively compare the efficiency of the analog and hybrid sub-array based TTD arrays. A fully digital array architecture is included in the comparison as the benchmark. In particular, we use it to emulate TTD-based beam training and thus highlight the advantages and disadvantages of TTD arrays. All considered array architectures are described in the reminder of this section.

An analog uniform linear TTD array with a single RF chain and N_{R} antennas is presented in Fig. 2.1. The n-th antenna branch has an analog PS with the phase tap $\phi_{\mathrm{A},n} = (n-1)\Delta\phi$ and an analog baseband TTD element with the delay tap $\tau_{\mathrm{A},n} = (n-1)\Delta\tau$, where $\Delta\phi$ and $\Delta\tau$ represent the phase and delay spacing between neighboring branches, respectively. Note that the PSs in the analog array can be implemented in the RF path, local oscillator path, or

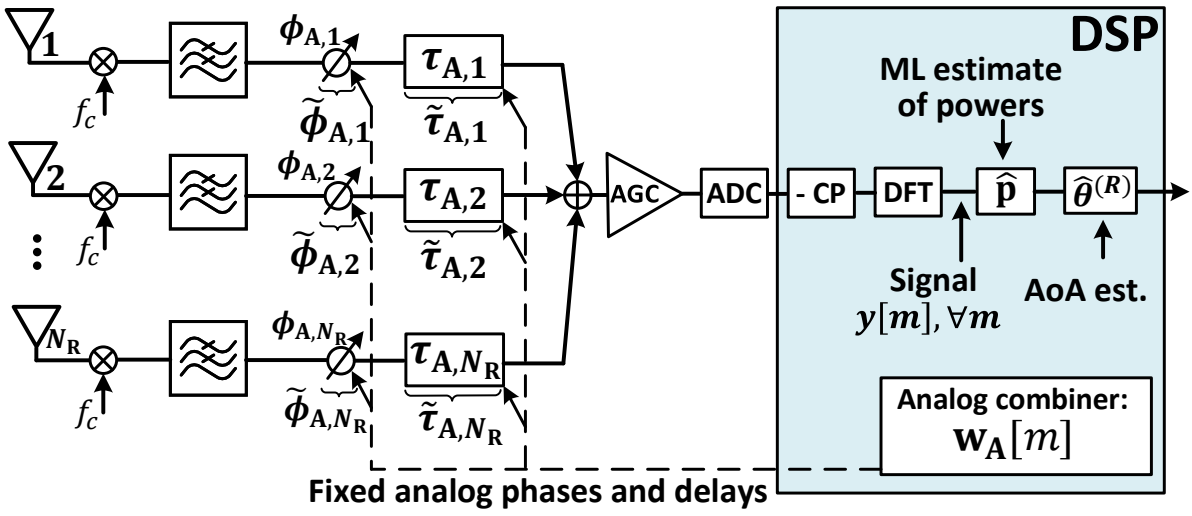


Figure 2.1: Architecture of analog TTD array with uniform delay spacing $\Delta\tau$ and phase spacing $\Delta\phi$ between antennas. The design of combiners and DSP algorithm is explained in Sec. 2.5.

baseband domain [PT12]. From mathematical perspective, these different implementations introduce the same phase taps in beam training algorithm design. Here we assume that the PSs are implemented in the baseband, as depicted in Fig. 2.1. In practice, the phase taps $\phi_{\mathrm{A},n}, n = 1, ..., N_{\mathrm{R}}$, can be distorted due to the errors in PSs, local oscillators, imbalance between in-phase and quadrature samples, or other hardware imperfections. Similarly, errors in TTD elements can distort the delay taps $\tau_{\mathrm{A},n}, n = 1, ..., N_{\mathrm{R}}$. In all antenna branches, we model the time-invariant distorted taps as independent zero-mean Gaussian random variables $\tilde{\phi}_{\mathrm{A},n} \sim \mathcal{N}(\phi_{\mathrm{A},n}, \sigma_{\mathrm{P}}^2)$ and $\tilde{\tau}_{\mathrm{A},n} \sim \mathcal{N}(\tau_{\mathrm{A},n}, \sigma_{\mathrm{T}}^2)$, respectively. For a specific delay spacing $\Delta\tau$, TTD frequency-dependent antenna weight vector results in a fixed beam training codebook of pencil beams, where different frequency components of the signal are hard-coded in different angular directions. The frequency-flat PSs increase the flexibility by enabling codebook rotations and different frequency-to-angle mapping. The maximum delay in the N_{R}-th antenna branch is $\tau_{\mathrm{A},N_{\mathrm{R}}} = (N_{\mathrm{R}} - 1)\Delta\tau$, which becomes an implementation

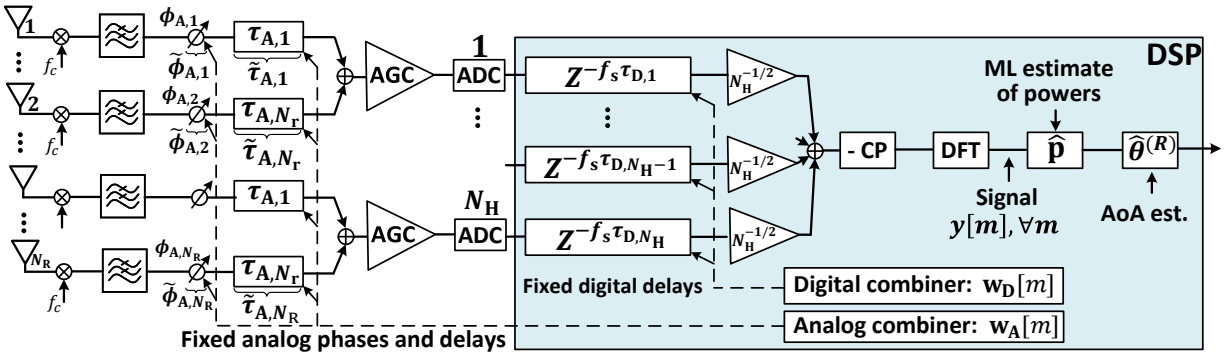


Figure 2.2: Architecture of hybrid sub-array based TTD array with uniform delay spacing $\Delta\tau$ and phase spacing $\Delta\phi$ between antennas. The design of combiners and DSP algorithm is explained in Sec. 2.5.

bottleneck for large antenna arrays. The state-of-the-art TTD delay range is in the order of 15 ns [GSR19], which can be insufficient for wideband beam training with a moderate number of antenna elements N_R, e.g., $N_\mathrm{R} = 32$, as previously discussed in [BYG20].

To alleviate the delay range requirement and improve the scalability of analog TTD arrays, we introduce a hybrid analog-digital architecture with N_H sub-arrays, each controlled by one distinct RF chain, as illustrated in Fig. 2.2. The hybrid array uses a combination of analog and digital signal delaying, where first all the sub-arrays of N_r antennas introduce the same delays $\tau_{\mathrm{A},n'} = (n'-1)\Delta\tau,\ n' = 1, ..., N_\mathrm{r}$, in the analog domain. The relative delay difference among antennas is compensated in the digital domain by introducing the fixed digital taps $\tau_{\mathrm{D},h} = (h-1)N_\mathrm{r}\Delta\tau,\ h = 1, ..., N_\mathrm{H}$, i.e., digital delays $f_\mathrm{s}\tau_{\mathrm{D},h}$, where f_s is the sampling frequency. As in the analog TTD array, the distorted phase taps $\tilde{\phi}_{\mathrm{A},n},\ n = 1, ..., N_\mathrm{R}$, and delay taps $\tilde{\tau}_{\mathrm{A},n},\ n = 1, ..., N_\mathrm{R}$, are modeled as independent Gaussian random variables.

We also introduce a fully connected hybrid TTD architecture illustrated in Fig. 2.3. As in the sub-array based architecture, the fully connected architecture has a delay element in each antenna branch. However, there are two notable differences between the two architectures.

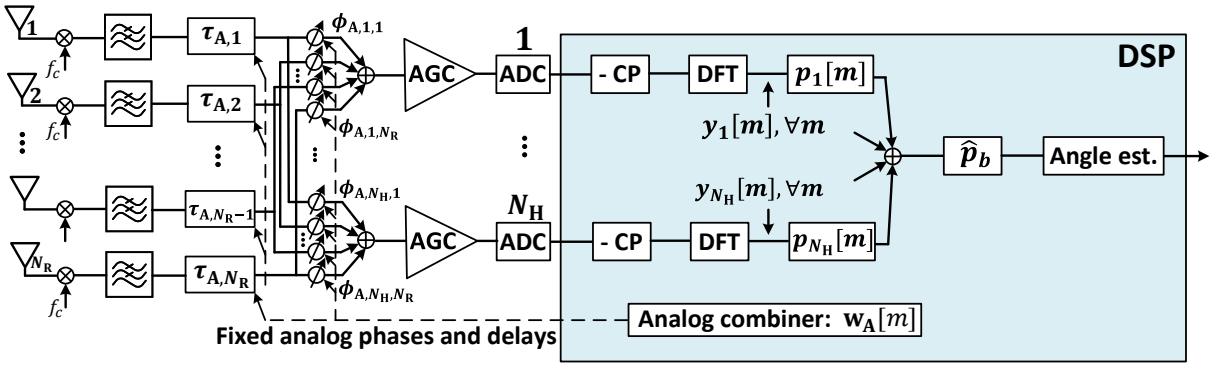


Figure 2.3: Architecture of fully connected hybrid TTD array with uniform delay spacing $\Delta\tau$ and phase spacing $\Delta\phi$ between antennas in each RF chain. This architecture is used later in joint beam training in Sec. 2.7.

Firstly, the fully connected architecture cannot split the signal delaying between the analog and digital domains. Thus, the delay taps are designed as in the analog TTD architecture, i.e., $\tau_{\mathrm{A},n} = (n-1)\Delta\tau,\ \forall n$. Secondly, each RF chain in the fully connected architecture controls all N_{R} antennas using N_{R} PSs. Consequently, the combined signal in each RF chain has the full beamforming gain. In addition, phase shifters allow the fully connected architecture to rotate frequency-dependent codebooks and achieve different frequency-to-angle mapping in N_{H} RF chains. The rotations increase the codebook diversity, which improves the beam training performance, as we discuss in more details in Sec. 2.7. Assuming uniformly spaced phase taps in each RF chain, the phase tap in the n-th antenna and r-th RF chain is mathematically defined as $\phi_{\mathrm{A},r,n} = (r-1)(n-1)\Delta\phi$.

A fully digital array, used as the benchmark, is illustrated in Fig. 2.4. The digital array can emulate a TTD array through DSP by using the fixed digital taps $\tau_{\mathrm{D},n} = (n-1)\Delta\tau,\ n = 1, ..., N_{\mathrm{R}}$, i.e., digital delays $f_{\mathrm{s}}\tau_{\mathrm{D},n}$ in the corresponding antenna branches. We assume phase-only beamforming without magnitude control in order to create a codebook of pencil beams as with both analog and hybrid TTD arrays. The ability to control the

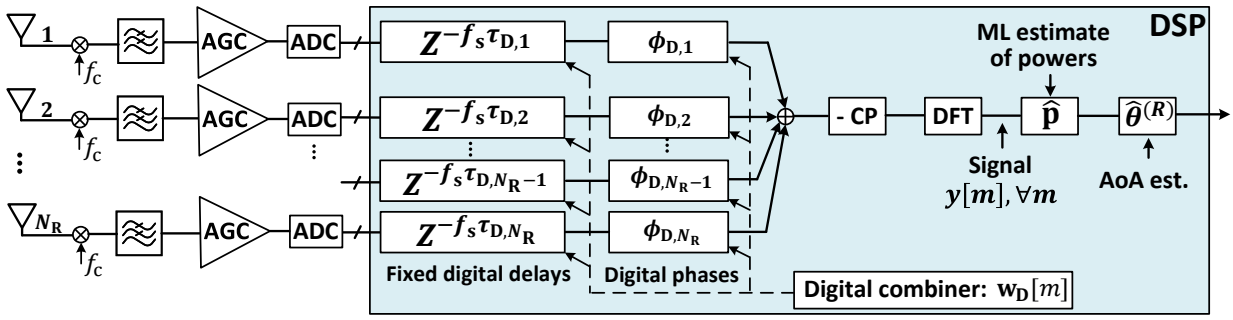


Figure 2.4: Architecture of the benchmark fully digital array that is used to emulate TTD-based beam training by introducing digital delays. The design of combiners and DSP algorithm is explained in Sec. 2.5.

digital phases $\phi_{D,n}$, $n = 1, ..., N_R$, in DSP, allows the signal frequency components to be independently steered/combined in any angular direction, which provides high flexibility in the beam training design. The digital array does not have the analog PSs and TTD elements before the ADCs, and it is assumed to be insensitive to hardware errors. However, each antenna element has a dedicated RF chain, which significantly affects the array power efficiency, as discussed in details in [BYL21].

In the next section, we explain how $\Delta\tau$ and $\Delta\phi$ are set up in the analog, hybrid sub-array based and digital architectures to obtain a beam training codebook robust to frequency-selective channels. We also introduce a DSP algorithm that exploits this codebook. Based on the designed $\Delta\tau$, Sec. 2.6 discusses the dependency of the UE beam training performance on the basic system parameters, TTD hardware constraints, and TTD hardware impairments.

2.4 System Model for UE Beam Training

We consider downlink beam training between the BS and UE, where the cyclic prefix (CP) based Orthogonal Frequency Division Multiplexing (OFDM) waveform is used as a training

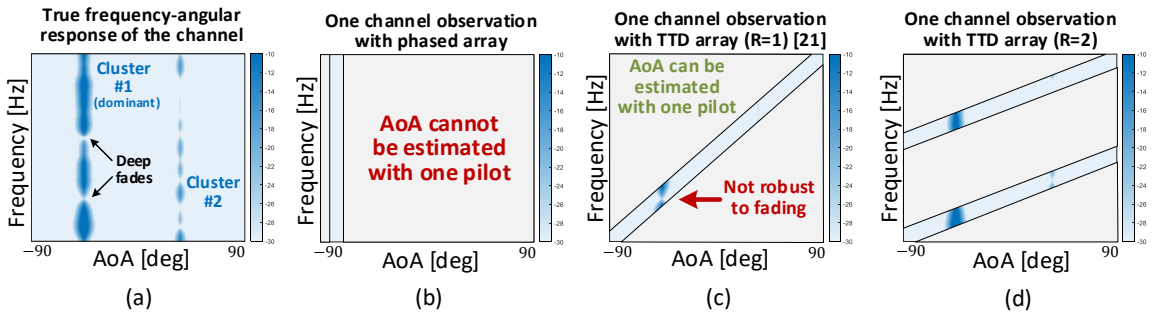


Figure 2.5: Beam training in clustered frequency-selective multipath channel: (a) An example of frequency-selective channel with two multipath clusters. Frequency-selectivity comes from intra- and inter-cluster delay spreads. The first cluster is dominant and its AoA needs to be estimated. (b) Channel observation of a phased array when only one symbol is used. Beam sweeping is necessary to cover all angles in the range $(-\pi/2, \pi/2)$. (c) Channel observation of a TTD array when only one symbol is used. Frequency components (subcarriers) are mapped into different angles to simultaneously probe the range $(-\pi/2, \pi/2)$. The angle estimation may fail in frequency-selective channels. (d) Enhanced TTD codebook with frequency diversity order $R = 2$.

symbol. The carrier frequency, bandwidth, and number of subcarriers are denoted as f_{c}, BW, and M_{tot}, respectively. The power-normalized training symbol uses M subcarriers from the predefined set $\mathcal{M}$, all loaded with binary phase shift keying (BPSK) modulated symbols. Both the BS and UE have half-wavelength spaced uniform linear arrays with N_{T} and N_{R} antennas, respectively.

2.4.1 Channel Model

we consider a mmW channel with L multipath clusters well-separated in angular domain. Due to the inter- and intra-cluster delay spread [3GP19b,JRB14], we assume that the channel is frequency-selective. The channel matrix $\mathbf{H}[m] \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ at the m-th subcarrier can be

expressed as

$$\mathbf{H}[m] = \sum_{l=1}^{L} g_l[m]\mathbf{a}_{\mathrm{R}}(\theta_l^{(\mathrm{R})})\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l^{(\mathrm{T})}), \tag{2.1}$$

where $\theta_l^{(\mathrm{R})}$ and $\theta_l^{(\mathrm{T})}$ are the AoA and AoD of the l-th cluster, defined with respect to the local coordinate systems at the UE and BS, respectively. We assume the array responses are frequency flat, i.e., $[\mathbf{a}_{\mathrm{R}}(\theta)]_n = N_{\mathrm{R}}^{-1/2}\exp(-j(n-1)\pi\sin(\theta)),\ n = 1, ..., N_{\mathrm{R}}$ and $[\mathbf{a}_{\mathrm{T}}(\theta)]_n = N_{\mathrm{T}}^{-1/2}\exp(-j(n-1)\pi\sin(\theta)),\ n = 1, ..., N_{\mathrm{T}}$. The complex gains at all subcarriers are modeled as complex Gaussian random variables, i.e., $g_l[m] \sim \mathcal{CN}(0, \sigma_l^2),\ \forall l, m$. The gains $g_l[m],\ \forall l, m$, are assumed to be independent across L multipath clusters and all well-separated subcarriers, e.g., separated by a coherence bandwidth, which is in the order of 10 MHz in mmW channels. The frequency-domain channel model in (2.1) can be approximated as [HGR16]

$$\mathbf{H}[m] \approx \mathbf{A}_{\mathrm{R}}\mathbf{\Lambda}[m]\mathbf{A}_{\mathrm{T}}^{\mathrm{H}}, \tag{2.2}$$

where $\mathbf{A}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times Q}$ and $\mathbf{A}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times Q}$ contain Q array responses $\mathbf{a}_{\mathrm{R}}(\xi_q)$ and $\mathbf{a}_{\mathrm{T}}(\xi_q)$ that correspond to Q uniformly spaced angles $\xi_q,\ q = 1, ..., Q$, in the range $(-\pi/2, \pi/2)$. The square matrix $\mathbf{\Lambda}[k] \in \mathbb{C}^{Q \times Q}$ has only L non-zero elements that correspond to the gains $g_l[m],\ \forall l$. Commonly, $Q \gg L$ and the approximation error in (2.2) can be neglected.

2.4.2 Received Signal Model in UE Beam Training

In general, AoDs evolve slower than AoAs over time in mmW channels. Since BSs have fixed orientation of antenna arrays, the evolution of AoDs is determined by the gradual birth and death of channel clusters [JRB14, JRB19]. On the other hand, UEs are prone to swift rotations in antenna orientations, which can lead to significant changes of AoAs, even in low mobility environments [JRB14, JRB19]. Additionally, mmW BSs are likely to be equipped with fully digital antenna arrays [Yan19], which enable the dominant AoD to be estimated using a single symbol by probing all angular directions at once [BHM16]. Thus, here we assume that the slowly-changing AoD $\theta^{(\mathrm{T})}$ at the BS has already been estimated

and used to design a fixed frequency-flat beam defined by the precoder vector $\mathbf{v} \in \mathbb{C}^{N_\mathrm{T}}$. The real challenge arises at the UE side where the dynamic AoA changes require frequent beam training to be performed in a fast and power-efficient manner. We propose the UE to be equipped with a TTD array and exploit its frequency-dependent beamforming to achieve a single-shot estimation of the AoA $\theta^{(\mathrm{R})}$. Therefore, the received signal $y[m]$ at the m-th subcarrier of the used OFDM symbol is

$$y[m] = \mathbf{w}^{\mathrm{H}}[m]\mathbf{H}[k]\mathbf{v}s[m] + \mathbf{w}^{\mathrm{H}}[m]\mathbf{n}[m],\ m \in \mathcal{M}, \tag{2.3}$$

where $s[m]$ is a BPSK symbol and $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_{\mathrm{N}}^2 \mathbf{I}_{N_\mathrm{R}}\right)$ is white Gaussian noise. The UE TTD combiner $\mathbf{w}[m] \in \mathbb{C}^{N_\mathrm{R}}$ of the m-th subcarrier can be decomposed as an element-wise Hadamard product of the analog combiner $\mathbf{w}_\mathrm{A}[m] \in \mathbb{C}^{N_\mathrm{R}}$ and digital combiner $\mathbf{w}_\mathrm{D}[m] \in \mathbb{C}^{N_\mathrm{R}}$, i.e., $\mathbf{w}[m] = \mathbf{w}_\mathrm{A}[m] \odot \mathbf{w}_\mathrm{D}[m] = \left[[\mathbf{w}_\mathrm{A}[m]]_1 [\mathbf{w}_\mathrm{D}[m]]_1, ..., [\mathbf{w}_\mathrm{A}[m]]_{N_\mathrm{R}} [\mathbf{w}_\mathrm{D}[m]]_{N_\mathrm{R}}\right]^{\mathrm{T}}$. Both the analog and digital combiners depend on the underlying array architecture. In an analog TTD array, $\mathbf{w}_\mathrm{D}[m] = \mathbf{1}_{N_\mathrm{R}}$, i.e., $\mathbf{w}[m] = \mathbf{w}_\mathrm{A}[m]$, since there is no digital combining and both the phases $\phi_{\mathrm{A},n},\ \forall n$, and delays $\tau_{\mathrm{A},n},\ \forall n$, are introduced in the analog domain. On the other hand, with a fully digital array, $\mathbf{w}_\mathrm{A}[m] = \mathbf{1}_{N_\mathrm{R}}$, i.e., $\mathbf{w}[m] = \mathbf{w}_\mathrm{D}[m]$, as the array is insensitive to hardware impairments and the signal is combined in the digital domain after applying the phases $\phi_{\mathrm{D},n},\ \forall n$, and delays $\tau_{\mathrm{D},n},\ \forall n$. In general, the n-th elements of $\mathbf{w}_\mathrm{A}[m]$ and $\mathbf{w}_\mathrm{D}[m]$ are given as

$$[\mathbf{w}_\mathrm{A}[m]]_n = \exp\left[-j\left(2\pi(f_m - f_\mathrm{c})\tilde{\tau}_{\mathrm{A},n} + \tilde{\phi}_{\mathrm{A},n}\right)\right] \tag{2.4}$$

$$[\mathbf{w}_\mathrm{D}[m]]_n = \exp\left[-j\left(2\pi(f_m - f_\mathrm{c})\tau_{\mathrm{D},n} + \phi_{\mathrm{D},n}\right)\right] \tag{2.5}$$

where $f_m = f_\mathrm{c} - \mathrm{BW}/2 + (m-1)\mathrm{BW}/(M_\mathrm{tot} - 1)$. Note that there is no magnitude, but only phase and delay control in (2.5), since the digital array is used here to emulate TTD-based beam training with pencil beams.

The expressions (2.4) and (2.5) indicate that the beam pointing direction depends on the subcarrier frequency, phases, and delays. With a proper configuration of the phase and delay

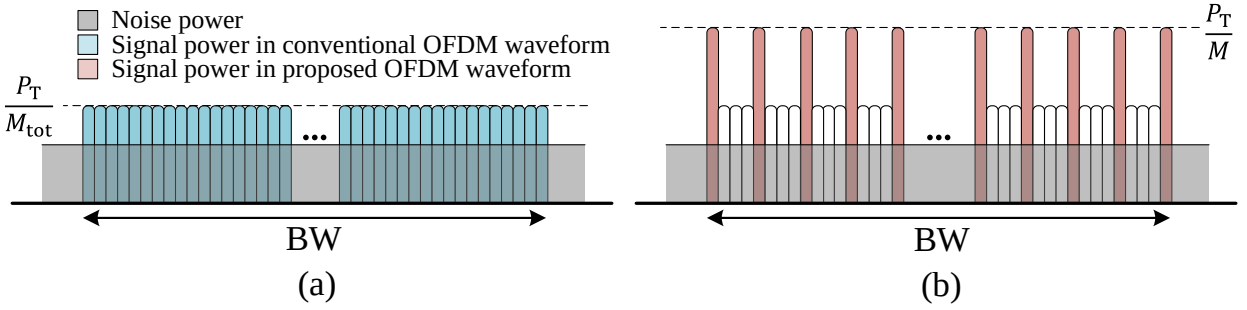


Figure 2.6: Illustration of OFDM power allocation in (a) conventional beam training and (b) TTD beam training.

taps in the analog and/or digital domain, it is possible to set up a codebook of combiners that covers all angular directions, as we discuss in the next section.

2.4.3 SNR and PAPR in OFDM-based Beam Training with TTD Arrays

The proposed beam training with frequency-dependent beams uses an OFDM based waveform where only a subset of M out of M_{tot} ($M < M_{\text{tot}}$) subcarriers is loaded at the BS. This allows the transmit signal power to be allocated to a lower number of subcarriers, which can increase the signal-to-noise ratio (SNR) per subcarrier compared to that in conventional beam training, as illustrated in Fig. 2.6. Specifically, when the total transmit power P_{T} is divided among M subcarriers, the SNR per subcarrier is given by the following expression

$$\text{SNR}_{\text{sc}} = \frac{20\log_{10}(N_{\text{T}})\,20\log_{10}(N_{\text{R}})\,\lambda^2}{(4\pi d)^2}\frac{P_{\text{T}}}{\Delta\text{BW}\,N_0 M}, \tag{2.6}$$

assuming a free-space path loss model. The terms $20\log_{10}(N_{\text{T}})$, $20\log_{10}(N_{\text{R}})$, λ, and d represent the transmit beamforming gain, receive beamforming gain, wavelength, and distance between the BS and UE. Power spectral density of the noise is denoted as N_0, while $\Delta\text{BW} = \text{BW}/(M_{\text{tot}} - 1)$ represents the subcarrier spacing. Note that the SNR per subcarrier is M_{tot}/M times larger than with a fully loaded OFDM waveform. Since the proposed DSP algorithm considers only M loaded subcarriers with high SNR, angle estimation is not

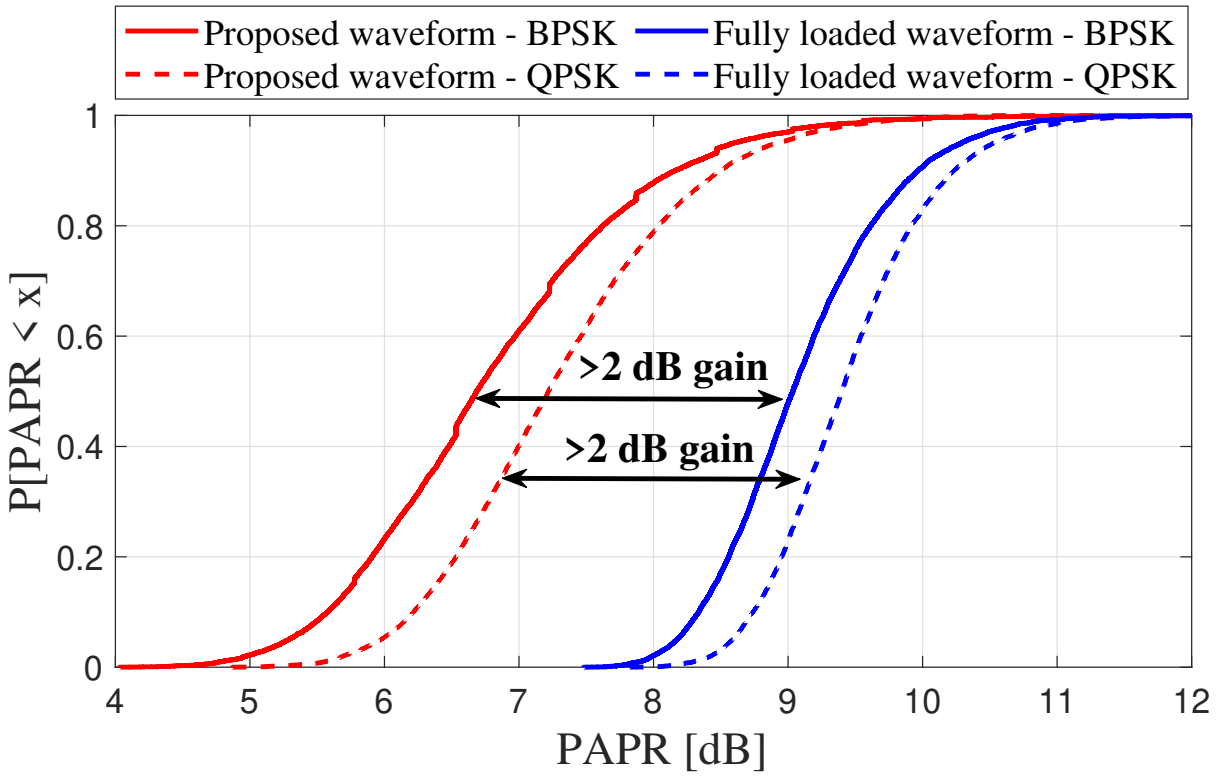


Figure 2.7: PAPR comparison between proposed and fully loaded OFDM waveforms, assuming BPSK or QPSK symbols.

noise-limited. Further, due to a lower number of used subcarriers, the proposed OFDM waveform for TTD beam training results in a more than 2dB lower peak-to-average power ratio (PAPR) than a fully loaded OFDM waveform, as presented in Fig. 2.7, where we assumed the same simulation parameters as in the previous subsection. We used a cyclic prefix of 128 samples and assumed that subcarriers are loaded either with BPSK or quadrature phase shift keying (QPSK) symbols.

2.5 Proposed Beam Training for UEs

2.5.1 Codebook Design for TTD-based UE Beam Training

As illustrated in Fig. 2.5(b), conventional phased arrays cannot estimate the AoA of the dominant cluster with one training symbol, and thus they require exhaustive beam sweeping. On the other hand, we have demonstrated in [YBC19] that D spatial directions in the angular range $(-\pi/2, \pi/2)$ can be simultaneously probed using an analog TTD array and a single OFDM symbol by mapping one subcarrier per direction, as illustrated in Fig. 2.5(c). We have shown that this can be achieved by setting the delay spacing to be $\Delta\tau = 1/\text{BW}$. The resulting codebook is, however, sensitive to frequency-selective channels since certain subcarriers can experience deep fades and thus miss to detect the incoming signal. The codebook can be enhanced by increasing its frequency diversity order R, i.e., by mapping R distinct subcarriers in each probed direction [BYG20]. Note that this enhancement requires $M = DR$ $(M \leq M_{\text{tot}})$ subcarriers to be used in beam training. The benefit of the enhanced codebook is illustrated in Fig. 2.5(d) for $R = 2$, where two subcarriers detect the dominant cluster. To increase the diversity, we define D distinct sets $\mathcal{M}_d,\ 1 \leq d \leq D$, of R subcarriers, where each set is associated with a different direction $d,\ 1 \leq d \leq D$. Mathematically, the R subcarriers from the set $\mathcal{M}_d$ have the same combiner $\mathbf{f}_d$, i.e., $\mathbf{w}[m] = \mathbf{f}_d,\ \forall m \in \mathcal{M}_d$, where the n-th element of $\mathbf{f}_d$ is defined as

$$[\mathbf{f}_d]_n = \exp[-j2\pi(n-1)(d-1-D/2)/D], \quad d \leq D. \tag{2.7}$$

The subcarriers in $\mathcal{M}_d$, however, should experience different channels, and thus we choose them uniformly across the bandwidth. The subcarriers in $\mathcal{M}_d$ see different channels if the distance between them is larger than the coherence bandwidth. This codebook can be created for an analog TTD array by setting the n-th phase and delay taps as follows

$$\phi_{\text{A},n} = (n-1)[\pi \sin(\theta_{\text{s}}) - \psi], \tag{2.8}$$

$$\tau_{\text{A},n} = (n-1)R/\text{BW}, \tag{2.9}$$

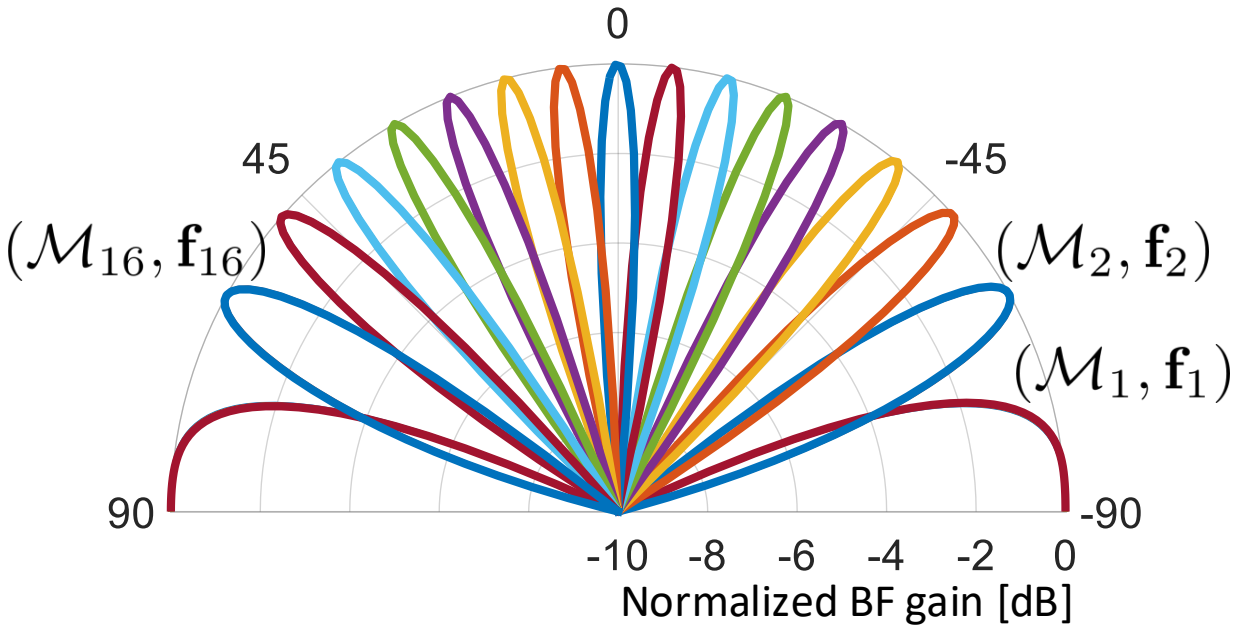


Figure 2.8: An example of robust TTD codebook for $N_{\mathrm{R}} = 16$, $D = 16$, and $R = 4$. All $D = 16$ directions are probed simultaneously. Direction d, $1 \leq d \leq D$, is associated with set of subcarriers $\mathcal{M}_d$ and combiner $\mathbf{f}_d$.

where $\psi = \mathrm{mod}(2\pi R(f_1 - f_{\mathrm{c}})/\mathrm{BW} + \pi, 2\pi) - \pi$, and mod() is the modulo operator. To ensure that $\mathbf{w}[m] = \mathbf{f}_d, \forall m \in \mathcal{M}_d$, for $d = 1, ..., D$, we set the steering angle θ_{s} to be $\theta_{\mathrm{s}} = -\pi/2$. An example of the resulting codebook with $N_{\mathrm{R}} = 16$, $D = 16$, and $R = 4$ is provided in Fig. 2.8. Different values of θ_{s} in (2.8) result in different codebook rotations, while the changes in (2.9) enable the adjustment of the range of probed angles. Note that the same enhanced codebook can be created for the hybrid sub-array based TTD array or fully digital array without the need to implement a fractional ADC sampling since $\Delta\tau$ is proportional to the Nyquist sampling period, i.e., $\Delta\tau = R/\mathrm{BW}$. Analog and digital delay taps of the hybrid sub-array based TTD array introduced in Sec. 2.3, can be expressed with respect to the indices of all antenna elements in the array $n = 1, .., N_{\mathrm{R}}$, as $\tau_{\mathrm{A},n} = (n - 1 - \lfloor (n-1)/N_{\mathrm{r}} \rfloor N_{\mathrm{r}})\,\Delta\tau$, and $\tau_{\mathrm{D},n} = \lfloor (n-1)/N_{\mathrm{r}} \rfloor N_{\mathrm{r}} \Delta\tau$, respectively. The operator $\lfloor x \rfloor$ rounds x to the nearest lower integer. Thus, the hybrid TTD array can create the enhanced codebook by setting the n-th

Table 2.1: Phase and delay tap settings for robust codebook design

Array arch.	$\mathbf{w}[m]$	$\phi_{\mathrm{A},n}$	$\tau_{\mathrm{A},n}$	$\phi_{\mathrm{D},n}$	$\tau_{\mathrm{D},n}$
Analog TTD	$\mathbf{w}_{\mathrm{A}}[m]$	(2.8)	(2.9)	N/A	N/A
Hybrid TTD	$\mathbf{w}_{\mathrm{A}}[m] \odot \mathbf{w}_{\mathrm{D}}[m]$	(2.10)	(2.11)	N/A	(2.12)
Digital	$\mathbf{w}_{\mathrm{D}}[m]$	N/A	N/A	(2.13)	(2.14)

taps of its analog and digital combiners in the following way

$$\phi_{\mathrm{A},n} = (n-1)[\pi \sin(\theta_{\mathrm{s}}) - \psi], \tag{2.10}$$

$$\tau_{\mathrm{A},n} = (n - 1 - \lfloor (n-1)/N_{\mathrm{r}} \rfloor N_{\mathrm{r}})\, R/\mathrm{BW}, \tag{2.11}$$

$$\tau_{\mathrm{D},n} = \lfloor (n-1)/N_{\mathrm{r}} \rfloor N_{\mathrm{r}} R/\mathrm{BW}, \tag{2.12}$$

where $\theta_{\mathrm{s}} = -\pi/2$ and ψ is defined as earlier. The result in (2.12) suggests that the h-th sub-array needs to introduce a digital delay of $2(h-1)N_{\mathrm{r}}R$ time samples, assuming the Nyquist sampling frequency $f_{\mathrm{s}} = 2\mathrm{BW}$. The considered hybrid sub-array based array in Fig. 2.2 does not apply the phase changes in the digital domain. The digital array can create the enhanced codebook by using the following digital taps

$$\phi_{\mathrm{D},n} = (n-1)\Delta\phi, \quad \Delta\phi \in \mathbb{R} \tag{2.13}$$

$$\tau_{\mathrm{D},n} = (n-1)R/\mathrm{BW}. \tag{2.14}$$

The phase tap in (2.13) implies that the digital array can leverage the DSP to introduce any phase spacing $\Delta\phi$. With $f_{\mathrm{s}} = 2\mathrm{BW}$, the n-th antenna branch will introduce the digital delay of $2(n-1)R$ time samples according to (2.14).

The phase and delay taps required for the design of a robust codebook are summarized in Table. 2.1 for all three arrays.

We note that the analog and hybrid sub-array based TTD architectures have the same limited flexibility of receive combining in beam training. Namely, once their corresponding analog combiners $\mathbf{w}_{\mathrm{A}}[m],\ m \in \mathcal{M}$, and digital combiners $\mathbf{w}_{\mathrm{D}}[m],\ m \in \mathcal{M}$ are set up, they cannot be further changed or manipulated in DSP. In both architectures, this happens

because the signals from different antenna branches are completely or partially combined before passing through ADCs. Thus, the inability to rotate the combiners limits the number of sounded directions to D in both arrays. The diversity order R is also limited, but not necessarily the same in both arrays, as discussed later in the chapter. On the other hand, the digital array can exploit digitized signals in all antenna branches and combine them from many different directions in DSP by changing the phases $\phi_{\mathrm{D},n},\ \forall n$. Different phases $\phi_{\mathrm{D},n}$ introduces angular shifts of the entire codebook, and enable scanning more angles and/or higher diversity.

2.5.2 DSP Algorithm for UE Beam Training

We use the designed beam training codebook to develop a non-coherent power-based DSP angle estimation algorithm with high accuracy [BYG20]. Non-coherent algorithms are preferred in mmW beam training as they do not require measurements in (2.3) to include the phase information, and thus they can avoid complex joint synchronization and beam training receiver processing.

Since the subcarriers from $\mathcal{M}_d,\ \forall d$, experience different channels, we can consider the received signal in all D probed directions as random. In a clustered multipath channel, the vector of expected powers in D directions $\mathbf{p} = [p_1, p_2, ..., p_D]^{\mathrm{T}}$ can be expressed as

$$\mathbf{p} = \mathbf{B}\mathbf{g} + N_{\mathrm{R}}\sigma_{\mathrm{N}}^2\mathbf{1}, \tag{2.15}$$

where $\mathbf{B} \in \mathbb{R}^{D\times Q}$ is a known dictionary obtained by generalizing the UE beamforming (BF) gains in Q angles $\xi_q,\ q = 1, ..., Q$, for all D combiners. The (d, q)-th element of $\mathbf{B}$ is defined as $[\mathbf{B}]_{d,q} = |\mathbf{f}_d^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\xi_q)|^2$, where $\mathbf{a}_{\mathrm{T}}(\xi_q)$ is the receive spatial response introduced in Sec. 2.4.1. The vector $\mathbf{g} \in \mathbb{R}^Q$ has only one non-zero element. For a detailed derivation of (2.15), please refer to Appendix A.1.

During beam training, the estimates of $p_d,\ \forall d$, are obtained by averaging out the powers

of all subcarriers from the corresponding set $\mathcal{M}_d$, $\forall d$, as follows

$$\hat{p}_d = \frac{1}{R} \sum_{m \in \mathcal{M}_d} |y[m]|^2. \tag{2.16}$$

In fact, it can be shown that the sample mean in (2.16) is the maximum likelihood (ML) estimator of p_d, $\forall d$. The vector of all power estimates is denoted as $\hat{\mathbf{p}}$, which approximate $\mathbf{p}$ in (2.15). Note that $\hat{\mathbf{p}}$ is estimated using $M = DR$ frequency-domain measurements $y[m], \forall m$, in (2.3) of only one OFDM symbol. Based on the power measurement model in (2.15), AoA estimation can be solved based on the ML criterion using simple linear algebra operations. The AoA $\theta^{(\mathrm{R})}$ estimate is obtained by finding the index of the column in $\mathbf{B}$ which has the highest correlation with $\hat{\mathbf{p}}$, which is mathematically expressed as

$$\hat{\theta}^{(\mathrm{R})} = \xi_{q^\star}, \text{ where } q^\star = \underset{q}{\operatorname{argmax}} \frac{\hat{\mathbf{p}}^{\mathrm{T}} [\mathbf{B}]_{:,q}}{||[\mathbf{B}]_{:,q}||}. \tag{2.17}$$

The proposed algorithm can achieve high AoA estimation accuracy by increasing Q, i.e., the number of the columns in the dictionary matrix $\mathbf{B}$. Although this increases the DSP complexity, the proposed beam training scheme can still be performed with a single OFDM symbol. Note that the accuracy can be negatively affected by hardware impairments, which distort the combiners and thus the elements $[\mathbf{B}]_{d,q}, \forall d, q$, which might not correspond to the expected beamforming gains. In this chapter, we use the root mean square error (RMSE) of AoA estimation as main metric for the comparison of the proposed TTD architectures. The AoA RMSE closely describes the beam training performance and it can be directly converted to an alternative metric in other applications, including the spectral efficiency in mmW data communication and position error in localization.

2.6 Comparison of TTD Architectures in UE Beam Training

In this section, we study the impact of limited TTD delay range in both architectures on beam training performance and we explain the interplay between the number of antenna elements N_{R}, bandwidth BW, and diversity order R. We also numerically evaluate the impact of

hardware impairments and ADC quantization error on the AoA estimation accuracy.

2.6.1 Impact of Limited TTD Delay Range on UE Beam Training

In this subsection, we assume that the analog and hybrid sub-array based architectures have TTD elements with the same state-of-the-art maximum delay compensation of $T_{\text{C-max}} = 15$ ns, or equivalently the same interleaving factor [BYL21].

To realize the proposed beam training algorithm, $\tau_{\text{A},N_{\text{R}}} \leq T_{\text{C-max}}$ needs to be satisfied for the analog, and $\tau_{\text{A},N_{\text{r}}} \leq T_{\text{C-max}}$ for the hybrid sub-array based TTD array. Based on these conditions, it is straightforward to show that the achievable diversity order R is limited as

$$1 \leq R \leq \frac{T_{\text{C-max}}}{N_{\text{R}} - 1}\text{BW} \quad \text{and} \quad 1 \leq R \leq \frac{T_{\text{C-max}}}{N_{\text{r}} - 1}\text{BW}, \tag{2.18}$$

for the analog and hybrid sub-array based array, respectively. Note that with $R < 1$, the beam training algorithm cannot be realized with a single OFDM symbol. On the other hand, a large R provides more precise ML estimates in (2.16) due to better averaging. The expressions in (2.18) describe the dependency of R on the basic system parameters N_{R}, N_{r}, and BW. In the remainder of this subsection, we numerically evaluate the interplay among them.

We study the beam training performance of different architectures in terms of AoA estimation accuracy, assuming that R is constrained to be maximal power of 2. We consider a system with carrier frequency $f_{\text{c}} = 60$ GHz, bandwidth values in the range 0.5 GHz $\leq$ BW $\leq$ 4.5 GHz, and $M_{\text{tot}} = 4096$ subcarriers for any bandwidth. The transmitter array size is $N_{\text{T}} = 128$, while the receive array size can take values $N_{\text{R}} = \{16, 32\}$. There are $N_{\text{r}} = 4$ antennas in each sub-array in hybrid TTD architecture, regardless of the total number of antennas. The number of probed directions in beam training is assumed to be $D = 2N_{\text{R}}$ and the dictionary size is $Q = 1024$. The channel consists of $L = 3$ clusters, where one is 10 dB stronger than the other two. Fading is simulated by 20 rays within each cluster with up to 10 ns spread. There is no intra-cluster angular spread. Pre-beamforming SNR is defined as

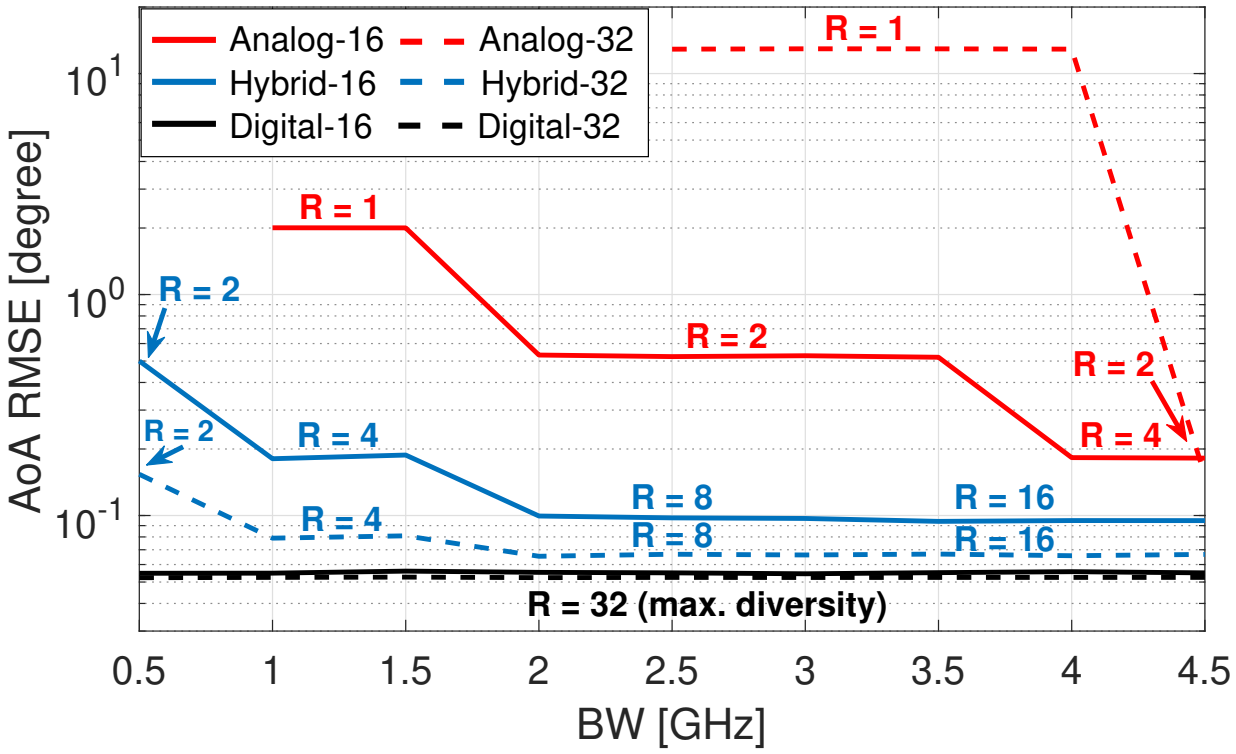


Figure 2.9: Beam training performance comparison of the three considered architectures and the interplay of R, N_{R}, and BW.

SNR $\triangleq \sum_{l=1}^{L} \sigma_l^2/\sigma_{\mathrm{N}}^2$, and it is assume to be SNR $= -20$ dB.

In Fig. 2.9, we present the results for the beam training performance and the interplay of the considered parameters. In both cases $N_{\mathrm{R}} = 16$ and $N_{\mathrm{R}} = 32$, the analog TTD array architecture has the highest RMSE of AoA estimation due to low achievable diversity order R. As discussed earlier, analog arrays have large delay range requirements, and thus better estimation accuracy (equivalently, higher R) requires larger BW. Similarly, increasing the array size N_{R} can have a positive effect on the performance. However, if BW is not large enough and there is no diversity ($R = 1$), larger arrays do not improve the estimation accuracy in frequency-selective channels. The analog arrays do not have the results for the values of BW for which the proposed single-shot beam training cannot be realized ($R < 1$).

In hybrid sub-array based TTD arrays, higher diversity orders can be utilized since $N_{\mathrm{r}} < N_{\mathrm{R}}$, which leads to better estimation accuracy compared to analog arrays. Increase in the number of antenna elements does not change achievable R in hybrid sub-array based arrays since we assume that $N_{\mathrm{r}} = 4$ remains constant. It does, however, improve the estimation accuracy of hybrid arrays, which approaches the sub-degree performance of fully digital arrays. Since R can be maximized through DSP in digital arrays, their performance is independent of BW. The floor of the AoA RMSE is determined by the dictionary size $Q = 1024$. Based on described results in Fig. 2.9, one can predict the diversity order R and beam training performance for any considered array architecture, given the system parameters BW, N_{R}, and $T_{\text{C-max}}$.

2.6.2 Impact of TTD Hardware Impairments on UE Beam Training

Next, we study the impact of practical TTD hardware impairments and ADC quantization errors on beam training in analog and hybrid sub-array based architectures. Here we keep AoA RMSE as the performance metric and use the same system parameters as in the previous subsection. We consider a specific case with $N_{\mathrm{R}} = 16$ and BW $= 2$ GHz.

In Fig. 2.10, we study the beam training performance under the phase and delay errors. Unlike analog and hybrid TTD arrays, fully digital array is not sensitive to these hardware impairments and we include its performance with the maximum $R = 32$ as the benchmark. With the considered system parameters, analog TTD array has the diversity order $R = 2$, which limits its angle estimation accuracy and robustness to hardware errors. We can see that the beam training algorithm can tolerate phase errors with the standard deviation of up to $\sigma_{\mathrm{P}} = 10^{\circ}$ and delay errors with the standard deviation of up to $\sigma_{\mathrm{T}} = 50$ ps. Hybrid sub-array based TTD array achieves a lower estimation accuracy and greater robustness to delay and phase errors than analog TTD array since it leverages the diversity order $R = 8$ in beam training. It can tolerate large phase errors and delay errors with the standard deviation of around $\sigma_{\mathrm{T}} = 200$ ps. It is worth noting that the delay errors in hybrid sub-array based

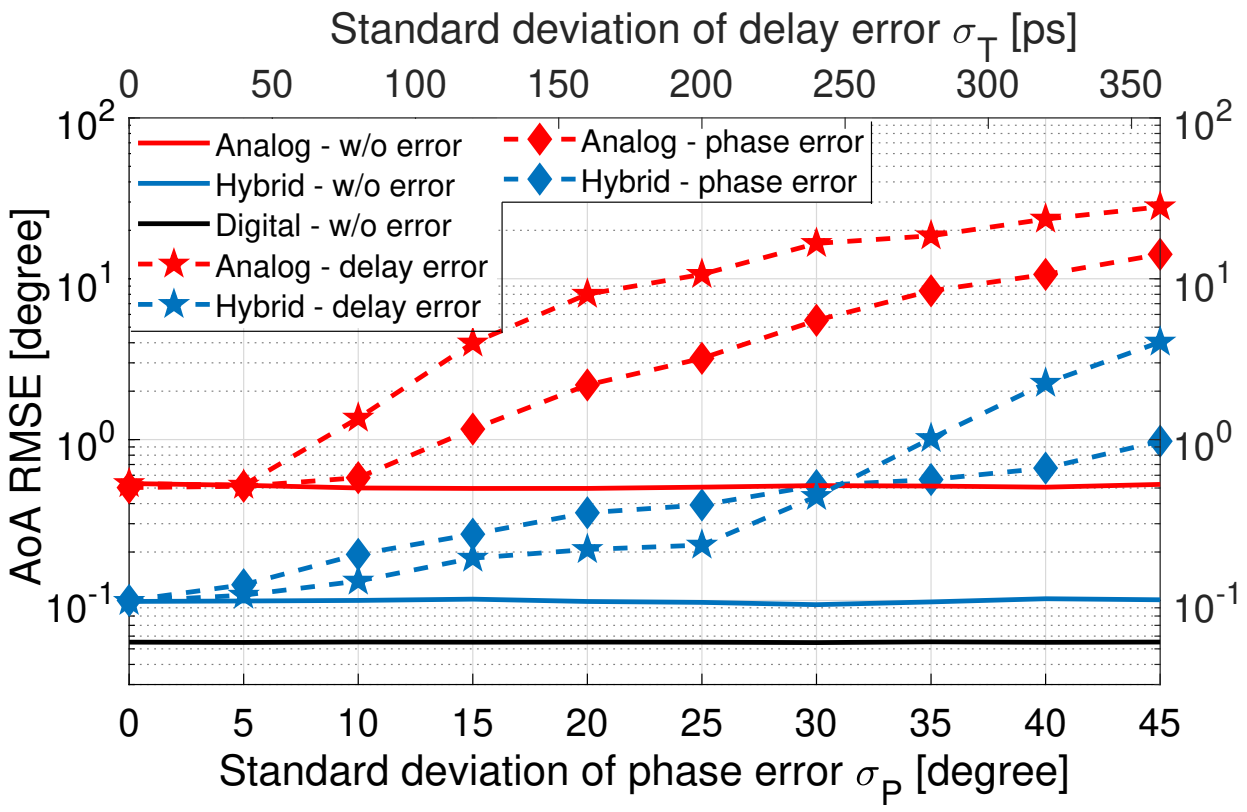


Figure 2.10: Beam training performance comparison of the three considered architectures under the distorted delay taps $\tilde{\tau}_n \sim \mathcal{N}(\tau_n, \sigma_{\mathrm{T}}^2)$, $\forall n$, and phase taps $\tilde{\phi}_n \sim \mathcal{N}(\phi_n, \sigma_{\mathrm{P}}^2)$, $\forall n$. The curves with the delay error (dashed with stars) and phase error (dashed with diamonds) are associated with the upper and lower x-axis, respectively.

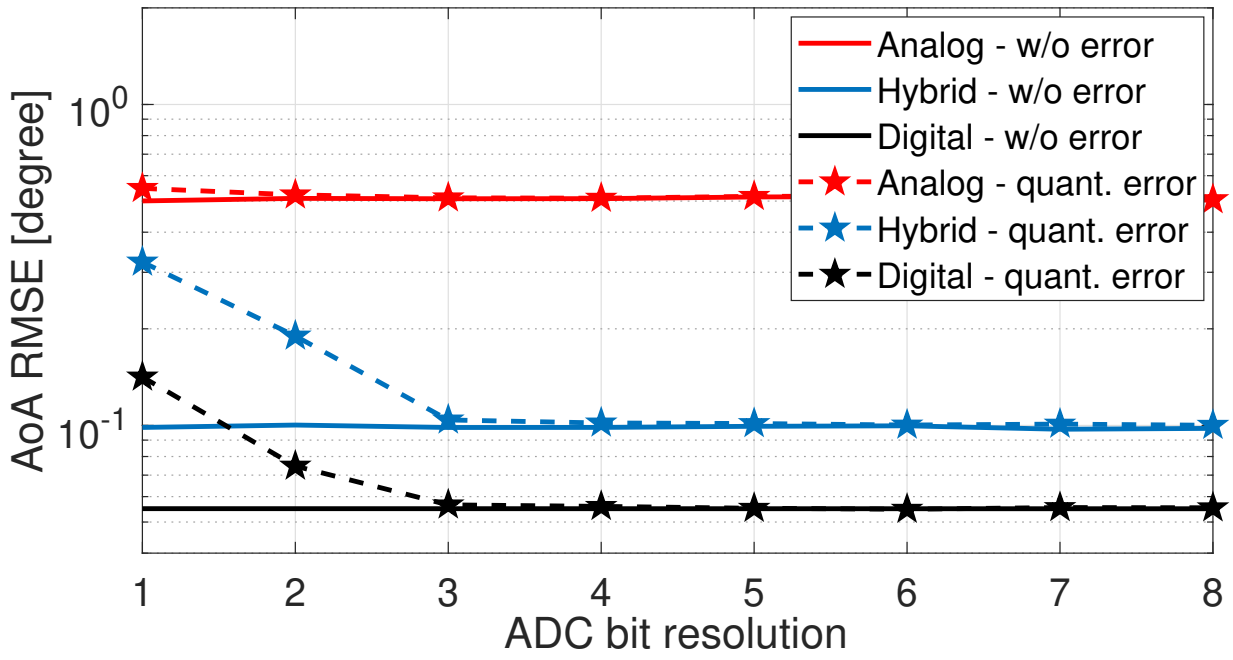


Figure 2.11: Beam training performance comparison of the three considered architectures under different ADC resolutions.

arrays are independent of the reduced delay taps in the corresponding TTD elements.

In Fig. 2.11, we present how finite ADC resolution affects the beam training performance with different array architectures. For fair comparison, we assume that the automatic gain control (AGC) outputs a unit-variance signal in all architectures. We can observe that the AoA estimation accuracy of the analog TTD array with a single RF chain is marginally affected by low ADC resolution. On the other hand, low resolution ADCs have a noticeable impact on beam training with the hybrid sub-array basedTTD and fully digital arrays, as combined quantization errors from different RF chains deteriorate the estimation accuracy. We note, however, that the deteriorated accuracy is still within the sub-degree range and lower than that of the analog array. Our results indicate that practical mmW and sub-THz transceivers may require ADCs with only a few bits of resolution for effective beam training. For example, with only 3-bit resolution, the performance loss is negligible in any array. In

addition, low-resolution ADCs have a positive impact on the overall power efficiency of the considered TTD architectures.

2.7 Joint Beam Training Between BS and UE

In this section, we describe a joint beam training algorithm between the BS and UE that requires only one OFDM symbol.

2.7.1 System Model for Joint Beam Training

We a consider a similar channel model as in Sec. 2.4.1, but we treat the channel gains $g_l[m]$. $\forall l, m$, and thus channel matrices $\mathbf{H}[m]$, $\forall m$, as constants. Equivalently, the time domain channel matrix is assumed to be constant during the beam training procedure, regardless of its duration (one or multiple OFDM symbols). We assume that the BS is equipped with a fully digital antenna array, while the UE is assumed to be equipped with a fully connected hybrid TTD array described in Sec. 2.3. Both arrays are assumed to be unaffected by hardware impairments.

Let $\mathbf{v}[m] \in \mathbb{C}^{N_\mathrm{T}}$, $\|\mathbf{v}[m]\|_2 = 1$, be a normalized BS digital precoder for the m-th subcarrier. Unlike in beam training for the UE only, the AoD is not known in advance and the precoder $\mathbf{v}[m]$ is not the same for all subcarriers. Similarly, let $\mathbf{w}_r[m] \in \mathbb{C}^{N_\mathrm{R}}$, $\|\mathbf{w}_r[m]\|_2 = 1$, be a normalized UE analog TTD combiner for the m-th subcarrier in the r-th RF chain. We omit the subscript 'A' in $\mathbf{w}_r[m]$ for brevity. With the corresponding channel in (2.1), the received signal $y_r[m]$ at the m-th subcarrier in the r-th RF chain can be expressed as

$$y_r[m] = \mathbf{w}_r^\mathrm{H}[m]\mathbf{H}[m]\mathbf{v}[m]s[m] + \mathbf{w}_r^\mathrm{H}[m]\mathbf{n}[m], \ m \in \mathcal{M}, \tag{2.19}$$

where $s[m]$, $|s[m]|^2 = 1/M$, is a BPSK pilot at the m-th subcarrier and $\mathbf{n}[m] \sim \mathcal{CN}\left(0, \sigma_\mathrm{N}^2 \mathbf{I}_{N_\mathrm{R}}\right)$ is white Gaussian noise. The n-th element of $\mathbf{w}_r[m]$ can be expressed as

$$[\mathbf{w}_r[m]]_n = \exp\left[-j\left(2\pi(f_m - f_\mathrm{c})\tau_n + \phi_{r,n}\right)\right]/\sqrt{N_\mathrm{R}}, \tag{2.20}$$

where f_m is the frequency of the m-th subcarrier, defined as $f_m = f_c - \text{BW}/2 + (m-1)\text{BW}/(M_{\text{tot}}-1)$.

2.7.2 Codebook Design for TTD-based Joint Beam Training

Joint beam training algorithm requires frequency-dependent codebooks to be designed both at the BS and UE side. Compared to the UE codebooks in [YBC19, BYG20, BYL21], the codebooks for joint training are based on different subcarrier selection and mapping schemes. With N_{T} and N_{R} antennas at the BS and UE, respectively, there are at least $N_{\text{T}}N_{\text{R}}$ beam pairs to be considered in the training. Instead of mapping the subcarriers into D different UE beams as in the previous sections, we aim to map them into different beam pairs. Thus, the set of used subcarriers $\mathcal{M}$ has $M = N_{\text{T}}N_{\text{R}}$ elements. We propose an OFDM waveform with non-uniform subcarrier selection for joint BS-UE training. The total of M_{tot} subcarriers is divided into N_{R} groups, and in each group, the first N_{T} subcarriers are selected and loaded with BPSK pilots. Mathematically, the set $\mathcal{M}$ is defined as

$$\mathcal{M} = \Big\{ m \mid m = m_{\text{T}} + (m_{\text{R}}-1)\lfloor M_{\text{tot}}/(N_{\text{R}})\rfloor,\ m_{\text{T}} = 1, \cdots, N_{\text{T}},\ m_{\text{R}} = 1, \cdots, N_{\text{R}} \Big\}, \quad (2.21)$$

where $\lfloor x \rfloor$ rounds x to the nearest lower integer. At the BS side, we design a codebook where in each of the N_{R} groups, the N_{T} subcarriers are assigned N_{T} discrete Fourier transform (DFT) precoders $\mathbf{u}_{m_{\text{T}}},\ m_{\text{T}} = 1, ..., N_{\text{T}}$, that cover the entire angular range $(-\pi/2, \pi/2)$. The k-th element of $\mathbf{u}_{m_{\text{T}}},\ \forall m_{\text{T}}$, is defined as

$$[\mathbf{u}_{m_{\text{T}}}]_k = \exp[-j2\pi(k-1)(m_{\text{T}} - 1 - N_{\text{T}}/2)/N_{\text{T}}]/\sqrt{N_{\text{T}}}. \quad (2.22)$$

Therefore, the subcarriers from the set $\mathcal{M}^{(\text{T})}_{m_{\text{T}}} = \{m \mid m = m_{\text{T}} + (m_{\text{R}}-1)\lfloor M_{\text{tot}}/(N_{\text{R}})\rfloor,\ m_{\text{R}} = 1, ..., N_{\text{R}}\}$ are assigned the m_{T}-th DFT precoder. In other words, the precoders $\mathbf{v}[m],\ m \in \mathcal{M}^{(\text{T})}_{m_{\text{T}}}$ are designed such that

$$\mathbf{v}[m] = \mathbf{u}_{m_{\text{T}}},\ m \in \mathcal{M}^{(\text{T})}_{m_{\text{T}}}. \quad (2.23)$$

At the UE side, we focus on the first RF chain and design a codebook where the N_{T} subcarriers from the set $\mathcal{M}^{(\text{R})}_{m_{\text{R}}} = \{m \mid m = m_{\text{T}} + (m_{\text{R}}-1)\lfloor M_{\text{tot}}/(N_{\text{R}})\rfloor,\ m_{\text{T}} = 1, ..., N_{\text{T}}\}$ are

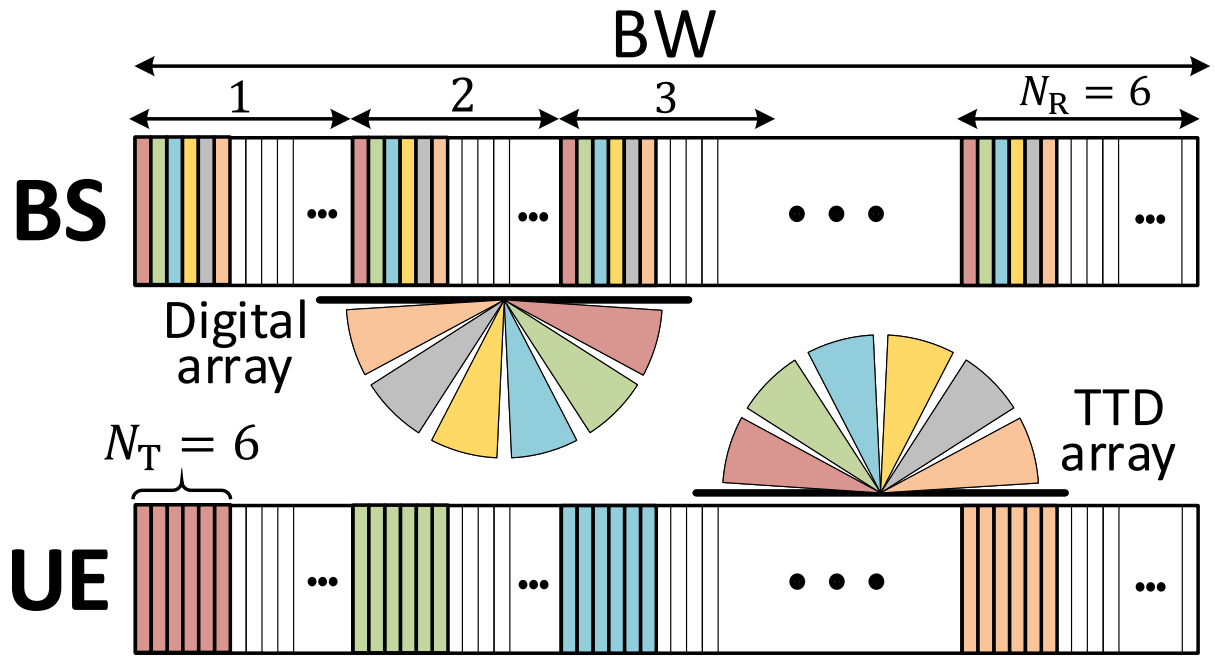


Figure 2.12: Illustration of subcarrier selection and codebook design at the BS and UE, assuming $N_T = 6$ and $N_R = 6$.

assigned the same DFT combiner $\mathbf{f}_{m_R}$, whose n-th element is defined as

$$[\mathbf{f}_{m_R}]_n = \exp[-j2\pi(n-1)(m_R - 1 - N_R/2)/N_R]/\sqrt{N_R}. \tag{2.24}$$

Thus, the TTD combiners $\mathbf{w}_1[m],\ m \in \mathcal{M}^{(R)}_{m_R}$, in the first RF chain are designed to satisfy the following equality

$$\mathbf{w}_1[m] = \mathbf{f}_{m_R},\ m \in \mathcal{M}^{(R)}_{m_R}. \tag{2.25}$$

The subcarrier selection and codebook design at both BS and UE are illustrated on a small example in Fig. 2.12.

Since the BS is equipped with a digital array, its codebook in (2.23) can be easily designed through frequency-domain DSP. Thus, the precoders for all subcarriers from the set $\mathcal{M}^{(T)}_{m_T}$ can be perfectly aligned in the m_T-th angular direction. An example of the resulting codebook is provided in Fig. 2.13(a). On the other hand, the UE codebook in (2.25) is cre-

ated by setting the delay taps $\tau_n,\ \forall n$, and phase taps $\phi_{r,n},\ \forall r, n$. Similarly as in [YBC19], we set the delay taps to be

$$\tau_n = (n-1)/\mathrm{BW},\ 1 \leq n \leq N_\mathrm{R}. \tag{2.26}$$

Clearly, the delay difference between neighboring antennas is $\Delta\tau = 1/\mathrm{BW}$ in (2.26). Unlike in [YBC19], there are N_T subcarriers that need to be mapped into the same direction according to (2.25). The N_T subcarriers from the set $\mathcal{M}^{(\mathrm{R})}_{m_\mathrm{R}}$ correspond to different frequencies and thus they experience the beam squint effect, i.e., slightly different combining angles. Consequently, the BF gains at the N_T subcarriers are different. An intuitive way to reduce the gain difference is to make the beam squint symmetric around the pointing direction of the desired precoders $\mathbf{f}_{m_\mathrm{R}},\ \forall m_\mathrm{R}$. This is achieved by aligning the codebook using the phase shifters. The squint alignment phase tap is the same in all RF chains and for the n-th antenna it is given as

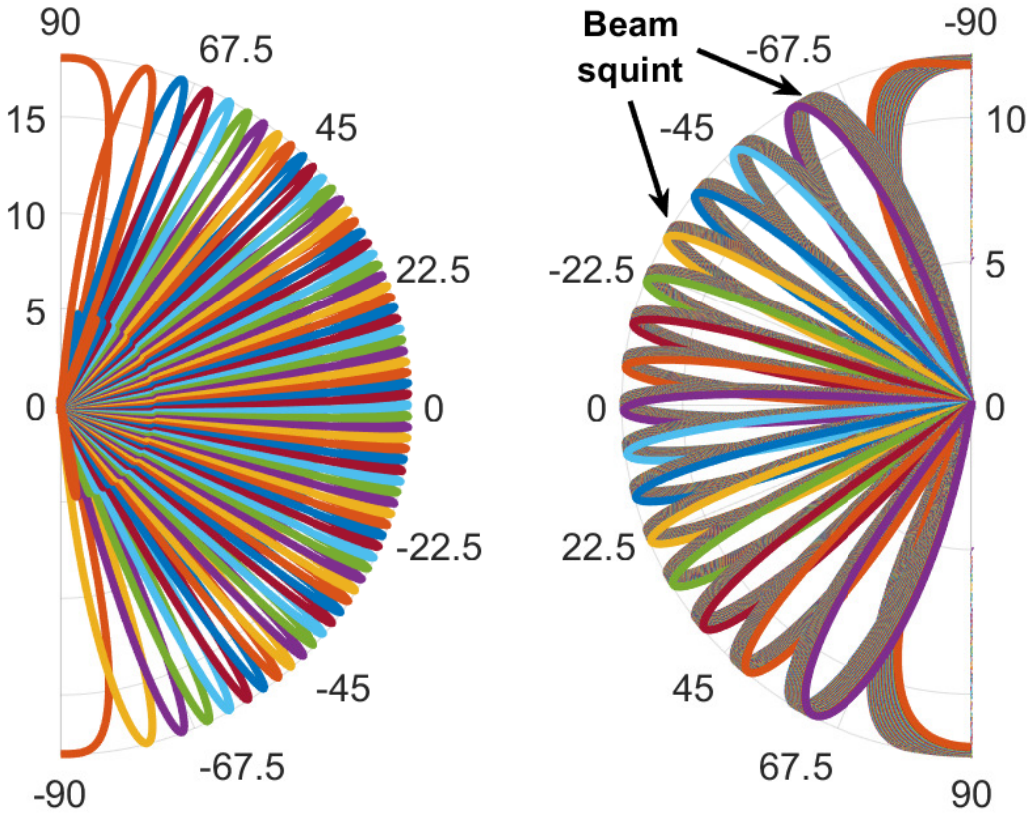
$$\phi_{\mathrm{sq},n} = -(n-1)\mathrm{mod}\left(2\pi(f_\mathrm{mid} - f_\mathrm{c})/\mathrm{BW} + \pi, 2\pi\right), \tag{2.27}$$

where $f_\mathrm{mid} = f_{N_\mathrm{T}/2} + \mathrm{BW}/(2M_\mathrm{tot} - 2)$ is the "middle" frequency of the subcarriers in $\mathcal{M}^{(\mathrm{R})}_1$ and mod() is the modulo operator. Since the alignment taps in (2.27) are frequency-flat, they are applied to all sets $\mathcal{M}^{(\mathrm{R})}_{m_\mathrm{R}},\ \forall m_\mathrm{R}$. To increase the robustness to frequency-selective channels through frequency diversity, we rotate the UE codebook in different RF chains by using the following rotation phase taps

$$\phi_{\mathrm{rot},r,n} = (r-1)(n-1)2\lfloor N_\mathrm{R}/N_\mathrm{H}\rfloor\pi/N_\mathrm{R},\quad \forall r, n. \tag{2.28}$$

The taps in (2.28) ensure that the codebook diversity is $R = N_\mathrm{H}$, i.e., that each BS-UE beam pair is probed by N_H different subcarriers in different RF chains. For example, the beam pair that is probed by the first subcarrier from $\mathcal{M}^{(\mathrm{R})}_{m_\mathrm{R}}$ in the first RF chain, is also probed by the first subcarrier from $\mathcal{M}^{(\mathrm{R})}_{m_\mathrm{R}+\lfloor N_\mathrm{R}/N_\mathrm{H}\rfloor}$ in the second RF chain, the first subcarrier from $\mathcal{M}^{(\mathrm{R})}_{m_\mathrm{R}+2\lfloor N_\mathrm{R}/N_\mathrm{H}\rfloor}$ in the third RF chain, etc. Given $\phi_{\mathrm{sq},n},\ \forall n$, and $\phi_{\mathrm{rot},r,n},\ \forall r, n$, we design the overall phase taps as

$$\phi_{r,n} = \phi_{\mathrm{sq},n} + \phi_{\mathrm{rot},r,n},\quad 1 \leq r \leq N_\mathrm{H}, 1 \leq n \leq N_\mathrm{R}. \tag{2.29}$$



(a) (b)

Figure 2.13: The resulting codebooks (a) at the BS with the fully digital array, and (b) at the UE with the TTD array, assuming $N_{\mathrm{T}} = 64$ and $N_{\mathrm{R}} = 16$. The beams are multiplied by the antenna size and they are plotted in the log scale.

The phase difference between the neighboring antennas in the r-th RF chain in $\Delta\phi_r = (r-1)2\lfloor N_{\mathrm{R}}/N_{\mathrm{H}} \rfloor \pi/N_{\mathrm{R}} - \mathrm{mod}\,(2\pi(f_{\mathrm{mid}} - f_{\mathrm{c}})/\mathrm{BW} + \pi, 2\pi)$. An example of the resulting UE codebook is provided in Fig. 2.13(b).

In the next subsection, we study the limitations of the designed codebooks. We first explain how the proposed design scales with the basic system parameters. Then we derive the maximum BF gain loss due to the beam squint at the UE side. Lastly, we explain why the proposed codebooks may require a larger transmit power or a denser network deployment than the codebooks in the conventional EBS.

2.7.3 Limitations of Codebook for TTD-based Joint Beam Training

The designed codebooks experience three main challenges: 1) scalability of the design training parameters when the numbers of antenna elements at the BS and UE increase; 2) BF gain loss at the UE side due to the beam squint in each probed direction; 3) required transmit power in beam training and/or distance between the BS and UE (network density).

1) *Scalability of design parameters*: The codebooks based on frequency-to-beam-pair mapping are intrinsically limited by the total number of subcarriers M_{tot} in the OFDM system. As discussed in Sec. 2.7.2, the design requires $M = N_{\text{T}}N_{\text{R}}$ subcarriers to be used in the training. Therefore, the condition $M \leq M_{\text{tot}}$ need to be satisfied to realize a joint BS-UE beam training with a single OFDM symbol and codebook diversity $R = N_{\text{H}}$. In the case of large array sizes when $M > M_{\text{tot}}$, the training can be realized by making a trade-off between the required number of training symbols and diversity R. A single-symbol beam training is possible so long as $\lceil M/(N_{\text{H}}M_{\text{tot}})\rceil \leq 1$. The idea is to leverage sectorized beam probing at the UE, where each RF chain probes N'_{R}, $N_{\text{R}}/N_{\text{H}} \leq N'_{\text{R}} < N_{\text{R}}$, directions using all M subcarriers. This requires a different design of the delay and rotation phase taps in (2.26) and (2.28), respectively, and it comes at the cost of a lower codebook diversity R ($< N_{\text{H}}$), as the beam pairs cannot be probed by N_{H} subcarriers. On the other hand, realizing beam training with $R = N_{\text{H}}$ when $M > M_{\text{tot}}$ requires the use of at least $\lceil M/M_{\text{tot}}\rceil$ wideband training symbols to probe all beam pairs. In the consecutive symbols, the UE needs to adjust the rotation phase shifts in (2.28). For the rest of this chapter, we will assume that $M < M_{tot}$, i.e., beam training can be realized with a single OFDM symbol and $R = N_{\text{H}}$.

2) *BF gain loss at UE side due to beam squint*: As shown in an example of the UE codebook in Fig. 2.13(b), the subcarriers from $M_{\text{R},m_{\text{R}}}$, $\forall m_{\text{R}}$, are combined from slightly different angles and thus can experience different BF gains. The phase taps in (2.27) reduce the BF gain loss by aligning the middle frequencies in $\mathcal{M}^{(\text{R})}_{m_{\text{R}}}$, $\forall m_{\text{R}}$, with the desired probing directions. Thus, maximal gain loss occurs for subcarriers with the lowest and highest frequencies

in $\mathcal{M}_{m_{\mathrm{R}}}^{(\mathrm{R})}$, $\forall m_{\mathrm{R}}$. For $\mathcal{M}_{1}^{(\mathrm{R})}$, it can be defined as the inner product of the impairment-free combiners that correspond to frequencies f_{mid} and f_1:

$$\Delta G_{\max}^{(\mathrm{R})} = N_{\mathrm{R}}^2 - \left| \sum_{n=1}^{N_{\mathrm{R}}} e^{j(2\pi f_{\mathrm{mid}}\tau_n + \phi_{r,n} - 2\pi f_1 \tau_n - \phi_{r,n})} \right|^2 \tag{2.30}$$

Since $f_{\mathrm{mid}} - f_1 = (N_{\mathrm{T}} - 1)\mathrm{BW}/(2M_{\mathrm{tot}} - 2)$, (2.30) can be simplified as follows

$$\begin{aligned} \Delta G_{\max}^{(\mathrm{R})} &= N_{\mathrm{R}}^2 - \left| \sum_{n=1}^{N_{\mathrm{R}}} e^{j(n-1)\pi(N_{\mathrm{T}}-1)/(M_{\mathrm{tot}}-1)} \right|^2 \\ &= N_{\mathrm{R}}^2 - \left| \frac{\sin(\pi N_{\mathrm{R}}(N_{\mathrm{T}} - 1)/(2M_{\mathrm{tot}} - 2))}{\sin(\pi(N_{\mathrm{T}} - 1)/(2M_{\mathrm{tot}} - 2))} \right|^2 . \end{aligned} \tag{2.31}$$

The maximum gain loss in (2.31) is obtained when $M = N_{\mathrm{T}}N_{\mathrm{R}} = M_{\mathrm{tot}}$ and it can be approximated as follows:

$$\Delta G_{\max}^{(\mathrm{R})} \approx N_{\mathrm{R}}^2 - \frac{1}{|\sin(\pi/(2N_{\mathrm{R}}))|^2} \tag{2.32}$$

As shown in Fig. 2.13(b), the closer the DFT combiners are to 0°, the smaller beam squint they experience. However, their beam widths are also smaller around 0° and thus the maximum gain loss in (2.31) and (2.32) applies to all probed directions, i.e, all sets $\mathcal{M}_{m_{\mathrm{R}}}^{(\mathrm{R})}$, $\forall m_{\mathrm{R}}$.

3) *Transmit power and network density*: Since the transmit power p_{T} is divided among $M = N_{\mathrm{T}}N_{\mathrm{R}}$ OFDM subcarriers, which probe $N_{\mathrm{T}}N_{\mathrm{R}}$ different beam pairs, roughly L/M of the transmit power is combined at the receiver using a single RF chain in a sparse mmW channel with L propagation clusters. With N_{H} UE RF chains, $N_{\mathrm{H}}L/M$ of the transmit power is combined at the receiver. Thus, assuming a free-space path loss model, the total received signal power at the UE in TTD beam training can be expressed as follows

$$p_{\mathrm{R}}[\mathrm{dB}] = p_{\mathrm{T}}[\mathrm{dB}] + G^{(\mathrm{T})}[\mathrm{dBi}] + G^{(\mathrm{R})}[\mathrm{dBi}] + 20\log_{10}\left(\frac{\lambda}{4\pi}\right) - 20\log_{10}(d) + 10\log_{10}\left(\frac{N_{\mathrm{H}}L}{M}\right), \tag{2.33}$$

where $G^{(\mathrm{T})}$ and $G^{(\mathrm{R})}$ are the BF gains at the BS and UE, respectively, λ is the wavelength, and d is the distance between the BS and UE. Unlike in conventional EBS, the received power in (2.33) has the penalty term $10\log_{10}\left(\frac{N_{\mathrm{H}}L}{M}\right)$, which depends on system parameters and channel

geometry. For this reason, the proposed TTD beam training requires a higher transmit power p_{T} at the BS than conventional EBS to support the same distance d. Alternatively, in the case when the transmit power budget is limited and cannot be increased, TTD beam training requires a denser network deployment, i.e., smaller distances between the BSs and UEs.

2.7.4 DSP Algorithm for Joint Beam Training

We use the designed BS and UE beam training codebooks to develop a DSP algorithm for joint AoD and AoA estimation. We propose a non-coherent power-based algorithm that does not rely on phase information in samples in (2.19). We also consider a coherent power-based algorithm and include it in the analysis of the misalignment probability as the benchmark.

Let b be the index of the BS-UE beam pair defined by the precoder $\mathbf{u}_{m_{\mathrm{T}}}$ and combiner $\mathbf{f}_{m_{\mathrm{R}}}$. We denote the set of $R = N_{\mathrm{H}}$ subcarriers that probe the b-th beam pair as $\mathcal{M}_b^{(\mathrm{B})}$. To define $\mathcal{M}_b^{(\mathrm{B})}$, we first find the subcarrier m_b from $\mathcal{M}$ in (2.21) that is mapped to b in the first UE RF chain, and then account for $R-1$ subcarriers in other RF chains by using the rotation shift multiplier in (2.28):

$$\mathcal{M}_b^{(\mathrm{B})} = \left\{ m | m = \mathrm{mod}\left(m_b + (r-1)\left\lfloor \frac{N_{\mathrm{R}}}{N_{\mathrm{H}}} \right\rfloor \left\lfloor \frac{M_{\mathrm{tot}}}{N_{\mathrm{R}}} \right\rfloor,\ M \right),\ r = 1, ..., R \right\}. \tag{2.34}$$

We vectorize the transmitted symbols $s[m],\ \forall m \in \mathcal{M}_b^{(\mathrm{B})}$ for the beam pair b and we denote the resulting vector $\mathbf{s}_b$. Similarly, we vectorize the corresponding received signal samples $y_r[m],\ \forall (r,m), m \in \mathcal{M}_b^{(\mathrm{B})}$, and we denote that vector $\mathbf{y}_b \in \mathbb{C}^{N_{\mathrm{H}}}$. A non-coherent power measurement $\hat{p}_b^{(\mathrm{nc})}$ for the beam pair b is then defined as follows

$$\hat{p}_b^{(\mathrm{nc})} = \frac{2}{\sigma_{\mathrm{N}}^2} \mathbf{y}_b^{\mathrm{H}} \mathbf{y}_b = \frac{2}{\sigma_{\mathrm{N}}^2} \sum_{(r,m), m \in \mathcal{M}_b^{(\mathrm{B})}} |y_r[m]|^2, \tag{2.35}$$

where $2/\sigma_{\mathrm{N}}^2$ is the scaling term. Note that $\hat{p}_b^{(\mathrm{nc})}$ includes powers of R frequency-domain samples (subcarriers). On the other hand, a benchmark coherent measurement $\hat{p}_b^{(\mathrm{c})}$, which

requires complex synchronization, is defined in the following way

$$\hat{p}_b^{(\mathrm{c})} = \frac{2}{\|\mathbf{s}_b\|_2^2 \, \sigma_{\mathrm{N}}^2} |\mathbf{y}_b^{\mathrm{H}} \mathbf{s}_b|^2 = \frac{2}{\|\mathbf{s}_b\|_2^2 \, \sigma_{\mathrm{N}}^2} \left| \sum_{(r,m), m \in \mathcal{M}_b^{(\mathrm{B})}} y_r^*[m] s[m] \right|^2, \tag{2.36}$$

where $2/(\|\mathbf{s}_b\|_2^2 \, \sigma_{\mathrm{N}}^2)$ is the scaling term. Coherent power measurements were previously studied in [LLH17].

The AoD and AoA estimates are based on the beam pair index $\hat{b}_{\max}^{(\mathrm{nc})}$ that corresponds to the maximum measured power in (2.35). The maximum power measurement $\hat{p}_{\max}^{(\mathrm{nc})}$ and index $\hat{b}_{\max}^{(\mathrm{nc})}$ are found as follows

$$\hat{p}_{\max}^{(\mathrm{nc})} = \max_{\hat{p}_b^{(\mathrm{nc})}} \hat{p}_b^{(\mathrm{nc})}, \qquad \hat{b}_{\max}^{(\mathrm{nc})} = \operatorname*{argmax}_b \hat{p}_b^{(\mathrm{nc})}. \tag{2.37}$$

Let $\xi_{m_{\mathrm{T}}^*}^{(\mathrm{T})}$ and $\xi_{m_{\mathrm{R}}^*}^{(\mathrm{R})}$ be the steering angles that correspond to $\hat{b}_{\max}^{(\mathrm{nc})}$ at the BS and UE side, respectively. Then the on-grid AoD and AoA estimates are

$$\hat{\theta}^{(\mathrm{T})} = \xi_{m_{\mathrm{T}}^*}^{(\mathrm{T})}, \qquad \hat{\theta}^{(\mathrm{R})} = \xi_{m_{\mathrm{R}}^*}^{(\mathrm{R})} \tag{2.38}$$

Note that the angle estimation accuracy of the proposed algorithm is limited by the number of antenna elements at the BS N_{T} and at the UE N_{R}. The use of larger antenna arrays can increase the accuracy, but it also imposes the beam training scalability challenges as discussed earlier. The benchmark AoD and AoA estimates can be obtained using the coherent power measurements $\hat{p}_b^{(\mathrm{c})}$, $\forall b$, in 2.37.

2.8 Beam Pair Misalignment Probability in Joint Beam Training

In lows SNRs, the performance of the proposed joint beam training can be affected by noise. In this section, we mathematically describe the beam pair misalignment probability $P_{\mathrm{miss}}^{(\mathrm{nc})}$ in the presence of Gaussian noise. We also demonstrate that the misalignment probability is lower in the proposed wideband beam training than in the fast EBS with a single-carrier waveform.

Without loss of generality, let the index $b = 1$ correspond to the optimal beam pair with the maximum received power, i.e., $p_{\max}^{(\mathrm{nc})} = p_1^{(\mathrm{nc})}$ and $b_{\max}^{(\mathrm{nc})} = 1$. The beam pair misalignment probability describes the probability that the maximum power measurement $\hat{p}_{\max}^{(\mathrm{nc})}$ and index $\hat{b}_{\max}^{(\mathrm{nc})}$ are not equal to $\hat{p}_1^{(\mathrm{nc})}$ and 1, respectively. Mathematically, this can be written as follows [LLH17]

$$P_{\mathrm{miss}}^{(\mathrm{nc})} = \mathbb{P}[\hat{b}_{\max}^{(\mathrm{nc})} \neq 1] = \mathbb{P}\left[\bigcup_{b=2}^{M} \hat{p}_1^{(\mathrm{nc})} < \hat{p}_b^{(\mathrm{nc})}\right] \tag{2.39}$$

Due to the independence of noise samples across different subcarriers and orthogonality of DFT beams in different RF chains, the power measurements $\hat{p}_b^{(\mathrm{nc})},\ \forall b$, are independent. However, the events $\hat{p}_1^{(\mathrm{nc})} < \hat{p}_b^{(\mathrm{nc})},\ \forall b$ are not independent, which makes it hard to determine the probability $P_{\mathrm{miss}}^{(\mathrm{nc})}$ exactly. For this reason, we calculate the upper bound on the misalignment probability. Using the union bound rule, the upper bound $P_{\mathrm{up}}^{(\mathrm{nc})}$ can be expressed in the following way [LLH17]

$$P_{\mathrm{miss}}^{(\mathrm{nc})} \leq \sum_{b=2}^{M} \mathbb{P}[\hat{p}_1^{(\mathrm{nc})} < \hat{p}_b^{(\mathrm{nc})}] = P_{\mathrm{up}}^{(\mathrm{nc})}. \tag{2.40}$$

To calculate $P_{\mathrm{up}}^{(\mathrm{nc})}$ in (2.40), we need to know the distributions of $\hat{p}_b^{(\mathrm{nc})},\ \forall b$. Since the received signal samples $y_r[m],\ \forall (r,m), m \in \mathcal{M}_b^{(\mathrm{B})}$, in (2.19) are complex Gaussian random variables with distribution $\mathcal{CN}(\mathbf{w}_r^{\mathrm{H}}[m]\mathbf{H}[m]\mathbf{v}[m]s[m], \sigma_{\mathrm{N}}^2)$, the power measurements $\hat{p}_b^{(\mathrm{nc})},\ \forall b$, have a non-central chi-squared distribution $\mathcal{X}^2(\lambda_b, 2R)$ with $2R$ degrees of freedom and non-centrality parameter λ_b given as

$$\lambda_b = \frac{2}{\sigma_{\mathrm{N}}^2}\frac{1}{M} \sum_{(r,m)\in\mathcal{M}_b^{(\mathrm{B})}} |\mathbf{w}_r^H[m]\mathbf{H}[m]\mathbf{v}[m]|^2. \tag{2.41}$$

The benchmark coherent power measurements $\hat{p}_b^{(\mathrm{c})},\ \forall b$, have a non-central chi-squared distribution $\mathcal{X}^2(\lambda_b, 2)$ with 2 degrees of freedom and non-centrality parameter $\lambda_b = \mu_b^*\mu_b$, where $\mu_b = \frac{\sqrt{2}}{\sqrt{RM\sigma_{\mathrm{N}}^2}} \sum_{(r,m)\in\mathcal{M}_b^{(\mathrm{B})}} \mathbf{w}_r^H[m]\mathbf{H}[m]\mathbf{v}[m]$.

The expression for $P_{\mathrm{up}}^{(\mathrm{nc})}$ in (2.40) can we rewritten as follows

$$P_{\mathrm{up}}^{(\mathrm{nc})} = \sum_{b=2}^{M} \mathbb{P}\left[\frac{\hat{p}_1^{(\mathrm{nc})}}{\hat{p}_b^{(\mathrm{nc})}} < 1\right]. \tag{2.42}$$

The ratio $\hat{p}_1^{(\mathrm{nc})}/\hat{p}_b^{(\mathrm{nc})}$ of two non-central chi-squared random variables gives a random variable with a doubly non-central F distribution, denoted as $F(n_1, n_2, \eta_1, \eta_2)$. Based on the chi-squared distributions, the degrees of freedom in the F distribution are $n_1 = 2R$ and $n_2 = 2R$, while the non-centrality parameters are given as $\eta_1 = \lambda_1$ and $\eta_2 = \lambda_b,\ b > 1$. Thus, the probability $\mathbb{P}\left[\hat{p}_1^{(\mathrm{nc})}/\hat{p}_b^{(\mathrm{nc})} < 1\right]$ represents the cumulative distribution function of the doubly non-central F random variable, which can be expressed as

$$\mathbb{P}\left[\frac{\hat{p}_1^{(\mathrm{nc})}}{\hat{p}_b^{(\mathrm{nc})}} < 1\right] = F(1|n_1, n_2, \eta_1, \eta_2) = F(1|2R, 2R, \lambda_1, \lambda_b). \tag{2.43}$$

The upper bound on the beam pair misalignment probability in (2.40) can be then written as

$$P_{\mathrm{up}}^{(\mathrm{nc})} = \sum_{b=2}^{M} F(1|2R, 2R, \lambda_1, \lambda_b). \tag{2.44}$$

In the following subsection, we compare the proposed wideband joint beam training with fast EBS that uses a single-carrier waveform. The two approaches are compared in terms of the SNR per measurement, required training overhead, and misalignment probability in a simple LoS channel.

2.8.1 Comparison with Fast Single-Carrier based EBS

The proposed joint beam training can be realized using only one OFDM symbol. However, with a larger number of subcarriers, the duration of the OFDM symbol can be significantly longer than that of a symbol with a single-carrier based waveform. Therefore, there is a question if the proposed joint beam training has an advantage over the single-carrier based EBS in terms of the required overhead and beam training performance. Here we compare the two approaches and demonstrate the benefits of the proposed joint beam training.

We first compare the proposed beam training and EBS in terms of the SNR per sample. In the proposed OFDM-based beam training, the measurements include powers of R samples from different subcarriers due to the frequency-dependent codebooks at both the BS and UE. As assumed in the system model, the signal power and noise power per subcarrier are

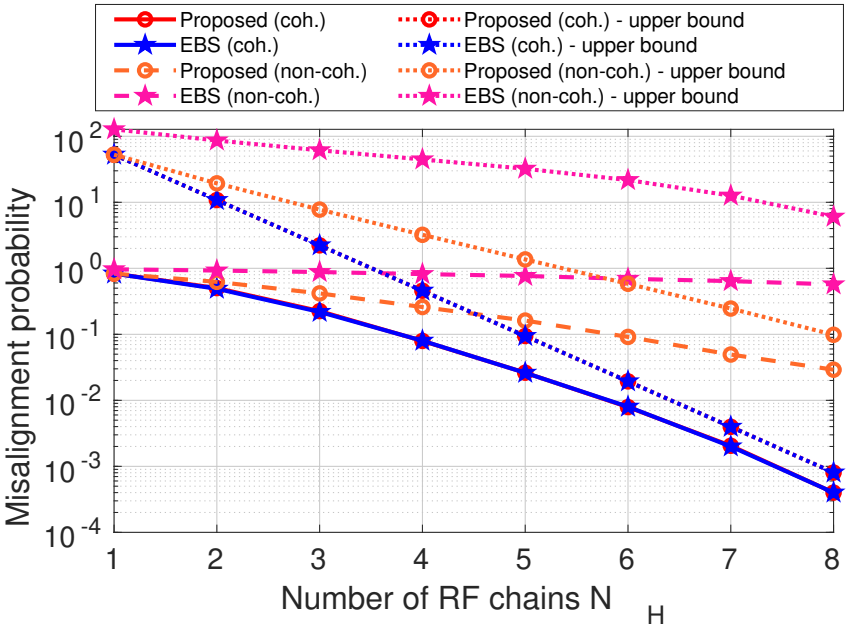


Figure 2.14: Misalignment probability in a simple LoS channel.

$|s[m]| = 1/M$ and σ_{N}^2, respectively. Thus, the SNR per sample in the proposed beam training is

$$\mathrm{SNR} = \frac{1}{M\sigma_{\mathrm{N}}^2}. \tag{2.45}$$

In the single-carrier based EBS, power measurements are made across multiple time-domain samples (symbols) and the entire bandwidth is used for each sample. Thus, the SNR per sample can be expressed using the definition in (2.45) and assuming that $M = M_{\mathrm{tot}}$ as follows

$$\mathrm{SNR} = \frac{1}{M_{\mathrm{tot}}\sigma_{\mathrm{N}}^2}\frac{\mathrm{BW}_{\mathrm{TTD}}}{\mathrm{BW}_{\mathrm{EBS}}}, \tag{2.46}$$

where the ratio $\mathrm{BW}_{\mathrm{TTD}}/\mathrm{BW}_{\mathrm{EBS}}$ accounts for a potential difference in the bandwidths used in the proposed beam training and EBS. Based on the SNRs in (2.45) and (2.46), we see that if $\mathrm{BW}_{\mathrm{TTD}} = \mathrm{BW}_{\mathrm{EBS}}$, the proposed OFDM-based beam training has M_{tot}/M times larger SNR per sample that the single-carrier based EBS. Clearly, the SNRs are the same when the EBS uses M_{tot}/M narrower bandwidth than the proposed beam training.

Next, we compare the two approaches in terms of the beam pair misalignment probability in a simple LoS channel. For simplicity, all beams are assumed to have a uniform beamforming gain, similar as in sector beams. We consider $N_{\mathrm{T}} = 32$ antennas at the BS and $N_{\mathrm{R}} = 16$ antennas at the UE. The bandwidth in both approaches is $\mathrm{BW}_{\mathrm{TTD}} = \mathrm{BW}_{\mathrm{TTD}} = 1$ GHz. There are $M_{\mathrm{tot}} = 4096$ subcarriers in the OFDM system, $M = 512$ of which are used in the proposed beam training. The pre-combining SNR in the proposed beam training is set to be $\mathrm{SNR} = -22$ dB. It is assumed that the total duration of all transmitted symbols is the same in both approaches and it is equal to $M_{\mathrm{tot}}/\mathrm{BW}$ (one OFDM symbol without CP). With such total duration and $R = N_{\mathrm{H}}$, the number of samples per beam pair power measurement is R and RM_{tot}/M in the proposed beam training and single-carrier based EBS, respectively. The misalignment probability is presented in Fig. 2.14 as a function of the number of RF chains N_{H}. We present both the simulated curves and the calculated upper bounds. The results indicate that the coherent power measurements lead to a lower misalignment probability than non-coherent power measurements in both the proposed beam training and EBS. Additionally, the proposed beam training is shown to have the same misalignment probability as the EBS when coherent power measurements are used in simple LoS channels. However, when non-coherent power measurements are used, the proposed joint beam training outperforms the EBS. The main reason for this is a higher SNR per sample in the proposed beam training with an OFDM waveform and frequency-dependent beams. Similar as in energy detection algorithms in spectrum sensing, the required number of samples in non-coherent beam training highly depends on the SNR. Thus, the EBS needs more samples to achieve the same misalignment probability as the proposed beam training, numerically studied in Fig. 2.15. The simulated curves in Fig. 2.15 indicate that the EBS needs an $N_{\mathrm{S}} = 3$ times or $N_{\mathrm{S}} = 4$ times higher number of samples for a comparable performance.

One of the main advantages of the proposed beam training over the fast single-carrier based EBS is a lower total required overhead. The overhead in the proposed beam training is equivalent to the duration of a single OFDM symbol. Assuming a 7% CP, the overhead

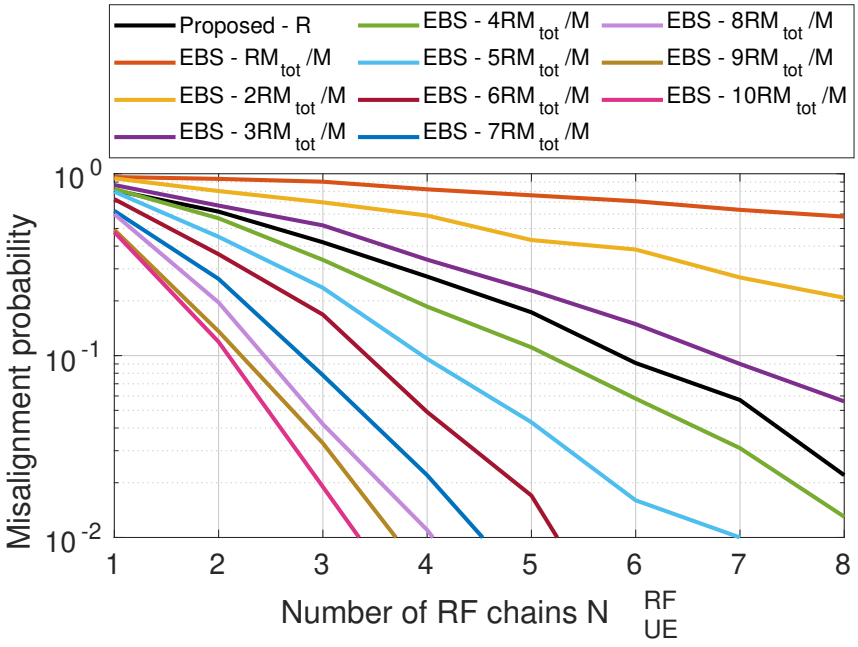


Figure 2.15: Misalignment probability in the EBS when more samples (symbols) are used.

in the proposed training can be expressed as follows

$$T_{\mathrm{TTD}} = 1.07 M_{\mathrm{tot}} \frac{1}{\mathrm{BW}_{\mathrm{TTD}}}. \tag{2.47}$$

On the other hand, the overhead of the EBS depends on the total duration of transmitted single-carrier symbols. During the one OFDM symbol in the proposed beam training, a total of $M_{\mathrm{tot}}\mathrm{BW}_{\mathrm{EBS}}/\mathrm{BW}_{\mathrm{TTD}}$ single-carrier symbols are transmitted in the EBS. However, as discussed earlier, the EBS needs to increase the number of symbols (samples) to achieve the same performance as the proposed beam training. In addition, the EBS requires the BS and UE to set up and switch the beams for each probed beam pair. Since the BS is equipped with a digital array, its beam can be set up and switched in DSP. On the other hand, the hybrid array at the UE needs to reconfigure its phase shifters when a different receive beam (or a set of receive beams) is probed. Clearly, the reconfiguration can slow down the beam sweeping process, especially when the UE has a large antenna array and

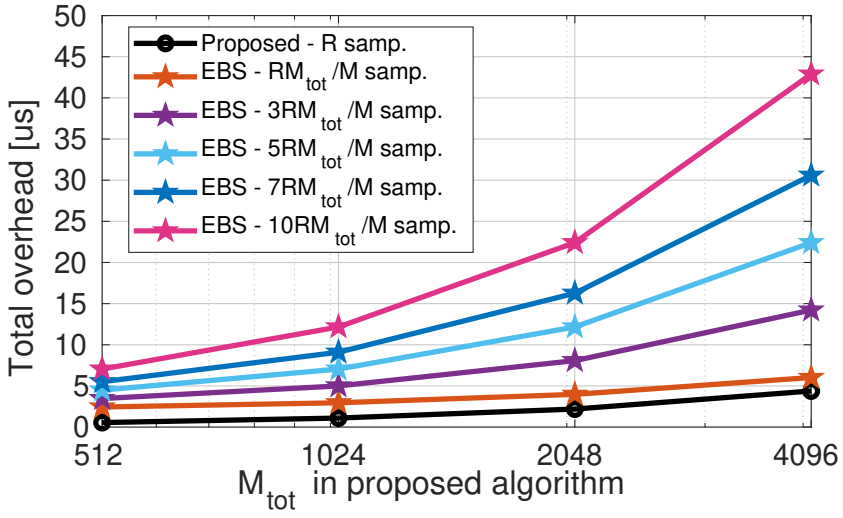


Figure 2.16: Total required overhead for beam training.

many beam candidates. Thus, the total training overhead in the EBS can be expressed as

$$\begin{aligned} T_{\text{EBS}} &= N_{\text{S}} M_{\text{tot}} \frac{\text{BW}_{\text{EBS}}}{\text{BW}_{\text{TTD}}} \frac{1}{\text{BW}_{\text{EBS}}} + N_{\text{R}}(T_{\text{setup}} + T_{\text{switch}}) \\ &= N_{\text{S}} M_{\text{tot}} \frac{1}{\text{BW}_{\text{TTD}}} + N_{\text{R}}(T_{\text{setup}} + T_{\text{switch}}) \end{aligned} \tag{2.48}$$

where T_{setup} and T_{switch} are the set up and switching times. Based on the state-of-the-art work on fast-switching antenna arrays, these values are $T_{\text{setup}} = 120$ ns and $T_{\text{switch}} = 8$ ns [SPL22]. Using the expressions in (2.47) and (2.48), we compare the total training overhead in the proposed beam training and EBS in Fig. 2.16. The results show that the proposed beam training has a significantly lower overhead. This comes as a consequence of the fact that the EBS needs to set up and switch the UE beams and to use N_{S} times more symbols to achieve the same performance as the proposed approach.

2.9 Evaluation of Joint Beam Training in Realistic Channels

In this section, we evaluate the performance of the proposed joint beam training and EBS in realistic LoS and non-line-of-sight (NLoS) mmW channels generated in Quadriga channel

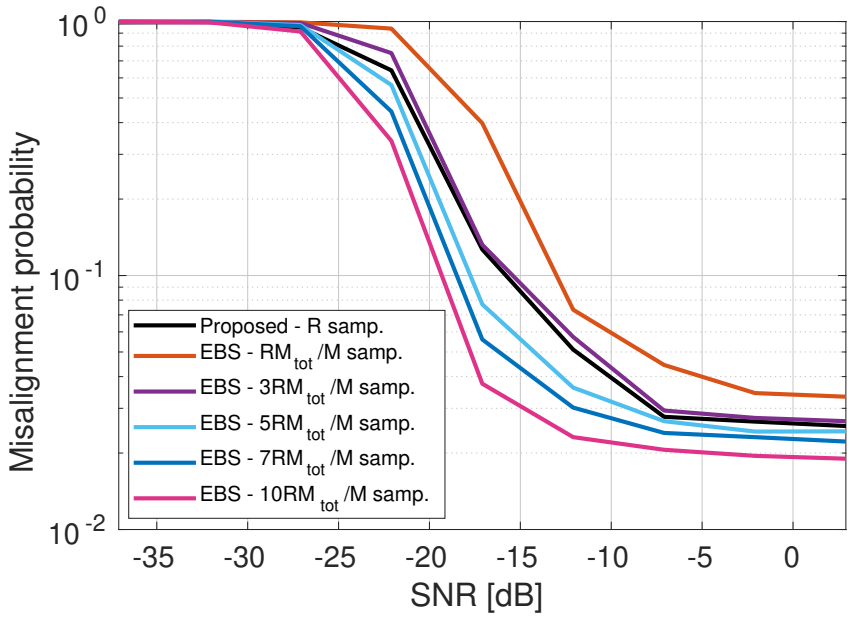
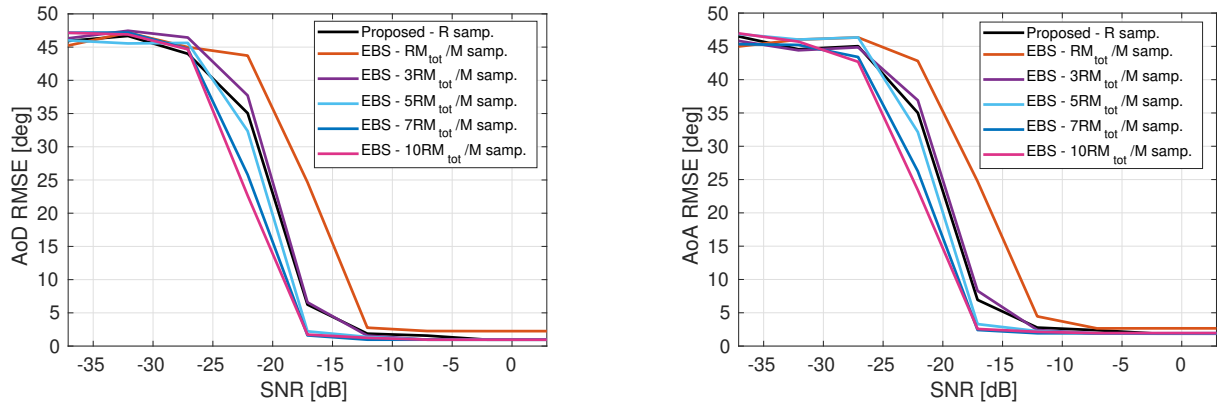


Figure 2.17: Misalignment probability in realistic LoS channels.

simulator [JRB19,JRB14]. We use the beam pair misalignment probability and angle estimation accuracy as the main metrics and we evaluate them for different pre-beamforming SNR values. The simulation parameters are the same as in the previous section. In this section, we do not use the simplifying assumption that all beams have a uniform beamforming gain.

In Fig. 2.17, we compare the misalignment probability in LoS channels in the proposed beam training and EBS. Similar as in the simplified LoS channel in Sec. 2.8.1, the EBS requires a larger number of samples to achieve the same performance as the proposed beam training in low SNRs. Without the assumption of uniform beamforming gains, the misalignment probability experiences a floor in high SNR in both approaches. Nevertheless, the probability floor is still fairly low, in the order of 10^{-2}. The comparison in terms of the RMSE of angle estimation in LoS channels is presented in Fig. 2.18. The results indicate that both the proposed beam training and EBS have accurate angle estimation in medium to high pre-beamforming SNRs. It is worth noting that the RMSEs of AoD and AoA estimation experience floors due to the limited numbers of antenna elements and probed beams



(a) (b)

Figure 2.18: RMSE of (a) AoD and (b) AoA estimation in realistic LoS channels.

at the BS and UE.

The comparison in terms of the misalignment probability in NLoS channels is presented in Fig. 2.19. Unlike in LoS channels, there are multiple comparably strong propagation clusters in NLoS channels and it is easier to miss the optimal beam pair, even in high SNRs. Thus, the floor beam pair misalignment probability is higher in NLoS than in LoS channels for both the proposed beam training and EBS. In addition, NLoS channels have significantly larger delay spreads than LoS channels. The proposed beam training, which uses a long OFDM waveform, is resistant to inter-symbol interference and it can capture the entire energy of NLoS channels with large delay spreads. On the other hand, short symbols in the single-carrier based EBS are susceptible to inter-symbol interference and they cannot capture the time-spread channel energy. This leads to a lower received signal power and a higher misalignment probability in the EBS. The performance gap between the two approaches can be reduced by increasing the number of samples, and thus the total overhead, in the EBS. The RMSE of angle estimation in NLoS channels is presented in Fig. 2.20. Although the misalignment probability is relatively high in NLoS channels, the RMSE of AoD and AoA estimation indicates that wrong beam pair estimates correspond to the beams that are close to the optimal beam pair. However, the RMSE floors are higher than in LoS channels due

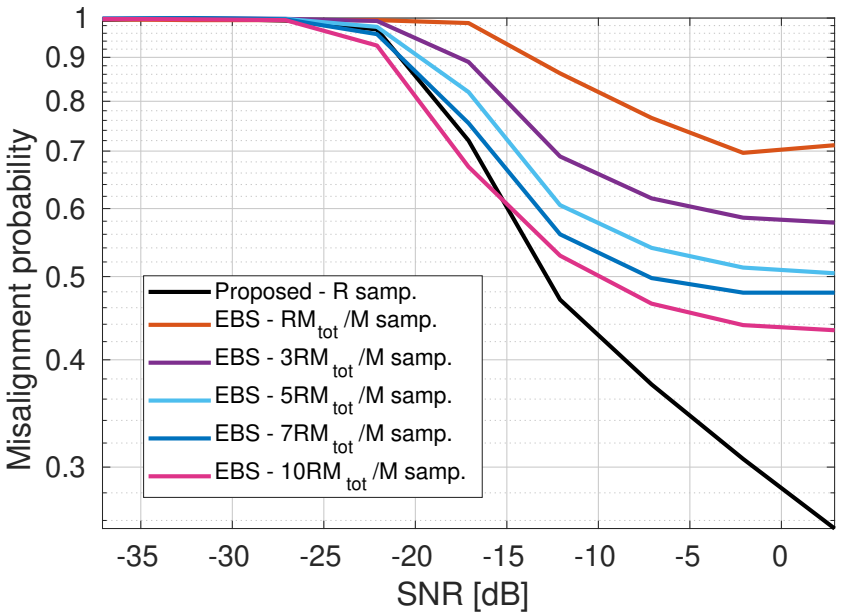
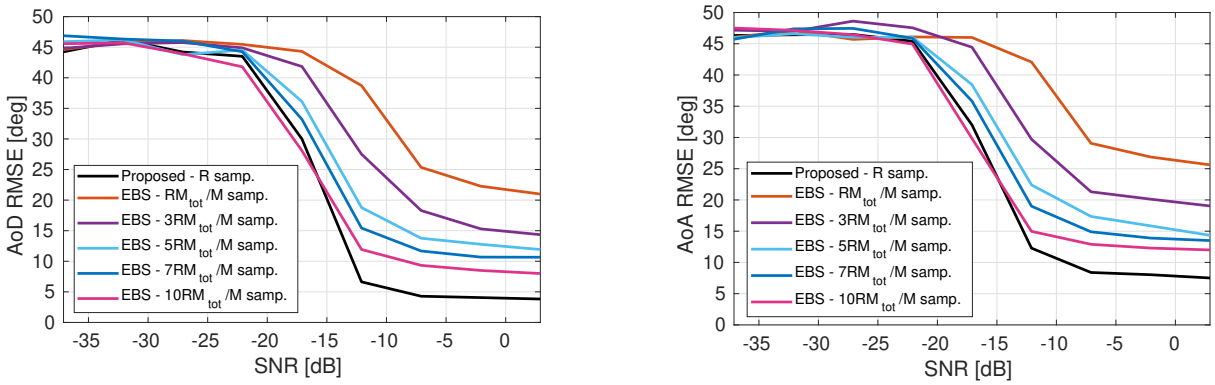


Figure 2.19: Misalignment probability in realistic NLoS channels.

to the presence of multiple propagation clusters and large delay spreads.

2.10 Conclusions

This chapter introduced three TTD architectures with baseband delay elements as potential candidates for mmW beam training. We demonstrated that a high AoA estimation accuracy can be achieved with both analog and hybrid TTD architectures using a power measurement based beam training scheme, which requires only one wideband training symbol. The dependency of the codebook design and beam training performance on system parameters, including the bandwidth, number of antenna elements, and maximum TTD delay compensation, was analyzed and numerically evaluated in a practical multipath fading channels. Numerical simulations revealed the angle estimation accuracy and robustness to hardware impairments of the proposed TTD architectures in UE beam training when benchmarked against the fully digital array. We extended the idea of beam training with frequency-dependent beams to



(a) (b)

Figure 2.20: RMSE of (a) AoD and (b) AoA estimation in realistic NLoS channels.

joint beam training between the BS and UE. The beam pair misalignment probability and required overhead in the proposed joint beam training were analyzed and compared to their counterparts in the EBS. The results indicate that the misalignment probability is at least as low as in the EBS, while the required overhead is lower than in the EBS.

CHAPTER 3

Channel Estimation using True-Time-Delay Arrays

3.1 Introduction

Due to large available bandwidth, mmW frequency bands are the key enabler of high data rates in the fifth generation of wireless communication systems [ABC14]. However, radio signals experience a higher propagation loss at mmW than at sub-6GHz frequencies [RXM17]. To compensate for this loss, the BS and UE need to be equipped with large antenna arrays and establish a directional communication link. The link can be established either through beam training [HGR16], which only provides information of the dominant propagation direction, or channel estimation procedure, which enables the acquisition of the entire channel between the BS and UE. Having the full channel knowledge provides a series of benefits, including the ability to optimize the BS and UE beamforming vectors, boost the data rate through spatial multiplexing, perform optimal power allocation, determine the backup links to be used in the case of a link failure, and others. For this reason, many researchers have been actively working on efficient mmW channel estimation techniques over the last several years.

3.1.1 Prior Work on mmW Channel Estimation

Initial channel estimation algorithms were designed for narrowband frequency-flat mmW channels [AEL14, BAN14, RVM12, MRM16, LGL16, SR17, MSH18]. In order to avoid exhaustive beam sweeping, the authors in [AEL14] proposed an iterative hierarchical approach to

estimate the channel parameters. More commonly, previous work aimed to exploit the sparsity of mmW channels and formulate channel estimation as a CS problem, which is often solved using matching pursuit based algorithms [LGL16, MRM16, SR17, MSH18]. However, mmW bandwidths are wide in practice and that inspired the development of channel estimation algorithms for frequency-selective channels [VAG17, RGV18, GHD16, WXZ19, KC21, WJG19, LGW19, VAT19]. Similar as in the narrowband case, the wideband algorithms often rely on the mmW channel sparsity and CS problem formulation in the time or frequency domain [VAG17, RGV18, GHD16, WXZ19, KC21, WJG19]. In multi-carrier systems, the problem is commonly defined in the frequency domain and the measurements are processed per-subcarrier. Previous work proposed various greedy reconstruction algorithms to estimate the sparse vector of channel parameters, including the path gains and angles.

Some of the previous work has pointed out that the conventional on-grid CS-based algorithms have the basis mismatch problem, which can affect the channel estimation accuracy. To address this issue, researchers have formulated channel estimation as an atomic norm minimization problem [TZW18, ZWT19, CZW19, CZW18, DTS18]. The channel parameters in atomic norm minimization (ANM) problems have continuous representations, which enables super-resolution estimation accuracy. Additionally, ANM-based channel estimation problems are often expressed as semidefinite programs and they can be efficiently solved using commercial solvers.

The majority of the existing algorithms was developed for arrays based on PSs. In particular, analog and hybrid analog-digital arrays were commonly considered, as they are believed to be promising architecture candidates for mmW systems due to their power efficiencies. However, unlike fully digital arrays, analog and hybrid arrays cannot probe all angular directions simultaneously and thus may require a significant overhead for channel estimation. Previous work also studied the impact of hardware impairments on the channel estimation performance when analog [MH17, VGH19] and hybrid [PGW19, WGW19, ZNY17] arrays are used. Impairments may be caused by imperfect phase shifters, local oscillators, imbalance

between in-phase and quadrature samples, antenna misplacement, or other hardware problems.

Recent work explored the use of different array architectures capable of steering frequency-dependent beams for fast channel probing [GYS20, YBC19, BYL21]. In [GYS20], the authors investigated fast beam training with a leaky wave antenna, which can probe all angular directions simultaneously by using different frequency components of the signal. In our recent work in [YBC19] and [BYL21], we proposed and analyzed TTD array architectures for fast UE beam training. Compared to phased arrays, TTD arrays have delay elements along with phase shifters in all antenna branches, which allows them to synthesize frequency-dependent beams [BYL21, RTY16, CSC18, JLJ18, Jan19, GSR19, GPB20, GG21, LPG21]. In multi-carrier systems, the subcarriers can probe different angular directions simultaneously even with an analog TTD array. Frequency-dependent beams, however, have not been extensively studied in the context of mmW channel estimation. In fact, a number of existing wideband algorithms rely on frequency-flat precoders and combiners to reduce the complexity burden of per-subcarrier processing. Additionally, it is unknown how practical hardware impairments might affect the channel estimation performance when TTD array architectures and frequency-dependent beams are used.

3.1.2 Contributions

We propose a DSP algorithm for mmW channel estimation that leverages TTD frequency-dependent beams, and then we analyze how it is affected by hardware impairments in the array. The contributions of this chapter can be summarized as follows:

- We design a TTD training codebook and a frequency-domain CS-based DSP algorithm to estimate sparse mmW channels. Unlike the majority of existing approaches, the algorithm is based on per-sub-band processing to reduce the complexity.
- Assuming antenna arrays without hardware errors, the proposed DSP algorithm is

compared with related state-of-the-art approaches designed for antenna arrays based on phase shifters in terms of the required number of training frames, channel estimation accuracy, and computational complexity.

- The performance of the proposed and state-of-the-art algorithms is evaluated in the presence of hardware impairments in the UE array. Since the proposed algorithm is designed for a TTD array, its performance is evaluated in the presence of the delay and phase errors. The performance of the state-of-the-art algorithms is evaluated in the presence of phase errors.

- We linearize the received signal model with hardware impairments and derive the lower bound for the variance of the channel parameter estimators. In particular, we derive the CRLB for the parameters of LoS mmW channels, including the AoD, AoA, and phase of the complex channel gain.

- A gradient descent based refinement is proposed to improve the estimation of channel parameters at the cost of a higher computational complexity. The performance of the algorithm with the proposed refinement is numerically evaluated and compared with a benchmark approach and derived CRLBs.

3.1.3 Organization and Notation

The rest of the chapter is organized as follows. In Sec. 3.2, we present the system model with considered hardware impairments and describe the CS-based problem formulation. In Sec. 3.3, we explain how the proposed TTD channel estimation codebook and DSP algorithm are designed. The comparison with existing approaches in an impairment-free setup is presented in Sec. 3.4. In Sec. 3.5, the impact of hardware impairments on channel estimation algorithms is evaluated, CRLBs are derived, and parameter refinement approach is described. Finally, Sec. 3.6 highlights the conclusions from this chapter.

Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-

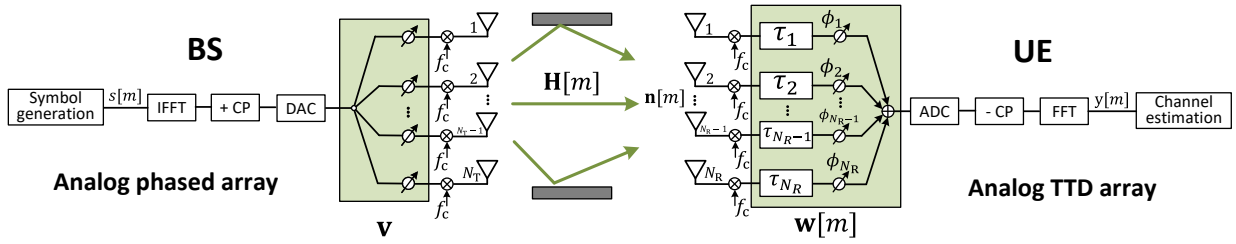


Figure 3.1: Considered system model, where the BS is equipped with an analog phased array and the UE with an analog TTD array.

case letters, respectively. The (i, j)-th element and j-th column of $\mathbf{A}$ are denoted by $[\mathbf{A}]_{i,j}$ and $[\mathbf{A}]_{:,j}$, respectively. Conjugate, transpose, Hermitian transpose are denoted by $(.)^*$, $(.)^{\mathrm{T}}$, and $(.)^{\mathrm{H}}$, respectively. The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is defined as $\mathbf{A} \otimes \mathbf{B}$.

3.2 System Model

In this chapter, we study downlink mmW channel estimation between a BS and a UE, which are synchronized in time. They use the bandwidth BW and communicate at the carrier frequency f_{c}. The BS is equipped with an analog phased array with N_{T} antennas, while the UE uses an analog TTD array that has N_{R} antenna elements, as illustrated in Fig. 3.1. Analog TTD architecture is suitable for UEs because it offers fast and power-efficient channel probing using frequency-dependent beams [BYL21]. The considered system uses a CP based OFDM waveform with M_{tot} subcarriers. Each subcarrier is assumed to be loaded with a BPSK symbol. When TTD arrays are used in an OFDM system, the CP needs to be longer than the cumulative delay introduced by the channel multipath components and TTD circuits, as we previously proved in [YBC19]. Here we assume that this requirement is satisfied and that the received signal can be modeled in the frequency-domain, after the analog-to-digital conversion and CP removal.

Let $s[m]$ be a BPSK symbol transmitted at the m-th subcarrier. Since the phased

array at the BS has frequency-flat beams, symbols $s[m], \forall m$, are precoded with the same vector $\mathbf{v} \in \mathbb{C}^{N_{\mathrm{T}}}$. On the other hand, the UE combines the signal from its antennas using a frequency-dependent combiner $\mathbf{w}[m] \in \mathbb{C}^{N_{\mathrm{R}}}$. Therefore, the complex baseband received signal at the m-th subcarrier can be expressed as follows

$$y[m] = \mathbf{w}^{\mathrm{H}}[m]\mathbf{H}[m]\mathbf{v}s[m] + \mathbf{w}^{\mathrm{H}}[m]\mathbf{n}[m], \tag{3.1}$$

where $\mathbf{H}[m] \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ and $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_{\mathrm{N}}^2 \mathbf{I}_{N_{\mathrm{R}}}\right)$ are the channel matrix and vector of complex white additive Gaussian noise at the m-th subcarrier, respectively.

Since the UE uses an analog TTD array, the n-th element of the frequency-dependent combiner $\mathbf{w}^{(t)}[m]$ is defined as follows [BYL21]

$$[\mathbf{w}[m]]_n = \exp\left(-j\left(2\pi(f_m - f_{\mathrm{c}})\bar{\tau}_n + \bar{\phi}_n\right)\right), \tag{3.2}$$

where $f_m = f_{\mathrm{c}} - \mathrm{BW}/2 + (m-1)\mathrm{BW}/(M_{\mathrm{tot}} - 1)$, and $\bar{\tau}_n$ and $\bar{\phi}_n$ are distorted delay and phase taps in the n-th antenna branch respectively. The distortion comes from various hardware impairments that can occur in the array, including errors in delay elements, phase shifters, local oscillators, or other. We assume that hardware impairments are time-invariant and independent across different antennas. The distorted delay tap is modeled as $\bar{\tau}_n = \tau_n + \tilde{\tau}_n$, where τ_n is the desired delay tap and $\tilde{\tau}_n$ is the delay error modeled as a Gaussian random variable $\tilde{\tau}_n \sim \mathcal{N}(0, \sigma_{\mathrm{T}}^2)$. Assuming uniform delay spacing $\Delta\tau$ between neighboring antennas, the desired tap is defined as $\tau_n = (n-1)\Delta\tau$. Similarly, the distorted phase tap is modeled as $\bar{\phi}_n = \phi_n + \tilde{\phi}_n$, with ϕ_n being the desired delay tap and $\tilde{\phi}_n$ being the phase error modeled as a Gaussian random variable $\tilde{\phi}_n \sim \mathcal{N}(0, \sigma_{\mathrm{P}}^2)$. The desired phase tap is defined as $\phi_n = (n-1)\Delta\phi$, where $\Delta\phi$ is the phase spacing. It is worth noting that time-invariant hardware errors can be estimated and corrected or reduced through a calibration process. In this chapter, however, our main goal is to get a better understanding of their impact on the channel estimation performance. We leave the problem of error estimation and correction in TTD arrays for future work.

3.2.1 Frequency-Selective Channel Model

Due to a high propagation loss, mmW channels have only a few propagation paths. The channel sparsity has also been proved experimentally through multiple measurement campaigns [XR21, RMS15, RGB13]. We model the channel as frequency-selective, consisting of L propagation paths and N_{tap} delay taps.

In the frequency domain, the channel at the m-th subcarrier can be expressed as follows

$$\mathbf{H}[m] = \sum_{l=1}^{L} G_l[m]\mathbf{a}_{\text{R}}(\theta_l^{(\text{R})})\mathbf{a}_{\text{T}}^{\text{H}}(\theta_l^{(\text{T})}), \tag{3.3}$$

where $G_l[m] = \sum_{d=1}^{N_{\text{tap}}} g_l p(dT_{\text{s}} - \Gamma_l)e^{-j\frac{2\pi}{M}md}$ is the complex gain in the frequency domain. The function $p(\Gamma)$ models pulse shaping and other analog filtering, T_{s} is the sampling time, $g_l \in \mathbb{C}$ and $\Gamma_l \in \mathbf{R}$ are the complex gain and delay of the l-th path, respectively. Assuming $f_m/f_{\text{c}} \approx 1,\ \forall m$, the n-th element of the frequency-flat array response $\mathbf{a}_{\text{R}}(\theta)$ is defined as $[\mathbf{a}_{\text{R}}(\theta)]_n = \exp(-j(n-1)\pi\sin(\theta))/\sqrt{N_{\text{R}}}$. The response $\mathbf{a}_{\text{T}}(\theta)$ is defined in a similar way. The angles $\theta_l^{(\text{R})} \in [-\pi/2, \pi/2)$ and $\theta_l^{(\text{T})} \in [-\pi/2, \pi/2)$ are the AoA and AoD of the l-th path. By exploiting the "column-row" property of matrix multiplication [Str09], the channel in (3.3) can be written in a compact form as follows

$$\mathbf{H}[m] = \mathbf{A}_{\text{R}}\mathbf{\Lambda}[m]\mathbf{A}_{\text{T}}^{\text{H}}, \tag{3.4}$$

where $\mathbf{\Lambda}[m] \in \mathbb{C}^{L\times L}$ is a diagonal matrix whose non-zero values correspond to the gains $G_l[m],\ \forall l$, and $\mathbf{A}_{\text{T}} \in \mathbb{C}^{N_{\text{T}}\times L}$ and $\mathbf{A}_{\text{R}} \in \mathbb{C}^{N_{\text{R}}\times L}$ are matrices of the spatial responses $\mathbf{a}_{\text{T}}(\theta_l^{(\text{T})}),\ \forall l$, at the BS and $\mathbf{a}_{\text{R}}(\theta_l^{(\text{R})}),\ \forall l$, at the UE, respectively. Let $\bar{\mathbf{A}}_{\text{T}} \in \mathbb{C}^{N_{\text{T}}\times Q_{\text{T}}}$ and $\bar{\mathbf{A}}_{\text{R}} \in \mathbb{C}^{N_{\text{R}}\times Q_{\text{R}}}$ be the dictionaries of spatial responses at the BS and UE side, respectively. The numbers of spatial responses Q_{T} and Q_{R} depend on the corresponding number of antenna elements and it usually assumed that $Q_{\text{T}} \gg N_{\text{T}}$ and $Q_{\text{R}} \gg N_{\text{R}}$. Unlike in an oversampled DFT dictionary, where the spatial responses have equidistant phases, the dictionaries $\bar{\mathbf{A}}_{\text{T}} = [\mathbf{a}_{\text{T}}(\xi_1^{(\text{T})}), ..., \mathbf{a}_{\text{T}}(\xi_{Q_{\text{T}}}^{(\text{T})})]$ and $\bar{\mathbf{A}}_{\text{R}} = [\mathbf{a}_{\text{R}}(\xi_1^{(\text{R})}), ..., \mathbf{a}_{\text{R}}(\xi_{Q_{\text{R}}}^{(\text{R})})]$ include the spatial responses with uniformly spaced angles $\xi_q^{(\text{T})} = -\pi/2 + (q-1)\pi/Q_{\text{T}},\ q = 1, ..., Q_{\text{T}},$

and $\xi_q^{(\mathrm{R})} = -\pi/2 + (q-1)\pi/Q_{\mathrm{R}},\ q = 1, ..., Q_{\mathrm{R}}$, respectively. Using these dictionaries, the expression in (3.4) can be approximated in the following way

$$\mathbf{H}[m] \approx \bar{\mathbf{A}}_{\mathrm{R}} \bar{\mathbf{\Lambda}}[m] \bar{\mathbf{A}}_{\mathrm{T}}^{\mathrm{H}}, \tag{3.5}$$

where the matrix $\bar{\mathbf{\Lambda}}[m] \in \mathbb{C}^{Q_{\mathrm{R}} \times Q_{\mathrm{T}}}$ has L non-zero elements that correspond to the channel gains $G_l[m],\ \forall l$. It is common to assume that $Q_{\mathrm{T}} \gg L$ and $Q_{\mathrm{R}} \gg L$, which makes the approximation error in (3.5) negligible.

3.2.2 Problem Formulation

Existing CS-based algorithms for wideband mmW channel estimation often consider a symmetric precoding/combining setup where both the BS and UE use frequency-flat pseudo-random beams [VAG17, RGV18]. With frequency-flat beams, compressive measurements at all subcarriers can be modeled with a single measurement matrix, which reduces the computational complexity and relaxes the memory storage requirements. Here we propose an *asymmetric* precoder/combiner design, where the BS uses frequency-flat pseudo-random beams, while the UE relies on frequency-dependent TTD beams.

Based on the described system model, the BS transmits a BPSK symbol $s^{(t)}[m]$ at the m-th subcarrier, in the t-th training frame. The transmitted symbol is precoded with a pseudo-random beamforming vector $\mathbf{v}^{(t)}$, whose n-th element is defined as $[\mathbf{v}^{(t)}]_n = e^{j\alpha_n}$. The phase α_n is assumed to be drawn randomly from the uniform distribution $\mathcal{U}(0, 2\pi)$ in each antenna branch. At the receive side, the UE uses a frequency-dependent vector $\mathbf{w}^{(t)}[m]$ to combine the signal at the m-th subcarrier. Thus, the received signal at the m-th subcarrier in the t-th training frame can be expressed as

$$y^{(t)}[m] = \mathbf{w}^{(t)\mathrm{H}}[m] \mathbf{H}[m] \mathbf{v}^{(t)} s^{(t)}[m] + \tilde{n}^{(t)}[m], \tag{3.6}$$

where $\tilde{n}^{(t)}[m] = \mathbf{w}^{(t)\mathrm{H}}[m] \mathbf{n}^{(t)}[m]$. The UE can remove the contribution of the BPSK symbol in (3.6) by multiplying the received samples with $(s^{(t)}[m])^{-1}$. With approximation in (3.5),

the received signal in (3.6) can be vectorized as follows

$$y^{(t)}[m] = \mathbf{f}^{(t)}[m]\mathbf{A}\boldsymbol{\lambda}[m] + \tilde{n}^{(t)}[m], \tag{3.7}$$

where $\mathbf{f}^{(t)}[m] \in \mathbb{C}^{1\times N_{\mathrm{T}}N_{\mathrm{R}}}$ is a sensing vector defined as $\mathbf{f}^{(t)}[m] = \mathbf{v}^{(t)\mathrm{T}} \otimes \mathbf{w}^{(t)\mathrm{H}}[m]$, $\mathbf{A} \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}}\times Q_{\mathrm{T}}Q_{\mathrm{R}}}$ is the dictionary defined as $\mathbf{A} = \bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}$, and $\boldsymbol{\lambda}[m] \in \mathbb{C}^{Q_{\mathrm{T}}Q_{\mathrm{R}}}$ is a sparse vector of channel gains obtained by stacking the columns of $\bar{\boldsymbol{\Lambda}}[m]$, i.e., $\boldsymbol{\lambda}[m] = \mathrm{vec}(\bar{\boldsymbol{\Lambda}}[m])$. It is important to note that the vector $\boldsymbol{\lambda}[m]$ has the same support, i.e., the same indices of non-zero elements, for any m, since the channel AoD-AoA pairs $(\theta_l^{(\mathrm{T})}, \theta_l^{(\mathrm{R})})$, $\forall l$, are common for all subcarriers. Assuming that the coherence time is longer than the total time needed for channel probing, the received signal after T training frames can be vectorized in the following way

$$\mathbf{y}[m] = \mathbf{F}[m]\mathbf{A}\boldsymbol{\lambda}[m] + \tilde{\mathbf{n}}[m], \tag{3.8}$$

where $\mathbf{y}[m] \in \mathbb{C}^{T}$, $\mathbf{y}[m] = [y^{(1)}[m], ..., y^{(T)}[m]]^{\mathrm{T}}$, $\mathbf{F}[m] \in \mathbb{C}^{T\times N_{\mathrm{T}}N_{\mathrm{R}}}$, $\mathbf{F}[m] = [\mathbf{f}^{(1)\mathrm{T}}[m], ...,$ $\mathbf{f}^{(T)\mathrm{T}}[m]]^{\mathrm{T}}$, and $\tilde{\mathbf{n}}[m] \in \mathbb{C}^{T}$, $\tilde{\mathbf{n}}[m] = [\tilde{n}^{(1)}[m], ..., \tilde{n}^{(T)}[m]]^{\mathrm{T}}$. Based on the model in (3.8), a sparse recovery problem for the m-th subcarrier can be defined as

$$\min \|\boldsymbol{\lambda}[m]\|_1 \quad \text{s.t.} \quad \|\mathbf{y}[m] - \mathbf{F}[m]\mathbf{A}\boldsymbol{\lambda}[m]\|_2^2 < \epsilon \tag{3.9}$$

where ϵ is the maximum error power.

Our main goal is to design a low-complexity sparse recovery algorithm that solves (3.9). Specifically, we aim to design frequency-dependent TTD combiners $\mathbf{w}^{(t)}[m], \forall t, m$, and a DSP algorithm that enable accurate channel estimation with a low computational burden and reduced training overhead compared to the state-of-the-art. Further, we want to provide an insight into how practical hardware impairments in the UE's TTD array affect the channel estimation performance.

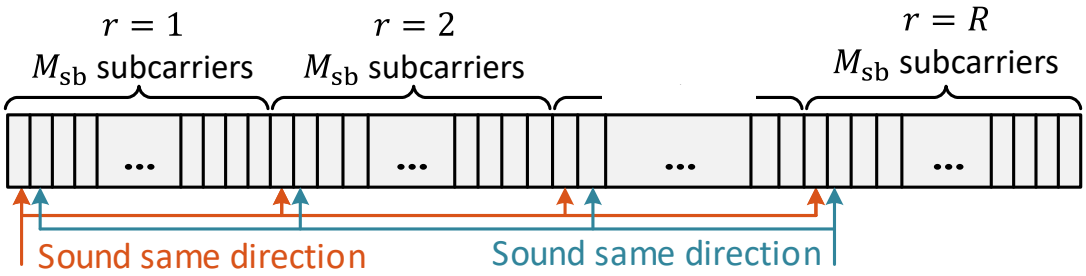


Figure 3.2: An illustration of sub-bands. In each sub-band, M_{sb} subcarriers sound the entire angular range $[-\pi/2, \pi/2)$.

3.3 Proposed Channel Estimation Algorithm

In this section, we describe the design of the proposed frequency-dependent TTD combiners $\mathbf{w}^{(t)}[m]$, $\forall t, m$. Additionally, we develop a DSP algorithm that has a low computational complexity despite the asymmetric design of precoders and combiners.

3.3.1 Design of UE TTD Codebook

Recently, the use of TTD codebooks for fast mmW beam training was explored in [YBC19, BYL21]. The key idea was to leverage frequency-dependent TTD combiners and probe all angular directions in the range $[-\pi/2, \pi/2)$ at once. For example, in an OFDM-based system, each angular direction can be probed by DFT beams associated with R uniformly spaced subcarriers [BYL21]. Here we propose a similar TTD codebook design for channel estimation. It is worth noting that the beam training codebook design in [YBC19] and [BYL21] considered only a subset of subcarriers. On the other hand, the channel estimation codebook in channel estimation must consider *all* M_{tot} OFDM subcarriers.

In an analog TTD array, the entire codebook of frequency-dependent DFT beams can be created with a proper configuration of the delay and phase spacings. The delay spacing controls how much the DFT beams are spread across the angular range, while the phase spacing determines how much they are rotated. Here we propose a design where the delay

spacing $\Delta\tau$ is time-invariant, while the phase spacing $\Delta\phi^{(t)}$ changes in each training frame to allow different frequency-to-angle mapping. As in our previous work, the delay taps τ_n, $\forall n$, are designed such that R subcarriers are mapped to each probed direction in any training frame. The bandwidth is first divided into R sub-bands of $M_{\text{sb}} = M_{\text{tot}}/R$ subcarriers, as illustrated in Fig. 3.2. Within each sub-band, M_{sb} subcarriers should be associated with M_{sb} distinct DFT beams that probe the entire angular range. In [BYL21], we showed that such DFT beams can be created by using $\Delta\tau = R/\text{BW}$, i.e., designing the delay taps as follows

$$\tau_n = (n-1)R/\text{BW}, \quad n = 1, ..., N_{\text{R}}. \tag{3.10}$$

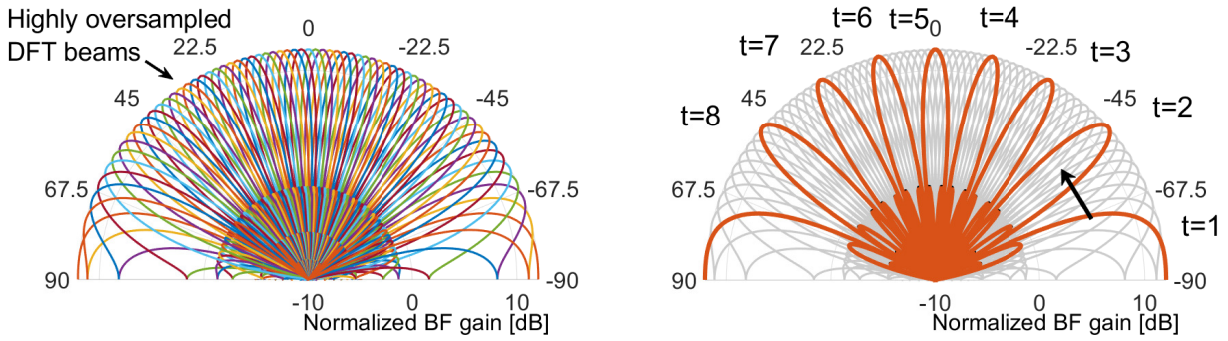
The taps in (3.10) ensure that all first subcarriers in R sub-bands probe the same direction, all second subcarriers probe the same, etc. An example of the resulting UE codebook with $\Delta\phi^{(t)} = 0$ in the t-th training frame is provided in Fig. 3.3(a). Note that if $M_{\text{sb}} > N_{\text{R}}$, the codebook consists of spatial DFT beams oversampled by the factor of $O = M_{\text{sb}}/N_{\text{R}}$.

The UE codebook can be rotated in different training frames to provide additional diversity and enable each subcarrier to sound multiple directions. This is achieved by changing the phase spacing $\Delta\phi^{(t)}$ between the frequency-flat phase shifters. For a uniform rotation of $2\pi/T$ in each training frame, the spacing is designed as $\Delta\phi^{(t)} = (t-1)2\pi/T$. Consequently, the phase taps are set as follows

$$\phi_n^{(t)} = (n-1)(t-1)2\pi/T, \ n = 1, ..., N_{\text{R}}, \ t = 1, ..., T. \tag{3.11}$$

Note that the phase taps in (3.11) are frequency-flat, i.e., they are applied to all subcarriers equally. An example of codebook rotations is illustrated in Fig. 3.3(b) for the first subcarriers in all R sub-bands.

In the next subsection, we describe a DSP algorithm that leverages the designed frequency-dependent TTD codebook to solve the sparse recovery problem in (3.9).



(a) (b)

Figure 3.3: An example of the designed UE codebook with $N_{\mathrm{R}} = 16$, BW = 2GHz, $M_{\mathrm{tot}} = 1024$, $R = 16$, and $T = 8$. (a) Complete codebook with $\Delta\phi^{(t)} = 0$ and $O = 4$. (b) Rotations for first subcarriers in all sub-bands.

3.3.2 OMP-based DSP Algorithm

In previous work, a number of greedy algorithms were proposed to solve (3.9). In particular, researchers explored different variations of the Orthogonal Matching Pursuit (OMP) [VAG17, RGV18], an algorithm that iteratively estimates the channel angles and gains. The majority of OMP-related algorithms is based on per-subcarrier processing, which can impose a significant computational burden. Here we propose a two-step OMP-based algorithm with a lower computational complexity. In the first step, the algorithm iteratively identifies the AoDs and AoAs using per-sub-band processing. With the identified channel support, the algorithm estimates the channel gains in a single iteration in the second step.

Assuming that M_{sb} subcarriers within each sub-band r have the same channel gains, the problem in (3.9) can be defined for all R sub-bands. Let $\boldsymbol{\lambda}_r \in \mathbb{C}^{Q_{\mathrm{T}}Q_{\mathrm{R}}}$ be an L-sparse vector of channel gains in the r-th sub-band. The vectors $\boldsymbol{\lambda}_r$, $\forall r$, have the same support as $\boldsymbol{\lambda}[m]$, $\forall m$. Let $\mathbf{F}^{(t)} \in \mathbb{C}^{M_{\mathrm{sb}} \times N_{\mathrm{T}} N_{\mathrm{R}}}$ be the sensing matrix for each of the R sub-bands in the t-th training frame, defined as $\mathbf{F}^{(t)} = [\mathbf{f}^{(t)\mathrm{T}}[1], ..., \mathbf{f}^{(t)\mathrm{T}}[M_{\mathrm{sb}}]]^{\mathrm{T}}$. Note that all sub-bands can have a common sensing matrix $\mathbf{F}^{(t)}$ because of the codebook design in Section 3.3.1, which ensured that the subcarriers $m = 1, ..., M_{\mathrm{sb}}$, from the first sub-band ($r = 1$) sound

the same directions as their counterparts from other sub-bands ($r > 1$). Let $\tilde{\mathbf{n}}_r^{(t)} \in \mathbb{C}^{M_{\text{sb}}}$ be the noise vector at the subcarriers in the r-th sub-band. Then, after T training frames and vectorization, the signal measurement vector $\mathbf{y}_r \in \mathbb{C}^{TM_{\text{sb}}}$ in the r-th sub-band can be expressed as

$$\mathbf{y}_r = \mathbf{F}\mathbf{A}\boldsymbol{\lambda}_r + \tilde{\mathbf{n}}_r, \tag{3.12}$$

where $\mathbf{F} \in \mathbb{C}^{TM_{\text{sb}} \times N_{\text{T}} N_{\text{R}}}$ is $\mathbf{F} = [\mathbf{F}^{(1)\text{T}}, ..., \mathbf{F}^{(T)\text{T}}]^{\text{T}}$, $\tilde{\mathbf{n}}_r \in \mathbb{C}^{TM_{\text{sb}}}$ is $\tilde{\mathbf{n}}_r = [\tilde{\mathbf{n}}_r^{(1)\text{T}}, ..., \tilde{\mathbf{n}}_r^{(T)\text{T}}]^{\text{T}}$. For conciseness, we introduce the effective measurement matrix $\boldsymbol{\Phi} \in \mathbb{C}^{TM_{\text{sb}} \times Q_{\text{T}} Q_{\text{R}}}$, defined as $\boldsymbol{\Phi} = \mathbf{F}\mathbf{A}$. Similar as in (3.9), a sparse recovery problem for sub-band r can be formulated as follows

$$\min \|\boldsymbol{\lambda}_r\|_1 \quad \text{s.t.} \quad \|\mathbf{y}_r - \boldsymbol{\Phi}\boldsymbol{\lambda}_r\|_2^2 < \epsilon. \tag{3.13}$$

The problem in (3.13) needs to be solved for all R sub-bands, given the corresponding measurement vectors $\mathbf{y}_r$, $\forall r$. Previous work on CS-based mmW channel estimation suggested that the required number of measurements to solve (3.13) with high probability scales as $L \log(Q_{\text{T}} Q_{\text{R}}/L)$ when $\boldsymbol{\Phi}$ is designed using pseudo-random beamforming vectors at both the BS and UE [AEL14, BAN14, RVM12, MRM16]. Here, this scaling represents a loose lower bound because the BS uses pseudo-random precoders, while the UE uses oversampled DFT combiners. Finding a tighter lower bound requires further analysis, but this is out of the scope of this chapter.

We propose an iterative OMP-based algorithm to solve (3.13). In each iteration, the algorithm identifies the index of one non-zero value in $\boldsymbol{\lambda}_r$, $\forall r$, i.e., AoD-AoA pair of one propagation path, and then it refines all approximate channel gains jointly. In the first iteration, an AoD-AoA pair index is estimated by jointly considering all R sub-bands and calculating the correlation between $\boldsymbol{\Phi}$ and residuals $\mathbf{z}_r = \mathbf{y}_r$, $\forall r$, as follows

$$c^* = \underset{c}{\text{argmax}} \sum_{r=1}^{R} \left| \left[\boldsymbol{\Phi}^{\text{H}} \mathbf{z}_r \right]_c \right|^2. \tag{3.14}$$

Let the set $\Omega = \{c^*\}$ be the current estimate of the channel support. The approximated

channel gains $\boldsymbol{\lambda}_r,\ \forall r$, are then estimated using least squares in the following way

$$\hat{\boldsymbol{\lambda}}_r = \left([\boldsymbol{\Phi}]_{:,\Omega}^{\mathrm{H}} [\boldsymbol{\Phi}]_{:,\Omega} \right)^{-1} [\boldsymbol{\Phi}]_{:,\Omega}^{\mathrm{H}} \mathbf{y}_r. \tag{3.15}$$

The channel support Ω and gains $\hat{\boldsymbol{\lambda}}_r,\ \forall r$, are used to subtract the contribution of the estimated propagation path from the measurements $\mathbf{y}_r,\ \forall r$, and obtain the corresponding measurement residuals for the next iteration. Mathematically, the residuals $\mathbf{z}_r,\ \forall r$, are calculated as

$$\mathbf{z}_r = \mathbf{y}_r - [\boldsymbol{\Phi}]_{:,\Omega} \hat{\boldsymbol{\lambda}}_r. \tag{3.16}$$

The algorithm iterates until a stopping criterion is satisfied. Using a predetermined number of iterations is impractical because the number of propagation paths in the channel is usually unknown. Thus, we propose the use of a stopping criterion based on the average subcarrier power E in the residuals $\mathbf{z}_r,\ \forall r$, calculated as follows [VAG17, RGV18]

$$E = \frac{1}{TM_{\mathrm{tot}}} \sum_{r=1}^{R} \|\mathbf{z}_r\|_2^2 . \tag{3.17}$$

When power in (3.17) falls below the threshold ϵ, the algorithm stops. It was shown in [RGV18] that the optimal threshold for algorithms based on per-subcarrier processing is $\epsilon = \sigma_{\mathrm{N}}^2$, assuming that the noise variance σ_{N}^2 is known at the receiver. Unlike in [RGV18], our proposed algorithm yields approximate channel gains $\boldsymbol{\lambda}_r,\ \forall r$, and consequently different residuals $\mathbf{z}_r,\ \forall r$, than algorithms with per-subcarrier processing, which creates the need for a different ϵ. At the subcarrier level, the current estimate of $\mathbf{y}[m]$ in (3.8) can be expressed as

$$\hat{\mathbf{y}}[m] = [\boldsymbol{\Phi}]_{\Psi,\Omega} \hat{\boldsymbol{\lambda}}_r + \underbrace{[\boldsymbol{\Phi}]_{\Psi,\Omega} \boldsymbol{\lambda}_e[m]}_{\text{Approx. error}}, \tag{3.18}$$

where $\Psi = \{m' \mid m' = \mathrm{mod}(m, M_{\mathrm{sb}}) + (t-1)M_{\mathrm{sb}},\ t = 1, ..., T\}$, $\hat{\boldsymbol{\lambda}}_r$ is the estimate of channel gains in the corresponding sub-band, and $\boldsymbol{\lambda}_e[m]$ is the channel gain error vector at the m-th subcarrier. For mathematical tractability, we treat $\boldsymbol{\lambda}_e[m]$ as a zero-mean random Gaussian vector that is independent and identically distributed across different subcarriers.

Algorithm 1 Proposed channel estimation algorithm

Inputs: $\mathbf{y}[m], \forall m, \quad \mathbf{y}_r, \forall r, \quad \mathbf{\Phi}, \quad \gamma, \quad \sigma_{\mathrm{N}}^2$

Initialize: $\mathbf{z}_r = \mathbf{y}_r, \forall r, \quad \Omega = \{\emptyset\}$

Calculate: $E = \frac{1}{TM_{\mathrm{tot}}} \sum_{r=1}^{R} \|\mathbf{z}_r\|_2^2$

Calculate: $E_{\mathrm{s}} = \frac{1}{TM_{\mathrm{tot}}} \sum_{r=1}^{R} \|\mathbf{y}_r\|_2^2 - \sigma_{\mathrm{N}}^2$

Calculate: $\epsilon = \gamma E_{\mathrm{s}} + \sigma_{\mathrm{N}}^2$

while $E > \epsilon$ **do**

  Estimate: $c^* = \mathrm{argmax}_c \sum_{r=1}^{R} \left| \left[\mathbf{\Phi}^{\mathrm{H}} \mathbf{z}_r \right]_c \right|^2$

  Update: $\Omega = \Omega \cup c^*$

  Estimate: $\hat{\boldsymbol{\lambda}}_r = \left([\mathbf{\Phi}]_{:,\Omega}^{\mathrm{H}} [\mathbf{\Phi}]_{:,\Omega} \right)^{-1} [\mathbf{\Phi}]_{:,\Omega}^{\mathrm{H}} \mathbf{y}_r, \quad \forall r$

  Update: $\mathbf{z}_r = \mathbf{y}_r - [\mathbf{\Phi}]_{:,\Omega} \hat{\boldsymbol{\lambda}}_r, \quad \forall r$

  Update: $E = \frac{1}{TM_{\mathrm{tot}}} \sum_{r=1}^{R} \|\mathbf{z}_r\|_2^2$

end while

Est.: $\hat{\boldsymbol{\lambda}}[m] = \left([\mathbf{\Phi}]_{\Psi,\Omega}^{\mathrm{H}} [\mathbf{\Phi}]_{\Psi,\Omega} \right)^{-1} [\mathbf{\Phi}]_{\Psi,\Omega}^{\mathrm{H}} \mathbf{y}[m], \quad \forall m$

where $\Psi = \{m' \mid m' = \mathrm{mod}(m, M_{\mathrm{sb}}) + (t-1)M_{\mathrm{sb}}, \ t = 1, ..., T\}$

Outputs: $\hat{\boldsymbol{\lambda}}[m], \forall m, \quad \Omega$

Let σ_{A}^2 be the variance of each element in the approximation error $[\mathbf{\Phi}]_{\Psi,\Omega} \boldsymbol{\lambda}_e[m]$. Then it is straightforward to show that the stopping threshold should be set as $\epsilon = \sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{N}}^2$. However, the distribution of channels gains is unknown in general, which makes the estimation of σ_{A}^2 challenging. Thus, we estimate σ_{A}^2 numerically in Section 3.4 as a fraction γ of the initial *useful* signal power E_{s} in a multipath channel, i.e., $\sigma_{\mathrm{A}}^2 = \gamma E_{\mathrm{s}}$, where E_{s} is defined as

$$E_{\mathrm{s}} = \frac{1}{TM_{\mathrm{tot}}} \sum_{r=1}^{R} \|\mathbf{y}_r\|_2^2 - \sigma_{\mathrm{N}}^2. \tag{3.19}$$

After stopping, the iterative algorithm outputs the channel support estimate Ω and dismisses the approximate channel gain estimates $\boldsymbol{\lambda}_r, \ \forall r$. Using Ω and (3.8), the channel gains are estimated for each subcarrier in a single iteration as follows

$$\hat{\boldsymbol{\lambda}}[m] = \left([\mathbf{\Phi}]_{\Psi,\Omega}^{\mathrm{H}} [\mathbf{\Phi}]_{\Psi,\Omega} \right)^{-1} [\mathbf{\Phi}]_{\Psi,\Omega}^{\mathrm{H}} \mathbf{y}[m], \tag{3.20}$$

where the set Ψ is defined as earlier. The proposed channel estimation algorithm is summarized in Algorithm 1.

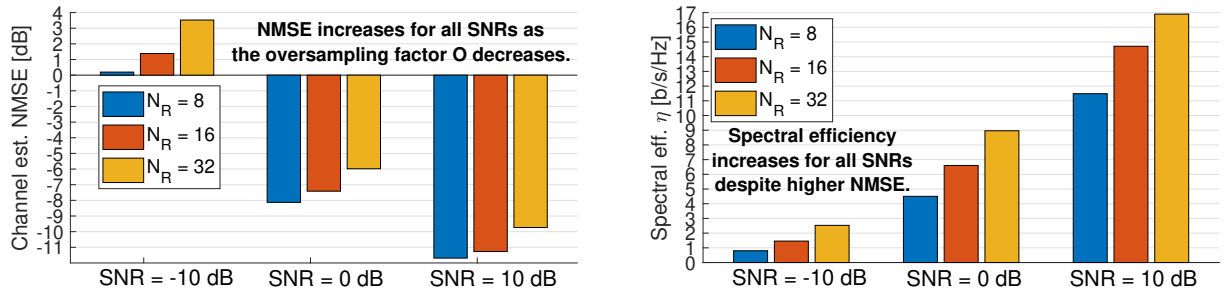
3.4 Comparison with State-of-the-Art

In this section, we assume a hardware impairment free setup, i.e., $\sigma_{\mathrm{T}} = 0$s and $\sigma_{\mathrm{P}} = 0°$s, to evaluate the performance of the proposed OMP-based algorithm. We also compare the proposed algorithm with related existing algorithms designed for phase shifter based arrays, including the classical OMP algorithm in [VAG17] and Simultaneous Weighted OMP (SW-OMP) algorithm proposed in [RGV18]. Unlike the classical OMP, which solves the channel estimation problem for each subcarrier individually, the SW-OMP uses measurement vectors at all subcarriers in each iteration to improve the angle and residual power estimation accuracy. The three algorithms are compared in terms of required number of training frames, channel estimation accuracy, and computational complexity.

We assume the carrier $f_{\mathrm{c}} = 28$ GHz, bandwidth BW $= 1$ GHz, $M_{\mathrm{tot}} = 1024$ subcarriers, and $N_{\mathrm{T}} = 64$ antennas at the BS. Both dictionaries $\bar{\mathbf{A}}_{\mathrm{R}}$ and $\bar{\mathbf{A}}_{\mathrm{T}}$ have $Q_{\mathrm{T}} = Q_{\mathrm{R}} = 256$ spatial responses. The UE assumes $R = 16$ sub-bands and designs its codebook accordingly. The channel is assumed to have $N_{\mathrm{tap}} = 4$ taps and $L = 3$ paths with identically distributed gains $g_l \sim \mathcal{CN}(0, \sigma_g^2),\ \forall l$. The SNR is defined as SNR $\triangleq L\sigma_g^2/\sigma_{\mathrm{N}}^2$.

First, we assume $N_{\mathrm{R}} = 16$ and $T = 35$ and evaluate the performance of the proposed algorithm for different values of the power fraction parameter $\gamma \in [0.05, 0.15]$. The numerical results indicate that the channel estimation accuracy increases with higher γ and thus it achieves the best performance with $\gamma = 0.15$. It is worth noting that the choice of γ is significantly impacted by the path gains $g_l,\ \forall l$. In certain non-line-of-sight scenarios, where multiple paths have comparable path gains, a further increase of γ may severely decrease the performance by causing the algorithm to miss some significant paths. Thus, for the remainder of this chapter, we assume the power fraction of $\gamma = 0.15$ to avoid potential

performance degradation.



(a) (b)

Figure 3.4: Performance of the algorithm for different sizes of the UE TTD array in terms of: (a) NMSE and (b) post-estimation spectral efficiency η.

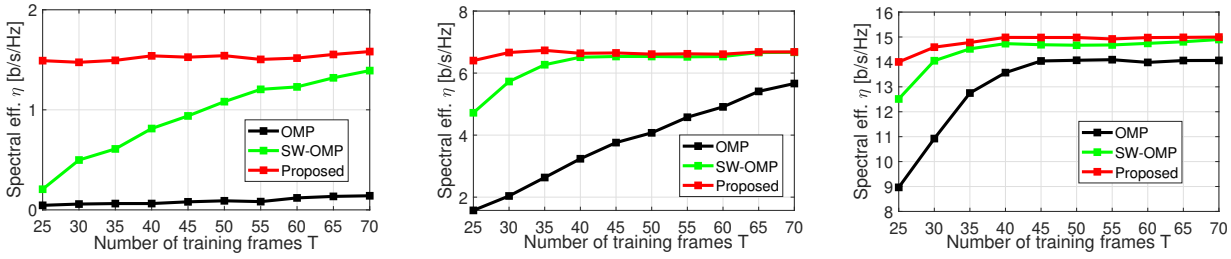
Next, we assume $T = 35$ and evaluate how the number of UE antennas N_{R} affect the performance of the proposed TTD-based channel estimation algorithm. Two metrics are used in this study. The first one is the normalized mean square error (NMSE), which is defined as

$$\mathrm{NMSE} = \frac{\sum_{m=1}^{M_{\mathrm{tot}}} ||\hat{\mathbf{H}}[m] - \mathbf{H}[m]||_F^2}{\sum_{m=1}^{M_{\mathrm{tot}}} ||\mathbf{H}[m]||_F^2}, \tag{3.21}$$

where $\hat{\mathbf{H}}[m]$ is a channel estimate at the m-th subcarrier. The second metrics is the average spectral efficiency and it is defined as

$$\eta = \frac{1}{M_{\mathrm{tot}}} \sum_{m=1}^{M_{\mathrm{tot}}} \sum_{k=1}^{K_m} \log_2 \left(1 + \frac{|\hat{\mathbf{w}}_k^{\mathrm{H}}[m] \mathbf{H}[m] \hat{\mathbf{v}}_k[m]|^2}{\sigma_{\mathrm{N}}^2} \right), \tag{3.22}$$

where K_m, $\hat{\mathbf{w}}_k[m]$, and $\hat{\mathbf{v}}_k[m]$ are the rank, left singular vector, and right singular vector of the estimated matrix $\hat{\mathbf{H}}[m]$, respectively. As shown in Fig. 3.4(a), the NMSE increases with larger antenna arrays in all SNR regimes. The number of measurements taken by $\mathbf{\Phi}$ does not depend on N_{R}, but only on T and M_{sb}, which are fixed. However, a higher N_{R} reduces the beam width and oversampling factor $O = M_{\mathrm{sb}}/N_{\mathrm{R}}$ of TTD combiners, which affects the channel estimation accuracy. Despite a higher NMSE, larger antenna arrays result in a higher post-estimation spectral efficiency due to more spatial degrees of freedom, as supported by the results in Fig. 3.4(b).

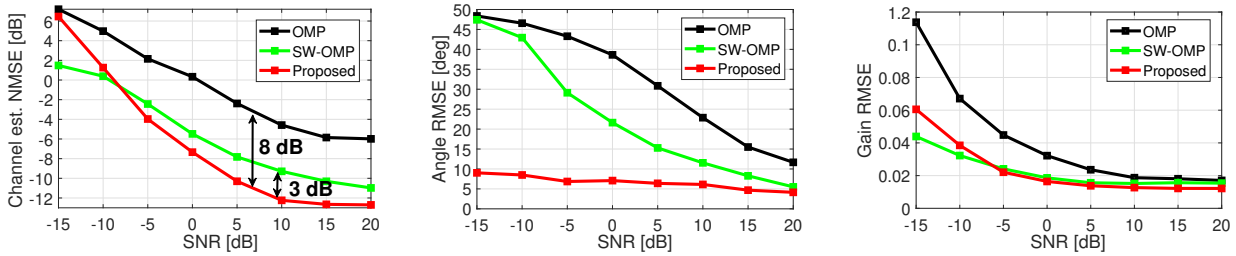


(a) (b) (c)

Figure 3.5: Comparison with state-of-the-art algorithms in terms of required number of training frames T for the same spectral efficiency η in different SNR scenarios: (a) SNR $= -10$ dB, (b) SNR $= 0$ dB, and (c) SNR $= 10$ dB.

A comparison between the proposed and state-of-the-art algorithms in terms the required overhead is presented in Fig. 3.5, assuming that the UE has $N_{\mathrm{R}} = 16$ antennas. The results indicate that the proposed algorithm requires a lower number of training frames T than the state-of-the-art for the same post-estimation spectral efficiency. The main reason for this is the fact that the measurement vector in the proposed algorithm has M_{sb} elements in each time frame due to per-sub-band processing. The difference in spectral efficiency is especially noticeable with low overhead, where the proposed algorithm requires 5, 10, and 50 training frames less than the SW-OMP in high, medium, and low SNR regimes, respectively. When compared with the OMP algorithm, the the proposed approach demonstrates an even bigger advantage in terms of the required overhead.

The three algorithms are compared in terms of the NMSE in Fig. 3.6(a). The performance is evaluated for $N_{\mathrm{R}} = 16$, $T = 35$, and different SNR values. In medium/high SNRs, the proposed algorithm outperforms the OMP and SW-OMP algorithms by 8 dB and 3 dB, respectively. On the other hand, in low SNR values, the channel estimation accuracy of the proposed algorithm deteriorates. Interestingly, even with a higher NMSE in low SNRs, the proposed algorithm has a higher spectral efficiency than the state-of-the-art, as previously shown for SNR $= -10$ dB in Fig. 3.5(a). In order to explain this result and get a better understanding of the spectral efficiency and NMSE curves, we further evaluate the



(a) (b) (c)

Figure 3.6: Comparison with state-of-the-art algorithms in terms of the (a) NMSE of channel estimation, (b) RMSE of angle estimation, and (c) RMSE of gain estimation, across different SNR values.

angle estimation accuracy and gain estimation accuracy separately. The RMSE of angle estimation is defined jointly for AoDs and AoAs as $\sqrt{\mathbb{E}\left[\frac{1}{J}\sum_{l=1}^{J}(\theta_l^{(\mathrm{R})}-\hat{\theta}_l^{(\mathrm{R})})^2+(\theta_l^{(\mathrm{T})}-\hat{\theta}_l^{(\mathrm{T})})^2\right]}$ across J estimated paths. If an algorithm identifies more than 3 paths, the RMSE calculation is done over the best 3 estimates. The RMSE of gain estimation is defined as $\sqrt{\mathbb{E}\left[\frac{1}{LM_{\mathrm{tot}}}\sum_{l=1}^{L}\sum_{m}\left|G_l[m]-\hat{G}_l[m]\right|^2\right]}$ across L paths. If an algorithm estimates less than 3 paths, the missing gain estimates are set to zero. The results of the angle and gain RMSEs are presented in Fig. 3.6(b) and Fig. 3.6(c), respectively. The proposed TTD-based algorithm has a significantly better angle estimation accuracy than the benchmark algorithms. Due to per-sub-band based processing, the columns of the measurement matrix $\mathbf{\Phi}$ in the proposed algorithm belong to a much larger vector space than in the OMP and SW-OMP algorithms, which consequently reduces the correlation among them and enables accurate angle estimation in (3.14). Since the singular vectors of a sparse mmW channel look like narrow pencil beams, the optimal precoders and combiners can be closely estimated by finding the best steering directions (channel AoDs and AoAs). Thus, due to more accurate angle estimates, the proposed algorithm leads to a higher spectral efficiency than the OMP and SW-OMP, as previously shown in Fig. 3.5. In terms of the RMSE of gain estimation, the proposed algorithm has a good performance in medium/high SNR, but the error increases in low SNRs. This increase explains why the overall NMSE of the proposed algorithm is

higher than that of the SW-OMP in a low SNR regime in Fig. 3.6(a).

Table 3.1: Complexity of proposed and state-of-the-art OMP-based algorithms

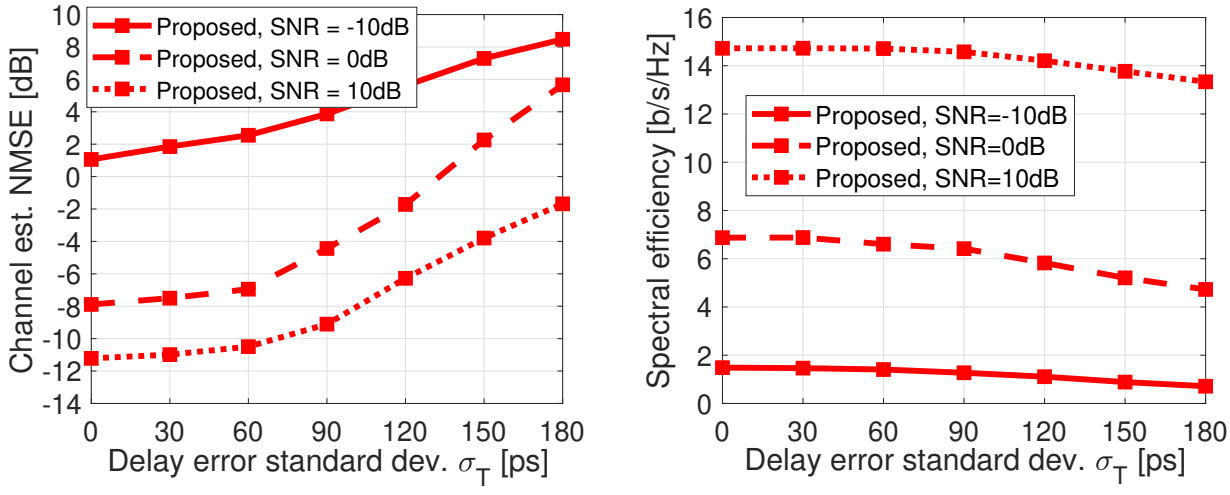
	Operation	Complexity
Proposed	Projection (j-th iter.)	$T_{\text{prop}} M_{\text{tot}} (Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Maximum proj. (j-th iter.)	$R(Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Gain calculation (j-th iter.)	$j^3 + 2j^2 T_{\text{prop}} M_{\text{sb}} + j T_{\text{prop}} M_{\text{tot}}$
	Residual update (j-th iter.)	$T_{\text{prop}} M_{\text{tot}}$
	Average power (j-th iter.)	$T_{\text{prop}} M_{\text{tot}}$
	Subcarrier gains (no iter.)	$J^3 M_{\text{sb}} + 2J^2 T_{\text{prop}} M_{\text{sb}} + J T_{\text{prop}} M_{\text{tot}}$
OMP from [VAG17]	Projection (j-th iter.)	$T_{\text{omp}} M_{\text{tot}} (Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Maximum proj. (j-th iter.)	$M_{\text{tot}} (Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Gain calculation (j-th iter.)	$j^3 M_{\text{tot}} + 2j^2 T_{\text{omp}} M_{\text{tot}} + j T_{\text{omp}} M_{\text{tot}}$
	Residual update (j-th iter.)	$T_{\text{omp}} M_{\text{tot}}$
	Average power (j-th iter.)	$T_{\text{omp}} M_{\text{tot}}$
SW-OMP from [RGV18]	Projection (j-th iter.)	$T_{\text{sw-omp}} M_{\text{tot}} (Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Maximum proj. (j-th iter.)	$M_{\text{tot}} (Q_{\text{T}} Q_{\text{R}} - (j-1))$
	Gain calculation (j-th iter.)	$j^3 + 2j^2 T_{\text{sw-omp}} + j T_{\text{sw-omp}} M_{\text{tot}}$
	Residual update (j-th iter.)	$T_{\text{sw-omp}} M_{\text{tot}}$
	Average power (j-th iter.)	$T_{\text{sw-omp}} M_{\text{tot}}$

Due to the iterative nature of the proposed, OMP, and SW-OMP algorithms, it is important to consider and compare their computational complexities. Table 3.1 summarizes the results of the comparison. In all three algorithms, the iterative part consists of the following steps: 1) the projection of the residuals on the measurement matrix, 2) search for the maximum projection, 3) gain calculation, 4) residual update, and 5) calculation of the average residual power per subcarrier. As discussed earlier in this section, the proposed algorithm requires a lower overhead than the OMP and SW-OMP, i.e., $T_{\text{prop}} < T_{\text{sw-omp}} < T_{\text{omp}}$, which

makes its complexity lower than that of the benchmark algorithms in the iterative part of the algorithm. The projection of the residuals on the measurement matrix is the most computationally expensive step. The proposed algorithm requires $(Q_{\mathrm{T}}Q_{\mathrm{R}} - (j-1))$ projections, each with $T_{\mathrm{prop}}M_{\mathrm{tot}}$ complex operations. Note that the complexity of the projections could be significantly reduced at the cost of a higher overhead if a sequential CS-based estimation of AoDs and AoAs was used. In that case, the complexity would roughly scale as $\mathcal{O}(Q_{\mathrm{T}} + Q_{\mathrm{R}})$ instead of $\mathcal{O}(Q_{\mathrm{T}}Q_{\mathrm{R}})$. The search for the maximum projection scales with the number of sub-bands R, instead of with the number of subcarriers as in the benchmark algorithms. A lower overhead leads to a lower complexity of the residual updates in (3.16) and calculation of the average residual power per subcarrier in (3.17). In both steps, the number of complex operations scales with $T_{\mathrm{prop}}M_{\mathrm{tot}}$. The proposed algorithm uses a larger measurement matrix $\mathbf{\Phi}$ than the benchmark algorithms, which increases the complexity of the calculation of approximate gains in (3.15). Specifically, the algorithm requires $j^3 + 2j^2T_{\mathrm{prop}}M_{\mathrm{sb}} + jT_{\mathrm{prop}}M_{\mathrm{tot}}$ operations, which is better than in the OMP, but worse than in the SW-OMP algorithm. After the iterative part of the algorithm, the OMP and SW-OMP terminate, while the proposed algorithm needs to calculate the channel gains of J estimated paths for each subcarrier, as defined in (3.20), which requires $J^3M_{\mathrm{sb}} + 2J^2T_{\mathrm{prop}}M_{\mathrm{sb}} + JT_{\mathrm{prop}}M_{\mathrm{tot}}$ operations.

3.5 Impact of Hardware Impairments

In this section, we study the impact of hardware impairments on channel estimation performance. Specifically, we first numerically evaluate the impact of the delay and phase errors in the analog TTD array on the proposed algorithm. As a benchmark, we evaluate the impact of phase errors on the OMP and SW-OMP algorithms, which were originally designed for antenna arrays based on phase shifters. Then we linearize the received signal model with hardware impairments and derive the CRLB for channel parameter estimators. Finally, we propose and evaluate a gradient descent based refinement which improves the estimation



(a) (b)

Figure 3.7: Performance of the proposed algorithm in the presence of delay errors in terms of: (a) NMSE and (b) spectral efficiency.

accuracy of the proposed algorithm.

3.5.1 Channel Estimation under Hardware Impairments

For evaluation of the impact of hardware impairments, we assumed $N_{\mathrm{R}} = 16$, $T = 35$, and the same system and channel parameters as in Sec. 3.4. We first evaluate individual and then joint impact of the delay and phase errors in the array.

Unlike the state-of-the-art algorithms which are designed for arrays based on phase shifters, the performance of the proposed algorithm might be affected by the delay errors in the TTD array. In Fig. 3.7, we considered delay errors with the standard deviation in the range $\sigma_{\mathrm{T}} \in [0, 180]$ps and numerically measured their impact on the channel estimation accuracy and post-training spectral efficiency of the proposed algorithm. Although the considered delay errors may seem small compared to the overall delay introduced in the TTD circuits, it is important to note two things: 1) The delay errors $\tilde{\tau}_n$, $\forall n$, with the standard deviation $\sigma_{\mathrm{T}} \in [0, 180]$ps are substantially higher than the resolution of the state-of-the-art TTD hardware, which is in the order of 5ps [GSR19]; 2) Even small delay errors can impact

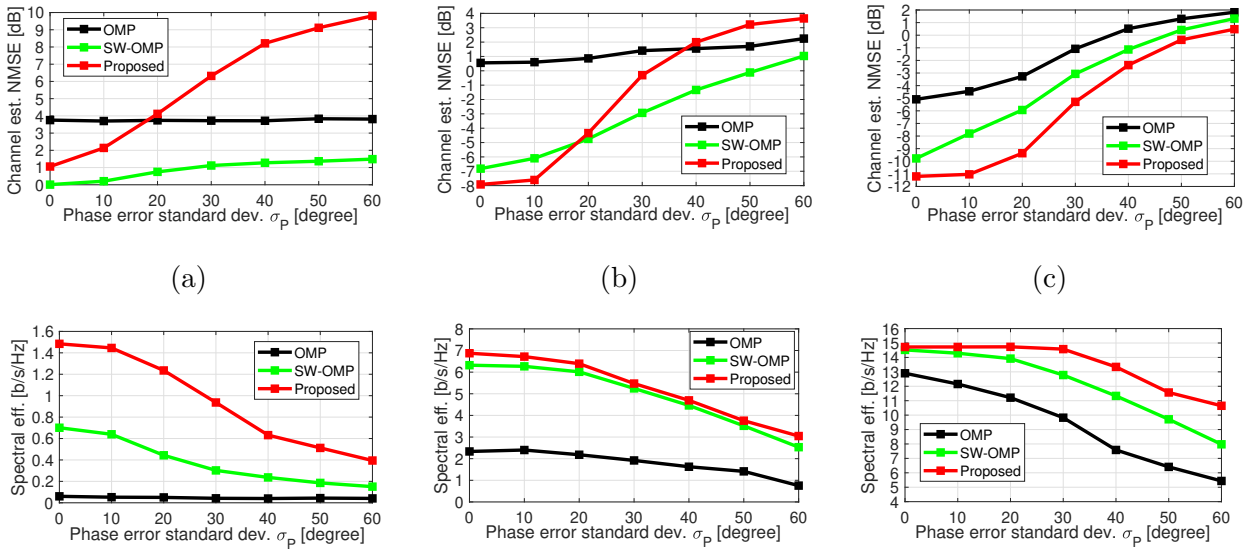


Figure 3.8: Performance of the proposed and state-of-the-art algorithms in the presence of phase errors in terms of: (a) NMSE for SNR $= -10$ dB, (b) NMSE for SNR $= 0$ dB, (c) NMSE for SNR $= 10$ dB, (d) spectral efficiency for SNR $= -10$ dB, (e) spectral efficiency for SNR $= 0$ dB, and (f) spectral efficiency for SNR $= 10$ dB.

the TTD combiners because the total phase change $2\pi(f_m - f_c)\tilde{\tau}_n$ induced by the delay error $\tilde{\tau}_n$ also depends on the considered bandwidth. As presented in Fig. 3.7(a), the channel estimation accuracy decreases as the standard deviation σ_T increases regardless of the SNR. However, the error is still relatively low (< 0dB) with significant delay errors, especially in moderate to high SNRs. In Fig. 3.7(b), we present the results for the post-estimation spectral efficiency. We can observe that even with large delay errors and degraded channel estimation accuracy, the spectral efficiency mainly depends on the SNR regime.

Next, we evaluate the performance of the proposed and state-of-the-art algorithms in the presence of phase errors in the UE's array. Similar as with delay errors, we considered different standard deviations σ_P of phase errors and numerically measured their impact on the channel estimation accuracy and post-training spectral efficiency. The results are presented in Fig. 3.8 for different SNR regimes. In all considered SNRs, the proposed algorithm sees a

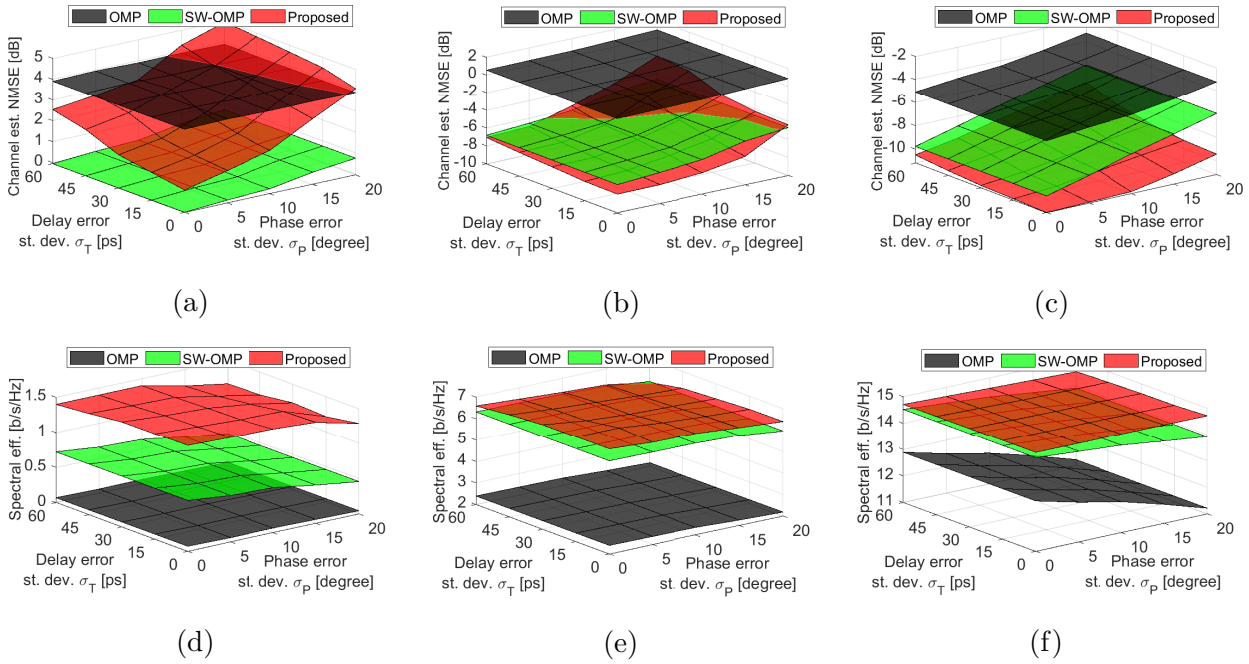


Figure 3.9: Performance of the proposed and state-of-the-art algorithms in the presence of delay and phase errors in terms of: (a) NMSE for SNR $= -10$ dB, (b) NMSE for SNR $= 0$ dB, (c) NMSE for SNR $= 10$ dB, (d) spectral efficiency for SNR $= -10$ dB, (e) spectral efficiency for SNR $= 0$ dB, and (f) spectral efficiency for SNR $= 10$ dB.

significant increase in channel estimation error as σ_{P} increases. This degradation translates into a lower post-estimation spectral efficiency for $\sigma_{\mathrm{P}} > 10°$, $\sigma_{\mathrm{P}} > 20°$, and $\sigma_{\mathrm{P}} > 30°$, at SNR $= -10$ dB, SNR $= 0$ dB, and SNR $= 10$ dB, respectively. Nevertheless, the spectral efficiency of the proposed algorithm remains higher than that of the state-of-the-art algorithms due to more accurate angle estimates, regardless of the SNR value.

In practice, antenna arrays are usually affected by multiple hardware impairments at the same time. For this reason, we evaluate a joint impact of the delay and phase errors on the channel estimation accuracy and post-estimation spectral efficiency of the proposed and state-of-the-art algorithms. Compared to the results of individual errors in Fig. 3.7 and Fig. 3.8, we consider smaller ranges of values for σ_{T} and σ_{P}. The results are shown in Fig. 3.9 for different SNR values. The performance of the state-of-the-art algorithms

is independent of delay errors and thus decreases only with an increase in σ_{P}. On the other hand, the performance of the proposed algorithm decreases when either of the errors increases. Similarly as in the results for individual errors, the proposed algorithm results in a higher spectral efficiency than the state-of-the-art, even though two errors affect the TTD array simultaneously.

3.5.2 CRLB of Parameter Estimation

In this subsection, we derive the lower bound for the variance of the channel parameter estimators in a LoS channel. We first derive the CRLB for AoD and AoA estimators and then we derive the CRLB for channel phase estimators.

Linearization of received signal model: In the proposed algorithm, the angles are estimated using the measurement model in (3.12) and per-sub-band processing. In a LoS channel, the vector $\boldsymbol{\lambda}_r$ in (3.12) has only one complex non-zero element G_r. For mathematical tractability, we assume that the gain G_r is normalized, i.e., $G_r = e^{j\psi_r}$, where ψ_r is a random phase of the complex gain. Assuming an infinitely large dictionary $\mathbf{A}$, the product $\mathbf{A}\boldsymbol{\lambda}_r$ can be simplified to $\mathbf{a}_{\mathrm{T}}^{*}(\theta^{(\mathrm{T})}) \otimes \mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})})e^{j\psi_r}$, where $\theta^{(\mathrm{T})}$ and $\theta^{(\mathrm{R})}$ are the true AoD and AoA, respectively. With this simplification, the model in (3.12) becomes

$$\mathbf{y}_r = \bar{\mathbf{x}}_r(\theta^{(\mathrm{T})}, \theta^{(\mathrm{R})}, \psi_r) + \tilde{\mathbf{n}}_r, \tag{3.23}$$

where $\bar{\mathbf{x}}_r(\theta^{(\mathrm{T})}, \theta^{(\mathrm{R})}, \psi_r) = \mathbf{F}(\mathbf{a}_{\mathrm{T}}^{*}(\theta^{(\mathrm{T})}) \otimes \mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})}))e^{j\psi_r}$. In order to keep the notation simpler, we use $\bar{\mathbf{x}}_r$ instead of $\bar{\mathbf{x}}_r(\theta^{(\mathrm{T})}, \theta^{(\mathrm{R})}, \psi_r)$ for the rest of this section, without the risk of ambiguity. The i-th element of the vector $\bar{\mathbf{x}}_r$ is given as follows

$$\begin{aligned}
[\bar{\mathbf{x}}_r]_i =& \mathbf{f}^{(t)}[m](\mathbf{a}_{\mathrm{T}}^{*}(\theta^{(\mathrm{T})}) \otimes \mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})}))e^{j\psi_r} \\
=& (\mathbf{v}^{(t)\mathrm{T}} \otimes \mathbf{w}^{(t)\mathrm{H}}[m])(\mathbf{a}_{\mathrm{T}}^{*}(\theta^{(\mathrm{T})}) \otimes \mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})}))e^{j\psi_r} \\
=& \mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\mathbf{w}^{(t)\mathrm{H}}[m]\mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})})e^{j\psi_r}
\end{aligned} \tag{3.24}$$

where $t = \lceil i/M_{\mathrm{sb}} \rceil$, $m = (r-1)M_{\mathrm{sb}} + \mathrm{mod}(i, M_{\mathrm{sb}})$.

Random hardware impairments distort the combiner $\mathbf{w}^{(t)}[m]$ and affect the received signal $[\mathbf{y}_r]_i$ non-linearly, which complicates the CRLB derivation. Thus, our first objective is to linearize the received signal model in (3.23). Using the expression for $\mathbf{w}^{(t)}[m]$ in (3.2), result in (3.24), $\bar{\tau}_n = \tau_n + \tilde{\tau}_n$, and $\bar{\phi}_n = \phi_n + \tilde{\phi}_n$, the received signal $[\mathbf{y}_r]_i$ can be expressed as

$$[\mathbf{y}_r]_i = e^{j\psi_r}\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\sum_{n=1}^{N_{\mathrm{R}}} e^{j(2\pi(f_m-f_c)\tau_n+\phi_n)}e^{-j(n-1)\pi\sin(\theta^{(\mathrm{R})})}e^{j(2\pi(f_m-f_c)\tilde{\tau}_n+\tilde{\phi}_n)} + [\tilde{\mathbf{n}}_r]_i. \tag{3.25}$$

Assuming that standard deviations of hardware impairments are relatively small, we can exploit the fact that $\exp(j(2\pi(f_m - f_c)\tilde{\tau}_n + \tilde{\phi}_n)) \approx 1 + j(2\pi(f_m - f_c)\tilde{\tau}_n + \tilde{\phi}_n)$ to rewrite (3.25) as follows

$$[\mathbf{y}_r]_i = [\mathbf{x}_r]_i + [\tilde{\mathbf{x}}_r]_i + [\tilde{\mathbf{n}}_r]_i, \tag{3.26}$$

where $[\mathbf{x}_r]_i$ and $[\tilde{\mathbf{x}}_r]_i$ are defined as follows

$$[\mathbf{x}_r]_i = e^{j\psi_r}\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\sum_{n=1}^{N_{\mathrm{R}}} e^{j(2\pi(f_m-f_c)\tau_n+\phi_n)}e^{-j(n-1)\pi\sin(\theta^{(R)})}, \tag{3.27}$$

$$[\tilde{\mathbf{x}}_r]_i = e^{j\psi_r}\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\sum_{n=1}^{N_{\mathrm{R}}} e^{j(2\pi(f_m-f_c)\tau_n+\phi_n)}e^{-j((n-1)\pi\sin(\theta^{(R)})+\pi/2)}(2\pi(f_m - f_c)\tilde{\tau}_n + \tilde{\phi}_n). \tag{3.28}$$

The deterministic value $[\mathbf{x}_r]_i$ represents an impairment-free sample of the signal. On the other hand, the complex random value $[\tilde{\mathbf{x}}_r]_i$ can be treated as a scaled coefficient of the Fourier transform of $2\pi(f_m - f_c)\tilde{\tau}_n + \tilde{\phi}_n$, $\forall n$, which represent random hardware impairments. The BS uses pseudo-random precoders $\mathbf{v}^{(t)}$ with quasi-omnidirectional beam patterns and therefore we can assume that the term $e^{j\psi_r}\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})$ does not change the variance, but only the phase, of $[\tilde{\mathbf{x}}_r]_i$. Since $2\pi(f_m - f_c)\tilde{\tau}_n + \tilde{\phi}_n$, $\forall n$, are zero-mean Gaussian random variables with variance $4\pi^2(f_m - f_c)^2\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{P}}^2$, it can be shown that $[\tilde{\mathbf{x}}_r]_i$ is a complex Gaussian random variable with zero-mean and variance $N_{\mathrm{R}}(4\pi^2(f_m - f_c)^2\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{P}}^2)$, i.e.,

$[\tilde{\mathbf{x}}_r]_i \sim \mathcal{CN}(0, N_{\mathrm{R}}(4\pi^2(f_m - f_c)^2\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{P}}^2))$. This random variable is independent of the noise $[\tilde{\mathbf{n}}_r]_i$. Thus, if we introduce a new variable $[\tilde{\tilde{n}}_r]_i = [\tilde{\mathbf{x}}_r]_i + [\tilde{\mathbf{n}}_r]_i$ for the sake of brevity, the received signal can be expressed in a vector form as follows

$$\mathbf{y}_r = \mathbf{x}_r + \tilde{\tilde{\mathbf{n}}}_r, \tag{3.29}$$

where $\tilde{\tilde{\mathbf{n}}}_r \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{TM_{\mathrm{sb}}})$, $\sigma^2 = N_{\mathrm{R}}(4\pi^2(f_m - f_c)^2\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{P}}^2 + \sigma_{\mathrm{N}}^2)$. We assume that the elements of $\tilde{\tilde{\mathbf{n}}}_r$ are independent and identically distributed for mathematical convenience.

CRLB for AoD and AoA: Let $\boldsymbol{\theta} = [\theta^{(\mathrm{T})}, \theta^{(\mathrm{R})}]^{\mathrm{T}}$. Based on the model in (3.29), the likelihood function is defined as $P(\mathbf{y}_r; \boldsymbol{\theta}) = \exp(-\|\mathbf{y}_r - \mathbf{x}_r\|_2^2/\sigma^2)/(\pi\sigma)^{TM_{\mathrm{sb}}}$. The log-likelihood function is $\mathcal{L}(\mathbf{y}_r; \boldsymbol{\theta}) = \log(P(\mathbf{y}_r; \boldsymbol{\theta}))$. The CRLB for AoD and AoA estimators is extracted from the inverse of the Fisher information matrix (FIM) $\boldsymbol{\Theta} \in \mathbb{R}^{2\times 2}$ defined in the following way

$$\begin{aligned}\boldsymbol{\Theta} &= \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\mathcal{L}(\mathbf{y}_r; \boldsymbol{\theta})\left(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\mathbf{y}_r; \boldsymbol{\theta})\right)^{\mathrm{H}}\right] \\ &= \begin{bmatrix} \Theta_{\theta^{(\mathrm{T})},\theta^{(\mathrm{T})}} & \Theta_{\theta^{(\mathrm{T})},\theta^{(\mathrm{R})}} \\ \Theta_{\theta^{(\mathrm{R})},\theta^{(\mathrm{T})}} & \Theta_{\theta^{(\mathrm{R})},\theta^{(\mathrm{R})}} \end{bmatrix},\end{aligned} \tag{3.30}$$

where $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\mathbf{y}_r; \boldsymbol{\theta})$ is the gradient of $\mathcal{L}(\mathbf{y}_r; \boldsymbol{\theta})$ with respect to the parameters in $\boldsymbol{\theta}$. By leveraging the fact that $\mathbb{E}[\mathbf{y}_r\mathbf{y}_r^{\mathrm{H}}] = \mathbf{x}_r\mathbf{x}_r^{\mathrm{H}} + \sigma^2\mathbf{I}_{TM_{\mathrm{sb}}}$ and $\mathbb{E}[\mathbf{y}_r\mathbf{y}_r^{\mathrm{T}}] = \mathbf{x}_r\mathbf{x}_r^{\mathrm{T}}$, it can be shown that the scalar Θ_{p_1,p_2} for parameters p_1 and p_2 is calculated as follows

$$\begin{aligned}\Theta_{p_1,p_2} &= \frac{1}{\sigma^2}\left(\frac{\partial \mathbf{x}_r^{\mathrm{H}}}{\partial p_1}\frac{\partial \mathbf{x}_r}{\partial p_2} + \frac{\partial \mathbf{x}_r^{\mathrm{T}}}{\partial p_1}\frac{\partial \mathbf{x}_r^{*}}{\partial p_2}\right) \\ &= \frac{2}{\sigma^2}\mathcal{R}\left(\frac{\partial \mathbf{x}_r^{\mathrm{H}}}{\partial p_1}\frac{\partial \mathbf{x}_r}{\partial p_2}\right).\end{aligned}$$

Therefore, the elements of $\boldsymbol{\Theta}$ in (3.30) are

$$\Theta_{\theta^{(\mathrm{T})},\theta^{(\mathrm{T})}} = \frac{2}{\sigma^2}\sum_{i=1}^{TM_{\mathrm{sb}}}\left|\mathbf{v}^{(t)\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{T}}(\theta^{(\mathrm{T})})\right|^2\left|\mathbf{w}_{\mathrm{ideal}}^{(t)\mathrm{H}}[m]\mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})})\right|^2, \tag{3.31}$$

$$\Theta_{\theta^{(\mathrm{R})},\theta^{(\mathrm{R})}} = \frac{2}{\sigma^2}\sum_{i=1}^{TM_{\mathrm{sb}}}\left|\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\right|^2\left|\mathbf{w}_{\mathrm{ideal}}^{(t)\mathrm{H}}[m]\dot{\mathbf{a}}_{\mathrm{R}}(\theta^{(\mathrm{R})})\right|^2, \tag{3.32}$$

$$\Theta_{\theta^{(\mathrm{T})},\theta^{(\mathrm{R})}} = \frac{2}{\sigma^2}\mathcal{R}\Big(\sum_{i=1}^{TM_{\mathrm{sb}}} \mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\mathbf{v}^{(t)\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{T}}(\theta^{(\mathrm{T})})\mathbf{w}_{\mathrm{ideal}}^{(t)\mathrm{H}}[m]\mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})})\mathbf{w}_{\mathrm{ideal}}^{(t)\mathrm{H}}[m]\dot{\mathbf{a}}_{\mathrm{R}}(\theta^{(\mathrm{R})})\Big), \quad (3.33)$$

$$\Theta_{\theta^{(\mathrm{R})},\theta^{(\mathrm{T})}} = \Theta_{\theta^{(\mathrm{T})},\theta^{(\mathrm{R})}}, \quad (3.34)$$

where the indices t and m depend on i as earlier, $\mathbf{w}_{\mathrm{ideal}}^{(t)}$ is an impairment-free combiner, and $\dot{\mathbf{a}}_{\mathrm{T}}(\theta^{(\mathrm{T})}) \in \mathbb{C}^{N_{\mathrm{T}}}$ and $\dot{\mathbf{a}}_{\mathrm{R}}(\theta^{(\mathrm{R})}) \in \mathbb{C}^{N_{\mathrm{R}}}$ are the first derivatives of the transmit and receive spatial responses, respectively. The n-th element of $\dot{\mathbf{a}}_{\mathrm{T}}(\theta^{(\mathrm{T})})$ is defined as $[\dot{\mathbf{a}}_{\mathrm{T}}(\theta^{(\mathrm{T})})]_n = -j(n-1)\pi\cos(\theta^{(\mathrm{T})})\exp(-j(n-1)\pi\sin(\theta^{(\mathrm{T})}))/\sqrt{N_{\mathrm{T}}}$. The n-th element of $\dot{\mathbf{a}}_{\mathrm{R}}(\theta^{(\mathrm{R})})$ is defined in a similar way. Based on the results from (3.31)-(3.34), we can calculate the inverse of $\boldsymbol{\Theta}$ and determine the lower bounds for the variances of AoD and AoA estimators as follows

$$\mathrm{Var}(\hat{\theta}^{(\mathrm{T})}) \geq [\boldsymbol{\Theta}^{-1}]_{1,1}, \quad (3.35)$$

$$\mathrm{Var}(\hat{\theta}^{(\mathrm{R})}) \geq [\boldsymbol{\Theta}^{-1}]_{2,2}. \quad (3.36)$$

CRLB for channel phases: Unlike the angles, the channel gains at all subcarriers are estimated using the measurement model in (3.8) and per-subcarrier processing. For mathematical tractability, we assume that the gain at the m-th subcarrier $G[m]$ is normalized, i.e., $G[m] = e^{j\psi[m]}$. With this assumption, estimation of the channel gain $G[m]$ reduces to estimation of the channel phase $\psi[m]$. Additionally, the assumption allows the same linearization method to be applied to simplify the per-subcarrier received signal model. The simplified model has the following form

$$\mathbf{y}[m] = \mathbf{x}[m] + \tilde{\tilde{\mathbf{n}}}[m], \quad (3.37)$$

where $\mathbf{x}[m] \in \mathbb{C}^T$ is defined similarly as in (3.27), but on a subcarrier level, and $\tilde{\tilde{\mathbf{n}}}[m] \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_T)$, $\sigma^2 = N_{\mathrm{R}}(4\pi^2(f_m - f_{\mathrm{c}})^2\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{P}}^2 + \sigma_{\mathrm{N}}^2)$. Based on (3.37), the likelihood function is $P(\mathbf{y}[m];\psi[m]) = \exp(-\|\mathbf{y}[m] - \mathbf{x}[m]\|_2^2/\sigma^2)/(\pi\sigma)^T$, while the log-likelihood is $\mathcal{L}(\mathbf{y}[m];\psi[m]) =$

$\log(P(\mathbf{y}[m];\psi[m]))$. The Fisher information for $\psi[m]$ is given by

$$\begin{aligned}\Theta_{\psi[m],\psi[m]} &= \frac{2}{\sigma^2}\frac{\partial \mathbf{x}^{\mathrm{H}}[m]}{\partial \psi[m]}\frac{\partial \mathbf{x}[m]}{\partial \psi[m]} \\ &= \frac{2}{\sigma^2}\sum_{t=1}^{T}\left|\mathbf{v}^{(t)\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta^{(\mathrm{T})})\right|^2\left|\mathbf{w}_{\mathrm{ideal}}^{(t)\mathrm{H}}[m]\mathbf{a}_{\mathrm{R}}(\theta^{(\mathrm{R})})\right|^2. \end{aligned} \tag{3.38}$$

Based on (3.38), the lower bound for the variance of channel phase is given as follows

$$\mathrm{Var}(\hat{\psi}[m]) \geq \Theta^{-1}_{\psi[m],\psi[m]}. \tag{3.39}$$

3.5.3 Gradient Descent for Parameter Refinement

Here we design a gradient descent based parameter refinement to improve the estimation accuracy of the proposed algorithm.

In the iterative part of the proposed algorithm, the angles are estimated using a predetermined dictionary. Such an approach yields the so called "on-grid" estimates, which limit the estimation accuracy. Increasing the dictionary size, i.e., increasing Q_{T} and Q_{R}, to improve the estimation accuracy is not a feasible solution because it comes at the cost of a prohibitively large computational complexity. An alternative way to improve the accuracy is to iteratively refine the initial estimates of channel parameters. A few refinement approaches were designed in the previous work, e.g., refinement based on the Newton method in [MRM16]. We propose a gradient descent based refinement to improve the initial channel estimates, including the AoD $\hat{\theta}_0^{(\mathrm{T})}$, AoA $\hat{\theta}_0^{(\mathrm{R})}$, and channel phase $\hat{\psi}_0[m]$.

Given the simplified received signal model in (3.29), we define a cost function $\mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta})$, $\boldsymbol{\theta} = [\theta^{(\mathrm{T})}, \theta^{(\mathrm{R})}]^{\mathrm{T}}$, jointly for *all* R sub-bands as follows

$$\mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta}) = \frac{1}{2RTM_{\mathrm{sb}}}\sum_{r=1}^{R}\|\mathbf{y}_r - \mathbf{x}_r\|_2^2. \tag{3.40}$$

The first derivative of $\mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta})$ with respect to a certain parameter θ in $\boldsymbol{\theta}$ is given by

$$\frac{\partial \mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta})}{\partial \theta} = -\frac{1}{RTM_{\mathrm{sb}}}\sum_{r=1}^{R}\mathcal{R}((\mathbf{y}_r - \mathbf{x}_r)^{\mathrm{H}}\frac{\partial \mathbf{x}_r}{\partial \theta}) \tag{3.41}$$

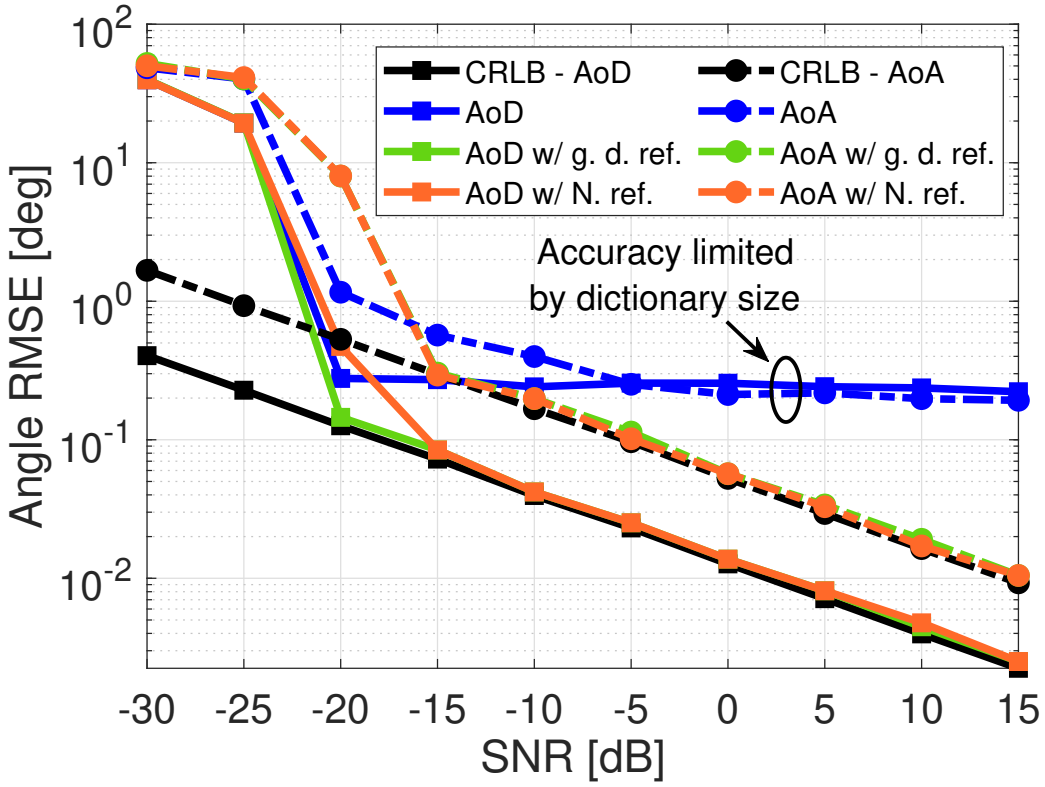


Figure 3.10: Comparison of the algorithms with and without parameter refinement against the lower bounds in terms of the angle estimation accuracy.

Based on (3.41), we can define the j-th iteration of the gradient descent based refinement of the AoD and AoA in the following way

$$\boldsymbol{\theta}_{j-1} = [\theta_{j-1}^{(\mathrm{T})}, \theta_{j-1}^{(\mathrm{R})}]^{\mathrm{T}} \tag{3.42}$$

$$\theta_{j}^{(\mathrm{T})} = \theta_{j-1}^{(\mathrm{T})} - \mu_{\theta^{(\mathrm{T})}} \frac{\partial \mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta}_{j-1})}{\partial \theta^{(\mathrm{T})}} \tag{3.43}$$

$$\theta_{j}^{(\mathrm{R})} = \theta_{j-1}^{(\mathrm{R})} - \mu_{\theta^{(\mathrm{R})}} \frac{\partial \mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta}_{j-1})}{\partial \theta^{(\mathrm{R})}} \tag{3.44}$$

were $\mu_{\theta^{(\mathrm{T})}}$ and $\mu_{\theta^{(\mathrm{R})}}$ and the learning rates for the AoD and AoA updates, respectively. The proposed refinement algorithm iterates over the steps (3.42)-(3.44) until the gradients $\partial \mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta}_{j-1})/\partial \theta^{(\mathrm{T})}$ and $\partial \mathcal{C}_{\mathrm{angle}}(\boldsymbol{\theta}_{j-1})/\partial \theta^{(\mathrm{R})}$ become sufficiently small, e.g., less than 10^{-3}.

The updated angles are used for the refinement of channel phases. Unlike the angles which

are common for all subcarriers, the channel phases need to be refined for each subcarrier individually. With the system model in (3.37), we define a cost function $\mathcal{C}_{\text{phase}}(\psi[m])$ for the m-th subcarrier as follows

$$\mathcal{C}_{\text{phase}}(\psi[m]) = \frac{1}{2T} \left\| \mathbf{y}[m] - \mathbf{x}[m] \right\|_2^2 . \tag{3.45}$$

The derivative of $\mathcal{C}_{\text{phase}}(\psi[m])$ with respect to $\psi[m]$ is

$$\frac{\partial \mathcal{C}_{\text{phase}}(\psi[m])}{\partial \psi[m]} = -\frac{1}{T} \mathcal{R}((\mathbf{y}[m] - \mathbf{x}[m])^{\text{H}} \frac{\partial \mathbf{x}[m]}{\partial \psi[m]}) \tag{3.46}$$

The j-th iteration in the channel phase refinement is

$$\psi_j[m] = \psi_{j-1}[m] - \mu_\psi \frac{\partial \mathcal{C}_{\text{phase}}(\psi_{j-1}[m])}{\partial \psi[m]}. \tag{3.47}$$

Similarly as with the angles, the algorithm iterates until the gradient becomes small.

The proposed refinement method increases the overall complexity of the estimation algorithm. The most computationally expensive part is the calculation of the gradient. Let J_{angle} and J_{phase} be the average number of iterations needed to refine the angles and channel phases, respectively. Since the angle refinement considers all R sub-bands simultaneously, the number of additional complex operations it requires scales as $\mathcal{O}(2TM_{\text{tot}}J_{\text{angle}})$. Similarly, the number of operations needed for channel phase refinement scales as $\mathcal{O}(2TM_{\text{tot}}J_{\text{phase}})$ since the phase needs to be refined at all M_{tot} subcarriers.

Finally, we compare the performance of the proposed channel estimation algorithm with and without gradient descent based refinement against the calculated CRLBs. In addition, we include the Newton-based parameter refinement as a benchmark [MRM16]. We assume a LoS channel, SNR $\triangleq 10\log_{10}(1/\sigma^2)$, and the same system parameters as in the previous simulations. The main metric is the RMSE and it is evaluated for different SNR values. For each SNR point, the noise power is split such that 10% of it comes from $2\pi(f_m - f_{\text{c}})\tilde{\tau}_n$ (delay error), 10% of it from $\tilde{\phi}_n$ (phase error), and 80% of it from the thermal noise. The results for the angle and channel phase estimation are presented in Fig. 3.10 and Fig. 3.11,

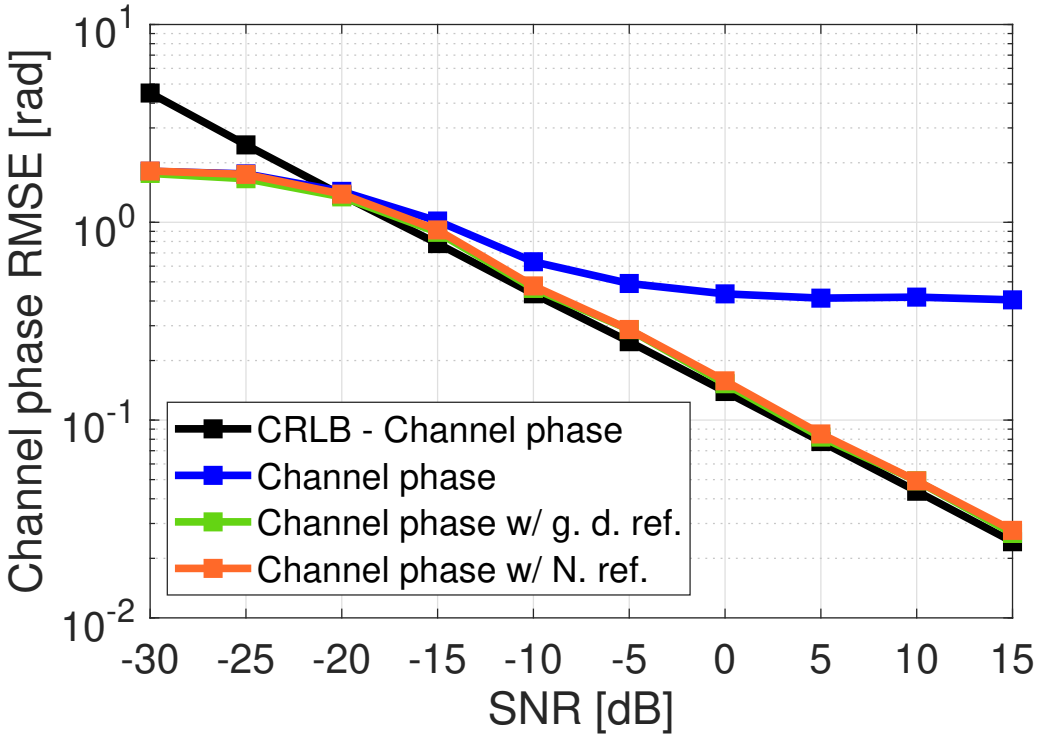


Figure 3.11: Comparison of the algorithms with and without parameter refinement against the lower bounds in terms of the channel phase estimation accuracy.

respectively. In Fig. 3.10, we observe that the angle estimation accuracy of the proposed algorithm without refinement is limited by the dictionary size, i.e., by the values Q_{T} and Q_{R}. The proposed gradient descent refinement and the benchmark Newton refinement have a similar performance. They both significantly improve the estimation accuracy and achieve the AoD and AoA CRLBs. However, it is worth noting that the Newton refinement requires the calculation of a Hessian and its inverse, which makes its computational complexity higher than that of the proposed gradient descent. In Fig. 3.11, the algorithm without refinement has a limited channel phase estimation accuracy. On the other hand, the algorithm with the gradient descent or Newton refinement exploits the refined angle estimates and iterative improvements of channel phase to achieve the corresponding CRLB.

3.6 Conclusions

In this chapter, we proposed and studied a CS-based channel estimation algorithm for analog TTD arrays that are affected by hardware impairments. The proposed algorithm was shown to achieve a good channel estimation accuracy while keeping the required overhead and computational complexity lower than in related benchmark approaches. Due to a particularly high angle estimation accuracy, the algorithm leads to a higher post-estimation spectral efficiency than the benchmark approaches. The proposed gradient descent refinement of channel parameters enables the algorithm to reach the estimation accuracy lower bound. The impact of time-invariant TTD hardware impairments was evaluated through extensive numerical simulations and the proposed algorithm was shown to have a high tolerance to the delay and phase errors.

CHAPTER 4

User association and Low-Interference Beam Scheduling

4.1 Introduction

In mmW systems, antenna arrays are commonly assumed to have a hybrid architecture with multiple RF chains [HIX15], which enables both the BS and UE to steer more than one beam at the time. Thus, the BS can exploit the hybrid architecture to serve multiple UEs simultaneously. Similarly, the UE can leverage its RF chains to increase its achievable data rate through the spatial multiplexing or connection with multiple BSs at the same time, which is often referred to as multi-connectivity [3GP18]. However, with an increased densification of mmW networks [GTM16] characterized by a small inter-site distance and a large number of UEs, the inter- and intra-site interference cannot be neglected. In particular, the main-lobe of highly directional beams can cause significant interference and prevent data-hungry UEs from satisfying their rate requirements.

The interference problem can be abated and the overall network performance can be improved by optimizing user association and/or beam scheduling in the network. A properly designed user association and/or beam scheduling framework aims to maximize a specific network utility function, while considering important service requirements, properties of directional communication, and features of mmW transceivers. For example, some of the important considerations include UEs' heterogeneous rate requirements, inter- and intra-cell interference, hybrid array architectures at the BSs and UEs. However, the existing opti-

mization frameworks are often mathematically too complex and not comprehensive enough. For example, the utility function is commonly non-convex and focused on maximization of the network sum rate, while the set of constraints do not necessarily lower the interference and capture important properties of mmW communication. In addition, with maximization of the network sum rate, available serving beams are scheduled to the UEs with good channels, while other UEs' rate requirements are often not satisfied. Therefore, there is a need to design and mathematically formulate a new comprehensive framework that maximizes the number of satisfied UEs, while suppressing the interference through intelligent beam scheduling.

In the following subsections, we review recent work on mmW user association and resource allocation, and then we highlight the key contributions of this chapter.

4.1.1 Prior Work

Prior work usually considered the problems of mmW user association [LY21, AV19, TAR21, ZRX21] and resource allocation [HWN17, SQL19, LE21, QLS22, SW19, SX21, Hu18, ZSL19, SPX19, AHC22, SCW22] separately. The frameworks that consider these problems jointly provide higher flexibility in terms of the network performance optimization, but they have rarely been studied [PWX21, BSC22b].

The work in [LY21] studied the user association problem in a network with enhanced mobile broadband (eMBB) UEs and ultra-reliable low-latency communication (URLLC) UEs that support multi-connectivity. The proposed framework aimed to maximize the rate of eMBB UEs, while guaranteeing the reliability constraints for URLLC UEs. In [AV19] and [TAR21], the authors proposed load balancing user association schemes and heuristic algorithms to solve the optimization problems in polynomial time. The work in [ZRX21] proposed a framework with a multi-criterion objective that minimizes both the link blockage probability and the maximum load at the BSs. The authors studied different scalarization techniques and demonstrated the advantage of considering multi-criterion over separate single-criterion

objectives. However, optimization frameworks in [LY21, AV19, TAR21, ZRX21] do not consider UEs' rate requirements and explicit interference management on a network level using a set of constraints. For example, interference suppression is performed only implicitly through load balancing in [AV19] and [TAR21].

In mmW networks, resource allocation includes the distribution of beams (space-time resources), frequencies, and power. A joint allocation of all of these resources represents a very complex optimization problem. Thus, many previously proposed frameworks aimed to either allocate only a subset of resources or provide a good sub-optimal solution to the joint allocation. The work in [HWN17, SQL19, LE21] studied beam allocation in a scenario with one BS that is equipped with a hybrid antenna array. Similarly, [QLS22] considered a single-cell beam allocation and it proposed angle-based precoding to suppress the residual interference. While the schemes in [HWN17, SQL19, LE21, QLS22] can eliminate the interference within the considered cell, the inter-cell interference remains a problem. In contrast, the authors of [SW19] studied the inter-cell interference between two BSs, and they introduced a low-complexity recursive algorithm that minimizes the number of inter-cell beam collisions. A more general case of inter-cell interference suppression with multiple base stations was considered in [SX21] and [Hu18]. In both [Hu18], the scheduling framework gives a higher priority to the UEs that cause lower inter-cell interference, i.e., that have higher SINR. Available system resources can also be allocated sequentially [ZSL19, SPX19, AHC22], e.g., the BSs first allocate beams and then frequencies and powers to all associated UEs. A recent work in [SCW22] proposed a beam scheduling scheme that suppresses the interference and minimizes the amount of unfulfilled UEs' requirements. However, the network scenario considered in [SCW22] is optimistic in terms of the beam interference since it assumes that all UEs have only one RF chain and that they do not support spatial multiplexing and/or multi-connectivity.

Optimization frameworks that jointly consider user association and beam scheduling in dense mmW networks were rarely studied in previous work. In [PWX21], the authors de-

signed a centralized algorithm to minimize the transmission time for all associated UEs. The proposed algorithm has multiple steps, including the UE clustering based on location, UE association and inter-cluster interference suppression, and power allocation. However, similar to [SCW22], the assumed setup is optimistic in terms of the beam interference because single-antenna UEs do not support spatial multiplexing and/or multi-connectivity. In addition, [PWX21] considers only LoS channels and neglects the impact of beam side-lobes on interference. In our recent work in [BSC22b], we introduced a new framework for mmW systems where both the BSs and UEs can steer multiple beams simultaneously. The proposed framework aims to maximize the number of UEs with satisfied rate requirements while minimizing the inter- and intra-cell main-lobe interference in the network. We also proposed a low-complexity greedy algorithm to solve the corresponding optimization problem and demonstrated that it outperforms the baseline user association approaches. Our previous work, however, did not address several important questions, including the distribution of any remaining serving beams, design of hybrid precoders and combiners to cancel the excess interference, and optimization of power allocation in the network.

4.1.2 Contributions

Given the shortcomings of the previous work, there is a need to design a more comprehensive optimization framework for joint user association and low-interference beam scheduling. In this chapter, we extend the work in [BSC22b] and we propose a new multi-step framework to optimize the network performance. The contributions can be summarized as follows:

- We propose a new framework where user association and beam scheduling are performed jointly in a centralized sub-network. The framework consists of three sequential steps: Step 1 - maximization of the number of UEs with fully satisfied rate requirements; Step 2 - maximization of the number of UEs with partially satisfied rate requirements using the remaining serving beams; Step 3 - design of hybrid precoders

and combiners, followed by power allocation on a sub-network level to boost the rates of partially satisfied UEs.

- We mathematically formulate linear optimization problems in all three steps. This includes the design of multi-criterion objectives using scalarization and the design of constraints that consider UEs' rate requirements and suppress the inter- and intra-cell interference.

- We explain that the optimization problems in Step 1 and Step 2 are NP-hard, and then we design a heuristic algorithm based on relaxation, rounding, and resource pruning to obtain sub-optimal solutions. We also analyze the complexity of the proposed algorithm and show that the solutions are obtained in polynomial time.

- Using realistic mmW channels generated in Quadriga [JRB19], we evaluate the impact of the interference management on the average SINR per associated link. We also compare the proposed optimization framework with existing baseline approaches, including the naive greedy association and the maximum sum rate association. The comparison is done in terms of average number of UEs with satisfied rate requirements, network sum rate, and transmit power usage per BS.

4.1.3 Organization and Notation

The rest of the chapter is organized as follows. In Sec. 4.2, we present the system model and problem formulation. Sec. 4.3 introduces a new joint optimization framework for user association and beam scheduling. In Sec. 4.4, the proposed low-complexity algorithm is described. The proposed framework is numerically evaluated and compared with benchmark approaches in Sec. 4.5. Finally, Sec. 4.6 concludes the chapter and discusses the ideas for future work.

Notation: Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. Sets are denoted by capital script letters. Hermitian

transpose and inverse are denoted by $(.)^{\mathrm{H}}$ and $(.)^{-1}$, respectively.

4.2 System Model

In this section, we introduce the considered system model. Important notation is summarized in Table 4.1.

We consider downlink communication in a mmW sub-network consisting of N_{BS} BSs from the set $\mathcal{B}$ and N_{UE} UEs from the set $\mathcal{U}$. The BSs are connected to a centralized processing unit (PU) with high computational power. An example of the considered sub-network is illustrated in Fig. 4.1. It is assumed that all downlink communication links operate at the same carrier frequency within the bandwidth BW. Each BS is equipped with a linear fully-connected hybrid array with $N_{\mathrm{BS}}^{\mathrm{A}}$ antennas and $N_{\mathrm{BS}}^{\mathrm{RF}}$ RF chains that enable steering of up to $N_{\mathrm{BS}}^{\mathrm{RF}}$ serving beams simultaneously. Similarly, each UE has a linear fully-connected hybrid array with $N_{\mathrm{UE}}^{\mathrm{A}}$ antennas and $N_{\mathrm{UE}}^{\mathrm{RF}}$ RF chains. The UEs can use their RF chains to achieve spatial multiplexing and/or multi-connectivity. It is assumed that $N_{\mathrm{UE}}^{\mathrm{RF}} \leq N_{\mathrm{BS}}^{\mathrm{RF}}$, so there can be up to $N_{\mathrm{UE}}^{\mathrm{RF}}$ links between one BS-UE pair. We assume that each UE u has already estimated the channel matrix between itself and E_{BS} closest BSs, where $E_{\mathrm{BS}} \leq N_{\mathrm{BS}}$. Using the control links, the channel estimates are reported to the corresponding BSs, which pass the estimates further to the PU. The channels and their estimates are assumed to be constant over a period of T slots. During these T slots, each associated UE u must have its specific data rate requirement R_u satisfied.

Let $\mathbf{F}_{b,t} \in \mathbb{C}^{N_{\mathrm{BS}}^{\mathrm{A}} \times N_{\mathrm{BS}}^{\mathrm{RF}}}$ be a precoding matrix with normalized columns that the BS b uses to serve its UEs in the time slot t. Given the hybrid architecture at each BS, the precoding matrix is calculated as $\mathbf{F}_{b,t} = \mathbf{F}_{\mathrm{RF}}^{b,t}\mathbf{F}_{\mathrm{BB}}^{b,t}$, where $\mathbf{F}_{\mathrm{RF}}^{b,t} \in \mathbb{C}^{N_{\mathrm{BS}}^{\mathrm{A}} \times N_{\mathrm{BS}}^{\mathrm{RF}}}$ is an RF precoding matrix consisting of normalized orthogonal DFT beams that correspond to different links and $\mathbf{F}_{\mathrm{BB}}^{b,t} \in \mathbb{C}^{N_{\mathrm{BS}}^{\mathrm{RF}} \times N_{\mathrm{BS}}^{\mathrm{RF}}}$ is a baseband precoding matrix. Similarly, let $\mathbf{W}_{u,t} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{A}} \times N_{\mathrm{UE}}^{\mathrm{RF}}}$ be a combining matrix with normalized with columns that the UE u uses to combine signal from one or

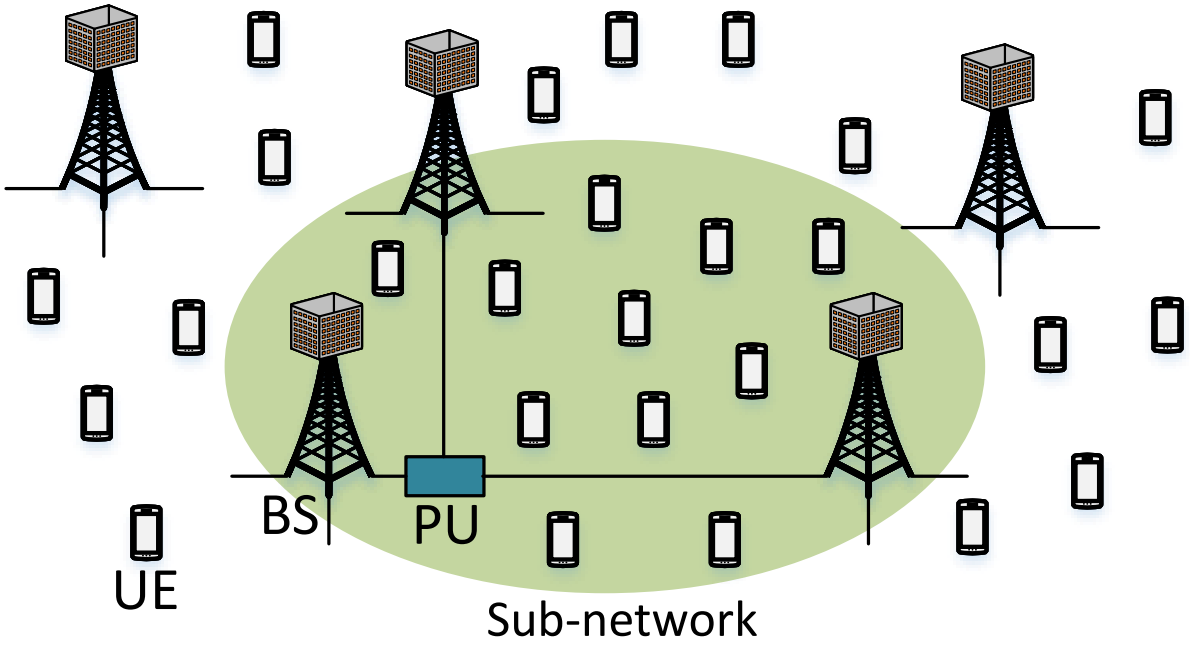


Figure 4.1: An example of the considered sub-network with $N_{\mathrm{BS}} = 3$ centralized BSs and $N_{\mathrm{UE}} = 12$ UEs. The sub-network can have irregular shape with different inter-site distances.

more BSs in the time slot t. The combining matrix is given as $\mathbf{W}_{u,t} = \mathbf{W}_{\mathrm{RF}}^{u,t}\mathbf{W}_{\mathrm{BB}}^{u,t}$, where $\mathbf{W}_{\mathrm{RF}}^{u,t} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{A}} \times N_{\mathrm{UE}}^{\mathrm{RF}}}$ is a matrix of normalized RF DFT combiners and $\mathbf{W}_{\mathrm{BB}}^{u,t} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{RF}} \times N_{\mathrm{UE}}^{\mathrm{RF}}}$ is a baseband combining matrix. Let the matrix $\mathbf{H}_{u,b} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{A}} \times N_{\mathrm{BS}}^{\mathrm{A}}}$ represent a channel estimate between the BS b and UE u. Using the estimate $\mathbf{H}_{u,b}$ and DFT beamforming matrices, the PU can determine $N_{\mathrm{UE}}^{\mathrm{RF}}$ distinct pairs $(\mathbf{d}_{\mathrm{BS}}^{\mathbf{u,b,l}}, \mathbf{d}_{\mathrm{UE}}^{\mathbf{u,b,l}})$, where $\mathbf{d}_{\mathrm{BS}}^{u,b,l} \in \mathbb{C}^{N_{\mathrm{BS}}^{\mathrm{A}}},\ l = 1, ..., N_{\mathrm{UE}}^{\mathrm{RF}}$, are *candidate* RF DFT precoders and $\mathbf{d}_{\mathrm{UE}}^{u,b,l} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{A}}},\ l = 1, ..., N_{\mathrm{UE}}^{\mathrm{RF}}$, are *candidate* RF combiners between b and u through digital processing. Importantly, we note that $\mathbf{d}_{\mathrm{BS}}^{u,b,l}$ and $\mathbf{d}_{\mathrm{UE}}^{u,b,l}$ remain the same for all T slots because it is assumed that the channel does not change. The pairs of candidate DFT beams $(\mathbf{d}_{\mathrm{BS}}^{\mathbf{u,b,l}}, \mathbf{d}_{\mathrm{UE}}^{\mathbf{u,b,l}})$ define a set $\mathcal{L}$ of $N_{\mathrm{UE}}^{\mathrm{RF}}$ independent links between b and u. The complexity of digitally estimating all candidate links in the sub-network scales as $\mathcal{O}(N_{\mathrm{UE}} N_{\mathrm{UE}}^{\mathrm{RF}} E_{\mathrm{BS}} N_{\mathrm{UE}}^{\mathrm{A}} N_{\mathrm{BS}}^{\mathrm{A}})$. With knowledge of all links, the PU can identify the interfering

(completely overlapping) DFT beam candidates at all BSs and UEs and then avoid scheduling them in the same time slot.

4.2.1 Pessimistic Estimates of Link Capacities

In dense mmW networks, link capacities are affected by the inter- and intra-cell interference. Before the process of user association and beam scheduling, the PU does not know which candidate links (beams) will be selected for communication. Thus, it is not possible to create the hybrid precoding and combining matrices, optimize power allocation, calculate interference power on each link, and determine the *true* link capacities in advance. Moreover, embedding the calculation of the link capacities in the user association and beam scheduling process makes the corresponding optimization problem non-convex and hard to solve. Recent work in [SCW22] considered an alternative approach where *pessimistic* estimates of interference powers are pre-calculated for all links. This enables the design of a convex and less complex optimization problem. Here we adopt a similar approach in which the PU uses the candidate RF precoders and combiners to estimate pessimistic interference powers and link capacities before user association and beam scheduling. Unlike in [SCW22], pessimistic interference power estimates here account for interference from E_{BS} BSs. The case when $E_{\mathrm{BS}} = N_{\mathrm{BS}}$ leads to the most pessimistic estimates of interference powers and capacities, but it also ensures that all associated UEs have their rate requirements certainly satisfied. Importantly, after user association and beam scheduling, the difference between the true capacities and pessimistic capacity estimates can be determined and exploited when power allocation is optimized in the sub-network, as discussed later in Sec. 4.3.3.

Let $\mathbf{d}_{\mathrm{BS}}^{u,b,l}$ and $\mathbf{d}_{\mathrm{UE}}^{u,b,l}$ be the RF DFT beams that correspond to the l-th link between b and u. Assuming that the transmit power budget P_{T} is split equally among $N_{\mathrm{BS}}^{\mathrm{RF}}$ RF chains, an estimate of the corresponding link capacity $\hat{c}_{u,b,l}$ is given as

$$\hat{c}_{u,b,l} = \mathrm{BW} \log_2 \left(1 + \frac{|\mathbf{d}_{\mathrm{UE}}^{u,b,l\mathrm{H}} \mathbf{H}_{u,b} \mathbf{d}_{\mathrm{BS}}^{u,b,l}|^2}{\hat{I}_{u,b,l} + P_{\mathrm{N}}} \frac{P_{\mathrm{T}}}{N_{\mathrm{BS}}^{\mathrm{RF}}} \right), \tag{4.1}$$

where P_{N} is the post-combining noise power. The noise power is defined as $P_{\mathrm{N}} = N_{\mathrm{UE}}^{\mathrm{A}} \mathrm{BW} N_0$, where N_0 is the noise power spectral density. If the channel between the BS b and UE u is not estimated, the capacity is $\hat{c}_{u,b,l} = 0, \ \forall l$. Note that $\hat{c}_{u,b,l}$ is the same in all slots. The interference power $\hat{I}_{u,b,l}$ in (4.1) is calculated as

$$\hat{I}_{u,b,l} = \sum_{b',u',l'} |\mathbf{d}_{\mathrm{UE}}^{u,b,l\mathrm{H}} \mathbf{H}_{u,b'} \mathbf{d}_{\mathrm{BS}}^{u',b',l'}|^2 \frac{P_{\mathrm{T}}}{N_{\mathrm{BS}}^{\mathrm{RF}}} \tag{4.2}$$

where $u' \neq u$, b' goes over E_{BS} BSs that are known the UE u, and l' goes over all links that belong to the UE u' and that are *not conflicted* with the link l. Thus, (4.2) accounts for the side-lobe interference from E_{BS} BSs and their candidate DFT beams that do not overlap with $\mathbf{d}_{\mathrm{BS}}^{u,b,l}$. The reason for not including overlapping beams will be clear in the following section, when we design a framework that suppresses the main-lobe interference.

4.2.2 Problem Formulation

Given the described system model, we want to design a sequential multi-step framework for joint user association and beam scheduling to achieve the following:

- The number of fully satisfied UEs is maximized.
- The number of partially satisfied UEs is maximized.
- Inter- and intra-cell interference is suppressed.
- Power allocation is optimized on a sub-network level.

4.3 Proposed Optimization Framework

The existing frameworks for user association and/or beam scheduling are often based on single-criterion objectives that maximize the network sum rate. With such objectives, the available system resources are mainly allocated to the UEs that experience very good channels and many UEs may be left out with unsatisfied rate requirements.

We address this issue by proposing a new three-step optimization framework for joint user association and low-interference beam scheduling in dense mmW networks. In the first step, the framework maximizes the number of UEs with fully satisfied rate requirements. In the second step, if there are any remaining resources, they are distributed such that the number of UEs with partially satisfied rate requirements is maximized. Finally, in the last step, hybrid precoders and combiners are designed and power allocation is optimized on all allocated communication links such that the partially satisfied UEs can further boost their rates.

In the following subsections, we mathematically formulate all three steps of the proposed framework.

4.3.1 Step 1 - Maximum Number of Fully Satisfied UEs

We first design a set of constraints that suppress the interference and consider UEs' rate requirements, and then we model a multi-criterion objective function that maximizes the number of satisfied UEs using a minimal number of beams.

Let $\mathcal{T}$ be a set of of T time slots over which the scheduling is done. Let [1] be a common label for all variables in Step 1. Let $\mathbf{s}^{[1]}$ be a vector of binary association variables $s_u^{[1]},\ u \in \mathcal{U}$, with $s_u^{[1]}$ being 1 if the rate requirement of the UE u is satisfied over the period of T slots, and 0 otherwise. Let $\mathbf{x}^{[1]}$ be a vector of $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}}N_{\text{BS}}$ binary association variables $x_{u,b,l,t}^{[1]},\ u \in \mathcal{U}, b \in \mathcal{B}, l \in \mathcal{L}, t \in \mathcal{T}$, for all links in the network, with $x_{u,b,l,t}^{[1]}$ being 1 if the UE u is served by the BS b using the link l in the time slot t, and 0 otherwise.

Due to the hybrid array architecture, each UE u can combine up to $N_{\text{UE}}^{\text{RF}}$ beams simultaneously in any time slot t. Given the variables $x_{u,b,l,t}^{[1]},\ \forall u, b, l, t$, this constraint is expressed as

$$\text{(C-1.1)}: \quad \sum_{b,l} x_{u,b,l,t}^{[1]} \leq N_{\text{UE}}^{\text{RF}} s_u^{[1]}, \quad \forall u \in \mathcal{U}, t \in \mathcal{T}. \tag{4.3}$$

The variable s_u in (C1) ensures that the beams are not allocated to the UEs whose rates

cannot be satisfied. The constraint (C-1.1) implies that a UE can get up to $TN_{\text{UE}}^{\text{RF}}$ links over T time slots.

Similarly, each BS can steer up to $N_{\text{BS}}^{\text{RF}}$ beams simultaneously in any time slot t, which is expressed as

$$\text{(C-1.2)}: \quad \sum_{u,l} x_{u,b,l,t}^{[1]} \leq N_{\text{BS}}^{\text{RF}}, \quad \forall b \in \mathcal{B}, t \in \mathcal{T}. \tag{4.4}$$

With E_{BS} known BSs, there are $E_{\text{BS}}N_{\text{UE}}^{\text{RF}}$ RF DFT beam candidates that can be used to serve a UE in any time slot. As explained in Sec. 4.2, the PU can identify the conflicted DFT beams at each BS. In particular, for each $u \in \mathcal{U}$, the PU can create an *interfering group*, i.e., a set $\mathcal{I}_u$ of the UEs that have at least one DFT beam candidate that is common with the candidate beams of the UE u. An illustration of an interfering group is presented in Fig. 4.2. Let u', $u' \neq u$, be a UE from the set $\mathcal{I}_u$. Let the BS index b' and the link index l' correspond to a DFT beam candidate that u' has in common with u. Highly directional interference from the BSs can be suppressed by not allowing DFT beam candidates of the UE u to be used by both the UE u and the UE u' in the same time slot t. To define this constraint mathematically, we introduce a new vector of binary association variables $\mathbf{a}^{[1]}$, whose element $a_{u,u',t}^{[1]}$ is 1 if the UE u' is allocated at least one of the DFT beam candidates of the UE u in the time slot t, and 0 otherwise. Given the variables $x_{u,b,l,t}^{[1]}$ and indices b' and l', the variable $a_{u,u',t}^{[1]}$ is modeled as a logical OR using the following set of constraints:

$$\text{(C-1.3)}: \quad a_{u,u',t}^{[1]} \leq \sum_{b',l'} x_{u',b',l',t}^{[1]}, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, t \in \mathcal{T}, \tag{4.5}$$

$$\text{(C-1.4)}: \quad a_{u,u',t}^{[1]} \geq x_{u',b',l',t}^{[1]}, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, t \in \mathcal{T}, b', l', \tag{4.6}$$

$$\text{(C-1.5)}: \quad 0 \leq a_{u,u',t}^{[1]} \leq 1, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, t \in \mathcal{T}. \tag{4.7}$$

Therefore, a constraint that suppresses the interference from the BSs can be defined as follows

$$\text{(C-1.6)}: \quad a_{u,u,t}^{[1]} + a_{u,u',t}^{[1]} \leq 1, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, u \neq u', t \in \mathcal{T}. \tag{4.8}$$

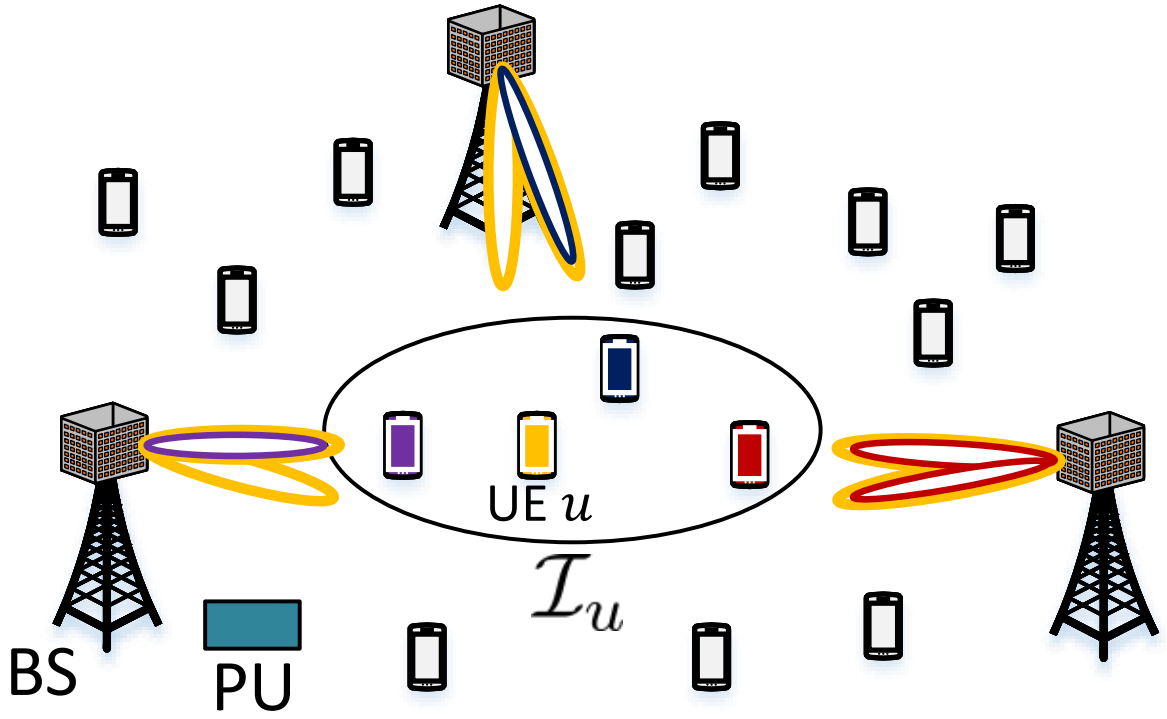


Figure 4.2: An illustration of the interfering group $\mathcal{I}_u$ with 4 UEs. The UE u has two DFT beam candidates from each BS. The red, blue, and magenta UEs have at least one DFT beam candidate in common with the UE u.

The PU can also identify the conflicted RF DFT beam candidates at each UE and it can use this information to suppress the combining interference in any time slot t. Let $\mathcal{J}_u$ be a set of DFT beam candidates that correspond to more than one link of the UE u. Let an arbitrary beam from $\mathcal{J}_u$ correspond to the links defined by the indices b'' and l''. The combining interference can be avoided by allowing only one link to use the beam, which is expressed as follows

$$\text{(C-1.7)}: \quad \sum_{b'',l''} x^{[1]}_{u,b'',l'',t} \leq 1, \quad \forall u \in \mathcal{U}, t \in \mathcal{T}. \tag{4.9}$$

Note that the total number of constraints in (C-1.3) - (C-1.7) depends on the number of conflicted beams at the BSs and UEs, and it may change from one association to another

depending on the network density and topology.

Each associated UE u needs to have its rate requirement R_u satisfied. This constraint can be expressed using the pessimistic capacity estimates in (4.1) as follows

$$\text{(C-1.8)}: \quad \sum_{b,l,t} \hat{c}_{u,b,l} x^{[1]}_{u,b,l,t} \geq R_u s^{[1]}_u, \quad \forall u \in \mathcal{U}. \tag{4.10}$$

Under the described constraints, our multi-criterion objective is to maximize the number of UEs with fully satisfied rate requirements using a minimal number of serving beams. If UEs have low rate requirements and all of them can be fully satisfied, reducing the number of serving beams reduces the overall side-lobe interference in the network. On the other hand, if UEs have high rate requirements and only a few of them can be fully satisfied, a minimal number of serving beams ensures that as many beams as possible are left for the UEs in Step 2. The number of UEs with fully satisfied rate requirements and the number of serving beams can be expressed as $\sum_u s^{[1]}_u$ and $\sum_{u,b,l,t} x^{[1]}_{u,b,l,t}$, respectively. Thus, the objective can be formulated as follows

$$\max_{\mathbf{s}^{[1]},\mathbf{x}^{[1]},\mathbf{a}^{[1]}} \sum_u s^{[1]}_u - \lambda_1 \sum_{u,b,l,t} x^{[1]}_{u,b,l,t}. \tag{4.11}$$

The parameter λ_1 can be obtained through scalarization [BV04] and its proper design in (4.11) is given by the following proposition.

Proposition 1. *The parameter* $\lambda_1 = \frac{1}{K_1+1}$, *where* $K_1 = N_{UE}(TN^{RF}_{UE}-1)+1$ *if* $TN_{UE}N^{RF}_{UE} \leq TN_{BS}N^{RF}_{BS}$, *and* $K_1 = TN_{BS}N^{RF}_{BS}$ *otherwise, guarantees that the number of fully satisfied UEs is maximized using a minimal number of serving beams.*

Proof. See Appendix B.1. □

Finally, with the designed constraints and multi-criterion objective, the linear optimiza-

tion problem in Step 1 can be formulated as follows

$$\begin{aligned} \max_{\mathbf{s}^{[1]},\mathbf{x}^{[1]},\mathbf{a}^{[1]}} \quad & \sum_{u} s_u^{[1]} - \lambda_1 \sum_{u,b,l,t} x_{u,b,l,t}^{[1]} \\ \text{s.t.} \quad & \text{(C-1.1)} - \text{(C-1.8)}, \\ & s_u^{[1]},\ x_{u,b,l,t}^{[1]},\ a_{u,u',t}^{[1]} \in \{0,1\}, \quad \forall u,u',b,l,t. \end{aligned} \tag{4.12}$$

The total number of optimization variables in (4.12) is $V_1 = N_{\text{UE}} + TN_{\text{UE}}N_{\text{UE}}^{\text{RF}}N_{\text{BS}} + T\sum_u |\mathcal{I}_u|$. This number can be reduced to $V_1 = N_{\text{UE}} + TN_{\text{UE}}N_{\text{UE}}^{\text{RF}}E_{\text{BS}} + T\sum_u |\mathcal{I}_u|$ if only non-zero-capacity links are considered.

4.3.2 Step 2 - Maximum Number of Partially Satisfied UEs

In Step 1, the UEs whose rate requirements are not fully satisfied are not associated with any BS. These UEs are allocated the remaining resources in Step 2 in a fair way, proportional to their rate requirements. This can be achieved using a similar optimization problem as in Step 1, but the set of constraints needs to be updated based on the solutions $\mathbf{s}^{[1]}$, $\mathbf{x}^{[1]}$, and $\mathbf{a}^{[1]}$. Let $\mathbf{s}^{[2]}$, $\mathbf{x}^{[2]}$, and $\mathbf{a}^{[2]}$ be vectors of Step 2 variables with the same definition as their counterparts from Step 1. Clearly, the number of variables in Step 2 is the same as in Step 1, i.e., $V_2 = V_1$.

In Step 2, the constraint (C-1.1) needs to be updated to ensure that that the fully satisfied UEs from Step 1 are not given any additional serving beams as follows

$$\text{(C-2.1)}: \quad \sum_{b,l} x_{u,b,l,t}^{[2]} \leq (1 - s_u^{[1]}) N_{\text{UE}}^{\text{RF}} s_u^{[2]}, \quad \forall u \in \mathcal{U}, t \in \mathcal{T}. \tag{4.13}$$

After Step 1, the number of available beams at the BSs, defined in (C-1.2), is updated in the following way

$$\text{(C-2.2)}: \quad \sum_{u,l} x_{u,b,l,t}^{[2]} \leq N_{\text{BS}}^{\text{RF}} - \sum_{u,l} x_{u,b,l,t}^{[1]}, \quad \forall b \in \mathcal{B}, t \in \mathcal{T}. \tag{4.14}$$

The UEs in Step 2 can be served only by the beams that do not violate the precoding interference suppression achieved in Step 1. Therefore, the variables $a_{u,u',t}^{[2]}$ in Step 2 need to

account the beams that were already allocated in Step 1. This is achieved by updating the constraints (C-1.3) and (C-1.4):

$$\text{(C-2.3)}: \quad a^{[2]}_{u,u',t} \leq \sum_{b',l'} x^{[2]}_{u',b',l',t} + \sum_{b',l'} x^{[1]}_{u',b',l',t}, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, t \in \mathcal{T}, \tag{4.15}$$

$$\text{(C-2.4)}: \quad a^{[2]}_{u,u',t} \geq x^{[2]}_{u',b',l',t} + x^{[1]}_{u',b',l',t}, \quad \forall u \in \mathcal{U}, u' \in \mathcal{I}_u, t \in \mathcal{T}, b', l'. \tag{4.16}$$

The constraints (C-1.5) - (C-1.7) do not require an update, i.e., (C-2.5) - (C-2.7) are the same as (C-1.5) - (C-1.7).

In Step 2, the UEs cannot have their rate requirements fully satisfied. However, each associated UE is required to have at least a fraction ϵ of its requirement satisfied, where $0 < \epsilon < 1$. Mathematically, the rate constraint is expressed as

$$\text{(C-2.8)}: \quad \sum_{b,l,t} \hat{c}_{u,b,l} x^{[2]}_{u,b,l,t} \geq \epsilon R_u s^{[2]}_u, \quad \forall u \in \mathcal{U}. \tag{4.17}$$

The constraint based on a partial rate satisfaction makes the distribution of the remaining resources proportional to the UEs' rate requirements.

Similar as in Step 1, we design a multi-criterion objective function whose primary goal is to maximize the number of partially satisfied UEs. Since the UEs are only partially satisfied, we want to at least boost their data rates as much as possible. For this reason, the objective's secondary goal is to schedule the beams such that the sum rate of partially satisfied UEs is maximal. The objective is defined as

$$\max_{\mathbf{s}^{[2]},\mathbf{x}^{[2]},\mathbf{a}^{[2]}} \sum_u s^{[2]}_u + \lambda_2 \sum_{u,b,l,t} \hat{c}_{u,b,l} x^{[2]}_{u,b,l,t}. \tag{4.18}$$

The parameter λ_2 is given by the following proposition.

Proposition 2. *For* $\hat{c}_{u,b,l}(1 - x^{[1]}_{u,b,l,t}) > 0, \ \forall u, b, l, t,$ *let* $c_{\min} = \min_{u,b,l,t} \hat{c}_{u,b,l}(1 - x^{[1]}_{u,b,l,t})$ *and* $c_{\max} = \max_{u,b,l,t} \hat{c}_{u,b,l}(1 - x^{[1]}_{u,b,l,t})$. *The parameter* $\lambda_2 = \frac{1}{K_2+1}$, *where* $K_2 = (N_{UE} - 1)TN^{RF}_{UE}c_{\max} - N_{UE}c_{\min}$ *if* $TN_{UE}N^{RF}_{UE} \leq TN_{BS}N^{RF}_{BS}$, *and* $K_2 = TN_{BS}N^{RF}_{BS}c_{\max}$ *otherwise,*

guarantees that the number of partially satisfied UEs is maximized and that the sum rate of partially satisfied UEs is maximal.

Proof. See Appendix B.2. □

The optimization problem in Step 2 is then formulated as

$$\begin{aligned} \max_{\mathbf{s}^{[2]},\mathbf{x}^{[2]},\mathbf{a}^{[2]}} \quad & \sum_{u} s_u^{[2]} + \lambda_2 \sum_{u,b,l,t} \hat{c}_{u,b,l} x_{u,b,l,t}^{[2]} \\ \text{s.t.} \quad & \text{(C-2.1)} - \text{(C-2.8)}, \\ & s_u^{[2]},\ x_{u,b,l,t}^{[2]},\ a_{u,u',t}^{[2]} \in \{0,1\}, \quad \forall u, u', b, l, t. \end{aligned} \tag{4.19}$$

4.3.3 Step 3 - Hybrid Beamformers and Power Allocation

The number of links scheduled in Step 1 and Step 2 is lower or equal than the initial number of links candidates, i.e, $\sum_{u,b,l,t}(x_{u,b,l,t}^{[1]} + x_{u,b,l,t}^{[2]}) \leq T N_{\text{UE}} N_{\text{UE}}^{\text{RF}} E_{\text{BS}}$. Additionally, since the PU knows all scheduled links in the sub-network and the corresponding DFT beam candidates, it can determine the RF precoding matrices $\mathbf{F}_{\text{RF}}^{b,t},\ \forall b, t$, and the RF combining matrices $\mathbf{W}_{\text{RF}}^{u,t},\ \forall u, t$. Specifically, the columns of the RF precoding matrix $\mathbf{F}_{\text{RF}}^{b,t}$ are vectors $\mathbf{d}_{\text{BS}}^{u',b,l'}$, where u' and l' go over the UEs and links that the BS b is serving in the time slot t. The matrix $\mathbf{W}_{\text{RF}}^{u,t}$ is determined in a similar way. The PU can then use RF precoding and combining matrices to calculate the matrices $\mathbf{F}_{\text{BB}}^{b,t},\ \forall b, t$, $\mathbf{W}_{\text{BB}}^{u,t},\ \forall u, t$, $\mathbf{F}_{b,t},\ \forall b, t$, and $\mathbf{W}_{u,t},\ \forall u, t$, as explained in Appendix B.3. The digital precoding matrices $\mathbf{F}_{\text{BB}}^{b,t},\ \forall b, t$, and digital combining matrices $\mathbf{W}_{\text{BB}}^{u,t},\ \forall u, t$, are designed to remove any excess side-lobe interference after RF precoding and combining. Therefore, with a lower number of links and digital precoders and combiners, the true interference $I_{u,b,l,t}$ experienced by the l-th link between the BS b and UE u in the time slot t is lower or equal than the interference $\hat{I}_{u,b,l}$ in (4.2). Note that the true interference $I_{u,b,l,t}$ can vary over different time slots because it depends on currently scheduled beams. A lower interference implies that the true time-varying link capacity $c_{u,b,l,t}$ is larger or equal than the pessimistic capacity $\hat{c}_{u,b,l}$ in (4.1) used for user association and

beam scheduling. This gap between the capacities provides the opportunity for optimization of power allocation on a sub-network level in every time slot. In our framework, the main objective of power allocation is to reduce the amount of power that is allocated to the fully satisfied UEs, while increasing the amount power that is allocated to the partially satisfied UEs. With a proper set of constraints, such an objective ensures that the fully satisfied UEs are provided the minimum required power, while the partially satisfied UEs are provided with the opportunity to boost their data rates as much as possible. It is important to note that the optimization of power allocation is possible only when the PU can calculate the interference for all BS-UE pairs, i.e., when $E_{\mathrm{BS}} = N_{\mathrm{BS}}$. Additionally, to simplify the notation in this subsection, we assume that the indices u, b, l, and t go over the used links (non-zero values in $x^{[1]}_{u,b,l,t}$ and $x^{[2]}_{u,b,l,t}$) only.

Let $\hat{\gamma}_{u,b,l}$ be the pessimistic SINR on the l-th link between the BS b and UE u as defined in (4.1), i.e., $\hat{\gamma}_{u,b,l} = |\mathbf{d}^{u,b,l\mathrm{H}}_{\mathrm{UE}} \mathbf{H}_{u,b} \mathbf{d}^{u,b,l}_{\mathrm{BS}}|^2 P_{\mathrm{T}} / ((\hat{I}_{u,b,l} + P_{\mathrm{N}}) N^{\mathrm{RF}}_{\mathrm{BS}})$. Let $p_{u,b,l,t}$, $\mathbf{f}_{u,b,l,t}$, and $\mathbf{w}_{u,b,l,t}$ be the allocated power, hybrid precoder, and hybrid combiner on the l-th link between the BS b and UE u in the time slot t. Let $g_{u,b,l,t} = |\mathbf{w}^{\mathrm{H}}_{u,b,l,t} \mathbf{H}_{u,b} \mathbf{f}_{u,b,l,t}|^2$ be the effective beamforming gain for the considered link. Let $g^{u,b,l,t}_{u\prime,b',l',t} = |\mathbf{w}^{\mathrm{H}}_{u,b,l,t} \mathbf{H}_{u,b'} \mathbf{f}_{u',b',l',t}|^2$ be the effective beamforming gain for the interfering link l' between the BS b' and UE u', $u' \neq u$, in the same time slot t.

Each BS has a limited power budget P_{T}. Thus, a sum of allocated powers must satisfy the following constraint

$$\text{(C-3.1)}: \quad \sum_{u,l} p_{u,b,l,t} \leq P_{\mathrm{T}}, \ b \in \mathcal{B}, t \in \mathcal{T}. \tag{4.20}$$

Clearly, the individual powers are constrained as follows

$$\text{(C-3.2)}: \quad 0 \leq p_{u,b,l,t} \leq (x^{[1]}_{u,b,l,t} + x^{[2]}_{u,b,l,t}) P_{\mathrm{T}}, \quad \forall u \in \mathcal{U}, b \in \mathcal{B}, l \in \mathcal{L}, t \in \mathcal{T}. \tag{4.21}$$

The use of $x^{[1]}_{u,b,l,t}$ and $x^{[2]}_{u,b,l,t}$ in (C-3.2) ensures that only selected links can have a non-zero power allocation.

In optimization of power allocation, the rate constraints have to remain satisfied for all associated UEs. For a fully satisfied UE u, the rate constraint can be mathematically

expressed using the true capacities $c_{u,b,l,t},\ \forall u,b,l,t$, as

$$\sum_{b,l,t} c_{u,b,l,t} \geq R_u. \tag{4.22}$$

After diving both sides by BW, the constraint becomes

$$\sum_{b,l,t} \log_2 \left(1 + \frac{g_{u,b,l,t} p_{u,b,l,t}}{\sum_{u',b',l'} g^{u,b,l,t}_{u',b',l',t} p_{u',b',l',t} + P_{\mathrm{N}}}\right) \geq \frac{R_u}{\mathrm{BW}}. \tag{4.23}$$

The constraints for partially satisfied UEs are formulated in a similar way after accounting for the fraction ϵ. The left-hand side of (4.23) is a highly non-linear function of the powers $p_{u,b,l,t},\ \forall u,b,l,t$. Since we aim to express power allocation as a linear optimization problem, (4.23) cannot be used as a constraint. The following proposition defines a set of linear constraints that represent a sufficient condition to satisfy (4.23).

Proposition 3. *The constraints defined as*

$$g_{u,b,l,t} p_{u,b,l,t} - \hat{\gamma}_{u,b,l} \sum_{u',b',l'} g^{u,b,l,t}_{u',b',l',t} p_{u',b',l',t} \geq \hat{\gamma}_{u,b,l} P_N, \quad \forall u,b,l,t.$$

represent a sufficient condition to satisfy (4.23).

Proof. Consider the following inequalities for the UE u

$$\sum_{b,l,t} \frac{c_{u,b,l,t}}{\mathrm{BW}} \overset{\text{(a)}}{\geq} \sum_{b,l,t} \frac{\hat{c}_{u,b,l}}{\mathrm{BW}} \overset{\text{(b)}}{\geq} \frac{R_u}{\mathrm{BW}} \ \left(\text{or } \frac{\epsilon R_u}{\mathrm{BW}}\right). \tag{4.24}$$

The inequality (b) is the data rate constraint defined in (C-1.8) (or (C-2.8)) and thus it holds true for the associated UE. The inequality (a) comes as a consequence of having a lower number of scheduled than candidate links and applying digital precoders and combiners to the scheduled beams, as discussed earlier. To provide mathematical guarantees that (a) holds true, it is sufficient to ensure that the true link capacities of the UE u are larger or equal than their pessimistic counterparts, i.e., $c_{u,b,l,t} \geq \hat{c}_{u,b,l},\ \forall u,b,l,t$, or equivalently

$$\log_2 \left(1 + \frac{g_{u,b,l,t} p_{u,b,l,t}}{\sum_{u',b',l'} g^{u,b,l,t}_{u',b',l',t} p_{u',b',l',t} + P_{\mathrm{N}}}\right) \geq \log_2(1 + \hat{\gamma}_{u,b,l}), \quad \forall u,b,l,t \tag{4.25}$$

Since log() is a monotonic function, (4.25) can be written as

$$\frac{g_{u,b,l,t}p_{u,b,l,t}}{\sum_{u',b',l'} g^{u,b,l,t}_{u',b',l',t}p_{u',b',l',t} + P_{\mathrm{N}}} \geq \hat{\gamma}_{u,b,l}, \quad \forall u, b, l, t \tag{4.26}$$

After reorganizing (4.26), we get

$$\text{(C-3.3)}: \quad g_{u,b,l,t}p_{u,b,l,t} - \hat{\gamma}_{u,b,l} \sum_{u',b',l'} g^{u,b,l,t}_{u',b',l',t}p_{u',b',l',t} \geq \hat{\gamma}_{u,b,l}P_{\mathrm{N}}, \quad \forall u, b, l, t, \tag{4.27}$$

which concludes the proof. □

Given the power constraints, we aim to minimize the amount of power that is allocated to the fully satisfied UEs. At the same time, we want to maximize the amount of power that is allocated to the partially satisfied UEs to boost their achievable rates. Thus, with $\mathbf{p}$ being a vector of optimization variables $p_{u,b,l,t}, \ \forall u, b, l, t$, the objective can be defined as follows

$$\max_{\mathbf{p}} \quad -\sum_{u\in\mathcal{U}_1}\sum_{b,l,t} p_{u,b,l,t} + \sum_{u\in\mathcal{U}_2}\sum_{b,l,t} p_{u,b,l,t} \tag{4.28}$$

In (4.28), $\mathcal{U}_1$ and $\mathcal{U}_2$ denote the sets of the UEs associated in Step 1 and Step 2, respectively.

Finally, the optimization problem can be formulated as

$$\begin{aligned} \max_{\mathbf{p}} \quad & -\sum_{u\in\mathcal{U}_1}\sum_{b,l,t} p_{u,b,l,t} + \sum_{u\in\mathcal{U}_2}\sum_{b,l,t} p_{u,b,l,t} \\ \text{s.t.} \quad & \text{(C-3.1)} - \text{(C-3.3)} \end{aligned} \tag{4.29}$$

Based on the objective function and constraint (C-3.3), (4.29) aims to reduce the amount of power that is allocated to the fully satisfied UEs until their true link capacities become equal to the pessimistic capacities which were used for user association and beam scheduling. Conversely, the partially satisfied UEs are provided with more power and their rates are improved. It is important to note that certain partially satisfied UEs may be allocated enough transmit power to have their rate requirements fully satisfied. The number of variables in (4.29) is equal to the number of selected links, i.e., $V_3 = \sum_{u,b,l,t}(x^{[1]}_{u,b,l,t} + x^{[2]}_{u,b,l,t})$.

Power allocation in (4.29) is a linear optimization problem with continuous variables $p_{u,b,l,t}, \ \forall u, b, l, t$, and it can be solved in polynomial time using interior point methods. On

the other hand, the optimization problems in Step 1 and Step 2 are NP-hard combinatorial problems. In the next section, we propose an algorithm based on relaxation, rounding, and resource pruning to solve these problems in polynomial time.

4.4 Proposed Algorithm for Step 1 and Step 2

The optimization problems in (4.12) and (4.19) are binary integer programs (BIP), which are known to be NP-hard. Existing optimization solvers, including Gurobi [Gur22] and Mosek [MOS22], rely on branch and bound techniques to solve BIPs. However, the worst case complexity of these techniques scales exponentially, i.e., they may require an exhaustive search across all permutations of binary optimization variables to find the optimal solution. In a scenario with a few BSs and a moderate number of UEs and time slots, the exhaustive search has a prohibitive computational complexity. In this section, we proposed a low-complexity algorithm based on relaxation, rounding, and resource pruning to sub-optimally solve the BIPs in Step 1 and Step 2.

We first relax the Step 1 BIP in (4.12) and formulate the following program with continuous optimization variables

$$\begin{aligned} \max_{\mathbf{s}^{[1]},\mathbf{x}^{[1]},\mathbf{a}^{[1]}} \quad & \sum_{u} s_u^{[1]} - \lambda_1 \sum_{u,b,l,t} x_{u,b,l,t}^{[1]} \\ \text{s.t.} \quad & \text{(C-1.1)} - \text{(C-1.8)}, \\ & 0 \leq s_u^{[1]},\ x_{u,b,l,t}^{[1]},\ a_{u,u',t}^{[1]} \leq 1, \quad \forall u, u', b, l, t. \end{aligned} \tag{4.30}$$

The relaxed linear optimization problem can be solved in polynomial time using interior point methods [RTV05]. However, the solution to (4.30) is fractional, meaning that the elements of $\mathbf{s}^{[1]}$, $\mathbf{x}^{[1]}$, and $\mathbf{a}^{[1]}$ of the optimization variables are not necessarily equal to 0 or 1. For example, the rate requirement of the UE u that experiences good channels can be fully satisfied using only a fraction of the link capacities, i.e., the corresponding variables $x_{u,b,l,t}^{[1]}$, $\forall b, l, t$, of the selected links may be significantly lower than 1. However, the proposed

Algorithm 2 Rounding with resource pruning

Inputs: $\mathbf{s}$, $\mathbf{x}$, δ

$\bar{s}_u \leftarrow 1$ if $s_u \geq \delta$, $\bar{s}_u \leftarrow 0$ otherwise

$\bar{x}_{u,b,l,t} \leftarrow 1$ if $x_{u,b,l,t} \geq \delta$, $\bar{x}_{u,b,l,t} \leftarrow 0$ otherwise

$$\bar{a}_{u,u',t} = \begin{cases} 1, & \text{if } \sum_{b',l'} \bar{x}_{u',b',l',t} \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Initialize: $\hat{\mathbf{s}} \leftarrow \bar{\mathbf{s}}, \quad \hat{\mathbf{x}} \leftarrow \bar{\mathbf{x}}, \quad \hat{\mathbf{a}} \leftarrow \bar{\mathbf{a}}$

Check feasibility of $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{a}}$

while $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{a}}$ are infeasible **do**

  Count constraint violations for each link

  Determine link (u^*, b^*, l^*, t^*) with most violations

  $\hat{x}_{u^*,b^*,l^*,t^*} \leftarrow 0$

  if Requirement R_{u^*} is not satisfied anymore **then**

    $\hat{s}_{u^*} \leftarrow 0, \quad \hat{x}_{u^*,b,l,t} \leftarrow 0, \ \forall b, l, t$

  end if

  $$\hat{a}_{u,u',t} = \begin{cases} 1, & \text{if } \sum_{b',l'} \hat{x}_{u',b',l',t} \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

  Check feasibility of $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{a}}$

end while

for Each UE **do**

  Keep minimal number of best links (beams)

end for

Outputs: $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, $\hat{\mathbf{a}}$

user association and beam scheduling framework requires all variables to be rounded to either 0 or 1. Thus, it is necessary to design a rounding algorithm that would generate binary vectors $\hat{\mathbf{s}}^{[1]}$, $\hat{\mathbf{x}}^{[1]}$, and $\hat{\mathbf{a}}^{[1]}$, and pass them to Step 2.

Given the rounded solution from Step 1, the Step 2 BIP in (4.19) is relaxed as follows

$$\begin{aligned} \max_{\mathbf{s}^{[2]},\mathbf{x}^{[2]},\mathbf{a}^{[2]}} \quad & \sum_{u} s_u^{[2]} + \lambda_2 \sum_{u,b,l,t} \hat{c}_{u,b,l} x_{u,b,l,t}^{[2]} \\ \text{s.t.} \quad & \text{(C-2.1)} - \text{(C-2.8)}, \\ & 0 \leq s_u^{[2]},\ x_{u,b,l,t}^{[2]},\ a_{u,u',t}^{[2]} \leq 1, \quad \forall u, u', b, l, t. \end{aligned} \tag{4.31}$$

Similar as in Step 1, a fractional solution consisting of the vectors $\mathbf{s}^{[2]}$, $\mathbf{x}^{[2]}$, and $\mathbf{a}^{[2]}$ need to be rounded. After rounding, the resulting binary vectors from Step 1 and Step 2 are passed to Step 3 to perform power allocation on the selected links.

Next, we propose a polynomial-time algorithm to round the fractional solutions to (4.30) and (4.31). The algorithm includes a threshold-based rounding, followed by an iterative resource pruning that yields feasible solutions.

4.4.1 Rounding with Resource Pruning

We propose a rounding algorithm that is used after solving the relaxed problem in Step 1 and after solving the relaxed problem in Step 2. Since the proposed algorithm is used after both Step 1 and Step 2, we simplify the notation in this subsection by omitting the superscripts [1] and [2] in the fractional solutions $\mathbf{s}$, $\mathbf{x}$, and $\mathbf{a}$.

The algorithm start by rounding the fractional solutions based on the threshold δ. In particular, we obtain a binary vector of associated UEs $\bar{\mathbf{s}}$, where each element $\bar{s}_u$ is set to 1 if $s_u \geq \delta$, and to 0 otherwise. Similarly, we obtain a binary vector of selected links $\bar{\mathbf{x}}$ by setting $\bar{x}_{u,b,l,t}$ to 1 if $x_{u,b,l,t} \geq \delta$, and to 0 otherwise. The elements of the vector $\bar{\mathbf{a}}$ are calculated based on the elements of $\bar{\mathbf{x}}$ and expression in (4.5) as follows

$$\bar{a}_{u,u',t} = \begin{cases} 1, & \text{if } \sum_{b',l'} \bar{x}_{u',b',l',t} \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.32}$$

To avoid missing the links where only a fraction of the capacity is used, we assume that δ is a small constant value, e.g., $\delta = 0.1$. However, with a small δ, it is likely that too

many links will be considered as 'used' and that the vectors $\bar{\mathbf{s}}$, $\bar{\mathbf{x}}$, and $\bar{\mathbf{a}}$ will violate multiple constraints and thus represent an infeasible solution. To obtain feasible vectors $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{a}}$, we propose an iterative approach based on resource pruning.

The main idea of resource pruning is to dismiss the most conflicted resource, i.e., the one that violates the most constraints, in each iteration. We first initialize the vectors $\hat{\mathbf{s}} = \bar{\mathbf{s}}$, $\hat{\mathbf{x}} = \bar{\mathbf{x}}$, and $\hat{\mathbf{a}} = \bar{\mathbf{a}}$. Then in each iteration, we identify the communication link (u^*, b^*, l^*, t^*) that violates most constraints and we set its corresponding association variable $\hat{x}_{u^*,b^*,l^*,t^*} = 0$. If removing the link prevents the rate requirement R_{u^*} from being satisfied, we set exclude the UE u^* from further consideration, i.e., we set $\hat{s}_{u^*} = 0$ and $\hat{x}_{u^*,b,l,t} = 0,\ \forall b, l, t$. The vector $\hat{\mathbf{a}}$ is updated based on the current $\hat{\mathbf{x}}$, similarly as in (4.32). The iterations stop when a feasible solution is produced. Lastly, given the set of the links used in the feasible solution, we ensure that each associated UE uses the smallest number of links (beams), while ensuring the rate constraints. The proposed algorithm based on rounding with resource pruning is summarized in Algorithm 2.

After the proposed algorithm is used in Step 2, a number of non-allocated serving beams might still be available at the BSs. Based on the objective function in (4.19), which aims to provide a maximal sum rate, the remaining serving beams are distributed among the UEs with partially satisfied rate requirements. Specifically, we consider the links that correspond to the remaining beams and we sort them out in a descending order in terms of the capacity. Then we start from the links with the highest capacity and allocate them if they do not violate the optimization constraints (C-2.1) - (C-2.8). The overall processing flow of the proposed algorithm is summarized in Fig. 4.3.

4.4.2 Computational complexity

The complexity of solving the linear programs in (4.30), (4.31), and (4.29) scales as $\mathcal{O}(V_1^n)$, $\mathcal{O}(V_2^n)$, and $\mathcal{O}(V_3^n)$, respectively, where n is an algorithm-dependent exponent. The proposed rounding algorithm used after Step 1 and Step 2 always converges because the maximum

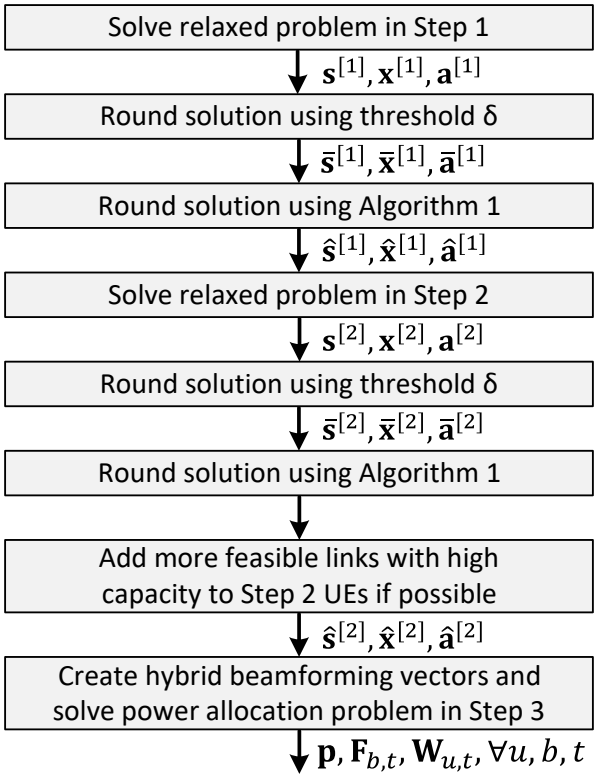


Figure 4.3: Processing flow in the proposed algorithm.

number of iterations is limited to $TN_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}}N_{\mathrm{BS}}$. This number represents the number of link candidates over time, i.e., the number of elements in the vector $\hat{\mathbf{x}}$. In each iteration, the algorithm counts how many constraints does each of the currently used links violate in (C-1.1), (C-1.2), and (C-1.6) - (C-1.8) (or in (C-2.1), (C-2.2), and (C-2.6) - (C-2.8)). Since $\hat{\mathbf{a}}$ is always updated based on the current $\hat{\mathbf{x}}$, the constraints (C-1.3) - (C-1.5) (or (C-2.3) - (C-2.5)) cannot be violated and thus they are not considered during the rounding algorithm. The complexity of counting scales as $\mathcal{O}(TN_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}}N_{\mathrm{BS}}(C_1 + C_2 + C_6 + C_7 + C_8))$, where the numbers of constraints are $C_1 = TN_{\mathrm{UE}}$, $C_2 = TN_{\mathrm{BS}}$, $C_6 = TN_{\mathrm{UE}}(N_{\mathrm{UE}} - 1)$ (worst case), $C_7 = TN_{\mathrm{UE}}N_{\mathrm{RF}}E_{\mathrm{BS}}/2$ (worst case), and $C_8 = N_{\mathrm{UE}}$. The complexity of finding the most conflicted link and checking the feasibility of $\hat{\mathbf{s}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{a}}$ scale as $\mathcal{O}(TN_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}}N_{\mathrm{BS}})$ and $\mathcal{O}(C_1N_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}} + C_2N_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}} + 2C_6 + C_7N_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}} + C_8TN_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}})$, respectively. After the

iterative part of the algorithm, it is ensured that all associated UEs use a minimal number of links. The worst case complexity of this step scales scales as $\mathcal{O}(N_{\mathrm{UE}}(TE_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}})^2)$. The complexity of adding more links with high capacity to the UEs associated in Step 2 scales as $\mathcal{O}(TN_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}}N_{\mathrm{BS}}(C_1N_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}} + C_2N_{\mathrm{UE}}N_{\mathrm{UE}}^{\mathrm{RF}} + 2C_6 + C_7N_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}} + C_8TN_{\mathrm{BS}}N_{\mathrm{UE}}^{\mathrm{RF}}))$ in the worst case.

4.5 Numerical Results

In this section, we numerically evaluate the proposed framework and low-complexity algorithm. Specifically, we first evaluate how interference management, i.e., intelligent beam scheduling and hybrid beamforming, affect the average SINR per associated link. Then we numerically compare the proposed algorithm with existing association schemes in terms of the average number of fully satisfied UEs, partially satisfied UEs, network sum rate, and transmit power usage per BS. Since all scheduled links are known after user association and beam scheduling, the calculation of the metrics in this section is based on the true interference powers and link capacities.

We consider a sub-network of $N_{\mathrm{BS}} = 3$ BSs with the inter-site distance of $d = 200$ m. There are N_{UE} UEs and their positions are generated randomly within the triangle formed by the BSs. We consider an urban micro scenario with realistic mmMAGIC channels between the BSs and UEs. We generate the channels using the Quadriga simulator [JRB19]. Based on the mmMAGIC channel model, which supports the spatial consistency, the multipath components of co-located UEs are highly correlated. Depending on the distance from the BSs, the UEs can experience either line-of-sight or non-line-of-sight scenarios. It is worth noting that in frequency-selective channels, the design of baseband precoders/combiners and power allocation are done per subcarrier if a multi-carrier system is used, which increases the computational complexity. In this chapter, our primary goal is to demonstrate the effectiveness of coordinated user association and beam scheduling, and thus for simplicity,

the numerical simulations are focused on the case when the delay spreads in the channels are negligible (smaller than the sampling rate). We assume the operating frequency $f = 28$ GHz, transmit power $P_{\mathrm{T}} = 30$ dBm, bandwidth BW $= 200$ MHz, noise power spectral density $N_0 = -174$ $\frac{\mathrm{dBm}}{\mathrm{Hz}}$, $N_{\mathrm{UE}}^{\mathrm{A}} = 8$ UE antennas, $N_{\mathrm{UE}}^{\mathrm{RF}} = 2$ UE RF chains, $N_{\mathrm{BS}}^{\mathrm{A}} = 32$ BS antennas, $N_{\mathrm{BS}}^{\mathrm{RF}} = 4$ BS RF chains. The rate requirements are drawn uniformly from $R_u \sim U(R_{\min}, R_{\max})$, $\forall u$, where $R_{\min} = 0.2$ Gbps and $R_{\max} = 1$ Gbps. The fraction ϵ in Step 2 is assumed to be $\epsilon = 0.1$. The proposed algorithm uses the threshold $\delta = 0.1$. The results are averaged across multiple Monte Carlo simulations to generate consistent results.

To demonstrate the benefits of beam scheduling and hybrid beamforming, we evaluate the performance of the proposed framework with and without the interference suppression constraints (C-1.3) - (C-1.7) and (C-2.3) - (C-2.7). The evaluation is done with RF precoders/combiners and hybrid precoders/combiners. The main metric in this study is the difference in the average SINR per link between the frameworks with and without interference management, i.e., the SINR gain per link when interference suppression constraints are used. For this evaluation, we consider a dense sub-network with $N_{\mathrm{UE}} = 50$ UEs, and $N_{\mathrm{BS}}^{\mathrm{RF}} = 12$ RF chains at each BS. For simplicity, we assume $T = 1$ time slot and that power allocation is not performed. The results are presented in Fig. 4.4. They indicate that the designed constraints, which suppress the main-lobe interference, lead to around 5 dB improvement in the average SINR per link when only analog RF precoders and combiners are used. Hybrid precoders and combiners can reduce the excess side-lobe interference and further increase the average SINR per link. As shown in Fig. 4.4, the average SINR also improves with a higher E_{BS}, i.e., when each UE knows multiple BSs. More global knowledge enables the PU to prevent main-lobe interference on a larger scale. In fact, with $E_{\mathrm{BS}} = N_{\mathrm{BS}} = 3$, interference from all BSs in the considered sub-network can be suppressed in a coordinated way.

Further, we solve the optimization problems in the proposed framework using the Gurobi solver in CVX [GB14] and using the proposed low-complexity algorithm to determine the average number of satisfied UEs in the network, network sum rate, and transmit power usage

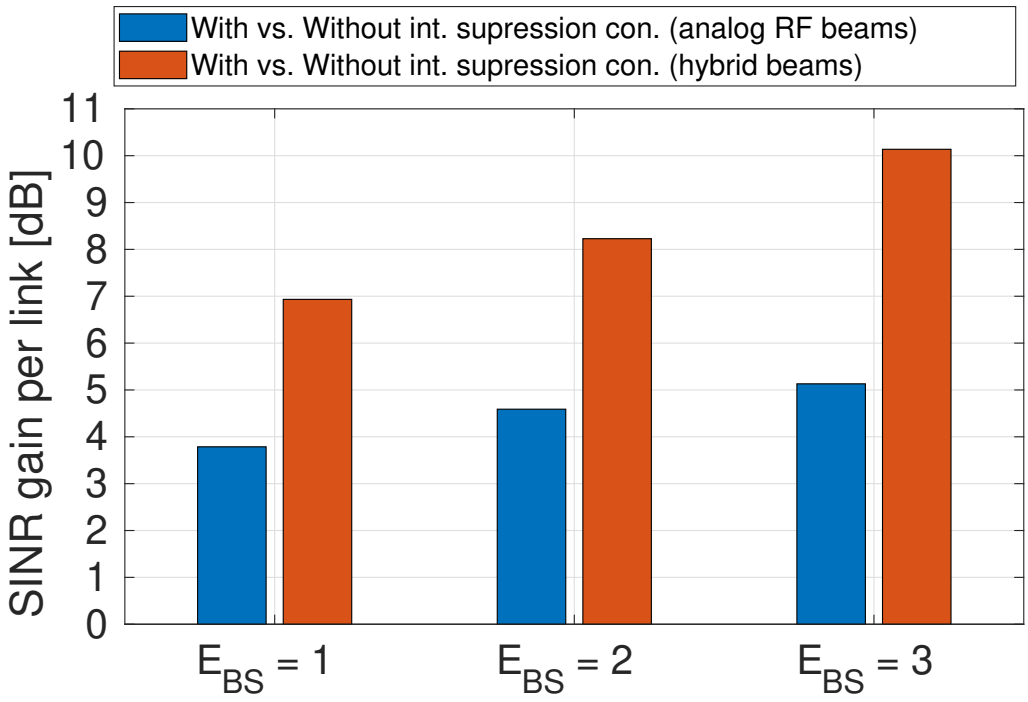


Figure 4.4: Evaluation of the proposed framework with versus without interference suppression constraints using RF and hybrid beamforming vectors.

per BS. Additionally, the CVX solution and the proposed algorithm are compared with a naive greedy user association approach. In the naive greedy approach, each UE selects the BS with the highest signal power, and then the BSs schedule their beams to the corresponding UE candidates with the best channels. For the rest of this section, we assume there are $T = 6$ time slots, $N_{\text{UE}} = 20$ UEs, and $N_{\text{BS}}^{\text{RF}} = 4$ RF chains at each BS.

In Fig. 4.5, we compared the three solutions in terms of the average number of fully satisfied, partially satisfied, and not served UEs. The CVX solution results in the highest number of satisfied UEs. However, obtaining this solution may take prohibitively long time, which makes it impractical. The proposed low-complexity algorithm has sub-optimal performance, but it leads to a significantly higher number of satisfied UEs than the naive greedy approach. Interestingly, when $E_{\text{BS}} < N_{\text{BS}}$, the true number of fully satisfied UEs (blue bar) is lower

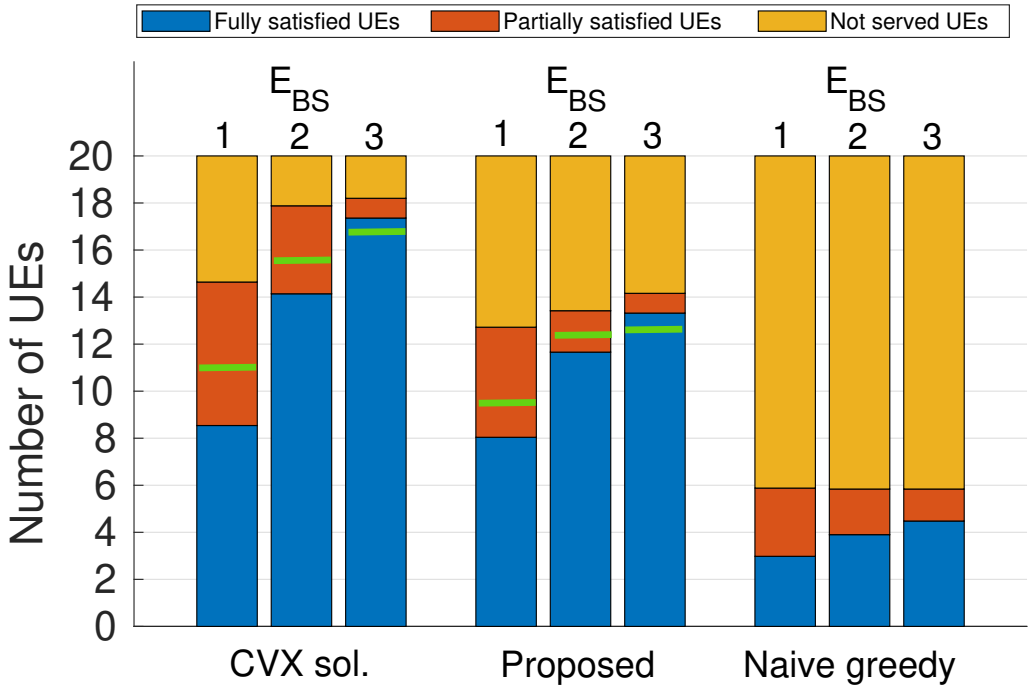


Figure 4.5: Average number of fully satisfied, partially satisfied, and not served UEs. The green line indicates the number of UEs associated in Step 1.

than the number of UEs associated in Step 1 (green line) for both the CVX solution and the proposed algorithm. This happens because the pessimistic estimates of interference power in (4.2) do not account for the main-lobe interference from unknown BSs, which cannot be suppressed. In other words, the estimates are not pessimistic enough, and some the UEs, which are considered to be fully satisfied after Step 1, could end up being only partially satisfied when the true interference and link capacities are calculated. On the other hand, when $E_{\mathrm{BS}} = N_{\mathrm{BS}} = 3$, this problem is not present and the true number of fully satisfied UEs is always greater or equal than the number of UEs associated in Step 1. As the number of fully satisfied UEs increases with a higher E_{BS}, i.e., with a higher level of strict coordination in beam scheduling, the number of partially satisfied UEs decreases because there is a lower number of non-interfering beams that can be scheduled for them. Additionally, after power

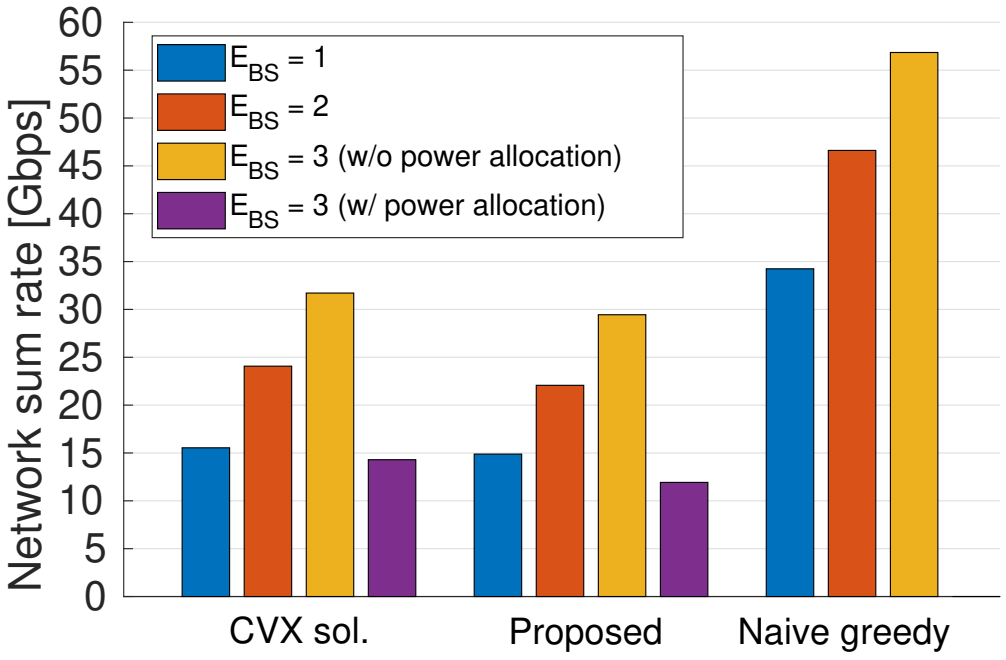


Figure 4.6: Average network sum rate.

allocation, the number of partially satisfied UEs can decrease further because some partially satisfied UEs can be allocated enough power to end up being fully satisfied.

The proposed framework is compared with the naive greedy approach in terms of the network sum rate and transmit power usage in Fig. 4.6 and Fig. 4.7, respectively. The naive greedy approach results in a higher network sum rate because it is focused on scheduling all available BS beams to the UEs with good channels, while using the entire budget of the transmit power at each BS in each time slot. Without power allocation, the network sum rate in the proposed framework increases with a higher E_{BS}, while the percentage of used transmit power decreases at the same time. This comes as a consequences of a higher number of associated UEs and a more coordinated beam scheduling with less interference. Additionally, without power allocation, the majority of transmit power is allocated to the UEs associated in Step 1. On the other hand, with power allocation and $E_{\mathrm{BS}} = N_{\mathrm{BS}} = 3$,

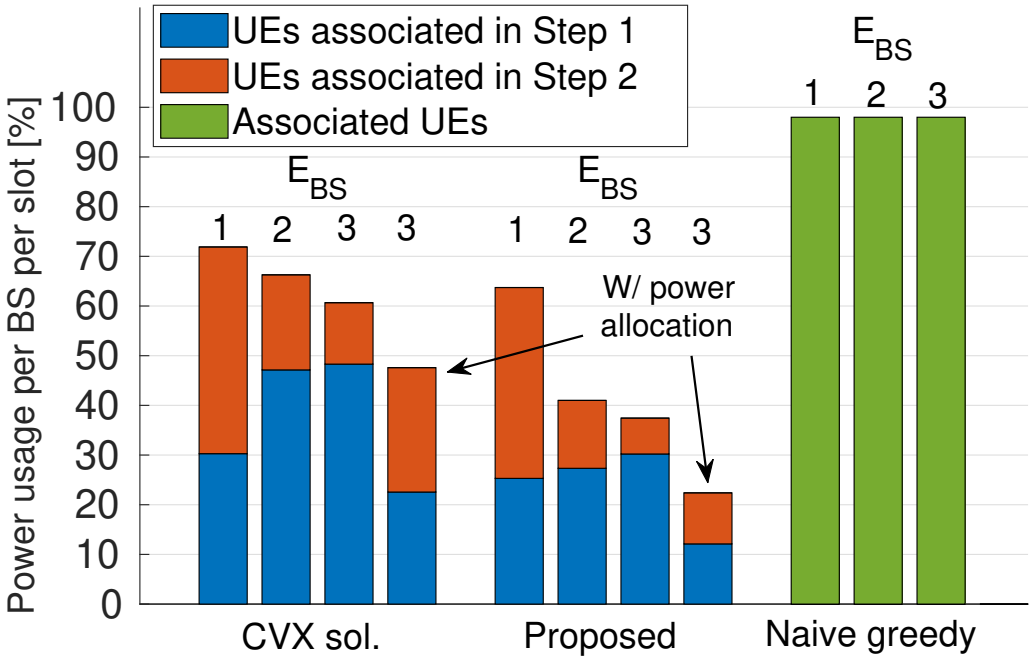
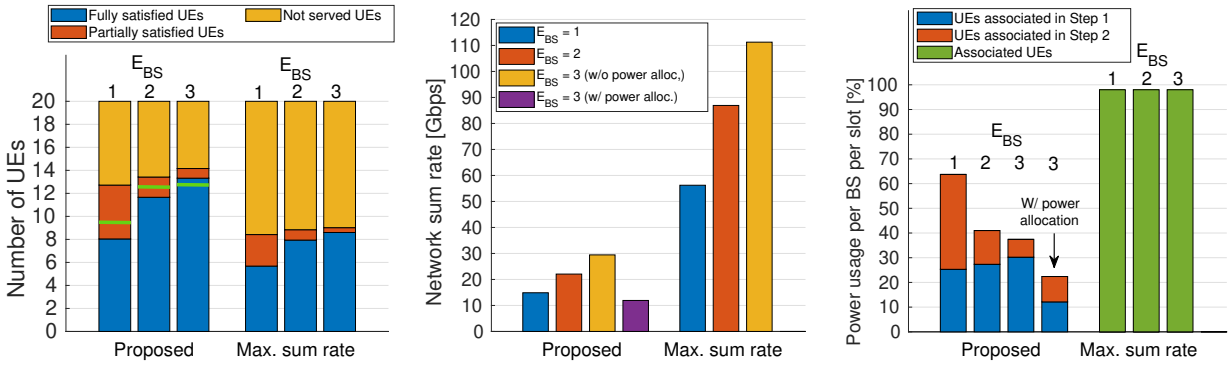


Figure 4.7: Average percentage of used transmit power per BS per time slot.

the network sum rate and the overall percentage of used power are reduced because the fully satisfied UEs are provided with the minimum required power to satisfy their requirements, as explained earlier in Section 4.3.3. Importantly, due to power allocation, the percentage of power allocated to the partially satisfied UEs is increased.

Besides with the naive greedy approach, we compare the proposed framework and low-complexity algorithm with a commonly considered optimization framework that aims to maximize the network sum rate. For this comparison, we use the same metrics as before and we present the results in Fig. 4.8. The maximum sum rate association jointly considers all known links in the sub-network and schedules them such that the overall data rate across all associated UEs is maximized. For this reason, the UEs with very good channels will be favored and they will be allocated all available BS beams and transmit power. Consequently, the rate requirements of many UEs remain unsatisfied. On the other hand, the proposed



(a) (b) (c)

Figure 4.8: Comparison between the proposed low-complexity algorithm and association based on the network sum rate maximization. (a) Average number of fully satisfied, partially satisfied, and not served UEs. (b) Average network sum rate. (c) Average percentage of used transmit power per BS per time slot.

algorithm leads to a lower network sum rate. However, due to different objectives, the proposed algorithm is able to schedule BS beams to a larger number of UEs and satisfy more of them using a lower transmit power per BS per time slot.

4.6 Conclusions

In this chapter, we proposed and mathematically formulated a sequential three-step optimization framework for user association in dense mmW networks. The framework was shown to maximize the number of UEs with satisfied rate requirements, while suppressing the inter- and intra-cell interference through intelligent beam scheduling and design of hybrid beamforming vectors. We also developed a low-complexity algorithm based on relaxation, rounding, and resource pruning to solve the formulated NP-hard optimization problem. Numerical results showed that the interference management in the proposed framework significantly increases the average SINR per associated link. Compared to the existing association approaches which are primarily focused on sum rate maximization, the proposed framework

was shown to lead to a higher number of satisfied UEs while requiring a lower transmit power per BS.

Table 4.1: Notation Summary

N_{BS}, $N_{\mathrm{BS}}^{\mathrm{A}}$, $N_{\mathrm{BS}}^{\mathrm{RF}}$	Number of BSs, BS antenna elements, and BS RF chains
N_{UE}, $N_{\mathrm{UE}}^{\mathrm{A}}$, $N_{\mathrm{UE}}^{\mathrm{RF}}$	Number of UEs, UE antenna elements, and UE RF chains
E_{BS}	Number of BSs that are known to each UE
P_{T}, P_{N}, N_0	Transmit power, noise power, and power spectral density of noise
f, BW, T	Carrier frequency, bandwidth, and number of time slots
d	Inter-site distance
ϵ	Fraction of rate requirements
δ	Threshold for initial rounding
λ_1, λ_2	Scaling parameters in multi-criterion objective functions in Step 1 and Step 2
V_1, V_3, V_3	Number of variables in Step 1 - Step 3
C_1, ..., C_8	Number of constraints (C-1.1), ..., (C-1.8), or equivalently in (C-2.1), ..., (C-2.8)
$\mathcal{U}$, $\mathcal{B}$, $\mathcal{L}$, $\mathcal{T}$	Sets of all UEs, all BSs, all links between pairs of UEs and BSs, and all time slots
u, b, l, t	Indices for UEs, BSs, links, and time slots
R_u, $R_{\min}$, $R_{\max}$	Rate requirement of UE u, minimum, and maximum possible requirement
$\mathbf{H}_{u,b}$	Channel matrix between UE u and BS b
$\mathbf{F}_{\mathrm{RF}}^{b,t}$, $\mathbf{F}_{\mathrm{BB}}^{b,t}$, $\mathbf{F}_{b,t}$	RF, digital, and hybrid precoding matrices at BS b in time slot t
$\mathbf{W}_{\mathrm{RF}}^{u,t}$, $\mathbf{W}_{\mathrm{BB}}^{u,t}$, $\mathbf{W}_{u,t}$	RF, digital, and hybrid combining matrices at UE u in time slot t
$\mathbf{d}_{\mathrm{BS}}^{u,b,l}$, $\mathbf{d}_{\mathrm{UE}}^{u,b,l}$	Candidate RF DFT precoder and combiner on link l between UE u and BS b
$\hat{\gamma}_{u,b,l}$	Pessimistic SINR on the l-th link between the UE u and BS b
$g_{u,b,l,t}$	Effective beamforming gain on link l between the UE u and BS b in time slot t
$\hat{I}_{u,b,l}$, $I_{u,b,l,t}$	Pessimistic and true interference on link l between UE u and BS b in time slot t
$\hat{c}_{u,b,l}$, $c_{u,b,l,t}$	Pessimistic and true capacities on link l between UE u and BS b
$\mathcal{I}_u$	Set of UEs that have at least one of their BS beams in common with UE u
$\mathcal{J}_u$	Set of DFT beam candidates that correspond to more than one link of the UE u
s_u	Var. that indicates if rate requirement of UE u is satisfied
$x_{u,b,l,t}$	Var. that indicates if link l between UE u and BS b is used in time slot t
$a_{u,u',t}$	Var. that indicates if UE u' is alloc. at least one BS beam of UE u in time slot t
$p_{u,b,l,t}$	Power alloc. var. on link l between UE u and BS b in time slot t
$\bar{s}_u$, $\bar{x}_{u,b,l,t}$, $\bar{a}_{u,u',t}$	Values of variables after initial threshold-based rounding
$\hat{s}_u$, $\hat{x}_{u,b,l,t}$, $\hat{a}_{u,u',t}$	Feasible values of variables after proposed algorithm

CHAPTER 5

Conclusions

5.1 Summary of Contributions

In this dissertation, we developed and studied approaches for fast beam training with TTD arrays, fast channel estimation with TTD arrays, and user association and low-interference beam scheduling in mmW networks.

In Chapter 2, we introduced three TTD architectures with baseband delay elements for fast mmW beam training. We developed a non-coherent power-based DSP algorithm for TTD-based beam training that achieves a high angle estimation accuracy using only one wideband training symbol. We studied how the codebook design and beam training performance in analog and hybrid TTD arrays depend on system parameters, including the bandwidth, number of antenna elements, and maximum TTD delay compensation. Numerical simulations revealed the angle estimation accuracy and robustness to hardware impairments of the proposed TTD architectures in UE beam training when benchmarked against the fully digital array. The idea of beam training with frequency-dependent beams was extended to joint beam training between the BS and UE. We analyzed the beam pair misalignment probability and required overhead in TTD-based joint beam training and we compared them to their counterparts in the EBS. The results indicate that the proposed TTD-based beam training can achieve a lower misalignment probability and required overhead than the EBS in realistic mmW channels.

In Chapter 3, we developed a frequecy-domain CS-based DSP algorithm to estimate

sparse mmW channels using an analog TTD array. The proposed algorithm was shown to achieve a good channel estimation accuracy while using sub-band based processing to reduce the required overhead and computational complexity compared to the related benchmark approaches. Due to a particularly high angle estimation accuracy, the algorithm leads to a higher post-estimation spectral efficiency than the benchmark approaches. The proposed algorithm is also evaluated in the presence of time-invariant errors in the analog TTD array and the results indicate a high robustness to hardware impairments. The analysis of the impact of hardware impairments was extended by linearizing the received signal model and deriving the CRLB for the parameters of LoS mmW channels, including the AoD, AoA, and phase of the complex channel gain. In order to improve the estimation accuracy of the proposed algorithm, we proposed a gradient descent based refinement of channel parameters, which was shown to reach the performance bound.

In Chapter 4, we proposed and mathematically formulated a new three-step optimization framework for user association and low-interference beam scheduling in a centralized mmW sub-network. The framework's objetive and constraints were designed to maximize the number of UEs with satisfied rate requirements, while suppressing the directional inter- and intra-cell interference through intelligent beam scheduling and design of hybrid beamforming vectors. Given that the formulated optimization problems are NP-hard, we designed a low-complexity heuristic algorithm based on relaxation, rounding, and resource pruning to obtain sub-optimal solutions in polynomial time. The performance of the proposed framework was evaluated in realistic mmW channels. Numerical results showed that the proposed scheduling and hybrid beamforming significantly increase the average SINR per associated link. We also compared the proposed framework with existing association approaches, which are primarily focused on sum rate maximization. The comparison revealed that the proposed framework leads to a higher number of satisfied UEs while requiring a lower transmit power per BS.

5.2 Future Work

In Chapter 2, the proposed beam training algorithms were focused on finding only the best steering direction in the channel. The future work could extend the algorithms to include the identification of multiple AoAs in the presence of strong directional interferers. Furthermore, practical mmW networks require beam training with planar arrays. For this reason, it would be interesting to design a procedure for joint beam training between the BS and UE with planar arrays. Perhaps the most important question related to TTD-based beam training, which is fundamentally different from beam sweeping, is its compatibility with the 5G New Radio Standard. Solving the compatibility problem would allow practical implementations of TTD-based beam training procedures.

In Chapter 3, we proposed a CS-based algorithm for low-complexity channel estimation with TTD arrays. The future work should study alternative designs of channel estimation algorithms. For example, a sequential estimation of the AoDs and AoAs has the potential to further reduce the DSP complexity. In addition, the potential benefits of the hardware error estimation and calibration on the channel estimation performance could be explored in the future. Similar as with TTD-based beam training, the future work should address the compatibility problem between the TTD-based channel estimation and 5G New Radio Standard.

Chapter 4 considered only data-hungry UEs, but future wireless networks will need to accommodate the UEs with various quality-of-service requirements. For example, a network have have the UEs with high data rate requirements and the UEs with low-latency and high-reliability requirements at the same time. In such networks, the optimization framework for user association and low-interference beam scheduling needs to have more sophisticated objectives and constraints compared to the ones considered in this chapter. In that context, a number of questions related to the framework modeling arise. What should the objectives maximize/minimize when the UEs have different requirements? Should the framework intro-

duce different priorities for the UEs? For example, the UEs with low-latency requirements have a higher priority than the UEs with high data rate requirements. If so, how should the available system resources, e.g., beams, frequencies, and power, be allocated to the UEs under the priority considerations? It will be interesting to study and answer these questions in the future work.

APPENDIX A

Appendix for Chapter 1

A.1 Derivation of Expected Powers in D Directions

As assumed in Sec. 2.4, the channel gains $g_l[k], \forall l, m$, are independent across different clusters and well-separated subcarriers. Thus, with $Q \gg L$ and a negligible approximation error, the channel in (2.2) can be considered as one frequency domain realization of the following channel matrix

$$\mathbf{H} = \mathbf{A}_{\mathrm{R}} \mathbf{\Lambda} \mathbf{A}_{\mathrm{T}}^{\mathrm{H}}. \tag{A.1}$$

The square matrix $\mathbf{\Lambda} \in \mathbb{C}^{Q \times Q}$ has only L non-zero elements that correspond to the cluster gains $g_l, \ \forall l$.

With the codebook design described in Sec. 2.5, the received signal in any probed direction d can be considered as a zero-mean complex Gaussian random variable and expressed as

$$Y_d = \mathbf{f}_d^{\mathrm{H}} \mathbf{H} \mathbf{v} s + \mathbf{f}_d^{\mathrm{H}} \mathbf{n}, \tag{A.2}$$

where $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_{\mathrm{N}}^2 \mathbf{I}_{\mathrm{R}}\right)$ is white Gaussian noise. The realizations of (A.2) are received symbols $Y[m], \ m \in \mathcal{M}_d$. The expected received signal power in direction d is $p_d = \mathbb{E}\left[|Y_d|^2\right] = \mathbb{E}[(\mathbf{f}_d^{\mathrm{H}} \mathbf{H} \mathbf{v} s + \mathbf{f}_d^{\mathrm{H}} \mathbf{n})^{\mathrm{H}} (\mathbf{f}_d^{\mathrm{H}} \mathbf{H} \mathbf{v} s + \mathbf{f}_d^{\mathrm{H}} \mathbf{n})]$. Based on the channel model in (A.1), it can be shown that

$$p_d = M^{-1} \mathbb{E}\left[\mathbf{v}^{\mathrm{H}} \mathbf{A}_{\mathrm{T}} \mathbf{\Lambda}^* \mathbf{A}_{\mathrm{R}}^{\mathrm{H}} \mathbf{f}_d \mathbf{f}_d^{\mathrm{H}} \mathbf{A}_{\mathrm{R}} \mathbf{\Lambda} \mathbf{A}_{\mathrm{T}}^{\mathrm{H}} \mathbf{v}\right] + \mathbb{E}\left[\mathbf{n}^{\mathrm{H}} \mathbf{f}_d \mathbf{f}_d^{\mathrm{H}} \mathbf{n}\right]. \tag{A.3}$$

We apply the trace operator Tr() to (A.3) and exploit its linearity and cyclic property to

obtain

$$p_d = M^{-1}\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{\Lambda}\mathbf{A}_{\mathrm{T}}^{\mathrm{H}}\mathbf{v}\mathbf{v}^{\mathrm{H}}\mathbf{A}_{\mathrm{T}}\mathbf{\Lambda}^{*}\mathbf{A}_{\mathrm{R}}^{\mathrm{H}}\mathbf{f}_d\mathbf{f}_d^{\mathrm{H}}\mathbf{A}_{\mathrm{R}}\right)\right] + N_{\mathrm{R}}\sigma_{\mathrm{N}}^2$$
$$= \mathrm{Tr}\left(\mathbf{G}\mathbf{A}_{\mathrm{R}}^{\mathrm{H}}\mathbf{f}_d\mathbf{f}_d^{\mathrm{H}}\mathbf{A}_{\mathrm{R}}\right) + N_{\mathrm{R}}\sigma_{\mathrm{N}}^2. \tag{A.4}$$

where $\mathbf{G} = \mathbb{E}\left[\mathbf{\Lambda}\mathbf{A}_{\mathrm{T}}^{\mathrm{H}}\mathbf{v}\mathbf{v}^{\mathrm{H}}\mathbf{A}_{\mathrm{T}}\mathbf{\Lambda}^{*}\right]$. Since $\mathbf{\Lambda}$ and $\mathbf{\Lambda}^{*}$ are sparse matrices, $\mathbf{\Lambda}\mathbf{A}_{\mathrm{T}}^{\mathrm{H}}\mathbf{v}\mathbf{v}^{\mathrm{H}}\mathbf{A}_{\mathrm{T}}\mathbf{\Lambda}^{*}$ yields another sparse $Q \times Q$ matrix with L^2 non-zero elements. There are L non-zero elements of the form $|g_l|^2|\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l^{(T)})\mathbf{v}|^2/M,\ \forall l$, on the main diagonal. The $L(L-1)$ off-diagonal elements are cross terms $g_{l_1}g_{l_2}^{*}\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_{l_1}^{(T)})\mathbf{v}\mathbf{v}^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta_{l_2}^{(T)})/M,\ \forall l_1, l_2$. Thus, $\mathbf{G}$ is a diagonal matrix with L non-zero elements $\sigma_l^2|\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l^{(T)})\mathbf{v}|^2/M,\ \forall l$, since $\mathbb{E}\left[g_{l_1}g_{l_2}^{*}\right] = 0,\ \forall l_1 \neq l_2$, and $\mathbb{E}\left[|g_l|^2\right] = \sigma_l^2,\ \forall l$. The product of $\mathbf{G}$ and the matrix of the UE BF gains $\mathbf{A}_{\mathrm{R}}^{\mathrm{H}}\mathbf{f}_d\mathbf{f}_d^{\mathrm{H}}\mathbf{A}_{\mathrm{R}}$ is a $Q \times Q$ matrix whose diagonal elements are equal to $|\mathbf{f}_d^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\xi_q)|^2[\mathbf{G}]_{q,q}$, so (A.4) becomes

$$p_d = \mathbf{b}_d^T\mathbf{g} + N_{\mathrm{R}}\sigma_{\mathrm{N}}^2 \tag{A.5}$$

where $\mathbf{b}_d = \left[|\mathbf{f}_d^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\xi_1)|^2, |\mathbf{f}_d^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\xi_2)|^2, ..., |\mathbf{f}_d^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\xi_Q)|^2\right]^T$ and $\mathbf{g} = \mathrm{diag}(\mathbf{G})$. By vectorizing the result in (A.5), we obtain

$$\mathbf{p} = \mathbf{B}\mathbf{g} + N_{\mathrm{R}}\sigma_{\mathrm{N}}^2\mathbf{1}, \tag{A.6}$$

where $\mathbf{p} = [p_1, p_2, ..., p_D]^T$ and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_D]^T$. Since the BS provides a large BF gain with the fixed precoder $\mathbf{v}$, we can assume that receiver array sees only one spatially filtered dominant cluster, e.g., the first one. Consequently, there is only one non-zero element in $\mathbf{g}$ equal to $|\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_1^{(\mathrm{T})})\mathbf{v}|^2\sigma_1^2/M$.

APPENDIX B

Appendix for Chapter 3

B.1 Proof of Proposition 1

Let $A = \sum_u s_u^{[1]}$ and $B = -\sum_{u,b,l,t} x_{u,b,l,t}^{[1]}$ for notation brevity. The set of all possible values for A and B includes the optimal trade-off curve that is a piece-wise linear (PWL) function, as illustrated in Fig. B.1(a). It is possible to find a hyperplane defined by $[\lambda_1, 1]$ which touches the optimal trade-off curve at the point where A is maximized and B is maximal (the red point in the figure). This is achieved by choosing λ_1 from the range $(0, \frac{1}{K_1})$, where $\frac{1}{K_1}$ is the lowest absolute value of non-zero slopes in the PWL trade-off curve. Note that $\frac{1}{K_1}$ is a theoretical lower bound and it might not exist in the actual optimal trade-off curve. Therefore, $\lambda_1 = \frac{1}{K_1+1} \in (0, \frac{1}{K_1})$ guarantees that the number of fully satisfied UEs is maximized using a minimal number of serving beams.

The constant K_1 can be interpreted as the highest change in price (number of serving beams) that the objective function in Step 1 is ready to pay so that the number of fully satisfied UEs is increased by 1. In the best case, $N-1$ users can be fully satisfied using $N-1$ serving beams. The highest change in price occurs if N UEs require the maximum number of serving beams to be satisfied, as illustrated in Fig. B.2(a). When $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}} \leq TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$, the constant K_1 is given as follows

$$K_1 = \max_{N \in [1, N_{\text{UE}}]} NTN_{\text{UE}}^{\text{RF}} - N + 1 = N_{\text{UE}}(TN_{\text{UE}}^{\text{RF}} - 1) + 1. \tag{B.1}$$

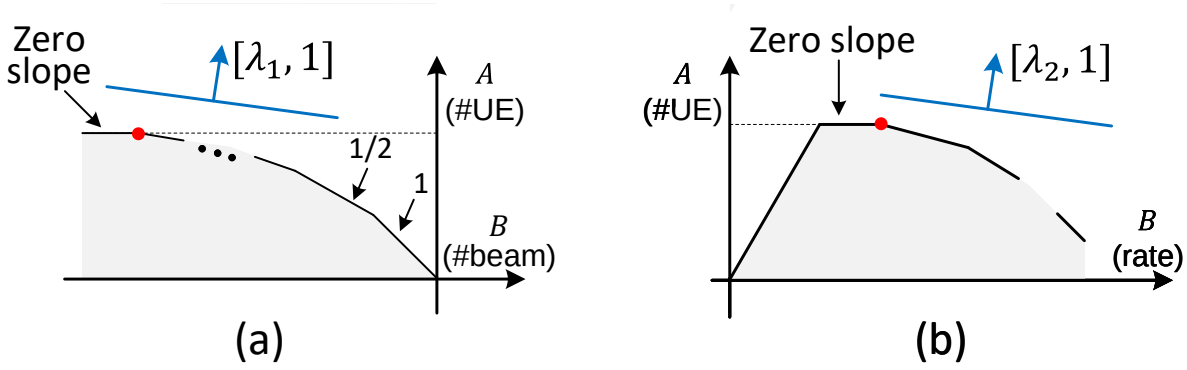


Figure B.1: PWL optimal trade-off curve in (a) Step 1, and in (b) Step 2.

On the other hand, when $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}} > TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$, K_1 is

$$K_1 = \max_{N \in [1, N_{\text{UE}}]} TN_{\text{BS}}N_{\text{BS}}^{\text{RF}} - N = TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}. \tag{B.2}$$

B.2 Proof of Proposition 2

Similar as in the proof for Proposition 1, it is possible to find a hyperplane defined by $[\lambda_2, 1]$ which touches the optimal PWL trade-off curve at the point where $A = \sum_u s_u^{[2]}$ is maximized and $B = \sum_{u,b,l,t} \hat{c}_{u,b,l} x_{u,b,l,t}^{[2]}$ is maximal, as illustrated in Fig. B.1(b). This is achieved by choosing λ_2 from the range $\left(0, \frac{1}{K_2}\right)$, i.e., $\lambda_2 = \frac{1}{K_2+1}$.

The constant K_2 can be interpreted as the highest change in reward (sum data rate) that the objective function declines to pay so that the number of partially satisfied UEs decreases by 1. Let $c_{\min} = \min_{u,b,l,t} \hat{c}_{u,b,l}(1 - x_{u,b,l,t}^{[1]})$ and $c_{\max} = \max_{u,b,l,t} \hat{c}_{u,b,l}(1 - x_{u,b,l,t}^{[1]})$ be the minimum and maximum non-zero link capacities in Step 2, respectively. Consider the case when each of N partially satisfied UEs is served using a link with the capacity $c_{\min}$. The highest change in reward, but with a negative impact on the number of partially satisfied UEs occurs when $N-1$ UEs are served using the maximum number of links that have the capacity $c_{\max}$, as illustrated in Fig. B.2(b). Thus, when $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}} \leq TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$,

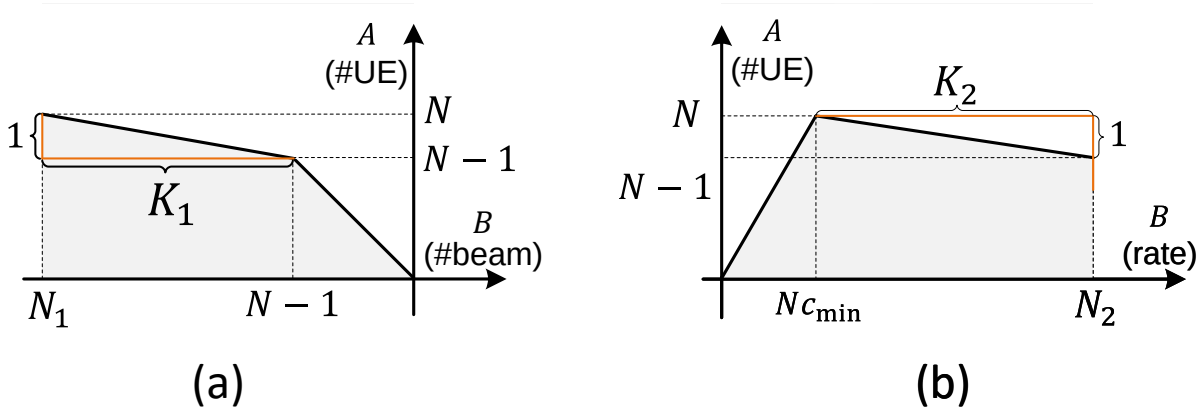


Figure B.2: Illustration of scalarization constants (a) K_1 in Step 1, and (b) K_2 in Step 2. The figures use N_1 and N_2 for notation brevity. If $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}} \leq TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$, $N_1 = NTN_{\text{UE}}^{\text{RF}}$ and $N_2 = (N-1)TN_{\text{UE}}^{\text{RF}}c_{\max}$, and $N_1 = TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$ and $N_2 = TN_{\text{BS}}N_{\text{UE}}^{\text{RF}}c_{\max}$ otherwise.

the constant K_2 is given as follows

$$\begin{aligned} K_2 &= \max_{N\in[1,N_{\text{UE}}]} (N-1)TN_{\text{UE}}^{\text{RF}}c_{\max} - Nc_{\min} \\ &= (N_{\text{UE}}-1)TN_{\text{UE}}^{\text{RF}}c_{\max} - N_{\text{UE}}c_{\min}. \end{aligned} \tag{B.3}$$

Conversely, when $TN_{\text{UE}}N_{\text{UE}}^{\text{RF}} > TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}$, K_2 is given as

$$K_2 = \max_{N\in[1,N_{\text{UE}}]} TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}c_{\max} - Nc_{\min} = TN_{\text{BS}}N_{\text{BS}}^{\text{RF}}c_{\max}. \tag{B.4}$$

B.3 Design of Hybrid Precoders and Combiners

Given the RF precoding matrices $\mathbf{F}_{\text{RF}}^{b,t},\ \forall b,t$, and RF combining matrices $\mathbf{W}_{\text{RF}}^{u,t},\ \forall u,t$, the PU first calculates the baseband precoding matrices $\mathbf{F}_{\text{BB}}^{b,t},\ \forall b,t$ and hybrid precoding matrices $\mathbf{F}_{b,t},\ \forall b,t$, and then baseband combining matrices $\mathbf{W}_{\text{BB}}^{u,t},\ \forall u,t$, and hybrid combining matrices $\mathbf{W}_{u,t},\ \forall u,t$.

Let $\mathcal{U}_{\text{BS},b,t} = \{u_1, u_2, ...\}$ be the set of UEs served by the BS b using the set of links $\mathcal{L}_{\text{BS},b,t} = \{l_1, l_2, ...\}$ in the time slot t. Let $\mathbf{H}_{\text{BS},b,t} \in \mathbb{C}^{|\mathcal{U}_{\text{BS},b,t}|\times|\mathcal{U}_{\text{BS},b,t}|}$ be the effective channel

matrix after the RF precoding from the BS b and the RF combining from the UEs from $\mathcal{U}_{\mathrm{BS},b,t}$. The (i,j)-th element of $\mathbf{H}_{\mathrm{BS},b,t}$ is

$$[\mathbf{H}_{\mathrm{BS},b,t}]_{i,j} = \mathbf{w}_{\mathrm{RF}}^{u_i,b,l_i,t\mathrm{H}} \mathbf{H}_{u_i,b} \mathbf{f}_{\mathrm{RF}}^{u_j,b,l_j,t}, \tag{B.5}$$

where $\mathbf{f}_{\mathrm{RF}}^{u_j,b,l_j,t}$ and $\mathbf{w}_{\mathrm{RF}}^{u_i,b,l_i,t}$ and the columns of $\mathbf{F}_{\mathrm{RF}}^{b,t}$ and $\mathbf{W}_{\mathrm{RF}}^{u,t}$, respectively. Specifically, $\mathbf{f}_{\mathrm{RF}}^{u_j,b,l_j,t}$ is the RF precoder for the UE u_j and its link l_j, while $\mathbf{w}_{\mathrm{RF}}^{u_i,b,l_i,t}$ is the RF combiner for the UE u_i and link l_i. Given $\mathbf{H}_{\mathrm{BS},b,t}$, the PU designs the zero forcing baseband precoding matrix, which removes the excess intra-cell side-lobe interference, as follows

$$\mathbf{F}_{\mathrm{BB}}^{b,t} = \mathbf{H}_{\mathrm{BS},b,t}^{\mathrm{H}} \left(\mathbf{H}_{\mathrm{BS},b,t} \mathbf{H}_{\mathrm{BS},b,t}^{\mathrm{H}} \right)^{-1}. \tag{B.6}$$

The hybrid precoding matrix is $\tilde{\mathbf{F}}_{b,t} = \mathbf{F}_{\mathrm{RF}}^{b,t} \mathbf{F}_{\mathrm{BB}}^{b,t}$. The columns $\tilde{\mathbf{f}}_{u_j,b,l_j,t},\ \forall u_j, l_j$, of $\tilde{\mathbf{F}}_{b,t}$ can be normalized using a diagonal normalization matrix $\mathbf{N}_{\mathrm{BS},b,t}$, whose non-zero element in the position (j,j) is defined as $1/\left\| \tilde{\mathbf{f}}_{u_j,b,l_j,t} \right\|_2$. Thus, the normalized hybrid precoding matrix is defined as

$$\mathbf{F}_{b,t} = \mathbf{F}_{\mathrm{RF}}^{b,t} \mathbf{F}_{\mathrm{BB}}^{b,t} \mathbf{N}_{\mathrm{BS},b,t}. \tag{B.7}$$

Let $\mathcal{B}_{\mathrm{UE},u,t} = \{b_1, b_2, ...\}$ be the set of BSs that serve the UE u using the set of links $\mathcal{L}_{\mathrm{UE},u,t} = \{l_1, l_2, ...\}$ in the time slot t. Let $\mathbf{H}_{\mathrm{UE},u,t} \in \mathbb{C}^{|\mathcal{B}_{\mathrm{UE},u,t}| \times |\mathcal{B}_{\mathrm{UE},u,t}|}$ be the effective channel matrix after the RF combining from the UE u and hybrid precoding from the BSs from $\mathcal{B}_{\mathrm{UE},u,t}$. The (i,j)-th element of $\mathbf{H}_{\mathrm{UE},u,t}$ is

$$[\mathbf{H}_{\mathrm{UE},u,t}]_{i,j} = \mathbf{w}_{\mathrm{RF}}^{u,b_i,l_i,t\mathrm{H}} \mathbf{H}_{u,b_j} \mathbf{f}_{u,b_j,l_j,t}, \tag{B.8}$$

where $\mathbf{f}_{u,b_j,l_j,t}$ is the hybrid precoder used at the BS b_j for the link l_j. Using $\mathbf{H}_{\mathrm{UE},u,t}$, the PU designs the zero forcing baseband combining matrix, which removes the excess side-lobe interference caused by multi-connectivity, as follows

$$\mathbf{W}_{\mathrm{BB}}^{u,t\mathrm{H}} = \left(\mathbf{H}_{\mathrm{UE},u,t}^{\mathrm{H}} \mathbf{H}_{\mathrm{UE},u,t} \right)^{-1} \mathbf{H}_{\mathrm{UE},u,t}^{\mathrm{H}}. \tag{B.9}$$

Alternatively, the minimum mean square error baseband combining matrix can be designed as

$$\mathbf{W}_{\mathrm{BB}}^{u,t\mathrm{H}} = \left(\mathbf{H}_{\mathrm{UE},u,t}^{\mathrm{H}} \mathbf{H}_{\mathrm{UE},u,t} + P_{\mathrm{N}} \mathbf{I}_{N_{\mathrm{UE}}^{\mathrm{RF}}} \right)^{-1} \mathbf{H}_{\mathrm{UE},u,t}^{\mathrm{H}}. \tag{B.10}$$

The hybrid combining matrix is $\tilde{\mathbf{W}}_{u,t} = \mathbf{W}_{\mathrm{RF}}^{u,t}\mathbf{W}_{\mathrm{BB}}^{u,t}$. Similar as the hybrid precoding vectors, the columns of $\tilde{\mathbf{W}}_{u,t}$ can be normalized using a diagonal normalization matrix $\mathbf{N}_{\mathrm{UE},u,t}$, whose non-zero element at the position (i,i) is $1/\left\|\tilde{\mathbf{w}}_{u,b_i,l_i,t}\right\|_2$. Therefore, the normalized combining matrix is defined as follows

$$\mathbf{W}_{u,t} = \mathbf{W}_{\mathrm{RF}}^{u,t}\mathbf{W}_{\mathrm{BB}}^{u,t}\mathbf{N}_{\mathrm{UE},u,t}. \tag{B.11}$$

The worst case complexities of calculating the baseband precoders $\mathbf{F}_{\mathrm{BB}}^{b,t},\ \forall b,t$, and baseband combiners $\mathbf{W}_{\mathrm{BB}}^{u,t},\ \forall u,t$, scale as $\mathcal{O}(TN_{\mathrm{BS}}(N_{\mathrm{BS}}^{\mathrm{RF3}}+2N_{\mathrm{BS}}^{\mathrm{RF2}}))$ and $\mathcal{O}(TN_{\mathrm{UE}}(N_{\mathrm{UE}}^{\mathrm{RF3}}+2N_{\mathrm{UE}}^{\mathrm{RF2}}))$, respectively.

REFERENCES

[3GP18] 3GPP. "E-UTRA and NR - Multi-connectivity, Stage-1, TS 37.340." December 2018.

[3GP19a] 3GPP. "NR - User Equipment (UE) procedures in Idle mode and RRC Inactive state, TS 38.304." September 2019.

[3GP19b] 3GPP. "Study on channel model for frequencies from 0.5 to 100 GHz, TR 38.901." June 2019.

[ABC14] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, **32**(6):1065–1082, June 2014.

[ABK17] Jeffrey G. Andrews, Tianyang Bai, Mandar N. Kulkarni, Ahmed Alkhateeb, Abhishek K. Gupta, and Robert W. Heath. "Modeling and Analyzing Millimeter Wave Cellular Systems." *IEEE Transactions on Communications*, **65**(1):403–430, 2017.

[AEL14] A. Alkhateeb, O. El Ayach, G. Leus, and R. 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.

[AHC22] Suyoung Ahn, Joonpyo Hong, Yunhee Cho, Jehyeon Na, and Jeongho Kwak. "Sequential Beam, User, and Power Allocation for Interference Management in 5G mmWave Networks." In *2022 International Conference on Information Networking (ICOIN)*, pp. 429–434, 2022.

[ALS14] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip. "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation." *IEEE Journal on Selected Areas in Communications*, **32**(6):1164–1179, 2014.

[AV19] Alireza Alizadeh and Mai Vu. "Load Balancing User Association in Millimeter Wave MIMO Networks." *IEEE Transactions on Wireless Communications*, **18**(6):2932–2945, 2019.

[Baj19] M. Bajor et al. "A Flexible Phased-Array Architecture for Reception and Rapid Direction-of-Arrival Finding Utilizing Pseudo-Random Antenna Weight Modulation and Compressive Sampling." *IEEE J. Solid-State Circuits*, **54**(5):1315–1328, May 2019.

[BAN14] Djamal E. Berraki, Simon M. D. Armour, and Andrew R. Nix. "Application of compressive sensing in sparse spatial channel recovery for beamforming in

mmWave outdoor systems." In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 887–892, 2014.

[BC22] V. Boljanovic and D. Cabric. "Millimeter-Wave Wideband Channel Estimation Using Analog True-Time-Delay Array Under Hardware Impairments." *Journal of Signal Processing Systems*, 2022.

[BHM16] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi. "Initial Access in Millimeter Wave Cellular Systems." *IEEE Transactions on Wireless Communications*, **15**(12):7926–7940, 2016.

[BHR14] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, and S. S. Panwar. "Directional cell search for millimeter wave cellular systems." In *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 120–124, 2014.

[BIB15] Robert Baldemair, Tim Irnich, Kumar Balachandran, Erik Dahlman, Gunnar Mildh, Yngve Selén, Stefan Parkvall, Michael Meyer, and Afif Osseiran. "Ultra-dense networks in millimeter-wave frequencies." *IEEE Communications Magazine*, **53**(1):202–208, 2015.

[BSC22a] V. Boljanovic, S. Sarkar, and D. Cabric. "Millimeter-wave user association and low-interference beam scheduling (Invited paper)." In *To Appear in 6th ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, New York, NY, USA, 2022. Association for Computing Machinery.

[BSC22b] Veljko Boljanovic, Shamik Sarkar, and Danijela Cabric. "Millimeter-wave user association and low-interference beam scheduling." In *Accepted for presentation at the 6th ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, New York, NY, USA, 2022. Association for Computing Machinery.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BYG20] V. Boljanovic, H. Yan, E. Ghaderi, D. Heo, S. Gupta, and D. Cabric. "Design of Millimeter-Wave Single-Shot Beam Training for True-Time-Delay Array." In *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2020.

[BYL21] V. Boljanovic, H. Yan, C.-C. Lin, S. Mohapatra, D. Heo, S. Gupta, and D. Cabric. "Fast Beam Training With True-Time-Delay Arrays in Wideband Millimeter-Wave Systems." *IEEE Transactions on Circuits and Systems I: Regular Papers*, **68**(4):1727–1739, 2021.

[CH10] T. Chu and H. Hashemi. "A true time-delay-based bandpass multi-beam array at mm-waves supporting instantaneously wide bandwidths." In *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, pp. 38–39, 2010.

[CM21] Ashok Kumar Reddy Chavva and Neelesh B. Mehta. "Millimeter-Wave Beam Selection in Time-Varying Channels With User Orientation Changes." *IEEE Transactions on Wireless Communications*, **20**(11):6987–7000, 2021.

[CSC18] M. Cho, I. Song, and J. D. Cressler. "A True Time Delay-based SiGe Bidirectional T/R Chipset for Large-Scale Wideband Timed Array Antennas." In *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 272–275, 2018.

[CZW18] Hongyun Chu, Le Zheng, and Xiaodong Wang. "Semi-Blind Millimeter-Wave Channel Estimation Using Atomic Norm Minimization." *IEEE Communications Letters*, **22**(12):2535–2538, 2018.

[CZW19] Hongyun Chu, Le Zheng, and Xiaodong Wang. "Super-Resolution mmWave Channel Estimation for Generalized Spatial Modulation Systems." *IEEE Journal of Selected Topics in Signal Processing*, **13**(6):1336–1347, 2019.

[DKS14] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb. "Initial beamforming for mmWave communications." In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 1926–1930, 2014.

[DPW17] D. De Donno, J. Palacios, and J. Widmer. "Millimeter-Wave Beam Training Acceleration Through Low-Complexity Hybrid Transceivers." *IEEE Transactions on Wireless Communications*, **16**(6):3646–3660, 2017.

[DTS18] Junquan Deng, Olav Tirkkonen, and Christoph Studer. "MmWave channel estimation via atomic norm minimization for multi-user hybrid precoding." In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2018.

[FCC16] FCC. "FACT SHEET: SPECTRUM FRONTIERS PROPOSAL TO IDENTIFY, OPEN UP VAST AMOUNTS OF NEW HIGH-BAND SPECTRUM FOR NEXT GENERATION (5G) WIRELESS BROADBAND.", July 2016.

[GB14] Michael Grant and Stephen Boyd. "CVX: Matlab Software for Disciplined Convex Programming, version 2.1." http://cvxr.com/cvx, March 2014.

[GG21] Erfan Ghaderi and Subhanshu Gupta. "A Four-Element 500-MHz 40-mW 6-bit ADC-Enabled Time-Domain Spatial Signal Processor." *IEEE Journal of Solid-State Circuits*, **56**(6):1784–1794, 2021.

[GHD16] Zhen Gao, Chen Hu, Linglong Dai, and Zhaocheng Wang. "Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding Over Frequency-Selective Fading Channels." *IEEE Communications Letters*, **20**(6):1259–1262, 2016.

[GPB20] E. Ghaderi, C. Puglisi, S. Bansal, and S. Gupta. "10.8 A 4-Element 500MHz-modulated-BW 40mW 6b 1GS/s Analog-Time-to-Digital-Converter-Enabled Spatial Signal Processor in 65nm CMOS." In *2020 IEEE International Solid-State Circuits Conference (ISSCC)*, February 2020.

[GSR19] Erfan Ghaderi, Ajith Sivadhasan Ramani, Arya A. Rahimi, Deukhyoun Heo, Sudip Shekhar, and Subhanshu Gupta. "An Integrated Discrete-Time Delay-Compensating Technique for Large-Array Beamformers." *IEEE Transactions on Circuits and Systems I: Regular Papers*, **66**(9):3296–3306, 2019.

[GTC14] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie. "Millimeter-Wave Enhanced Local Area Systems: A High-Data-Rate Approach for Future Wireless Networks." *IEEE Journal on Selected Areas in Communications*, **32**(6):1152–1163, 2014.

[GTM16] Xiaohu Ge, Song Tu, Guoqiang Mao, Cheng-Xiang Wang, and Tao Han. "5G Ultra-Dense Cellular Networks." *IEEE Wireless Communications*, **23**(1):72–79, 2016.

[Gur22] Gurobi Optimization, LLC. "Gurobi Optimizer Reference Manual.", 2022.

[GYS20] Yasaman Ghasempour, Chia-Yi Yeh, Rabi Shrestha, Daniel Mittleman, and Edward Knightly. "Single Shot Single Antenna Path Discovery in THz Networks." In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[HGR16] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed. "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems." *IEEE Journal of Selected Topics in Signal Processing*, **10**(3):436–453, 2016.

[HIX15] Shuangfeng Han, Chih-lin I, Zhikun Xu, and Corbett Rowell. "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G." *IEEE Communications Magazine*, **53**(1):186–194, 2015.

[HPR15] K. Hosoya, N. Prasad, K. Ramachandran, N. Orihashi, S. Kishimoto, S. Rangarajan, and K. Maruhashi. "Multiple Sector ID Capture (MIDC): A Novel Beamforming Technique for 60-GHz Band Multi-Gbps WLAN/PAN Systems." *IEEE Transactions on Antennas and Propagation*, **63**(1):81–96, 2015.

[Hu18] Anzhong Hu. "Beam Grouping Based User Scheduling in Multi-Cell Millimeter-Wave MIMO Systems." *IEEE Access*, **6**:55004–55012, 2018.

[HWN17] Shiwen He, Yongpeng Wu, Derrick Wing Kwan Ng, and Yongming Huang. "Joint Optimization of Analog Beam and User Scheduling for Millimeter Wave Communications." *IEEE Communications Letters*, **21**(12):2638–2641, 2017.

[iee21] "IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications." *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)*, pp. 1–4379, 2021.

[Jan19] S. Jang et al. "A 1-GHz 16-Element Four-Beam True-Time-Delay Digital Beamformer." *IEEE Journal of Solid-State Circuits*, **54**(5):1304–1314, 2019.

[JLJ18] S. Jang, R. Lu, J. Jeong, and M. P. Flynn. "A True Time Delay 16-Element 4-Beam Digital Beamformer." In *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 12–15, June 2018.

[JPY15] C. Jeong, J. Park, and H. Yu. "Random access in millimeter-wave beamforming cellular networks: issues and approaches." *IEEE Communications Magazine*, **53**(1):180–185, 2015.

[JRB14] Stephan Jaeckel, Leszek Raschkowski, Kai Börner, and Lars Thiele. "QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials." *IEEE Transactions on Antennas and Propagation*, **62**(6):3242–3256, 2014.

[JRB19] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein. "QuaDRiGa - Quasi Deterministic Radio Channel Generator, user manual and documentation." *Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.2.0*, 2019.

[JSR20] Christoph Jans, Xiaohang Song, Wolfgang Rave, and Gerhard Fettweis. "Frequency-Selective Analog Beam Probing for Millimeter Wave Communication Systems." In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2020.

[KC21] In-Soo Kim and Junil Choi. "Spatial Wideband Channel Estimation for mmWave Massive MIMO Systems With Hybrid Architectures and Low-Resolution ADCs." *IEEE Transactions on Wireless Communications*, **20**(6):4016–4029, 2021.

[LE21] Jihyun Lee and Eylem Ekici. "User Scheduling and Beam Alignment in mmWave Networks With a Large Number of Mobile Users." *IEEE Transactions on Wireless Communications*, **20**(10):6481–6492, 2021.

[LGL16] Junho Lee, Gye-Tae Gil, and Yong H. Lee. "Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications." *IEEE Transactions on Communications*, **64**(6):2370–2386, 2016.

[LGW19] Anwen Liao, Zhen Gao, Hua Wang, Sheng Chen, Mohamed-Slim Alouini, and Hao Yin. "Closed-Loop Sparse Channel Estimation for Wideband Millimeter-Wave Full-Dimensional MIMO Systems." *IEEE Transactions on Communications*, **67**(12):8329–8345, 2019.

[LLH17] Chunshan Liu, Min Li, Stephen V. Hanly, Iain B. Collings, and Philip Whiting. "Millimeter Wave Beam Alignment: Large Deviations Analysis and Design Insights." *IEEE Journal on Selected Areas in Communications*, **35**(7):1619–1631, 2017.

[LLZ20] Chunshan Liu, Min Li, Lou Zhao, Philip Whiting, Stephen V. Hanly, and Iain B. Collings. "Millimeter-Wave Beam Search With Iterative Deactivation and Beam Shifting." *IEEE Transactions on Wireless Communications*, **19**(8):5117–5131, 2020.

[LPG21] Chung-Ching Lin, Chase Puglisi, Erfan Ghaderi, Soumen Mohapatra, Deukhyoun Heo, Subhanshu Gupta, Han Yan, Veljko Boljanovic, and Danijela Cabric. "A 4-Element 800MHz-BW 29mW True-Time-Delay Spatial Signal Processor Enabling Fast Beam-Training with Data Communications." In *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC)*, pp. 287–290, 2021.

[LWC16] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo. "User Association in 5G Networks: A Survey and an Outlook." *IEEE Communications Surveys & Tutorials*, **18**(2):1018–1044, 2016.

[LY21] Rui Liu and Guanding Yu. "User Association for Millimeter-Wave Ultra-Reliable Low-Latency Communications." *IEEE Wireless Communications Letters*, **10**(2):315–319, 2021.

[MH17] Nitin Jonathan Myers and Robert W. Heath. "A compressive channel estimation technique robust to synchronization impairments." In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2017.

[MOS22] MOSEK ApS. "The MOSEK optimization toolbox for MATLAB manual. Version 9.0.", 2022.

[MRM16] Zhinus Marzi, Dinesh Ramasamy, and Upamanyu Madhow. "Compressive Channel Estimation and Tracking for Large Arrays in mm-Wave Picocells." *IEEE Journal of Selected Topics in Signal Processing*, **10**(3):514–527, 2016.

[MSH18] Jianhua Mo, Philip Schniter, and Robert W. Heath. "Channel Estimation in Broadband Millimeter Wave MIMO Systems With Few-Bit ADCs." *IEEE Transactions on Signal Processing*, **66**(5):1141–1154, 2018.

[NZL17] Song Noh, Michael D. Zoltowski, and David J. Love. "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment." *IEEE Transactions on Wireless Communications*, **16**(9):5689–5701, 2017.

[PGW19] Joan Palacios, Nuria González-Prelcic, and Joerg Widmer. "Managing Hardware Impairments in Hybrid Millimeter Wave MIMO Systems: A Dictionary Learning-Based Approach." In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 168–172, 2019.

[PT12] A. S. Y. Poon and M. Taghivand. "Supporting and Enabling Circuits for Antenna Arrays in Wireless Communications." *Proceedings of the IEEE*, **100**(7):2207–2218, 2012.

[PWX21] Prosanta Paul, Hongyi Wu, and ChunSheng Xin. "BOOST: A User Association and Scheduling Framework for Beamforming mmWave Networks." *IEEE Transactions on Mobile Computing*, **20**(10):2924–2935, 2021.

[QLS22] Nan Qu, Bingbing Li, Rubayet Shafin, Mingqian Liu, Fengkui Gong, and Lingjia Liu. "Angle-based Downlink Beam Selection and User Scheduling for Massive MIMO Systems." *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[RGB13] Theodore S. Rappaport, Felix Gutierrez, Eshar Ben-Dor, James N. Murdock, Yijun Qiao, and Jonathan I. Tamir. "Broadband Millimeter-Wave Propagation Measurements and Models Using Adaptive-Beam Antennas for Outdoor Urban Cellular Communications." *IEEE Transactions on Antennas and Propagation*, **61**(4):1850–1859, 2013.

[RGV18] Javier Rodríguez-Fernández, Nuria González-Prelcic, Kiran Venugopal, and Robert W. Heath. "Frequency-Domain Compressive Channel Estimation for Frequency-Selective Hybrid Millimeter Wave MIMO Systems." *IEEE Trans. Wireless Commun.*, **17**(5):2946–2960, 2018.

[RJ20] Wolfgang Rave and Christoph Jans. "On the Mapping between Steering Direction and Frequency of a Uniform Linear Array with Fixed True Time Delays." In *WSA 2020; 24th International ITG Workshop on Smart Antennas*, pp. 1–6, 2020.

[RMS15] Theodore S. Rappaport, George R. MacCartney, Mathew K. Samimi, and Shu Sun. "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design." *IEEE Transactions on Communications*, **63**(9):3029–3056, 2015.

[RMZ17] Maryam Eslami Rasekh, Zhinus Marzi, Yanzi Zhu, Upamanyu Madhow, and Haitao Zheng. "Noncoherent MmWave Path Tracking." In *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*, HotMobile '17, p. 13–18, New York, NY, USA, 2017. Association for Computing Machinery.

[RTV05] Cornelis Roos, Tamas Terlaky, and Jean-Philiipe Vial. *Interior Point Methods for Linear Optimization*. Springer New York, NY, 2005.

[RTY16] Ruth Rotman, Moshe Tur, and Lior Yaron. "True Time Delay in Phased Arrays." *Proceedings of the IEEE*, **104**(3):504–518, 2016.

[RVM12] Dinesh Ramasamy, Sriram Venkateswaran, and Upamanyu Madhow. "Compressive adaptation of large steerable arrays." In *2012 Information Theory and Applications Workshop*, pp. 234–239, 2012.

[RXM17] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang. "Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks—With a Focus on Propagation Models." *IEEE Transactions on Antennas and Propagation*, **65**(12):6213–6230, 2017.

[SCW22] Ziyuan Sha, Siyu Chen, and Zhaocheng Wang. "Near Interference-Free Space-Time User Scheduling for MmWave Cellular Network." *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[SPL22] Bodhisatwa Sadhu, Arun Paidimarri, Wooram Lee, Mark Yeck, Caglar Ozdag, Yujiro Tojo, Jean-Olivier Plouchart, Xiaoxiong Gu, Yusuke Uemichi, Sudipto Chakraborty, Yo Yamaguchi, Ning Guan, and Alberto Valdes-Garcia. "A 24-to-30GHz 256-Element Dual-Polarized 5G Phased Array with Fast Beam-Switching Support for ¿30,000 Beams." In *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 65, pp. 436–438, 2022.

[SPX19] Jia Shi, Haris Pervaiz, Pei Xiao, Wei Liang, Zan Li, and Zhiguo Ding. "Resource Management in Future Millimeter Wave Small-Cell Networks: Joint PHY-MAC Layer Design." *IEEE Access*, **7**:76910–76919, 2019.

[SQL19] Xuyao Sun, Chenhao Qi, and Geoffrey Ye Li. "Beam Training and Allocation for Multiuser Millimeter Wave Massive MIMO Systems." *IEEE Transactions on Wireless Communications*, **18**(2):1041–1053, 2019.

[SR17] Shu Sun and Theodore S. Rappaport. "Millimeter Wave MIMO channel estimation based on adaptive compressed sensing." In *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 47–53, 2017.

[Str09] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, MA, fourth edition, 2009.

[SW19] Ziyuan Sha and Zhaocheng Wang. "Least Pair-Wise Collision Beam Schedule for mmWave Inter-Cell Interference Suppression." *IEEE Transactions on Wireless Communications*, **18**(9):4436–4449, 2019.

[SX21] Yaxin Song and Shaoyi Xu. "Beam Management Based Multi-cell Interference Suppression for Millimeter Wave Communications." In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1–5, 2021.

[TAR21] Ehsan Moeen Taghavi, Alireza Alizadeh, Nandana Rajatheva, Mai Vu, and Matti Latva-aho. "User Association in Millimeter Wave Cellular Networks with Intelligent Reflecting Surfaces." In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1–6, 2021.

[TZW18] Yingming Tsai, Le Zheng, and Xiaodong Wang. "Millimeter-Wave Beamformed Full-Dimensional MIMO Channel Estimation Based on Atomic Norm Minimization." *IEEE Transactions on Communications*, **66**(12):6150–6163, 2018.

[VAG17] Kiran Venugopal, Ahmed Alkhateeb, Nuria González Prelcic, and Robert W. Heath. "Channel Estimation for Hybrid Architecture-Based Wideband Millimeter Wave Systems." *IEEE Journal on Selected Areas in Communications*, **35**(9):1996–2009, 2017.

[VAT19] Evangelos Vlachos, George C. Alexandropoulos, and John Thompson. "Wideband MIMO Channel Estimation for Hybrid Beamforming Millimeter Wave Systems via Random Spatial Sampling." *IEEE Journal of Selected Topics in Signal Processing*, **13**(5):1136–1150, 2019.

[VGH19] Kiran Venugopal, Nuria González-Prelcic, and Robert W. Heath. "Optimal Frequency-Flat Precoding for Frequency-Selective Millimeter Wave Channels." *IEEE Transactions on Wireless Communications*, **18**(11):5098–5112, 2019.

[WFW21] Hanyu Wang, Jun Fang, Peilan Wang, Guangrong Yue, and Hongbin Li. "Efficient Beamforming Training and Channel Estimation for Millimeter Wave OFDM Systems." *IEEE Transactions on Wireless Communications*, **20**(5):2805–2819, 2021.

[WGW19] Yue Wu, Yuantao Gu, and Zhaocheng Wang. "Efficient Channel Estimation for mmWave MIMO With Transceiver Hardware Impairments." *IEEE Transactions on Vehicular Technology*, **68**(10):9883–9895, 2019.

[WJG19] Bolei Wang, Mengnan Jian, Feifei Gao, Geoffrey Ye Li, and Hai Lin. "Beam Squint and Channel Estimation for Wideband mmWave Massive MIMO-OFDM Systems." *IEEE Transactions on Signal Processing*, **67**(23):5893–5908, 2019.

[WML16] K. Witrisal, P. Meissner, E. Leitinger, Y. Shen, C. Gustafson, F. Tufvesson, K. Haneda, D. Dardari, A. F. Molisch, A. Conti, and M. Z. Win. "High-Accuracy Localization for Assisted Living: 5G systems will turn multipath channels from foe to friend." *IEEE Signal Processing Magazine*, **33**(2):59–70, 2016.

[WXZ19] Yucheng Wang, Wei Xu, Hua Zhang, and Xiaohu You. "Wideband mmWave Channel Estimation for Hybrid Massive MIMO With Low-Precision ADCs." *IEEE Wireless Communications Letters*, **8**(1):285–288, 2019.

[XR21] Yunchou Xing and Theodore S. Rappaport. "Millimeter Wave and Terahertz Urban Microcell Propagation Measurements and Models." *IEEE Communications Letters*, **25**(12):3755–3759, 2021.

[Yan19] H. Yan et al. "Performance, Power, and Area Design Trade-Offs in Millimeter-Wave Transmitter Beamforming Architectures." *IEEE Circuits and Systems Magazine*, **19**(2):33–58, Secondquarter 2019.

[YBC19] Han Yan, Veljko Boljanovic, and Danijela Cabric. "Wideband Millimeter-Wave Beam Training with True-Time-Delay Array Architecture." In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 1447–1452, 2019.

[YC19] H. Yan and D. Cabric. "Compressive Initial Access and Beamforming Training for Millimeter-Wave Cellular Systems." *IEEE Journal of Selected Topics in Signal Processing*, **13**(5):1151–1166, 2019.

[YXX19] P. Yang, Y. Xiao, M. Xiao, and S. Li. "6G Wireless Communications: Vision and Potential Techniques." *IEEE Network*, **33**(4):70–75, 2019.

[ZLC20] D. Zhang, A. Li, H. Chen, N. Wei, M. Ding, Y. Li, and B. Vucetic. "Beam Allocation for Millimeter-Wave MIMO Tracking Systems." *IEEE Transactions on Vehicular Technology*, **69**(2):1595–1611, 2020.

[ZNY17] Lou Zhao, Derrick Wing Kwan Ng, and Jinhong Yuan. "Multi-User Precoding and Channel Estimation for Hybrid Millimeter Wave Systems." *IEEE Journal on Selected Areas in Communications*, **35**(7):1576–1590, 2017.

[ZRX21] Masoud Zarifneshat, Proteek Roy, and Li Xiao. "Multi-Objective Approach for User Association to Improve Load Balancing and Blockage in Millimeter Wave Cellular Networks." *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[ZSL19] Zhongling Zhao, Jia Shi, Zan Li, Long Yang, Yue Zhao, and Wei Liang. "Matching Theory Assisted Resource Allocation in Millimeter Wave Ultra Dense Small Cell Networks." In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2019.

[ZWT19] Yu Zhang, Yue Wang, Zhi Tian, Geert Leus, and Gong Zhang. "Super-Resolution Spatial Channel Covariance Estimation for Hybrid Precoding in mmWave Massive MIMO." In *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.

[ZXM19] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan. "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies." *IEEE Vehicular Technology Magazine*, **14**(3):28–41, 2019.