Title
Array Architectures and Physical Layer Design for Millimeter-Wave Communications Beyond 5G

Author
Yan, Han

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Array Architectures and Physical Layer Design for Millimeter-Wave Communications Beyond 5G

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical and Computer Engineering

by

Han Yan

2020

ABSTRACT OF THE DISSERTATION

Array Architectures and Physical Layer Design for Millimeter-Wave Communications Beyond 5G

by

Han Yan
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2020
Professor Danijela Cabric, Chair

Ever increasing demands in mobile data rates have resulted in exploration of millimeter-wave (mmW) frequencies for the next generation (5G) wireless networks. Communications at mmW frequencies is presented with two keys challenges. Firstly, high propagation loss requires base stations (BSs) and user equipment (UEs) to use a large number of antennas and narrow beams to close the link with sufficient received signal power. Consequently, communications using narrow beams create a new challenge in channel estimation and link establishment based on fine angular probing. Current mmW system use analog phased arrays that can probe only one angle at the time which results in high latency during link establishment and channel tracking. It is desirable to design low latency beam training by exploring both physical layer designs and array architectures that could replace current 5G approaches and pave the way to the communications for frequency bands in higher mmW band and sub-THz region where larger antenna arrays and communications bandwidth can be exploited. To this end, we propose a novel signal processing techniques exploiting unique properties of mmW channel, and show both theoretically, in simulation and experiments its advantages over conventional approaches. Secondly, we explore different array architecture design and analyze their trade-offs between spectral efficiency and power consumption and area. For comprehensive comparison, we have developed a methodology for optimal design of

system parameters for different array architecture candidates based on the spectral efficiency target, and use these parameters to estimate the array area and power consumption based on the circuits reported in the literature. We show that the hybrid analog and digital architectures have severe scalability concerns in radio frequency signal distribution with increased array size and spatial multiplexing levels, while the fully-digital array architectures have the best performance and power/area trade-offs.

The developed approaches are based on a cross-disciplinary research that combines innovation in model based signal processing, machine learning, and radio hardware. This work is the first to apply compressive sensing (CS), a signal processing tool that exploits sparsity of mmW channel model, to accelerate beam training of mmW cellular system. The algorithm is designed to address practical issues including the requirement of cell discovery and synchronization that involves estimation of angular channel together with carrier frequency offset and timing offsets. We have analyzed the algorithm performance in the 5G compliant simulation and showed that an order of magnitude saving is achieved in initial access latency for the desired channel estimation accuracy. Moreover, we are the first to develop and implement a neural network assisted compressive beam alignment to deal with hardware impairments in mmW radios. We have used 60GHz mmW testbed to perform experiments and show that neural networks approach enhances alignment rate compared to CS. To further accelerate beam training, we proposed a novel frequency selective probing beams using the true-time-delay (TTD) analog array architecture. Our approach utilizes different subcarriers to scan different directions, and achieves a single-shot beam alignment, the fastest approach reported to date. Our comprehensive analysis of different array architectures and exploration of emerging architectures enabled us to develop an order of magnitude faster and energy efficient approaches for initial access and channel estimation in mmW systems.

The dissertation of Han Yan is approved.

Babak Daneshrad

Gregory Pottie

Lieven Vandenberghe

Danijela Cabric, Committee Chair

University of California, Los Angeles

2020

Table of Contents

List of Figures

Acknowledgments

Graduate school has been quite an experience with its fair share of ups and downs. I have learned a lot along the way and the support of many people has made it a rewarding experience.

First and foremost, I would like to thank my family for being supportive and encouraging during graduate studies.

I would like to thank my advisor Prof. Danijela Cabric for her encouragement to work on a variety of difficult problems. Her feedback and criticism has been extremely useful in shaping my research. I would also like to thank the members of my committee Prof. Babak Daneshrad, Prof. Gregory Pottie, and Prof. Lieven Vandenberghe for their helpful comments.

I would like to thank the following collaborators, without whom I would not have been able to complete the my PhD research: Mr. Benjamin Domae, Dr. Curtis Ling, Dr. Sridhar Ramesh, Mr. Timothy Gallagher, and Mr. Veljko Boljanovic.

I am also grateful that I had the opportunity to work on some other interesting research projects with wonderful collaborators Prof. Deuk Hyoun Heo, Dr. Erfan Ghaderi, Mr. Enes Krijestorac, Mr. Kevin Balke, Dr. Riten Gupta, Mr. Shi Bu, Dr. Shailesh Chaudhari, Prof. Subhanshu Gupta, Mr. Samer Hanna, and Prof. Sudhakar Pamarti.

The PhD has not been just an academic adventure because of the company of many friends including some amazing lab mates: Paulo Urriza, Jun Wang, Chun-Hao Liu, Mihir Laghate, Ghaith Hattab, Yen-Chin Wang, Samurdhi Karunaratne and others have been great company in the lab and great explorers outside. I would also like to thank other friends I made at UCLA.

And finally, I would like to thank the National Science Foundation (NSF), Semiconductor Research Corporation (SRC), Defense Advanced Research Projects Agency (DARPA), Qualcomm, and the Graduate Division of UCLA for funding my research.

Curriculum Vitae

2009 – 2013	B.E. in Information and Communication Engineering, Zhejiang University, Hangzhou, P.R.China.
2013 – 2015	M.S. in Electrical Engineering, UCLA.
2013 – 2018	Graduate Student Researcher at UCLA ECE.
2017	Summer intern at Maxlinear, Calrsbad, CA.
2018	Dissertation Year Fellowship, Graduate Division, UCLA.
2019	Summer intern at Telefar Co, Los Angeles, CA.
2019	Qualcomm Innovation Fellowship, Qualcomm.
2020	Distinguished PhD Dissertation Research Award, UCLA ECE

Selected Publications

[1] H. Yan, B. Domae, and D. Cabric, "mmRAPID: machine learning assisted noncoherent compressive millimeter-wave beam alignment," submitted to *2020 the 4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems (mmNets),* Jun. 2020 .

[2] H. Yan, V. Boljanovic, and D. Cabric, "Wideband millimeter-wave beam training with true-time-delay array architecture," (*invited paper*) in *Proc. 2019 53rd Asilomar Conference on Signals, Systems, and Computers,* Nov. 2019 .

[3] H. Yan and D. Cabric, "Compressive initial access and beamforming training for millimeter-wave cellular systems," in *IEEE Journal of Selected Topics in Signal Processing,* vol. 13, no. 5, pp. 1151-1166, Sept. 2019 .

[4] H. Yan and D. Cabric, "DSP linearization for millimeter-wave all-digital receiver array with low-resolution ADCs," (*invited paper*) in *Proc. 2019 IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* Jul. 2019 .

[5] H. Yan, S. Ramesh, T. Gallagher, C. Lin, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," in *IEEE Circuits and Systems Magazine,* vol. 19, no. 2, pp. 33-58, Secondquarter. 2019 .

[6] H. Yan, S. Hanna, K. Balke, R. Gupta, and D. Cabric, "Software defined radio implementation of carrier and timing synchronization for distributed arrays," in *Proc. IEEE Aerospace Conference (AeroConf),* Mar. 2019 .

[7] H. Yan, V. Boljanovic, and D. Cabric, "Tracking sparse mmWave channel: performance analysis under intra-cluster angular spread," (*invited paper*) in *Proc. 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* Jun. 2018 .

[8] H. Yan, S. Chaudhari, and D. Cabric, "Wideband channel tracking for mmWave MIMO system with hybrid beamforming architecture," (*invited paper*) in *Proc. 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),* Dec. 2017 .

[9] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive MIMO systems," in *Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Mar. 2017 .

[10] H. Yan and D. Cabric, "Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems," in *Proc. 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP),* Dec. 2016 .

[11] H. Yan and D. Cabric, "Digitally enhanced inter-modulation distortion compensation in wideband spectrum sensing," in *Proc. 2016 50th Asilomar Conference on Signals, Systems and Computers,* Nov. 2016 .

CHAPTER 1

Introduction

1.1 Motivation

Global mobile data traffic is expected to exceed 160 exabytes per month by the year 2025, 4.2 times higher than that in 2019 [56]. More than 95% of data will be transmitted to and from 6 billion smart gadgets, enabling applications like high definition video streaming, virtual/augmented reality, automatic driving [45], etc. The the fifth generation mobile network new radio (5G-NR) is expected to deliver more than half of the mobile traffic [56] due to its capability to support multi-Gbps peak data rate, ultra-low latency, more reliability and connection density [48].

Although the legacy microwave frequency bands (sub-6GHz) is a key component in 5G-NR, they are extremely crowded and it has become difficult to provide significant capacity gains in these bands [49]. In order to further improve the network throughput for beyond 5G, one of the most promising proposals is to use higher frequencies where abundant spectrum is available. The millimeter-wave (mmW) frequency band[1] is one such example. In fact, 5G-NR is the first generation of cellular system to include mmW band. Operation in 26, 28, and 39 GHz has been standardized and deployed. Meanwhile, mmW communication has also become a key enabler for wireless personal/local area network. The IEEE standards 802.15.3c and 802.11ad have been developed in the unlicensed mmW band, i.e., 57 to 71 GHz [88, 89]. In the near future, it is expected that more mmW licensed band will be open to support the 5G evolution together with the unlicensed bands [8].

[1]Strictly speaking, the mmW band includes frequency 30 to 300 GHz since the associated wavelength is between 1 to 10 millimeters. However, mmW is commonly refer to frequency >24GHz.

In the past few years, there have been extensive testing, channel measurements and prototype developments for mmW band cellular communication by academia, industry and service providers, e.g., New York University [139], Samsung [96,131], Nokia [46], AT&T [79]. These studies have characterized and quantified key design challenges for cellular system design in mmW band. First of all, the channel propagation loss in mmW band is 20dB higher than in sub-6GHz band [137] and mmW signals undergo 20-35dB penetration loss from the human body [131]. To provide adequate service coverage, it is necessary to densify the mmW base stations with cell radius as small as a hundred meters [27,33]. Furthermore, both mmW base station (BS) and user equipment (UE) can equip with large number of antennas for transmission and reception due to the reduce wavelength, which enables forming of narrow beams with high beamforming gains. Due to the properties of mmW propagation and sparsity of multipath channel, beam steering can effectively improve link budget. Besides, to make radio chipsets power and cost efficient, the current 5G-NR mmW transceivers are designed with phased antenna array (PAA) [87] or subarray where multiple PAA and associated radio-frequency (RF)-chain are digitally connected [136]. As a consequence, signal processing techniques [83] and network protocols [159] for mmW mobile network are designed under constraints of PAA architectures.

1.2 Challenges and objectives

The future generations of mmW mobile network will operate in the higher mmW frequency band with more spectrum to meet the ever increasing demands [189,197]. Their realization will demand fundamental rethinking of radio architectures, signal processing and networking protocols to be scalable with the increased antenna number and bandwidths. This dissertation intends to address three major challenges in these areas.

Fast and scalable initial access: In cellular networks, the link between BS and UE is established during initial access (IA) where BSs transmit a synchronization sequence towards UEs so that UEs can identify BSs in their proximity. In conventional sub-6GHz networks, a BS omni-directionally transmits a synchronization sequence by a single antenna. However,

such transmission scheme from BS might not reach UEs in the mmW band due to the high propagation and penetration losses. A crude way of closing the link is to sweep directional beams at the BS and the UE, respectively, in all possible directions in search of a synchronization sequence. However, such exhaustive search based approach has poor scalability with the increased number of antennas and increased number of users. Novel signal processing techniques for mmW IA need to be developed in an overhead and latency efficient manner.

Fast and scalable beam training: Beam training is a key component of mmW cellular system that enables maximum gain of beamformed directional communications between BS and UE [72]. It is also referred as beam alignment as BSs and UEs needs to align narrow beams from antenna array towards angle-of-arrival (AoA) and angle-of-departure (AoD) of dominant propagation path. This physical layer procedure is not only needed as part of the IA, but also required whenever UE mobility occurs which causes the variation of channel propagation directions. The existing beam alignment with PAA uses extensive beam sweeping to estimate AoD and AoA, as it can only probe one steering direction at a time. Such PAA-based beam alignment has its associated overhead and latency linearly increases with array size and number of users. Challenge arise in handling the future evolution of mmW mobile network. Novel signal processing, both in analog or digital domain, is required to accelerate beam alignment for mmW radios.

Power and cost efficient array architecture for spatial multiplexing: In the current 5G-NR mmW, the support of multiple-input multiple-output (MIMO) spatial multiplexing by using PAA is highly limited. More advanced array architecture is certainly needed in the 5G evolution, particularly in the BS side. There are multiple trending candidate, including a hybrid analog and digital array architecture which digitally controls multiple analog PAA to conduct multi-steam communication [17, 23, 96]. Meanwhile, the current array architecture in sub-6GHz, namely fully digital array, is another promising candidate. However, the performance, power, and cost trade-off of these array architecture candidate have been fully understood.

1.3 Contributions and novelties of dissertation

The contribution and novelties of the dissertation is summarized as follows.

Compressive sensing based fast initial access for beam alignment

We start by addressing the scalability issue of mmW initial access with PAA. Instead of using discrete Fourier transform (DFT) based directional sounding beam during this procedure, we propose to use quasi-omni pseudorandom antenna weight vector (AWV). We have developed sparsity exploiting signal processing technique, which facilitates the system to jointly achieve initial cell discovery, synchronization, and fine resolution beam alignment. Cell discovery algorithm has been tailored for this novel sounding beam for non-compromised sensitivity. Further, impact of synchronization errors, i.e., carrier frequency offset (CFO) and sample timing offser (STO), have been incorporated in the design. The proposed beam alignment algorithm that re-processes the asynchronous IA signals can achieve highly accurate angular channel estimation. We have theoretically analyzed the proposed algorithm in terms of miss detection rate under timing synchronization errors, and further derived Cramér-Rao lower bound of angular estimation under frequency offset, considering the 5G-NR compliant IA procedure. To accommodate the ever increasing bandwidth for beam training in standard evolution beyond 5G, we designed the beam squint robust algorithm. We use simulation with realistic mmW channels model to show that the proposed approach is advantageous to existing solution. The proposed algorithm offers orders of magnitude access latency saving compared to existing solution, when the same discovery, post training signal-to-noise ratio (SNR), and overhead performance are targeted. This conclusion holds true in various propagation environments and 3D locations of a mmW pico-cell. Further, our results demonstrated that the proposed beam squint robust algorithm is able to retain unaffected alignment accuracy with increased system bandwidth.

To the best knowledge of the author, this is the first work that applies compressive sensing framework in the mmW initial access for link establishment.

Machine learning assisted noncoherent compressive beam alignment

Although compressive sensing is a promising framework to accelerate beam alignment, it suffers from model mismatch due to practical array hardware impairments. In this topic, we introduced a neural network assisted compressive beam alignment method that uses noncoherent received signal strength measured by a small number of pseudorandom sounding beams to infer the optimal beam steering direction. We have developed a proof-of-concept of the proposed approach with a 60 GHz 36-element PAA and tested in a suburban line-of-sight environment. Our results show that our approach requires only 5 channel probings to achieve post alignment beamforming gain within 1dB margin compared to exhaustive search with 64 channel probings (92.2% overhead reduction). Compared to purely model-based noncoherent compressive beam alignment, our method also has 75% overhead savings.

To the best knowledge of the author, this is the first work that implements and experimentally verifies the efficacy of machine learning assisted compressive noncoherent mmW beam alignment.

True-time-delay analog array based single shot beam alignment

Apart from advanced digital signal processing tailored for mmW PAA radio, novel analog array architecture and analog signal processing can also be leveraged to enhance the beam alignment. We present the theory and signal processing algorithm of an one-shot beam training scheme using true-time-delay (TTD) array. We have derived the TTD array based mmW wideband system model with cyclic-prefix (CP) based orthogonal frequency-division multiplexing (OFDM) waveforms. We show that TTD arrays can apply unique AWV to different subcarriers with a single RF-chain and analog array. Such beam is referred as rainbow beam. Moreover, we analyzed the properties of rainbow beam and derived the relationship between subcarrier frequencies and the corresponding steering directions in the closed form. The required delay tap spacing of TTD arrays and number of subcarriers to simultaneously scan the entire angular space were also analytically determined. Lastly, we

have designed a low-complexity signal processing algorithm that leverage rainbow beam for beam alignment using a single OFDM pilot symbol. The numerical results from realistic mmW system simulation show that the proposed approach can achieve accurate single-shot beam alignment in urban environment.

To the best knowledge of the author, this is the first work that develops the single-shot beam alignment approach using true-time-delay analog arrays.

Comparative study on mmW array architecture for beamforming and MIMO

we provide an overview of the state-of-the-art mmW massive antenna array designs and comparison among three array architectures, namely digital array, partially-connected hybrid array (sub-array), and fully-connected hybrid array. The comparison of performance, power, and area for these three architectures is performed for three representative mmW cellular system downlink use cases, which cover a range of pre-beamforming SNR and multiplexing regimes. This is the first study to comprehensively model and quantitatively analyze all design aspects and criteria including: 1) optimal linear precoder, 2) impact of quantization error in digital-to-analog converter (DAC) and phase shifters, 3) RF signal distribution network, 4) power and area estimation based on state-of-the-art mmW circuits including baseband digital precoding, digital signal distribution network, high-speed DACs, oscillators, mixers, phase shifters, RF signal distribution network, and power amplifiers. Our simulation results show that the fully-digital array architecture is the most power and area efficient compared against optimized designs for sub-array and hybrid array architectures. Our analysis shows that digital array architecture benefits greatly from multi-user multiplexing. The analysis also reveals that sub-array architecture performance is limited by reduced beamforming gain due to array partitioning, while the system bottleneck of the fully-connected hybrid architecture is the excessively complicated and power hungry RF signal distribution network.

To the best knowledge of the author, this is the first work that analyzes RF distribution scalability bottleneck in hybrid arrays.

1.4 Organization of dissertation

The remaining of the dissertation is organized as follows. In Chapter 2, we present initial access and beam training design that utilizes compressive sensing. This chapter is based on my previous publication [183]. In Chapter 3, we discusses the implementation concerns of compressive beam alignment. Further, we present a machine learning assisted proof-of-concept using mmW testbed. This chapter is based my manuscript submitted for publication [186]. In Chapter 4, the fast beam alignment design that leverage the true-time-delay analog array is presented. This chapter is based on my previous publication [180] In Chapter 5, the comparative study of array architecture for MIMO multiplexing is presented.This chapter is based on my previous publication [187]. Chapter 6 concludes the dissertation and highlight open research questions.

CHAPTER 2

Initial Access and Beam Training for Millimeter-Wave Cellular Systems

2.1 Introduction

The mmW communication is a promising technology for the future cellular network including the 5G-NR [21]. Due to abundant spectrum, it is expected that the mmW network will support ultra-fast data rate. As shown in both theory and prototypes, mmW system requires beamforming (BF) with large antenna arrays at both BS and UE to combat severe propagation loss [139].

Significant differences in propagation characteristics and hardware architectures for mmW band compared to microwave band require novel signal processing techniques [83] and physical layer procedures [5].

IA is the fundamental physical layer procedure that allows UE to discover and synchronize with nearby BS before further communication. However, IA for mmW networks brings new challenges and opportunities as compared to IA for sub-6GHz band networks. In mmW system, conventional omni-directional IA with single antenna can not be reliable, and as a result IA needs to leverage transmitter and receiver antenna array to exploit BF gain [30,72]. A key design challenge in mmW IA is the design of sounding beams for reliable discovery. In addition, beam training is required to achieve high BF gain enabled by large arrays and establish communication link. However, beam training now introduces additional access latency and signaling overhead due to repeated channel probing.

2.1.1 Related works

A number of works investigated various sounding beam designs and signal processing algorithms for mmW IA and beam training. Directional beams for IA and beam training are the most popular and extensively investigated in recent literature [15, 18, 30, 71, 72, 102, 103, 105, 114, 191]. Directional initial access (DIA) is first studied in [30] where a Generalized Likelihood Ratio Test is proposed to solve the cell discovery problem under unknown MIMO channel and synchronization parameters. The authors concluded that the directional IA signal improves discovery range as compared to omni-directional IA. The DIA is further investigated in [103] where overhead and access latency are analyzed. Works [102] and [191] study DIA and its access latency in large networks using stochastic geometry. Impact of beam-width of sounding beams in DIA is researched in [18]. The comparison between omni-directional and DIA is also discussed in [114]. IA using out-of-band information, e.g., location, sub-6GHz measurement, are discussed in [71, 72]. The aforementioned works mostly focused on the overhead and latency for the cell discovery, while beam training is either not discussed or assumed to have coarse resolution [103]. It is common that DIA is paired with directional beam training [15, 105] where hierarchical sounding beams are used in multiple stages to achieve fine angular resolution for each user individually. However, such user-specific hierarchical sounding beams introduce prohibitive latency when a BS is connected to large number of UEs.

The alternative approaches for beam training are based on parametric channel estimation [19, 74, 112, 128, 145, 161, 168, 172, 198]. Exploiting the mmW sparse scattering nature, compressive sensing (CS) approaches have been considered to effectively estimate channel parameters based on channel observations obtained via various sounding beams. Works [19,112] proposed a CS-based narrowband BF training with pseudorandom sounding beamformers in the downlink, and [145, 168] extended this approach for a wideband channel. Other related works include channel covariance estimation [128, 161, 172] which requires periodic channel observations, and UE centric uplink training [74, 198]. It is worth nothing that all recent works focus on channel estimation alone while assuming perfect cell discovery and synchro-

nization. The 5G-NR frame structure that supports IA is rarely considered, and further the feasibility of joint initial access and CS-based beam training has not been investigated.

There are also recent works that consider some practical aspects of IA. For example, frequency offset robust algorithms in narrowband mmW beam training are reported in [119, 126, 144, 181]. There are several hardware prototypes that consider a practical approach of using received signal strength (RSS) in CS-based beam training. Channel estimation problem without phase measurement is a challenging problem, which was solved via novel signal processing algorithms based on RSS matching pursuit [142], Hash table [81], and sparse phase retrieval [141]. Note that phase free measurements were associated with a particular testbed, and this constraint does not necessarily apply to mmW systems in general.

Last but not the least, the evolution of future mmW systems will certainly use wider bandwidths and larger array aperture. Therefore, the spatial wideband effect, a.k.a., beam squint, will play a vital role in the system design [41, 170]. Recent works reveal that a beam squint unaware array processing results in a compromised channel estimation performance in both sub-6GHz and mmW systems [92, 143, 169, 171]. However, the mmW beam training that involves both transmitter and receiver array has not been considered. Further, the impact of estimation error due to beam squint in data communication phase under frequency flat beam steering is not fully investigated. Overall, while IA and beam training algorithms have been extensively studied in the literature, there is a lack of understanding about the theoretical limits and performance of signal processing algorithms that jointly achieve cell discovery and beam training using asynchronous IA signal in mmW frequency selective channel.

2.1.2 Contributions

In this work, we propose to use quasi-omni pseudorandom sounding beams and novel signal processing algorithm to jointly achieve initial cell discovery, synchronization, and fine resolution beam training. More specifically, we provide answers to the following questions.

How to use pseudorandom sounding beams for IA? We propose an energy detection algorithm for initial discovery tailored for pseudorandom sounding beams. We derive the optimal

detection threshold, analyze the miss detection probability and the impact of synchronization errors, i.e., CFO and STO.

How to reuse received IA signal for beam training? We propose a novel CS-based beam training algorithm that re-processes the frequency asynchronous IA signals to provide well aligned beam pair. We derive the Cramér–Rao lower bound (CRLB) of asynchronous training in line-of-sight (LOS) channel. We show that proposed algorithm reaches CRLB in LOS and remains effective in non-line-of-sight (NLOS).

What are the benefits of compressive IA? We compare the proposed approach with DIA followed by hierarchical directional beam training. Key performance indicators for both approaches are numerically compared, including discovery rate, post beam training SNR, overhead and access latency. The simulation study based on 5G-NR frame structure and measurement-endorsed 3D 28 GHz channel shows that the proposed approach is advantageous to DIA for UEs across wide range of locations in a pico-cell.

How to design squint robust beam training with increased IA bandwidth for beyond 5G systems? We propose a dictionary adaptation based approach that facilitates the proposed compressive beam training to be robust to the spatial wideband effect. In particular, the non-identical array responses for different subcarriers are incorporated in both on-grid search and off-grid refinement. The enhanced beam training method estimates propagation angle more accurately and provides higher post-training array gain across a wideband frequency range compared to existing compressive sensing based approach under high beam squint regime.

2.1.3 Organizations and notations

The rest of the chapter is organized as follows. We start with a brief introduction of 5G-NR frame structure, IA and beam training in Section 2.2. In Section 2.3, we present the system model and problem statement. Section 2.4 includes the proposed algorithm for cell discovery and timing acquisition followed by associated performance analysis. In Section 2.5 we present the algorithm and analysis for initial beam training under CFO. The squint robust beam

training is presented in Section 2.6. The access latency, overhead, and complexity analysis is included in Section 2.7. The numerical results are presented in Section 2.8. Open research issues are summarized in Section 2.9. Finally, Section 2.10 concludes the chapter.

Notations: Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. The (i, j)-th element of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. Conjugate, transpose, Hermitian transpose, and pseudoinverse are denoted by $(.)^*$, $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$, and $(.)^{\dagger}$ respectively. The inner product is $\langle \mathbf{a}, \mathbf{b} \rangle \triangleq \mathbf{a}^{\mathrm{H}}\mathbf{b}$. The l_2-norm of $\mathbf{h}$ is denoted by $||\mathbf{h}||$. diag$(\mathbf{a})$ aligns vector $\mathbf{a}$ into a diagonal matrix. Kronecker and Hadamard product are denoted as $\otimes$ and $\circ$, respectively. $\Re(x)$ and $\Im(x)$ are the real and imaginary parts of x, respectively. Set $\mathcal{S} = [a, b]$ contains all integers between a and b.

2.2 Preliminaries: initial access and beam training

In this section, we introduce the mmW physical layer initial access procedure in 5G-NR cellular network. We briefly review the frame structure, synchronization sequences, and directional IA scheme as well as beam training. The reader is referred to work [72] for a more detailed survey.

Frame Structure: Figure 2.1 shows the frame structure of 5G-NR. We focus on two functional blocks, namely synchronization signal burst (SSB) and channel state information reference signal (CSI-RS). 5G-NR uses OFDM, and the subcarrier spacing is either 120 kHz or 240 kHz for mmW band. The SSB set is transmitted by a BS with period T_{F}, typically 20 ms. A SSB set consists of up to $M = 64$ burst blocks. In each one of the burst blocks of duration T_{B}, a specific sounding beam pair is used by BS and UE. The CSI-RS block with duration T_{r} is dedicated to specific UE(s) for fine beam training and tracking. CSI-RS can use all frequency resources, i.e., up to B_{tot}, and it has periodicity of T_{R}, an implementation dependent value.

Synchronization Signal: Referring to Figure 2.1, each SSB has 4 OFDM symbols, i.e., primary synchronization signal (PSS), physical broadcast channel (PBCH), and secondary

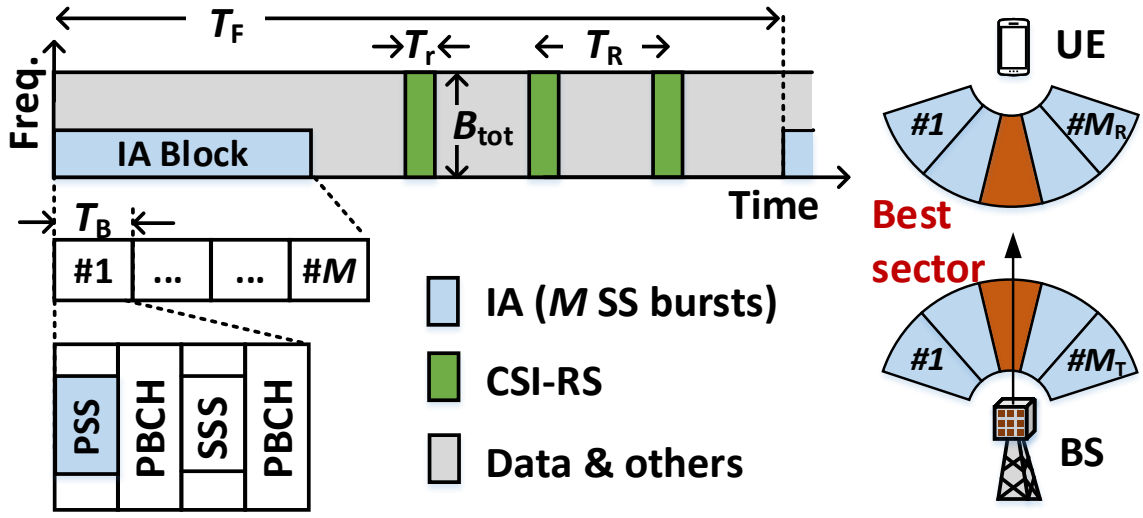


Figure 2.1: The 5G-NR mmW frame structure with emphasis in beam management function and the illustration for directional initial access.

synchronization signal (SSS), followed by another PBCH. PSS is used in cell detection and synchronization, and it is assigned to the middle $P = 128$ subcarriers of the first OFDM symbol. The PSS in 4G-LTE is based on Zadoff-Chu (ZC) sequences due to their perfect cyclic-autocorrelation property and their Fourier duals [101], while in 5G-NR PSS is replaced by Maximum Length Sequences (M-sequences) [7]. There are $N_{\text{PPS}} = 3$ and 336 unique sequences of PSS and SSS, respectively, and these 1008 combination define the cell identifier of BS. PBCH carries control information.

Beamformed Initial Access: The BS periodically transmits IA blocks and such signals are processed by UEs which desire to establish the initial access, reconnect after beam misalignment, and search for additional BSs for potential handover. The sounding beams in SSB are intended to facilitate multi-antenna processing in BS and UE when no a priori channel information is available. Referring to Figure 2.1, BS and UE in the DIA scheme use M_{T} and M_{R} transmitter and receiver beams to cover angular space at both ends. One transmitter and receiver beam pair is used at a time, for all $M = M_{\text{T}}M_{\text{R}}$ SSB.

Beam Training: The purpose of beam training is to identify and report the best beam

pair between BS and UE. The sounding beams in DIA typically have large beam-width and flat response inside angular sectors [160]. Such design covers the angular space of BS and UE within M bursts, but achieves coarse propagation directions estimation [103]. Thus DIA relies on directional beam training to refine angular resolution where BS and UE steer narrow sounding beams within the sectors of interest during CSI-RS periods.

2.3 System model

This section introduces the system model that adopts the 5G-NR frame structure and problem formulation. All important notations are summarized in Table 2.1.

2.3.1 Asynchronous received signal model in initial access

Consider a single cell system with a BS equipped with N_{T} antennas. The BS transmits beamformed IA signal over mmW sparse multipath channel to UEs. We focus on the IA and BF training procedure for a single UE[1]. The UE uses analog array architecture, i.e., phased array, with N_{R} antennas. We assume that a single stream of IA signal is transmitted by the BS regardless of its architecture.

We first consider the received signal model when a UE searches for BS to initialize the connection. In this procedure, UE follows a periodic SSB structure and uses predefined receiver beamformers to capture the signal according to [72]. As illustrated in Figure 2.2, when the signal is present, the received samples, sampled at T_{s}, is denoted as

$$y[n] = \sum_{d=0}^{N_{\mathrm{c}}-1} e^{j(\epsilon_{\mathrm{F}} n + \psi[n])} \mathbf{w}^{\mathrm{H}}[n]\mathbf{H}[d]\mathbf{v}[n-d-\epsilon_{\mathrm{T}}]s[n-d-\epsilon_{\mathrm{T}}] + \mathbf{w}^{\mathrm{H}}[n]\mathbf{z}[n], \quad n \in [0, N_{\mathrm{F}}-1]. \tag{2.1}$$

In the above equation, ϵ_{T} is the unknown integer sample STO within range[2] $0 \leq \epsilon_{\mathrm{T}} \leq$

[1]Since this downlink procedure does not have UE-specific precoding, it is straightforward to extend it to multiple UEs.

[2]We assume coarse timing synchronization is available with 10 µs level accuracy that corresponds to current 4G-LTE. Practically it is achievable via GPS clock or non-standalone mmW network [72].

Table 2.1: Nomenclature

Symbol	Explanations
p, P	Index and total number of subcarriers
m, M	Index and total number of SSB
l, L	Index and total number of multipaths
N_{T}, N_{R}	Number of antenna in BS and UE
$T_{\mathrm{B}}, N_{\mathrm{B}}$	Duration and sample number in each SSB
T_{s}, T_{F}	Sample duration of IA signal and period of SSB set
T_{R},T_{r}	Period and duration of CSI-RS
N_{c}, N_{cp}	Max. excess delay taps and length of CP
$N_{\mathrm{train}}, N_{\mathrm{U}}$	Required CSI-RS and UE number
Δf, ϵ_{F}	CFO in [Hz] and normalized in [rad/sample]
ϵ_{T}	Initial STO in UE (number of sample)
$\mathbf{H}[d]$	MIMO channel at d-th delay sample
$\mathbf{a}_{\mathrm{T}}(\theta)$, $\mathbf{a}_{\mathrm{R}}(\phi)$	Spatial responses of BS and UE
ϕ_l, θ_l, g_l, τ_l	Gain/AoA/AoD/delay of l-th multipath
α_l, β_l	Real and imaginary parts of g_l
$\mathbf{s}, \tilde{\mathbf{s}}$, $\tilde{s}[n]$	Frequency and time domain PSS vector and sequence
$\mathbf{v}_m, \mathbf{w}_m$	RF precoder/combiner of the m-th burst
$z[n]$, $\mathbf{z}_m$, σ_{n}^2	Noise sequence, vector, and power
$P_{\mathrm{FA}}^{\star}$	Target FA prob. in initial discovery
$P_{\mathrm{MD,PT}}$, $P_{\mathrm{MD,NT}}$	MD prob. w/ and w/o perfect timing
$\gamma_{\mathrm{PT}}, \eta_{\mathrm{PT}}$, $\gamma_{\mathrm{NT}}, \eta_{\mathrm{NT}}$	Detection statistic and threshold w/ and w/o perfect timing
$\mathbf{y}_m$	Received OFDM symbols at m-th burst
$\mathbf{d}$, $\mathbf{t}$, $\mathbf{r}$	Vectors with candidates delay/AoA/AoD
G_{D}, G_{T}, G_{R}	Parameter grid in delay/AoA/AoD est.
$\mathbf{Q}(\epsilon_{\mathrm{F}})$, $\mathbf{F}$	Phase error matrix and DFT matrix

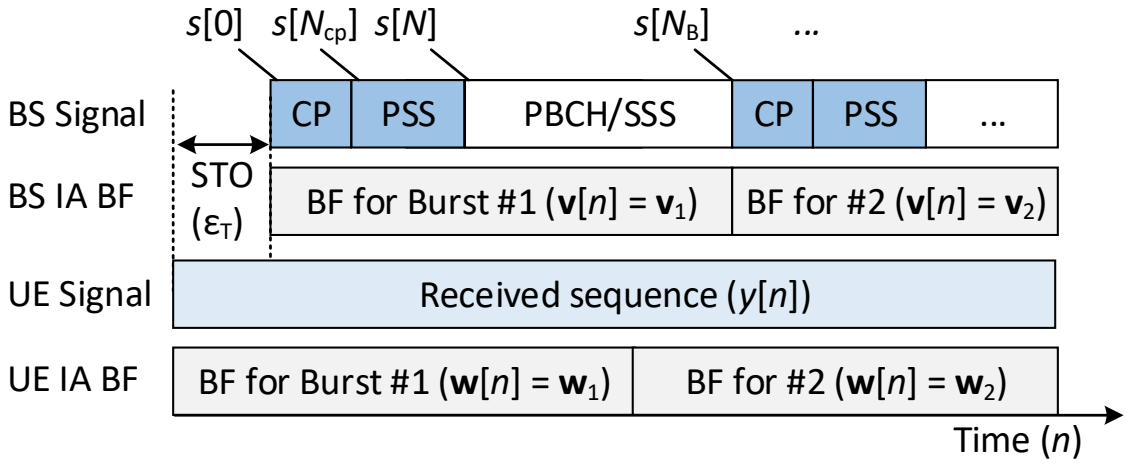


Figure 2.2: Illustration of the received signal model as time sequence.

$\epsilon_{\mathrm{T,max}} \leq N_{\mathrm{B}}$, where $\epsilon_{\mathrm{T,max}}$ is the largest offset known to the system and N_{B} is the number of samples in one SSB, i.e., $N_{\mathrm{B}} = T_{\mathrm{B}}/T_{\mathrm{s}}$. The phase measurement error $\exp[j(\epsilon_{\mathrm{F}} n + \psi[n])]$ comes from two sources. ϵ_{F} is the normalized initial CFO, i.e., $\epsilon_{\mathrm{F}} = 2\pi T_{\mathrm{s}} \Delta f$ where Δf is absolute CFO in Hz between BS and UE. $\psi[n]$ is the phase noise process in the UE receiver. N_{c} is the maximum excessive multipath delay in discrete time, based on which CP $N_{\mathrm{cp}} > N_{\mathrm{c}}$ for OFDM symbols is designed. $s[n]$ is the time domain signals of SSB. Referring to Figure 2.2, we focus on the PSS and treat other symbols as zero [30], i.e.,

$$s[n] = \begin{cases} s_{\mathrm{zc}}[n - (m-1)N_{\mathrm{B}} + P - N_{\mathrm{CP}}], & n \in \mathcal{S}_{\mathrm{CP},m} \\ s_{\mathrm{zc}}[n - (m-1)N_{\mathrm{B}} - N_{\mathrm{CP}}], & n \in \mathcal{S}_{\mathrm{PSS},m} \\ 0, & \text{otherwise} \end{cases},$$

where $\mathcal{S}_{\mathrm{CP},m} \triangleq [(m-1)N_{\mathrm{B}}, (m-1)N_{\mathrm{B}} + N_{\mathrm{CP}} - 1]$, $\mathcal{S}_{\mathrm{PSS},m} \triangleq [(m-1)N_{\mathrm{B}} + N_{\mathrm{CP}}, (m-1)N_{\mathrm{B}} + N - 1]$ are the sets with sample index corresponding to CP and PSS in the m-th burst, respectively. $|s_{\mathrm{zc}}[n]| = 1, n \in [0, P-1]$ is the Fourier dual of a known PSS sequence, and $N = P + N_{\mathrm{cp}}$ is the number of samples in PSS including CP. $\mathbf{z}[n]$ is the Additive White Gaussian noise (AWGN) and $\mathbf{z}[n] \sim \mathcal{CN}(0, \sigma_{\mathrm{n}}^2 \mathbf{I}_{N_{\mathrm{R}}})$. Vectors $\mathbf{v}[n]$ and $\mathbf{w}[n]$ are beamformers used by BS and UE at instance n, respectively, and they are from a predefined set of IA

beam codebook, i.e., $\mathbf{w}[n] \in \mathcal{W} \triangleq \{\mathbf{w}_1, \cdots, \mathbf{w}_M\}$ and $\mathbf{v}[n] \in \mathcal{V} \triangleq \{\mathbf{v}_1, \cdots, \mathbf{v}_M\}$. BS and UE sequentially use respective beamformers for an interval of N_B samples and switch to the next one in $\mathcal{W}$ and $\mathcal{V}$, i.e., $\mathbf{w}[n] = \mathbf{w}_m$, if $\lfloor n/N_\mathrm{B} \rfloor = m$ and $\mathbf{v}[n] = \mathbf{v}_m$, if $\lfloor n/N_\mathrm{B} \rfloor = m$. Beamformer switching is assumed not to introduce latency or phase offset in the transmission and reception. In this work, we focus on the system where each element of $\mathbf{v}_m$ and $\mathbf{w}_m$ is randomly and independently chosen from a set $\mathcal{S}_\mathrm{T} = \left\{\pm 1/\sqrt{N_\mathrm{T}}, \pm j\sqrt{N_\mathrm{T}}\right\}$, and $\mathcal{S}_\mathrm{R} = \left\{\pm 1\sqrt{N_\mathrm{R}}, \pm j\sqrt{N_\mathrm{R}}\right\}$. Such sounding beams require only 4-level phase quantization when steered by phased array and have randomized quasi-omnidirectional beam pattern.

The discrete time MIMO channel at delay d ($d < N_\mathrm{c}$) is denoted as $\mathbf{H}[d] \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$. Following the extended Saleh-Valenzuela (S-V) channel model in [83], we express $\mathbf{H}[d]$ as

$$\mathbf{H}[d] = \frac{1}{\sqrt{N_\mathrm{T} N_\mathrm{R}}} \sum_{l=1}^{L} \sum_{r=1}^{R} [g_{l,r} p_\mathrm{c}(dT_\mathrm{s} - \tau_{l,r}) \cdot \mathbf{a}_\mathrm{R}(\phi_{l,r}^{(\mathrm{az})}, \phi_{l,r}^{(\mathrm{el})}) \mathbf{a}_\mathrm{T}^\mathrm{H}(\theta_{l,r}^{(\mathrm{az})}, \theta_{l,r}^{(\mathrm{el})})],$$

where L and R are the number of multipath clusters (typically small, $L \leq 4$ [138]) and sub-paths (rays), respectively. Scalar $g_{l,r}$, $\tau_{l,r}$, $\theta_{l,r}^{(\mathrm{az})}, \theta_{l,r}^{(\mathrm{el})}$ and $\phi_{l,r}^{(\mathrm{az})}, \phi_{l,r}^{(\mathrm{el})}$ are the complex gain, excessive delay, AoD in azimuth and elevation plane, and AoA in two planes of the r-th sub-path within the l-th cluster, respectively. Function $p_\mathrm{c}(t)$ is the time domain response filter due to limited temporal resolution T_s. With antenna spacing being half of wavelength that corresponds to the carrier frequency f_c, the angular response vectors at the BS and UE are denoted as $\mathbf{a}_\mathrm{T}(\theta) \in \mathbb{C}^{N_\mathrm{T}}$ and $\mathbf{a}_\mathrm{R}(\phi) \in \mathbb{C}^{N_\mathrm{R}}$. In uniform planar array (UPA) with $N_\mathrm{R}^{(\mathrm{az})}$ by $N_\mathrm{R}^{(\mathrm{el})}$ element ($N_\mathrm{R}^{(\mathrm{az})} N_\mathrm{R}^{(\mathrm{el})} = N_\mathrm{R}$), the receiver array response is defined as

$$[\mathbf{a}_\mathrm{R}\left(\phi^{(\mathrm{az})}, \phi^{(\mathrm{el})}\right)]_{(k_\mathrm{v}-1)N_\mathrm{R}^{(\mathrm{az})}+k_\mathrm{h}} = \exp\left[j\pi(k_\mathrm{h}-1)\sin(\phi^{(\mathrm{az})})\sin(\phi^{(\mathrm{el})}) + (k_\mathrm{v}-1)\cos(\phi^{(\mathrm{el})})\right].$$

The transmitter array response is similarly defined.

Note that the above model aligns with measurement-endorsed mmMAGIC channel model [9] and is used for the system performance evaluation in Section 2.8. However, for the sake of tractable algorithm design and analysis, the following assumptions and definitions are made.

Assumption 1: Assuming BS and UE use uniform linear array (ULA) with omni-directional element pattern in the 2D environment, i.e., array response reduces to $[\mathbf{a}_\mathrm{R}\left(\phi^{(\mathrm{az})}\right)]_k =$

$\exp[j\pi(k-1)\sin(\phi^{(\mathrm{az})})]$ and $[\mathbf{a}_{\mathrm{R}}\left(\phi^{(\mathrm{az})}\right)]_k = \exp[j\pi(k-1)\sin(\phi^{(\mathrm{az})})]$. We further remove the superscript in angle θ and ϕ for clarity. Intra-cluster AoA, AoD, and delay offsets are zero, i.e., $\sum_{r=1}^{R} g_{l,r} \triangleq g_l, \phi_{l,r} = \phi_l,\ \theta_{l,r} = \theta_l,\ \tau_{l,r} = \tau_l, \forall r$. Index r is omitted in the rest of chapter for clarity. The phase error process is solely from CFO i.e., phase noise process is $\psi[n] = 0, \forall n$ in (2.1). The complex path gain g_l is deterministic complex value, i.e., $\sum_{l=1}^{L} |g_l|^2 = \sigma_{\mathrm{g}}^2$.

Definition 1: The pre-BF SNR is defined as $\mathrm{SNR} \triangleq \sigma_{\mathrm{g}}^2/\sigma_{\mathrm{n}}^2$.

2.3.2 Problem formulations

We intend to address the following three problems, and their connection to the existing works are remarked.

Problem 1 (Initial Discovery and Timing Acquisition): The UE needs to detect the SSB from in-band received samples (2.1). This problem is a binary hypothesis testing with unknown channel $\mathbf{H}[d]$ and synchronization errors ϵ_{T} and ϵ_{F}.

$$\begin{aligned} \mathcal{H}_0: &\quad y[n] = \mathbf{w}^{\mathrm{H}}[n]\mathbf{z}[n], \\ \mathcal{H}_1: &\quad y[n] = \sum_{d=0}^{N_{\mathrm{c}}-1} \left(e^{j\epsilon_{\mathrm{F}} n} \mathbf{w}^{\mathrm{H}}[n]\mathbf{H}[d]\mathbf{v}[n-d-\epsilon_{\mathrm{T}}]s[n-d-\epsilon_{\mathrm{T}}] \right) + \mathbf{w}^{\mathrm{H}}[n]\mathbf{z}[n]. \end{aligned} \tag{2.2}$$

In addition, the STO ϵ_{T} is estimated at this stage.

Problem 2 (Initial BF Training): The BF training is triggered once UE has detected IA signals. In this stage, UE re-uses the asynchronous signal samples (2.1) to estimate the AoD and AoA of a path with significant power, say $\theta^\star$ and $\phi^\star$, and they are then used in designing beamformers, $\mathbf{v}^\star = \mathbf{a}_{\mathrm{T}}(\theta^\star)$ and $\mathbf{w}^\star = \mathbf{a}_{\mathrm{R}}(\phi^\star)$, in data communications phase.

Remark 1: The above problems can be solved by DIA and directional beam training with the help of CSI-RS, while our solution relies on processing IA block only. In additional, although *Problem 2* has overlap with parametric channel estimation, approaches from this class are not directly comparable. In fact, [74, 145] estimate the entire wideband channel, which facilitates optimal MIMO processing, but the assumptions of perfect synchronization and equal channel bandwidths in beam training and data communication do not necessarily

apply to the 5G-NR compliant model considered in our work. Our goal is to provide well-aligned beam pair within IA block, i.e. without requiring CSI-RS slots. Finally, the cell identifier recognition and PBCH decoding are important tasks but are not studied in this work.

Problem 3 (Squint Robust Beam Training): The model in *Assumption 1* and *Problem 2* are formulated for 5G-NR compliant cellular system where the bandwidth and array size are not large enough to exhibit significant beam squint phenomenon. We extend our channel model and design beam training algorithm that is robust to squint due to increased channel bandwidth.

2.4 Initial discovery and timing synchronization

This section presents the proposed initial discovery and timing synchronization followed by their performance analysis.

2.4.1 Initial discovery and timing synchronization algorithm

The UE processes the received signal using the correlation filter with $s_{\text{zc}}[n]$, and obtains the detection statistics:

$$\tilde{y}[n] = \frac{1}{P}\sum_{k=0}^{P-1} y[n+k]s_{\text{zc}}^{*}[k]. \tag{2.3}$$

Intuitively, there are M correlation peaks across M SSB. The magnitude of the m-th peaks depends on the array gain of the m-th sounding beamformer, CFO, and TO. Our proposed detector combines energy from all M SSB and compares it with the threshold. In contrast to previous works [30, 102, 191] where the detection threshold is a fixed constant, we propose to use the optimal detection threshold based on Neyman-Pearson criterion that meets target false alarm (FA) rate $P_{\text{FA}}^{\star}$.

To understand the impact of timing synchronization error, we first consider a *Genie* scenario where the UE has perfect timing (PT) information, i.e., $\epsilon_{\text{T}} = 0$. In this case, the

proposed PSS detection scheme is an energy detector over all M bursts. In addition, a sample time window with N_{c} is used to collect energy from all multipaths. Specifically, the proposed hypothesis testing scheme is expressed as

$$\gamma_{\mathrm{PT}} \triangleq \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_{\mathrm{c}}-1} \left|\tilde{y}[k+mN_{\mathrm{b}}]\right|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{\mathrm{PT}}, \tag{2.4}$$

where the detection threshold η_{PT} is used to reach false alarm rate constraint such that $\Pr(\gamma_{\mathrm{PT}} > \eta_{\mathrm{PT}}|\mathcal{H}_0) = P^{\star}_{\mathrm{FA}}$.

In a practical scenario without initial timing information (NT), i.e., $\epsilon_{\mathrm{T}} \neq 0$, we propose to use the following detector

$$\gamma_{\mathrm{NT}} \triangleq \max_{0 \leq n < \epsilon_{\mathrm{T,max}}} \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_{\mathrm{c}}-1} \left|\tilde{y}[n+k+mN_{\mathrm{b}}]\right|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{\mathrm{NT}} \tag{2.5}$$

that searches all possible instances within STO window $\epsilon_{\mathrm{T}} \in [0, \epsilon_{\mathrm{T,max}}]$ and uses the highest energy collected for the hypothesis test. The sample index corresponding to the highest energy in (2.5) is the estimate of STO, namely

$$\hat{\epsilon}_{\mathrm{T}} = \underset{0 \leq n < \epsilon_{\mathrm{T,max}}}{\arg\max} \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_{\mathrm{c}-1}} \left|\tilde{y}[n+k+mN_{\mathrm{b}}]\right|^2. \tag{2.6}$$

2.4.2 Performance of initial discovery and timing acquisition

In this subsection, we analyze performance of the proposed discovery algorithm in terms of miss detection rate, and the impact of initial synchronization error ϵ_{F} and ϵ_{T}. The exact expression is challenging and tedious, if not impossible, and therefore we provide a tight closed-form approximation in the following proposition. To be concise, the subscripts of γ and η that indicate the timing information assumption are denoted as binary variable $\mathrm{E} \in \{\mathrm{NT}, \mathrm{PT}\}$.

Proposition 1. *The optimal threshold of (2.5) that reaches target FA rate* $\Pr(\gamma_E \geq \eta^{\star}_E|\mathcal{H}_0) = P^{\star}_{FA}$ *is approximately*[3]

$$\eta^{\star}_E = \sigma_n^2 \left[\frac{N_c}{P} + \sqrt{\frac{N_c}{MP^2}} \xi_E\left(\epsilon_{T,max}, P^{\star}_{FA}\right)\right], \tag{2.7}$$

[3] *Approximation is tight when STO search window size* $\epsilon_{T,max} \geq 100$.

where $\xi_E(\epsilon_{T,max}, P^\star_{FA})$ *is the threshold adjustment factor dependent on synchronization computed as*

$$\xi_E = \begin{cases} \mathrm{Q}^{-1}\left(P^\star_{FA}\right), & E = PT \\ \mathrm{Q}^{-1}\left(\frac{1}{\epsilon_{T,max}}\right) - \frac{0.78\ln\left(-\ln\left(1-P^\star_{FA}\right)\right)}{\mathrm{Q}^{-1}\left(\frac{1}{\epsilon_{T,max}}\right)}, & E = NT \end{cases}, \tag{2.8}$$

where $\mathrm{Q}(.)$ *and* $\mathrm{Q}^{-1}(.)$ *are Q-function and inverse Q-function, respectively. The associated miss detection (MD) rate* $P_{MD,E} \triangleq \Pr\left(\gamma_E < \eta^\star_E | \mathcal{H}_1\right)$ *using the optimal threshold* $\eta^\star_E$ *is*

$$P_{MD,E} = \mathrm{Q}\left(\frac{\kappa(\epsilon_T, \epsilon_F)\mathrm{SNR} - \sqrt{\frac{N_c}{MP^2}}\xi_E\left(\epsilon_{T,max}, P^\star_{FA}\right)}{\sqrt{\frac{2\kappa^2(\epsilon_T,\epsilon_F)\mathrm{SNR}^2}{M} + \frac{N_c}{P^2M}}}\right), \tag{2.9}$$

where the SNR degradation factor $\kappa(\epsilon_T, \epsilon_F)$ *is defined as*

$$\kappa(\epsilon_T, \epsilon_F) = \frac{2 - \Re\left(\exp\left[jK(\epsilon_T)\epsilon_F\right]\right) - \Re\left(\exp\left[j[P - K(\epsilon_T)]\epsilon_F\right]\right)}{P^2\left[1 - \Re\left(\exp\left[j\epsilon_F\right]\right)\right]}, \tag{2.10}$$

where $K(\epsilon_T)$ *is the number of samples during PSS reception that UE switches beamformer due to TO.*

$$K(\epsilon_T) = \begin{cases} N_B - \epsilon_T, & \text{if } N_B - P \le \epsilon_T < N_B \\ 0, & \text{otherwise} \end{cases}. \tag{2.11}$$

Proof. See Appendix A.1. □

Remark 2: $1 - P_{\mathrm{MD,NT}}$ is a close approximation of probability that UE detects IA and correctly estimates ϵ_{T}.

We gain two main insights from MD expressions (2.9) corresponding to threshold adjustment factor $\xi_{\mathrm{E}}(\epsilon_{\mathrm{T,max}}, P^\star_{\mathrm{FA}})$ and SNR degradation factor $\kappa(\epsilon_{\mathrm{F}}, \epsilon_{\mathrm{T}})$. Firstly, the CFO affects MD performance by effectively reducing SNR via term $\kappa(\epsilon_{\mathrm{F}}, \epsilon_{\mathrm{T}})$. Under maximum CFO at UE of ±5ppm and typical frame parameters P, M, N_{c} specified in Section 2.8, the SNR degradation is bounded by 4 dB, i.e., $10\log_{10}[\kappa(\epsilon_{\mathrm{F}}, \epsilon_{\mathrm{T}})] \ge -4\mathrm{dB}, \forall\epsilon_{\mathrm{T}}$. Secondly, the STO has impact on both factors. As seen in (2.10), the SNR in the detection problem degrades when severe STO exists. In fact, $K(\epsilon_{\mathrm{T}})$ in $\kappa(\epsilon_{\mathrm{F}}, \epsilon_{\mathrm{T}})$ models phenomenon that receiver sounding

beam switches during the reception of PSS, i.e., $K(\epsilon_{\mathrm{T}}) \neq 0$. In addition, the presence of STO forces system to use peak detection scheme (2.5) where system searches peak location over a sample window with length $\epsilon_{\mathrm{T,max}}$, i.e.,the worst case in (2.5). Under $\mathcal{H}_0$, the algorithm picks strongest noise realization over $\epsilon_{\mathrm{T,max}}$ samples and thus system needs to use higher threshold than in PT scenario, as seen in (2.7) and (2.8). Note that such degradation does not depend on the value of ϵ_{T}, and the degradation in (2.9) is not critical with practical maximum STO uncertainty $\epsilon_{\mathrm{T,max}} \leq N_{\mathrm{B}}$. In summary, synchronization offset does not severely affect discovery performance of the proposed scheme.

2.4.3 Benchmark approach: directional initial discovery

For completeness, we briefly introduce the benchmark approach using directional sounding beam in initial discovery [30]. The system model of DIA is similar to Section 2.3, except that sounding beamformers $\mathcal{W}$ and $\mathcal{V}$ are codebooks that steer directional sector beams, e.g., [15, 125]. Adapting the approach in [30] for the wideband channel and known PSS in SSB, the cell discovery in DIA uses the following detector

$$\gamma_{\mathrm{DIA}} \triangleq \max_{n} |\tilde{y}_{\mathrm{DIA}}[n]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{\mathrm{DIA}} \tag{2.12}$$

where γ_{DIA} and η_{DIA} are the detection statistic and threshold in DIA. Sequence $\tilde{y}_{\mathrm{DIA}}[n]$ is the correlation output in (2.3) that corresponds to directional sounding beams. Refer to Figure 2.1, the UE detects the burst with maximum power and denotes the index as $m^{\star}_{\mathrm{DIA}}$ which is used in directional beam training.

2.5 Compressive initial beam training

This section presents the proposed initial access based BF training. We start with signal rearrangement based on information obtained from successful cell discovery and timing acquisition. Then, we introduce the CS problem formulation followed by the proposed algorithm. Finally, we analyze the CRLB of AoA/AoD estimation in LOS.

2.5.1 Signal rearrangement after timing acquisition

The further processing requires correct detection and CP removal, and therefore we make a following assumption.

Assumption 2: In beam training, the received IA signal (2.1) is correctly detected and STO ϵ_{T} is correctly estimated.

The UE first removes CPs of P PSS samples from $y[n]$ corresponding to M bursts and rearranges them into vector

$$\begin{aligned} \mathbf{y} =& [\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_m^{\mathrm{T}}, \cdots, \mathbf{y}_M^{\mathrm{T}}]^{\mathrm{T}}, \\ \{\mathbf{y}_m\}_p =& y[\hat{\epsilon}_{\mathrm{T}} + N_{\mathrm{CP}} + (p-1) + (m-1)N_{\mathrm{B}}], p \leq P. \end{aligned} \tag{2.13}$$

For notation convenience, in the rest of subsection, we restate the received time domain signal after removing CP at the m-th SSB $\mathbf{y}_m \in \mathbb{C}^P$ according to the model in Section 2.3,

$$\mathbf{y}_m = \underbrace{\sum_{l=1}^{L} \tilde{g}_{m,l} \mathbf{Q}(\epsilon_{\mathrm{F}}) \mathbf{F}^{\mathrm{H}} \left[\mathbf{f}(\tau_l) \circ \mathbf{s}\right]}_{\mathbf{x}_m(\boldsymbol{\xi})} + \mathbf{z}_m, \tag{2.14}$$

In the above equation, deterministic vector $\mathbf{x}_m(\boldsymbol{\xi}) \in \mathbb{C}^P$ is observations model of unknown parameters $\boldsymbol{\xi} \triangleq [\epsilon_{\mathrm{F}}, \cdots, \theta_l, \phi_l, \tau_l, \alpha_l, \beta_l, \cdots]^{\mathrm{T}}$, where $\alpha_l = \Re(g_l)$ and $\beta_l = \Im(g_l)$. $\mathbf{z}_m \in \mathbb{C}^P$ is the vectorized random noise. We also define $\mathbf{x}(\boldsymbol{\xi}) = [\mathbf{x}_1^{\mathrm{T}}(\boldsymbol{\xi}), \cdots, \mathbf{x}_M^{\mathrm{T}}(\boldsymbol{\xi})]^{\mathrm{T}}$. Specifically, in (2.14) vector $\mathbf{s} \in \mathbb{C}^P$ contains PSS symbols assigned to P subcarriers. Vector $\mathbf{f}(\tau_l) \in \mathbb{C}^P$ is the frequency response corresponding to the excessive delay τ_l of a multipath, i.e., the contribution of τ_l on the p-th subcarrier is

$$[\mathbf{f}(\tau_l)]_p = \exp\left[(-j2\pi(p-1)\tau_l)/(PT_{\mathrm{s}})\right]. \tag{2.15}$$

Matrix $\mathbf{F} \in \mathbb{C}^{P \times P}$ is DFT matrix[4]. The effective channel gain is defined as

$$\tilde{g}_{m,l} = \exp\left[j\epsilon_{\mathrm{F}} N_{\mathrm{B}}(m-1)\right] g_l \mathbf{w}_m^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\phi_l) \mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l) \mathbf{v}_m,$$

[4]With absence of CFO, multiple DFT matrix $\mathbf{F}$ in $\mathbf{y}_m$ gives frequency domain symbols $\sum_{l=1}^{L} \tilde{g}_{m,l}(\mathbf{f}(\tau_l) \circ \mathbf{s}) + \mathbf{z}_m$.

and it includes the contribution of phase rotation across different SSB due to CFO and IA beamformers $\mathbf{v}_m$ and $\mathbf{w}_m$. Matrix $\mathbf{Q}(\epsilon_{\mathrm{F}}) = \mathrm{diag}\left(\left[1, e^{j\epsilon_{\mathrm{F}}}, \cdots, e^{j(P-1)\epsilon_{\mathrm{F}}}\right]^{\mathrm{T}}\right)$ contains phase rotations within an OFDM symbol.

2.5.2 Baseline compressiove sensing formulation

Directly estimating $\boldsymbol{\xi}$ from (2.14) via maximum-likelihood estimator (MLE) requires multi-dimensional search with prohibitive complexity. In the following subsections, we re-formulate *Problem 2* to facilitate sequential parameter estimation. With straightforward extension of the derivation in [83, Sec. V], the vector $[\tilde{\mathbf{g}}_l]_m = \tilde{g}_{m,l}$ in (2.14) can be re-formulated as

$$\tilde{\mathbf{g}}_l = \tilde{\mathbf{Q}}(\epsilon_{\mathrm{F}})\tilde{\mathbf{A}}^{\mathrm{H}}\mathrm{vec}(\tilde{\mathbf{H}}_l), \tag{2.16}$$

where $\tilde{\mathbf{A}} \in \mathbb{C}^{G_{\mathrm{T}}G_{\mathrm{R}} \times M}$ is defined by the Hermitian conjugate of its m-th column as

$$([\tilde{\mathbf{A}}]_m)^{\mathrm{H}} = (\mathbf{v}_m^{\mathrm{T}} \otimes \mathbf{w}_m^{\mathrm{H}})(\mathbf{A}_{\mathrm{T}}^{*} \otimes \mathbf{A}_{\mathrm{R}}).$$

Note that the above equation is different from [83, Sec. V] which requires M^2 sounding beam pairs. The matrix $\tilde{\mathbf{Q}}(\epsilon_{\mathrm{F}}) \in \mathbb{C}^{M \times M}$ contains the phase rotation in each SSB due to CFO.

$$\tilde{\mathbf{Q}}(\epsilon_{\mathrm{F}}) = \mathrm{diag}\left(\left[1, e^{jN_{\mathrm{B}}\epsilon_{\mathrm{F}}}, \cdots, e^{jN_{\mathrm{B}}(M-1)\epsilon_{\mathrm{F}}}\right]^{\mathrm{T}}\right). \tag{2.17}$$

In fact, matrices $\mathbf{A}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times G_{\mathrm{T}}}$ and $\mathbf{A}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times G_{\mathrm{R}}}$ are the dictionaries of angular responses with AoAs and AoDs from grids with G_{T} and G_{R} uniform steps from $-\pi/2$ to $\pi/2$, respectively. In order words, the k-th columns in $\mathbf{A}_{\mathrm{T}}$ and $\mathbf{A}_{\mathrm{R}}$ are $[\mathbf{A}_{\mathrm{R}}]_k = \mathbf{a}_{\mathrm{R}}([\mathbf{r}]_k)$ and $[\mathbf{A}_{\mathrm{T}}]_k = \mathbf{a}_{\mathrm{T}}([\mathbf{t}]_k)$, respectively, where $[\mathbf{r}]_k$ and $[\mathbf{t}]_k$ are the vectors that contain angle candidates.

$$[\mathbf{r}]_k = -\frac{\pi}{2} + (k-1)\Delta\phi, \quad [\mathbf{t}]_k = -\frac{\pi}{2} + (k-1)\Delta\theta. \tag{2.18}$$

Also note that the steps $\Delta\theta$ and $\Delta\phi$ depend on the desired resolution. In this work, G_{T} and G_{R} are used as number of steps and namely $\Delta\theta = 2\pi/G_{\mathrm{T}}$ and $\Delta\phi = 2\pi/G_{\mathrm{R}}$. Matrix $\tilde{\mathbf{H}}_l \in \mathbb{C}^{G_{\mathrm{R}} \times G_{\mathrm{T}}}$ contains the complex path gain of the l-th path, i.e., it has 1 non-zero element whose location depends on the AoA and AoD of the l-th cluster in the angular grids.

Remark 3: Assuming noisy observation of $\tilde{\mathbf{g}}_l$ and zero CFO, (2.16) reduces to the baseline problem in [83, Sec. V]. However, (2.14) implies that the former assumption is non-trivial unless $\mathbf{s} = \mathbf{1}, \tau_l = 0, \forall l$, e.g., [19]. Moreover, algorithm designed with latter assumption is sensitive to CFO [181]. Finally, the AoA/AoD estimators are commonly confined in $\mathbf{r}$ and $\mathbf{t}$ [19]. We address these challenges in the following three subsections.

2.5.3 Effective gain estimation

To address the challenge discussed in *Remark 3*, we propose the following approach. We treat $\mathbf{Q}(\epsilon_{\mathrm{F}})$ in (2.14) as identity matrix and estimate delay of dominant path and gain, say τ_l and $\tilde{g}_{m,l}$, by maximum likelihood approach. Actually, the proposed algorithm uses sparse impulse support $[\mathbf{d}]_q = q\Delta\tau$ to construct a dictionary, where $\Delta\tau = N_{\mathrm{c}}T_{\mathrm{s}}/G_{\mathrm{D}}$ is the step-size of delay candidates. Based on the knowledge of the model (2.15) and PSS signal $\mathbf{s}$, the delay estimation is implemented as

$$\hat{q} = \underset{1\le q\le G_{\mathrm{D}}}{\arg\max} \langle \mathbf{p}_q, \bar{\mathbf{y}} \rangle / \|\mathbf{p}_q\|^2 \text{ and } \hat{\tau} = [\mathbf{d}]_{\hat{q}}, \tag{2.19}$$

where $\bar{\mathbf{y}} = \sum_{m=1}^{M} \mathbf{y}_m/M$ is the received PSS samples averaged over M SSB. The vector $\mathbf{p}_q$ contains PSS samples when the true delay of dominant path is $[\mathbf{d}]_q$, i.e., $\mathbf{p}_q \triangleq \mathbf{F}^{\mathrm{H}}\left[\mathbf{f}\left([\mathbf{d}]_q\right) \circ \mathbf{s}\right]$, where $\mathbf{f}([\mathbf{d}]_q)$ is by plugging in $[\mathbf{d}]_q$ into (2.15). The estimated delay tap $\hat{\tau}$ enables estimating effective gain of a significant path by

$$\hat{\mathbf{g}} = \left(\mathbf{p}_{\hat{q}}^{\mathrm{H}} \otimes \mathbf{I}_M\right) \mathbf{y}, \tag{2.20}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix.

2.5.4 On-grid joint AoA and AoD estimation robust to CFO

The second step uses a modified matching pursuit to solve CS problem (2.16) from $\hat{\mathbf{g}}$ while incorporating the existence of CFO in $\tilde{\mathbf{Q}}$. In the conventional matching pursuit step, say the k-th, the anticipated effective channel response corresponding to an AoA and AoD pair, i.e., $[\tilde{\mathbf{A}}]_k$, is used to evaluate inner product with $\mathbf{g}$ [112]. The proposed heuristic treats AoA and

AoD as known in the k-th step, and uses the MLE of CFO $\hat{\epsilon}_{\mathrm{F},k}$ which is available in closed form. The modified matching pursuit is expressed as

$$\hat{k} = \underset{1\le k\le G_{\mathrm{R}}G_{\mathrm{T}}}{\arg\max} \langle \tilde{\mathbf{Q}}(\hat{\epsilon}_{\mathrm{F},k})\tilde{\mathbf{a}}_k, \hat{\mathbf{g}}\rangle / \|\tilde{\mathbf{a}}_k\|^2, \tag{2.21}$$

where $\tilde{\mathbf{a}}_k \triangleq [\tilde{\mathbf{A}}]_k$ from (2.16). The matrix $\tilde{\mathbf{Q}}(\hat{\epsilon}_{\mathrm{F},k})$ has structure as (2.17). The input $\hat{\epsilon}_{\mathrm{F},k}$ is the MLE of CFO when treating AoA/AoD as they correspond to ones in $[\tilde{\mathbf{A}}]_k$. Specifically, the CFO estimator relies on the estimator in [94] by treating $\bar{\mathbf{y}}_k = \tilde{\mathbf{a}}_k^* \circ \hat{\mathbf{g}}$ as a tone with frequency ϵ_{F}.

$$\hat{\epsilon}_{\mathrm{F},k} = \frac{1}{N_{\mathrm{B}}} \angle \left(\frac{1}{M-1} \sum_{m=1}^{M-1} [\bar{\mathbf{y}}_k]_m^* [\bar{\mathbf{y}}_k]_{m+1} \right). \tag{2.22}$$

Operation $\angle(x) = \tan^{-1}[\Im(x)/\Re(x)]$ evaluates angle based on complex samples. To get estimates of the AoA, AoD, and CFO, index $\hat{k}$ is used to select candidates from grids (2.18) after the following adjustment $\hat{k}_{\mathrm{R}} = \lfloor (\hat{k}-1)/G_{\mathrm{T}} \rfloor + 1$ and $\hat{k}_{\mathrm{T}} = \hat{k} - (\hat{k}_{\mathrm{R}} - 1)G_{\mathrm{T}}$,

$$\hat{\phi} = [\mathbf{r}]_{\hat{k}_{\mathrm{R}}}, \quad \hat{\theta} = [\mathbf{t}]_{\hat{k}_{\mathrm{T}}}, \quad \hat{\epsilon}_{\mathrm{F}} = \hat{\epsilon}_{\mathrm{F},\hat{k}}. \tag{2.23}$$

2.5.5 Off-grid refinement

The aformentioned heuristics provide estimates of delay, AoA, and AoD that are restricted to the grid, i.e., elements of $\mathbf{d}, \mathbf{r}$ and $\mathbf{t}$. Grid refinement is a technique to provide off-grid estimation accuracy. There are several approaches considered in the literature including multi-resolution refinement [109] and the Newtonized gradient refinement [110]. In this work, we propose to use first order descent approach. As initialization of refinement, the estimator from previous steps is saved into $\hat{\boldsymbol{\xi}}^{(k)}$ for $k = 1$. In the k-th iteration, the error vector is evaluated

$$\mathbf{e}^{(k)} = \mathbf{y} - \mathbf{x}\left(\hat{\boldsymbol{\xi}}^{(k)}\right), \tag{2.24}$$

where $\mathbf{y}$ is the received signal after rearrangement as (2.13), $\mathbf{x}(\hat{\boldsymbol{\xi}}^{(k)})$ is obtained by plugging in estimated parameters into parametric model (2.14). In other words, $\mathbf{e}^{(k)}$ is the error vector

between observed signal sequence and received signal model using current estimates, which is then used to update parameters. The complex gain in iteration k is computed as

$$\hat{g}^{(k+1)} = (\boldsymbol{\nabla}\mathbf{x}_g)^{\dagger}\mathbf{y}, \tag{2.25}$$

where $\boldsymbol{\nabla}\mathbf{x}_g = (\partial\mathbf{x}(\boldsymbol{\xi})/\partial g)|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^{(k)}}$ is the partial derivative of $\mathbf{x}(\boldsymbol{\xi})$ over parameter g in (2.14) evaluated at $\hat{\boldsymbol{\xi}}^{(k)}$. The refinement steps for delay, CFO, AoA, and AoD are moving towards the gradient of their estimators in the previous iterations. For concise notation, in the following equation and paragraph we use x to denote the parameter to be refined, i.e., $x = \{\tau, \epsilon_{\mathrm{F}}, \theta, \phi\}$. The refinement steps are

$$\hat{x}^{(k+1)} = \hat{x}^{(k)} + \mu_x \Re\left[(\boldsymbol{\nabla}\mathbf{x}_x)^{\dagger}\mathbf{e}^{(k)}\right], x = \{\tau, \epsilon_{\mathrm{F}}, \theta, \phi\}, \tag{2.26}$$

where μ_x is the step-size, vector $\boldsymbol{\nabla}\mathbf{x}_x = (\partial\mathbf{x}(\boldsymbol{\xi})/\partial x)|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^{(k)}}$ is the the partial derivative of $\mathbf{x}(\boldsymbol{\xi})$ in (2.14) over parameter of interest. The above approach iteratively runs by appending updated parameter into $\mathbf{x}(\hat{\boldsymbol{\xi}}^{(k+1)})$ for the next iteration until the error $\|\mathbf{e}^{(k)}\|^2$ converges or falls below threshold ϵ_0.

It is worth noting that the proposed approach can be extended to support multi-path training which has been covered by a variety of works in CS-based approaches [74, 112, 119, 126, 128, 145, 161]. However, the main motivation of this work is to showcase and analyze pseudorandom sounding beams in the initial access and initial beam training. Thus the only metric directly comparable to its counterparts [18, 30, 71, 72, 102, 103, 114, 191], namely single path training, is evaluated.

The algorithm is summarized in Algorithm 1.

2.5.6 Analysis of initial beam training in line-of-sight channel

In this subsection, we provide lower bound of AoA/AoD estimation variance in pure LOS[5] scenario, namely CRLB in joint estimating $\boldsymbol{\xi} = [\epsilon_{\mathrm{F}}, \theta_1, \phi_1, \tau_1, \alpha_1, \beta_1]^{\mathrm{T}}$. Based on (2.14), the

[5]In the NLOS environment, the *Assumption 1* facilitates intuitive algorithm design, but the failure to consider intra-cluster angular spread results in an inaccurate performance analysis.

Algorithm 1 Compressive Initial Access and Beam Training

Input: Received IA signal sequence $y[n]$

Output: Discovery decision; Beam pair $\mathbf{v}^\star, \mathbf{w}^\star$

% ——— Initial Discovery ———

PSS correlation (2.3).

Energy detection (2.5) and timing acquisition (2.6).

if PositiveDecision **then**

% —— Initial BF Training (Coarse) ——

Arrange sequence $y[n]$ into vector $\mathbf{y}$ as (2.13).

Estimate excessive delay as (2.19).

Estimate effective channel gain as (2.20).

Matching pursuit (2.21) with CFO estimation (2.22).

Get AoA, AoD, and CFO estimators in (2.23).

% —— Initial BF Training (Fine) ——

while $\|\mathbf{e}^{(k)}\| > \epsilon_0$ in (2.24) **do**

Use refinement steps (2.25) and (2.26); $k = k + 1$.

end while

Report beam pair $\mathbf{w}^\star = \mathbf{a}_{\mathrm{R}}(\phi_l^{(k)})$, $\mathbf{v}^\star = \mathbf{a}_{\mathrm{T}}(\theta_l^{(k)})$.

end if

likelihood function is $\Pr(\mathbf{y};\boldsymbol{\xi}) = (2\pi\sigma_{\mathrm{n}}^{2MP})^{-1}\exp\left(-(\|\mathbf{y}-\mathbf{x}(\boldsymbol{\xi})\|^2)/(\sigma_{\mathrm{n}}^2)\right)$. The log-likelihood function is $L(\mathbf{y};\boldsymbol{\xi}) \triangleq \ln[\Pr(\mathbf{y};\boldsymbol{\xi})]$. The lower bound of estimation variance is given in the following proposition.

Proposition 2. *The CRLB of AoA/AoD estimation in the compressive initial BF training stage in LOS environment is*

$$\mathrm{var}(\hat{\phi}_1) \geq [\mathbf{J}^{-1}]_{2,2}, \quad \mathrm{var}(\hat{\theta}_1) \geq [\mathbf{J}^{-1}]_{3,3} \tag{2.27}$$

where $\mathbf{J} \triangleq \partial^2 L(\mathbf{y};\boldsymbol{\xi})/\partial\boldsymbol{\xi}^2$ *is the Fisher Information Matrix whose expressions are listed in Appendix A.2.*

Proof. See Appendix A.2. □

2.5.7 Benchmark beam training 1: hierarchical directional search

The directional beams in SSB allow BS and UE to coarsely estimate the propagation directions [42]. Although approach in [42] is not tailored for wideband channel with synchronization offset, it relies on RSS measurement within burst and therefore it is robust to the model mismatch. Using the SSB index that corresponds to the maximum received power, the system uses the knowledge of directional sounding beams to infer channel propagation angles. Specifically, as illustrated in Figure 2.1, the estimated $\theta^\star$ and $\phi^\star$ are the centers of the $\hat{m}_{\mathrm{T}}$-th and $\hat{m}_{\mathrm{R}}$-th sounding beams in BS and UE [42], respectively. Note that the estimated angle sector indices $\hat{m}_{\mathrm{T}}$ and $\hat{m}_{\mathrm{R}}$ are computed from the SSB index $m^\star_{\mathrm{DIA}}$ in (2.12), i.e., $\hat{m}_{\mathrm{R}} = \lfloor (m^\star_{\mathrm{DIA}}-1)/M_{\mathrm{T}} \rfloor + 1$, and $\hat{m}_{\mathrm{T}} = m^\star_{\mathrm{DIA}} - (\hat{m}_{\mathrm{R}}-1)M_{\mathrm{T}}$. The large width of a sector beam results in poor angular resolution in DIA. In order to improve the resolution, hierarchical directional beam training scans narrower beams within the sector of interest. Such procedure occurs during CSI-RS bursts which are scheduled for individual UEs.

2.6 Squint robust beam training in beyond 5G

We have discussed the compressive sensing based beam training design in a 5G-NR compliant model without beam squint. In the beyond 5G-NR era, with the increased bandwidth as well as the increased array size, the beam squint becomes more significant. In this section, we present the enhanced version of beam training robust to beam squint.

2.6.1 Received signal with spatial wideband effect

By extending the channel model of [171] to multi-antenna UEs scenario, the channel in the p-th subcarrier is denoted as

$$\mathbf{H}^{(\mathrm{f})}[p] = \frac{1}{\sqrt{N_{\mathrm{T}}N_{\mathrm{R}}}} \sum_{l=1}^{L} g_l \exp\left[\frac{-j2\pi\tau_l(p-1)}{PT_{\mathrm{s}}}\right] \mathbf{a}_{\mathrm{R}}(\phi_l)\tilde{\mathbf{a}}_{\mathrm{T}}^{\mathrm{H}}(\theta_l, p). \tag{2.28}$$

Due to the fact that antenna array at BS is commonly larger than in mobile terminal, the beam squint is modeled at the transmitter end. Thus the p-th subcarrier experiences a unique transmitter array response, which is defined by its n-th element

$$[\tilde{\mathbf{a}}_{\mathrm{T}}(\theta, p)]_n = \exp\left(j\pi(n-1)\frac{f_p}{2f_{\mathrm{c}}}\sin(\theta)\right) = \exp\left[j\pi(n-1)\left(1+\frac{\tilde{p}(p)}{PT_{\mathrm{s}}f_{\mathrm{c}}}\right)\sin(\theta)\right].$$

The function $\tilde{p}(p) = p - P/2 - 1$ relates the index of subcarrier $p \in [1, P]$ with the radio frequency for this subcarrier $f_p = f_{\mathrm{c}} + \tilde{p}(p)/(PT_{\mathrm{s}})$ [75]. For notational convenience, we use $\tilde{p}$ instead of $\tilde{p}(p)$ in the rest of the chapter. Note that when the bandwidth in beam training is negligible as compared to the carrier frequency[6], i.e., $1/(T_{\mathrm{s}}f_{\mathrm{c}}) \ll 1$, the transmitter array response vector reduces to the one in Section 2.3, i.e., $\tilde{\mathbf{a}}_{\mathrm{T}}(\theta, p) \approx \mathbf{a}_{\mathrm{T}}(\theta), \forall p$. Also, when the array size is small or the true propagation angle is close to bore-sight, i.e., $\theta \approx 0$, the impact of beam squint is negligible [170].

In order to focus on the spatial wideband effect, we ignore the initial synchronization error and utilize the unit symbols $\mathbf{s} = \mathbf{1}$ when developing the enhanced training algorithm for beam squint regime. Following extension of Section 2.5.1, the received signal in the

[6]Such condition holds true in 5G-NR compliant frame structure.

frequency domain is

$$y_m^{(\mathrm{f})}[p] = \mathbf{w}_m^{\mathrm{H}} \mathbf{H}^{(\mathrm{f})}[p] \mathbf{v}_m + z_m^{(\mathrm{f})}[p], \tag{2.29}$$

where $z_m^{(\mathrm{f})}[p]$ is the post combining additive white Gaussian noise (AWGN) in the m-th SS-burst and the p-th subcarrier. Using the similar approach as in 2.5.2 and ignoring noise for notational clarity, the above received signal can be reformulated as

$$y_m^{(\mathrm{f})}[p] = (\mathbf{v}_m^{\mathrm{T}} \otimes \mathbf{w}_m^{\mathrm{H}})(\tilde{\mathbf{A}}_{\mathrm{T}}^{*}[p] \otimes \mathbf{A}_{\mathrm{R}} \otimes \mathbf{a}_{\mathrm{d}}^{\mathrm{T}}[p])\mathbf{g}.$$

In this expression, $\tilde{\mathbf{A}}_{\mathrm{T}}[p] \in \mathbb{C}^{N_{\mathrm{T}} \times G_{\mathrm{T}}}$ is defined by its k-th column as $\tilde{\mathbf{a}}_{\mathrm{T}}([\mathbf{t}]_k, p)$ where the on-grid AoD candidates are from (2.18), and $\mathbf{A}_{\mathrm{R}}$ is defined in Section 2.5. The vector $\mathbf{a}_{\mathrm{d}}[p] \in \mathbb{C}^{G_{\mathrm{D}}}$ is defined as by its k-th element as $e^{-j2\pi(p-1)[\mathbf{d}]_k/(PT_{\mathrm{s}})}$. $\mathbf{g} \in \mathbb{C}^{G_{\mathrm{T}} G_{\mathrm{R}} G_{\mathrm{D}}}$ is L-sparse vector that contains path gain that corresponds to a tuple of AoA, AoD, and delay on the grid.

2.6.2 Squint robust beam training

We propose to apply the matching pursuit based algorithm to estimate channel parameters from $\mathbf{y}^{(\mathrm{f})}[p] = [y_1^{(\mathrm{f})}[p], \cdots, y_M^{(\mathrm{f})}[p]]^{\mathrm{T}}$. To incorporate beam squint in estimation, the following components $\bar{\mathbf{A}}_{\mathrm{T}} \triangleq \sum_{p=0}^{P-1} \tilde{\mathbf{A}}_{\mathrm{T}}[p]/P \in \mathbb{C}^{N_{\mathrm{T}} \times G_{\mathrm{T}}}$ are pre-computed as it is then used in the squint aware dictionary. Specifically, $\left[\bar{\mathbf{A}}_{\mathrm{T}}\right]_{1,k} = 1, \forall k$, and

$$\left[\bar{\mathbf{A}}_{\mathrm{T}}\right]_{n,k} = \exp[j\pi(n-1)\sin([\mathbf{t}]_k)] \frac{\exp[-j\frac{\pi(n-1)\sin([\mathbf{t}]_k)}{2f_{\mathrm{c}}T_{\mathrm{s}}}] - \exp[j\frac{\pi(n-1)\sin([\mathbf{t}]_k)}{2f_{\mathrm{c}}T_{\mathrm{s}}}]}{P\left(1 - \exp[j\frac{\pi(n-1)\sin([\mathbf{t}]_k)}{Pf_{\mathrm{c}}T_{\mathrm{s}}}]\right)}, n \neq 1. \tag{2.30}$$

The proposed method first identifies the delay of dominant path $\hat{\tau}_l$. Then, the signal $\mathbf{y}[p]$ is filtered as

$$\bar{\mathbf{y}} = \frac{1}{P} \sum_{p=1}^{P} \mathbf{y}^{(\mathrm{f})}[p] e^{j2\pi \frac{\hat{\tau}_l (p-1)}{T_{\mathrm{s}} P}} \tag{2.31}$$

With an accurate excessive delay estimate $\hat{\tau}_l$, the compressive sensing problem is formulated

for the signal $\bar{\mathbf{y}}$ as

$$\bar{\mathbf{y}} = \underbrace{\begin{bmatrix} (\mathbf{v}_1^{\mathrm{T}} \otimes \mathbf{w}_1^{\mathrm{H}})(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \mathbf{A}_{\mathrm{R}}) \\ \vdots \\ (\mathbf{v}_M^{\mathrm{T}} \otimes \mathbf{w}_M^{\mathrm{H}})(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \mathbf{A}_{\mathrm{R}}) \end{bmatrix}}_{\boldsymbol{\Psi}} \bar{\mathbf{g}} \tag{2.32}$$

where $\bar{\mathbf{g}} = (\mathbf{1}_{G_{\mathrm{D}}}^{\mathrm{T}} \otimes \mathbf{I}_{G_{\mathrm{T}}G_{\mathrm{R}}})\mathbf{g} \in \mathbb{C}^{G_{\mathrm{T}}G_{\mathrm{R}}}$ is the sparse vector whose non-zero elements correspond to AoA and AoD of the the significant path whose delay is estimated as $\hat{\tau}_l$. The filtering (2.31) significantly reduce contribution of other paths in (2.32) when operating in wideband. The beam squint aware dictionary $\boldsymbol{\Psi} \in \mathbb{C}^{M \times G_{\mathrm{T}}G_{\mathrm{R}}}$ utilize the knowledge of $\bar{\mathbf{A}}_{\mathrm{T}}$ in (2.30), $\mathbf{A}_{\mathrm{R}}$, and sounding beamformer $\mathbf{w}_m$ and $\mathbf{v}_m$.

The on-grid estimates of AoA and AoD can be effectively achieved by applying matching pursuit in (2.32). Further, the off-grid accuracy can be achieved in a similar manner as Section 2.5.5. Specifically, iterative gradient descent algorithm is used based on an initial value from the on-grid estimates. The gradient for each of the parameter of interest is directly available by taking the derivative of (2.29). The residual error in the k-th iteration is defined by

$$e^{(\mathrm{k})} = \sum_{m,p} |y_m^{(\mathrm{f})}[p] - \mathbf{w}_m \hat{\mathbf{H}}^{(\mathrm{f})}[p]\mathbf{v}_m|^2$$

where $\hat{\mathbf{H}}^{(\mathrm{f})}[p]$ in the k-th iteration is computed by plugging the estimated channel parameters into (2.28). The algorithm is summarized in Algorithm 2.

The proposed approach provides AoA/AoD report robust to beam squint during training. In addition, the beamforming vector of BS in data communication phase also needs to accommodate squint. Although detailed discussion is beyond the scope of this work, we propose a simple heuristic approach for data communication phase, where frequency flat[7] beamformer $\mathbf{v}^{\star}$ are design such that it provides constant gain for all frequency range of data bandwidth, i.e., $[f_{\mathrm{c}} - B_{\mathrm{tot}}/2, f_{\mathrm{c}} + B_{\mathrm{tot}}/2]$. Given estimated AoD $\theta^{\star}$, the squint aware beam

[7]Frequency flat beamformers can be implemented by analog array architecture. Readers are referred to [143, 171] for approaches using frequency dependent beamformer design in handling squint after beam training, where digital or hybrid array architecture is used.

$\mathbf{v}^\star$ needs to have its beam-width large enough to cover $[\theta_{\min}, \theta_{\max}]$, which correspond to the squinted directions of the entire data channel. Specifically, the two critical directions are $\theta_{\min} = \sin^{-1}[(1 - \frac{B_{\text{tot}}}{2f_c})\sin(\theta^\star)]$ and $\theta_{\max} = \sin^{-1}[(1 + \frac{B_{\text{tot}}}{2f_c})\sin(\theta^\star)]$.

Algorithm 2 Squint Robust BF Training

Input: Received IA signal sequence $y[n]$

Output: AoA/AoD pair $\phi^\star, \theta^\star$

% —— Beam Training (Coarse) ——

1: Estimate excessive delay $\hat{\tau}_l$ as (2.19).

2: Convert $y[n]$ to frequency symbols $\mathbf{y}^{(\text{f})}[p]$

3: Filter frequency domain measurements as (2.31).

4: On-grid angle estimates via matching pursuit in (2.32) that utilize squint aware dictionary (2.30).

% —— BF Training (Fine) ——

5: **while** $e^{(k)} > \epsilon_0$ **do**

6: Use gradient refinement based on (2.29); $k = k + 1$.

7: **end while**

8: Report AoA/AoD pair $\phi^\star = \phi_l^{(k)}$, $\theta^\star = \theta_l^{(k)}$.

2.6.3 Benchmark beam training 2: squint non-aware compressive sensing

There are various wideband mmW channel parameter estimation approaches that use compressive sensing [74, 143, 145, 168, 171, 198]. To compare the performance of proposed squint robust beam training, we use [168] as the second benchmark method. In this method, the channel parameters are estimated for each subcarrier seperately, i.e., the problem is decoupled into multiple parallel parameter estimations, which are then solved via compressive sensing algorithm. Although [168] is originally designed for on-grid angle estimation, the off-grid accuracy can be achieved by narrowband refinements. Due to a squint non-aware nature of this approach, AoD estimated from different subcarriers has deviation under beam squint regime. We use the empirical average of AoD estimates over all subcarriers for comparison

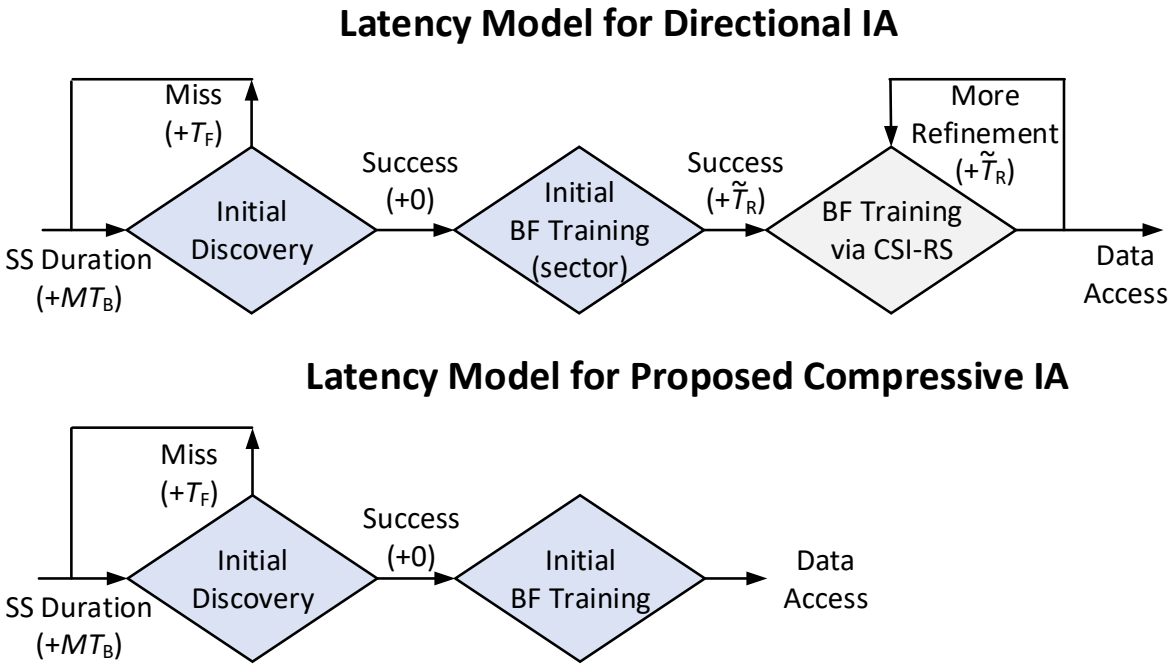


Figure 2.3: Initial access latency model for directional initial access and the proposed compressive initial access. The associated latency in each step is shown under bracket.
with our wideband AoD estimates.

2.7 Access latency, overhead and digital backend complexity

In this section, we present a model for analyzing three system performance indicators, namely access latency, overhead, and computational complexity. Note that this unified model applies to both directional scheme and the proposed approach.

Based on [72], we propose to use the latency model[8] for both SSB and CSI-RS as shown in Figure 2.3. In both IA schemes, the failure of cell discovery introduces penalty of T_{F} for a new IA block. When cell discovery occurs, the additional latency is required for scheduled

[8]Note that model is simplified to emphasize the topics discussed in this work. There are other types of latency including processing and feedback through beam reporting.

CSI-RS according to the required number N_{train}. Thus the access latency is

$$T_{\text{latency}} = MT_{\text{B}} + \sum_{k=0}^{+\infty} P_{\text{MD}}^{k}(1 - P_{\text{MD}})kT_{\text{F}} + \tilde{T}_{\text{R}}N_{\text{Train}} \tag{2.33}$$

where the first term includes latency for cell discovery. In the second term, $\tilde{T}_{\text{R}}$ is the average time for the UE to get the scheduled CSI-RS for beam training and it is expressed as

$$\tilde{T}_{\text{R}} = \frac{1}{N_{\text{U}}}\left[\sum_{k=0}^{K_{\text{F}}}\sum_{q=1}^{K_{\text{R}}}((k-1)T_{\text{F}} + qT_{\text{R}}) + \sum_{q=1}^{K_{\text{res}}}(K_{\text{F}}T_{\text{F}} + qT_{\text{R}})\right]$$

In the above equation, N_{U} denotes the number of UEs in the network. They share available CSI-RS in a time division manner to combat channel dynamic. Due to the limited number of CSI-RS $K_{\text{R}} = \lfloor (T_{\text{F}} - MT_{\text{B}})/T_{\text{R}} \rfloor$ within one IA period, more than one frame duration is required to meet scheduling of large number of UE N_{U}. Therefore, in (2.33) $K_{\text{F}} = \lfloor (N_{\text{U}} - 1)/K_{\text{R}} \rfloor$ is the number of frames required to assign all CSI-RS to UEs and $K_{\text{res}} = N_{\text{U}} - K_{\text{cyc}}K_{\text{R}}$ is the residual delay in the last frame. As shown in the next section, DIA and directional BF training typically require larger N_{train} than the proposed approach.

Following [72], the overhead (OH) ratio is modeled by counting the time-frequency resource in IA and CSI-RS

$$\text{OH} = \frac{MB_{\text{IA}}T_{\text{B}} + K_{\text{R}}B_{\text{tot}}T_{\text{r}}}{B_{\text{tot}}T_{\text{F}}} \times 100\% \tag{2.34}$$

where $B_{\text{IA}} = 1/T_{\text{s}}$ is the bandwidth in IA and the channel usage is MT_{s} every period T_{F}. We focus on varying CSI-RS density K_{R}. Note that with reduced K_{R} (increased T_{R}), the OH reduces with a cost of additional latency.

Although existing work shows that narrowband based IA provides SNR gain due to the low noise bandwidth [30], increasing bandwidth B_{IA} in IA and beam training offers improved latency. In this work, we assume that increasing B_{IA} is achieved by using fixed pilot length P and reduced T_{s}. This facilitates shorter OFDM symbol periods, and thus latency reduces with smaller T_{B} in (2.33). Furthermore, the overhead remains the same since $B_{\text{IA}}T_{\text{B}}$ in (2.34) remains constant.

Table 2.2: Digital Baseband Operations (complex multiplications)

Function Block	**Equation**	**Operations**
Initial Discovery		
PSS FIR corr.	(2.3)	PN_{B}
Detection and time sync.	(2.4) or (2.5)	N_{B}
Initial BF training (on-grid stage)		
Excess. delay est.	(2.19) and (2.20)	$PG_{\mathrm{D}} + PM$
AoA/AoD est.	(2.21)	$MG_{\mathrm{T}}G_{\mathrm{R}}$
CFO est.	(2.22)	$2MG_{\mathrm{T}}G_{\mathrm{R}}$
Initial BF training (off-grid stage)		
Alternative updates	(2.25) and (2.26)	$\mathcal{O}(K_{\mathrm{ite}}N_{\mathrm{T}}N_{\mathrm{R}}MP)$
Error norm evaluation	-	$K_{\mathrm{ite}}MP$

The required baseband operations of the proposed approach are summarized in Table 2.2, where only the complex multiplications are taken into account. To reach the on-grid accuracy, the existing compressive sensing based wideband channel estimation requires complexity $\mathcal{O}(PMG_{\mathrm{T}}G_{\mathrm{R}})$ when all P subcarriers are used [168], or $\mathcal{O}(P_{\mathrm{sel}}MG_{\mathrm{T}}G_{\mathrm{R}})$ where a selective number of P_{sel} subcarriers are used [145]. The proposed approach reduces this most computationally demanding steps into $\mathcal{O}(MG_{\mathrm{T}}G_{\mathrm{R}})$. Admittedly, the refinement stage involves higher complexity, since each iteration contains the computation of gradient and gradient based updates (2.26) and (2.25). Here, only scaling laws in terms of system parameters are provided for clarity. Refinements require K_{ite} iteration and quantitative analysis of this values is left as future work. The beam squint robust algorithm in Section 2.6 contains the same online computational complexity. Moreover, it is worth noting that the above analysis contains online computation, and assumes there is an offline pre-computation of all required dictionaries for matching pursuit, i.e., $\mathbf{p}_q$ in (2.19), $\tilde{\mathbf{a}}_k$ in (2.21). Lastly, directional IA involves the computation of (2.4) or (2.5) and its computational complexity is given in Table 2.2.

2.8 Results

This section presents the numerical comparison between the proposed approach and DIA with directional beam training.

2.8.1 Simulation settings

The simulations follow 5G-NR frame structure. We first evaluate performance in the simplified 2D S-V channel model. The maximum excess delay is set as $N_c = 4$ samples. As for the DIA, we use two approaches to design directional sector beams, i.e., least-squares based sector beamforming codebook [15] and frequency sampling method based sector beamforming codebook [125, C23.4]. Examples of beam patterns[9] are shown in Figure 2.4. In each of the Monte Carlo simulations, we generate an independent random realization of pseudorandom sounding beam codebook and channel parameters, unless otherwise mentioned.

The next evaluation focuses on the performance of the proposed algorithm in a realistic 3D mmW propagation environment where the sparsity is compromised, i.e., there are non-trivial angular and delay spreads within each multipath cluster. In Section 2.8.3 we simulate the system with QuaDRiGa simulator [90] based on mmMAGIC model [9] in 28 GHz urban-micro (UMi) environment. We remove *Assumptions 1, 2* from Section 2.3.1 and 2.5.1. UPA $N_T = 16 \times 4$, $N_R = 4 \times 4$ are used at BS and UE, respectively, to exploit the higher sparsity in the elevation plane. The proposed algorithm follows straightforward extension, namely, the estimated indices in dictionary (2.23) are mapped to AoA/AoD in both azimuth and elevation plane to fit into 3D environment. In the simulations, the transmit power is set to $P_{out} = 46$ dBm. The large scale channel model includes pathloss and shadowing. The AWGN on the receiver with 4 dB noise figure is added with power of $-170 + 10\log_{10}(\text{BW})$ dBm, where the noise bandwidth is $1/T_s$ and $B_{tot} = 400$ MHz [91] for IA and data stage, respectively. Moreover, the UE phase noise, $\psi[n]$ in (2.1), is modeled as Weiner process [50] that corresponds to

[9] We uses an optimistic DIA system where sector beams are synthesized by arrays with ideal phase and magnitude control.

oscillator with phase noise spectrum -114 dBc/Hz at 1 MHz offset [55]. The other detailed simulations setting in QuaDRiGa can be found in the supplementary material [178]. The DIA and beam training are also extended for UPA and 3D channel, i.e., frequency sampling method based sector beams are extended in both azimuth and elevation plane. During each one of N_{train} CSI-RS, BS and UE use 16 sounding beams pairs which bisect previous scanned azimuth and elevation angular regions. We use post-training SNR as performance indicator, which is evaluated by dividing channel gain $P_{\text{out}} \sum_{d=0}^{N_{\text{c}}-1} |(\mathbf{w}^{\star})^{\text{H}} \mathbf{H}[d] \mathbf{v}^{\star}|^2$ over noise power in B_{tot}.

Lastly, we evaluate the performance of the enhanced beam training under beam squint regime, where up to 2 GHz bandwidth for IA and beam training is considered. To demonstrate the impact of spatial wideband phenomenon, we assume that the transmitter uses linear arrangement with $N_{\text{T}} = 128$ antenna elements, and the AoD is chosen to be far from bore-sight, i.e., randomly drawn from $[35°, 45°]$. The UE uses $N_{\text{R}} = 16$ elements, and beam squint is not modeled at the receiver. The training SNR is 10 dB regardless of the bandwidth.

Unless otherwise mentioned, the simulation parameters are summarized in Table 2.3.

2.8.2 Performance in simplified Saleh-Valenzuela channel model

The miss detection rate[10] of the proposed approach for initial discovery is shown in Figure 2.5, and it is verified against the theoretical expressions (2.9). We have the following findings. Firstly, the lack of perfect timing synchronization introduces around 3 dB sensitivity loss as shown between the blue circled curve and red solid curve. However, this issue is unavoidable in practical systems. Secondly, less than 3 dB sensitivity loss occur when ± 5 ppm CFO is present in addition to STO, as shown by the light blue dashed and green dashed-and-dotted curves. Finally, the practical STO ($\leq$10 µs) is noncritical as shown by red solid and blue dashed curves. But when STO is large enough to cause transmitter and receiver burst beamforming window mismatch, e.g., 17 µs STO which corresponds to large $K(\epsilon_{\text{T}})$ in (2.11),

[10]Miss detection rate in simulation is evaluated by a generalized definition $\Pr(\gamma_{\text{NT}} > \eta_{\text{NT}}, \hat{\epsilon_{\text{T}}} = \epsilon_{\text{T}} | \mathcal{H}_1)$ in this proposed approach when $\epsilon_{\text{T}} \neq 0$.

Table 2.3: Summary of Simulation settings

Parameters	**Values in Simulations**
Frame Structure	
SSB Signal Bandwidth	$1/T_{\mathrm{s}} = 57.6\,\mathrm{MHz}$ [72]
Carrier and PSS Length	$P = 128$ [72]
Max Excessive Delay	$N_{\mathrm{c}} = \{4, 32\}$
Length of CP	$N_{\mathrm{cp}} = \{8, 32\}$
SSB Duration	$N_{\mathrm{B}} = 1024$ ($T_{\mathrm{B}} = 17.84\,\mu\mathrm{s}$) [72]
SSB Num.	$M = 64$ [72]; $M_{\mathrm{T}} = 16$, $M_{\mathrm{R}} = 4$
SSB Set Period	$T_{\mathrm{F}} = 20\,\mathrm{ms}$ [72]
Initial Synchronization Offset	
Freq. Offset at UE	Up to ± 5 ppm [3]
Timing Offset at UE	$\epsilon_{\mathrm{T}} = 170$ or 960, ($\Delta\tau = 3\,\mu\mathrm{s}$ or $\Delta\tau = 17\,\mu\mathrm{s}$)
STO Search Window	$\epsilon_{\mathrm{T,max}} = 1024$
Algorithm Design	
Target False Alarm	$P^{\star}_{\mathrm{FA}} = 0.01$
Dictionary Size	$G_{\mathrm{D}} = 500$, $G_{\mathrm{T}} = 2N_{\mathrm{T}}$, $G_{\mathrm{R}} = 2N_{\mathrm{R}}$

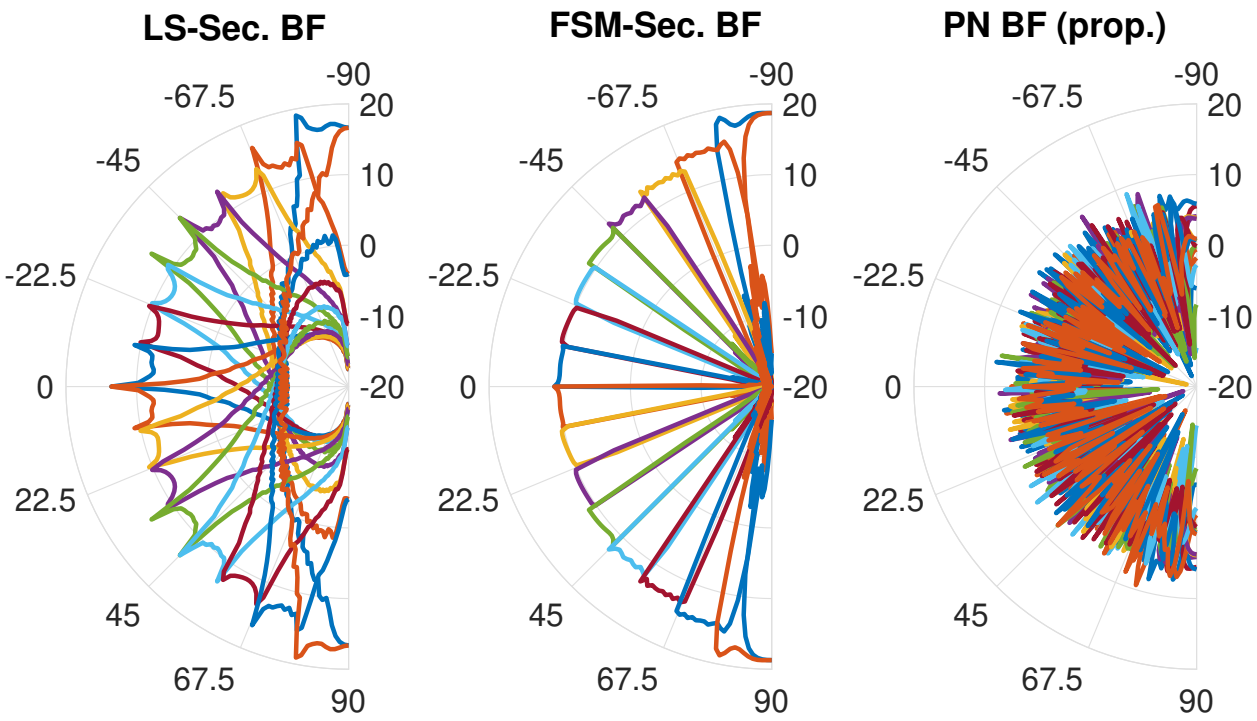


Figure 2.4: Beam patterns of two sector beam designs [15, 125] with $M_{\mathrm{T}} = 16$ transmit sectors and one realization of 16 pseudorandom beams. In the polar plot, the r-axis refers to the gain in decibel and the angular axis refers to steering angle in degrees. All patterns, i.e., least-square based sector beam (LS-Sec. BF), frequency sampling method based sector beam (FSM-Sec. BF), and pseudorandom sounding beam (PN BF), are simulated using uniform linear array with $N_{\mathrm{T}} = 128$ elements.

severe sensitivity loss is introduced as shown in grey dotted curves. In summary, these simulations verified the findings from Section 2.4 that practical initial synchronization error introduces up to few decibel sensitivity loss as compared to perfect synchronization scenario.

The comparison among proposed approach and benchmark DIA based discovery approaches is also presented in Figure 2.5. Although common sense may doubt the efficacy of the proposed approach since there is no significant angular gain for any beam pattern, as illustrated in Figure 2.4, the results show that there is only a couple of decibel difference among the proposed approach and DIA. However, such gap is less than the performance fluctuation of DIA with difference codebooks. The rationale behind this result is that the proposed scheme collects signal energy spread over all M SSB which in fact gives equivalent energy measurement as directional approach where energy collection occurs only when a sector beam aligns with true propagation direction.

The beam training performance of the proposed BF training algorithm in LOS is presented in Figure 2.6. The performance metrics are the residual root mean square error (RMSE) defined by $\mathrm{RMSE}_{\mathrm{AoA}} = \sqrt{\mathbb{E}|\hat{\phi}_1 - \phi_1|^2}$ and $\mathrm{RMSE}_{\mathrm{AoD}} = \sqrt{\mathbb{E}|\hat{\theta}_1 - \theta_1|^2}$. The simulations are conducted with *Assumption 2*. The same pseudorandom setting is used in both simulation and theoretical CRLB evaluation. The refinement steps are forced to terminate in up to 100 iterations. We have the following findings. Firstly, when the off-grid refinement are used, the proposed algorithm reaches CRLB in high SNR regime. Secondly, the coarse estimation in high SNR has a compromised performance as compared to CRLB. However coarse estimation (without refinement) has adequate accuracy for beam steering since RMSE is order of magnitude lower than 3 dB beam-width in steering, i.e., $102^\circ/N_{\mathrm{T}}$ and $102^\circ/N_{\mathrm{R}}$. Finally, Figure 2.5 and 2.6 reveal that in SNR region between -15 dB and -7.5 dB reliable detection occurs but beam training performance is poor. Admittedly, this implies a compromised experience for UEs at the cell edge, which is worth further investigation.

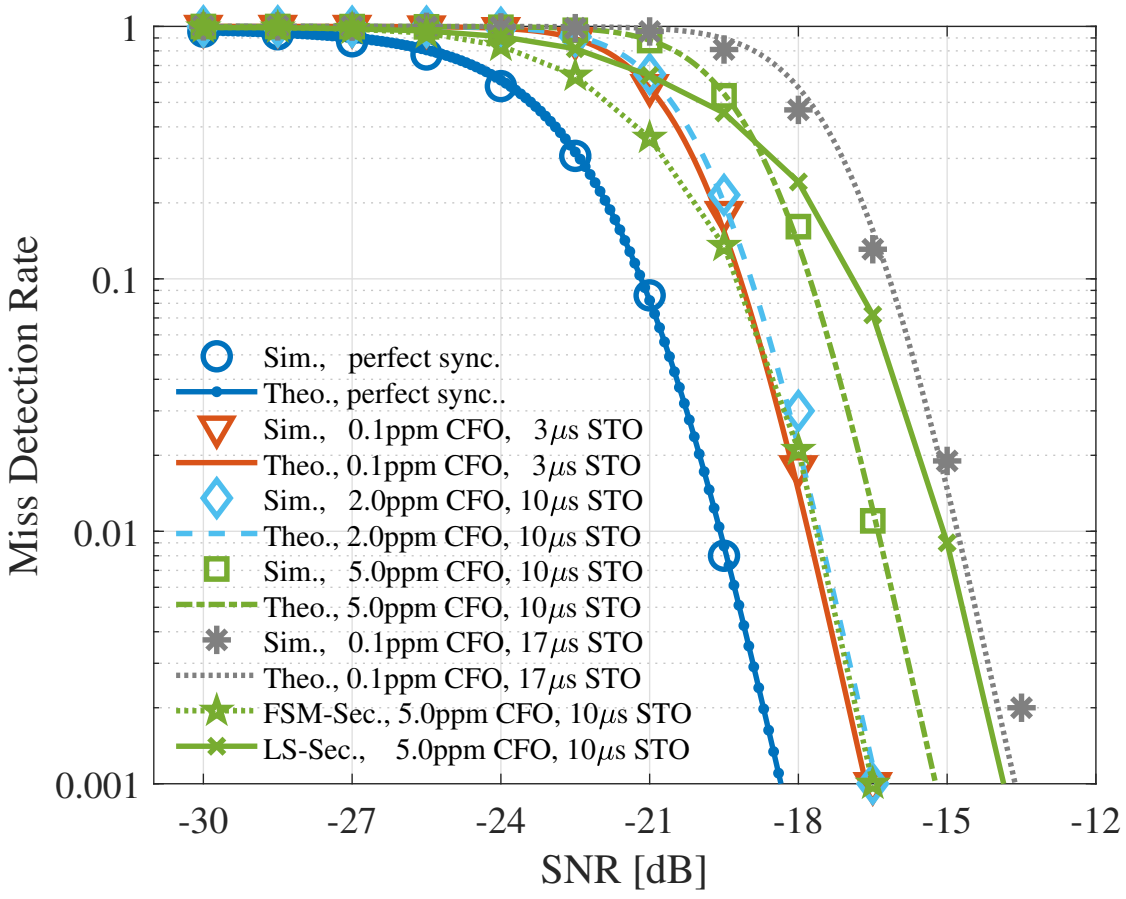


Figure 2.5: Simulated (Sim.) and theoretical (Theo.) results of the miss detection rate of the proposed initial discovery with various synchronization errors. The discovery rate of the directional initial access is also included as benchmark and both least square based sector beam (LS-Sec) and frequency sampling method based sector beams (FSM-Sec). The base station and user equipment have $N_{\mathrm{T}} = 128$ and $N_{\mathrm{R}} = 32$ uniform linear array and Saleh-Valenzuela channel has $L = 2$ multi-paths.

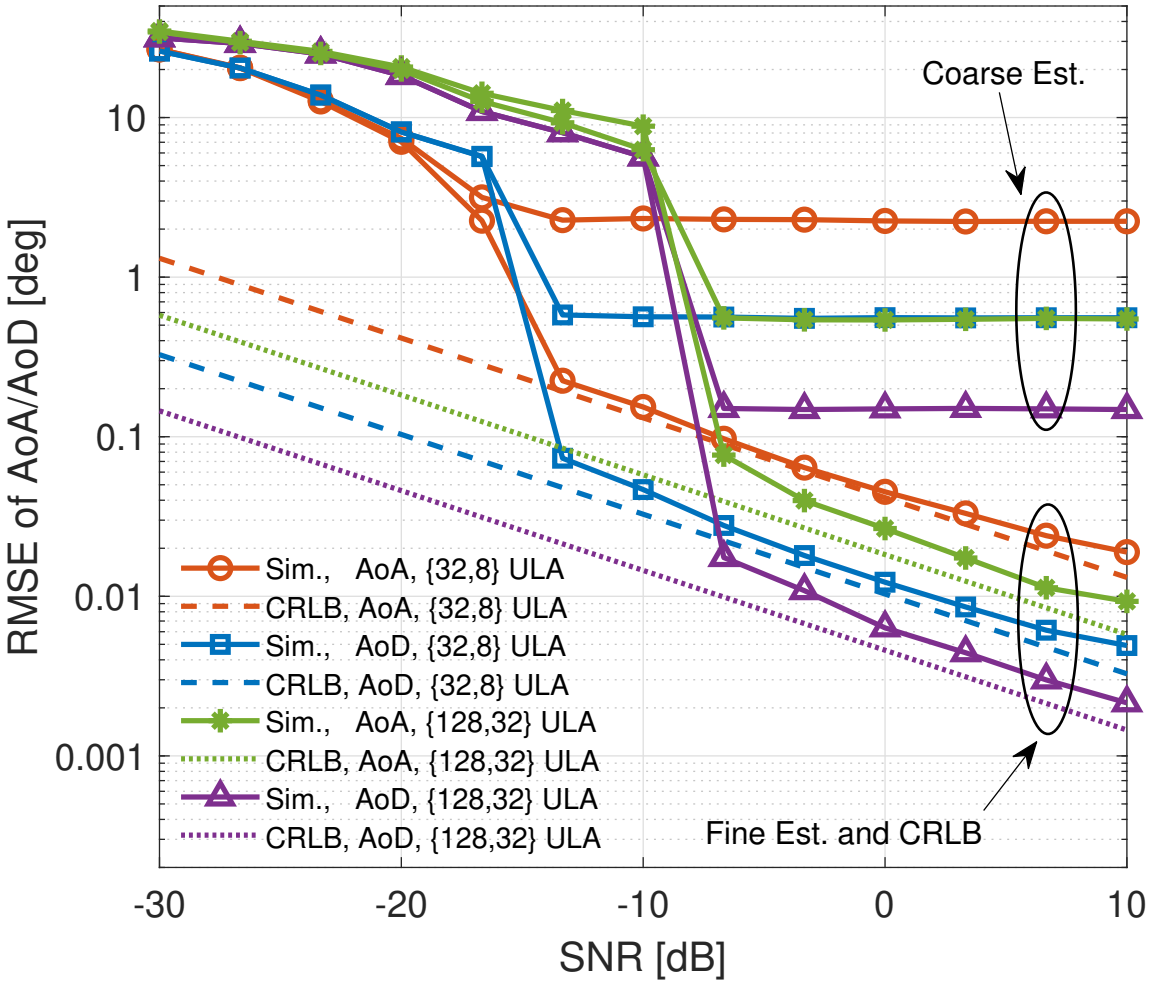
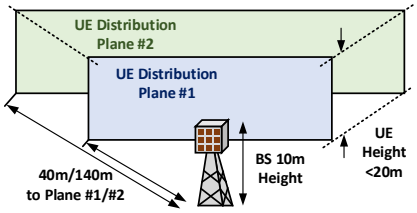


Figure 2.6: Simulated results of the proposed algorithm, with and without refinement steps, and theoretical bound of root mean square error of angle-of-arrival estimation in line-of-sight environment. System with different size of uniform linear arrays are evaluated. Firstly, base station and user equipment have $\{N_{\mathrm{T}}, N_{\mathrm{R}}\} = \{32, 8\}$ antennas. Secondly, they have $\{N_{\mathrm{T}}, N_{\mathrm{R}}\} = \{128, 32\}$ antennas. System has 5ppm carrier frequency offset (CFO).

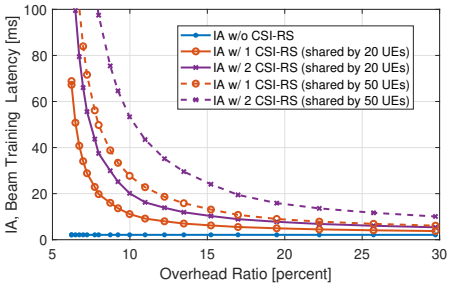
2.8.3 Performance in QuaDRiGa channel simulator

Figure 2.7 (a) illustrates the network setting implemented in QuaDRiGa. We simulate the performance of typical UEs distributed in two planes, with different distance towards the pico-cell mmW BS. We present the following findings based on Figure 2.7 (b), which shows the cumulative distribution function (CDF) of post-training beam steering SNR. Firstly, the proposed approach provides comparable performance to DIA with $N_{\text{train}} = 2$ CSI-RS. In fact, in LOS, both approaches closely achieve beam steering towards true LOS path. Although the SNR seems excessively high in LOS, this implies that the transmit power can be reduced to save power. Secondly,DIA with less than $N_{\text{train}} = 2$ CSI-RS has compromised SNR performance. This drawback is intuitive because wide sounding sector beam fails to extract precise angle information. The SNR improvement of using higher N_{train} is more significant in LOS. Thirdly, although the proposed approach is tailored for sparse channels and presence of phase measurement error due to CFO, it is robust in NLOS scenarios where channel sparsity is compromised and practical phase noise occurs. Admittedly, the algorithm has a certain chance to completely fail when NLOS UEs are distributed in the second plane. However, in these cases the counterparts based on DIA and CSI-RS training cannot do much better job either. In fact, they have lower probability to reach post-training beam steering SNR above 0 dB compared to the proposed approach.

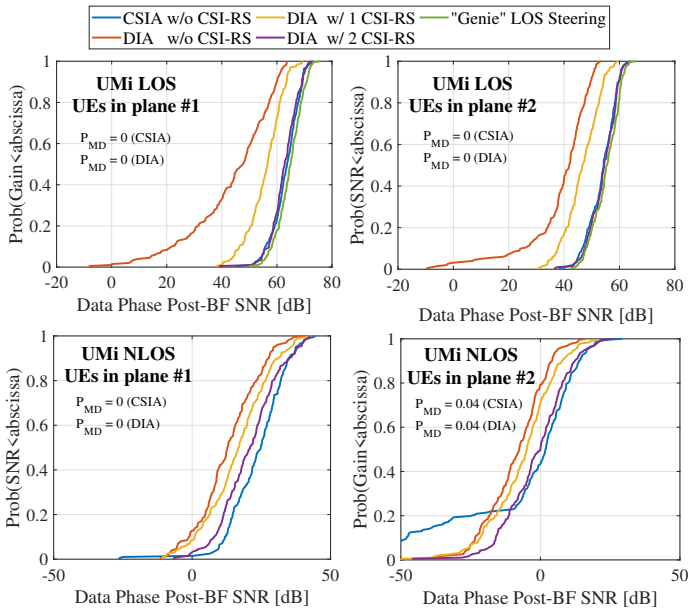
The overhead and initial access latency savings of the proposed approach are significant, since it does not require CSI-RS, as shown in Figure 2.7 (c). As explained in Section 2.7, for DIA based approaches when number of UEs in the network increases, the latency increases dramatically due to CSI-RS scheduling. Increasing the density of CSI-RS effectively reduces latency, but it results in increased overhead. The proposed approach relies on advanced signal processing to digitally conduct beam training and avoids requesting CSI-RS after initial access. In summary, up to two order of magnitudes saving in initial access latency is reached as compared to DIA.



(a) Network illustration where a user equipment (UE) is randomly distributed in the horizontal plane with height of UE within 20 m.



(c) Access latency versus overhead of both initial access scheme (2.33) and (2.34) with different number of UEs that share the scheduled CSI-RS.



(b) The cumulative distributed function of post-training beam steering signal-to-noise ratio in the data phase. For the directional initial access (DIA), different number of channel state information reference signal (CSI-RS) N_{train} are considered. The proposed compressive initial access (CSIA) does not required CSI-RS. The signal-to-noise ratio distribution corresponding to beam steering towards true line-of-sight path (when existing) is also included as benchmark. The miss detection rates are included in each plots.

Figure 2.7: Initial access and beam training of proposed and directional beam training evaluated in 3D outdoor urban micro (UMi) network using 28 GHz mmMAGIC channel model [9]. The trade off between post-training signal-to-noise ratio in the data phase, required overhead, and access latency are also studied.

2.8.4 Performance under beam squint phenomenon

Figure 2.8 shows the AoD estimation accuracy as a function of beam training channel bandwidth. The proposed approach retains non-compromised accuracy throughout an entire range of bandwidths since the squint tailored dictionary is accounted in the processing. The squint non-aware method (benchmark training 2) has increased error when bandwidth increases.

The post training BF gain in data phase is also important performance indicator. To better understand the impact of AoD estimation accuracy, we evaluate two beam design candidates in the transmitter, wide beams and pencil beams, that steer the beam in direction reported by corresponding beam training algorithm. The training occurs over 2 GHz bandwidth, and its AoD estimation accuracy can be inferred from Figure 2.8. We assume analog architecture, therefore a single precoding vector is applied to all subcarriers. The wide beam precoder is designed with critical beam width based on Section 2.6, with vector coefficients as in [125, Chapt 23.4]. The pencil beam uses conventional steering vector. The beamforming gain across a wideband range is presented in Figure 2.9. We have the following findings. First, utilizing the beam squint non-aware CS approach, the pointing direction of beams could be completely mis-aligned with the true propagation angle, thus resulting in significant loss of beamforming gain, for both pencil and wide beams. Second, by utilizing the wide beam width, our proposed squint aware algorithm achieves almost constant BF gain across 1500 MHz bandwidth. Lastly, RMSE of AoD $< 0.1^\circ$ is sufficient to achieve broadband BF gain when wide beams are used.

2.8.5 Baseband processing requirements

Using the simulation parameters in Table 2.3 to evaluate required operations in Table 2.2, the baseband resource of the proposed method are in the same order of magnitude with DIA, i.e., $(PN_{\mathrm{B}} + PG_{\mathrm{d}} + 3MG_{\mathrm{T}}G_{\mathrm{R}})/(PN_{\mathrm{B}}) \approx 7.2$. There are two reasons for this finding. Firstly, exhaustive PSS correlation filter (2.3) is extreme computational demanding in IA.

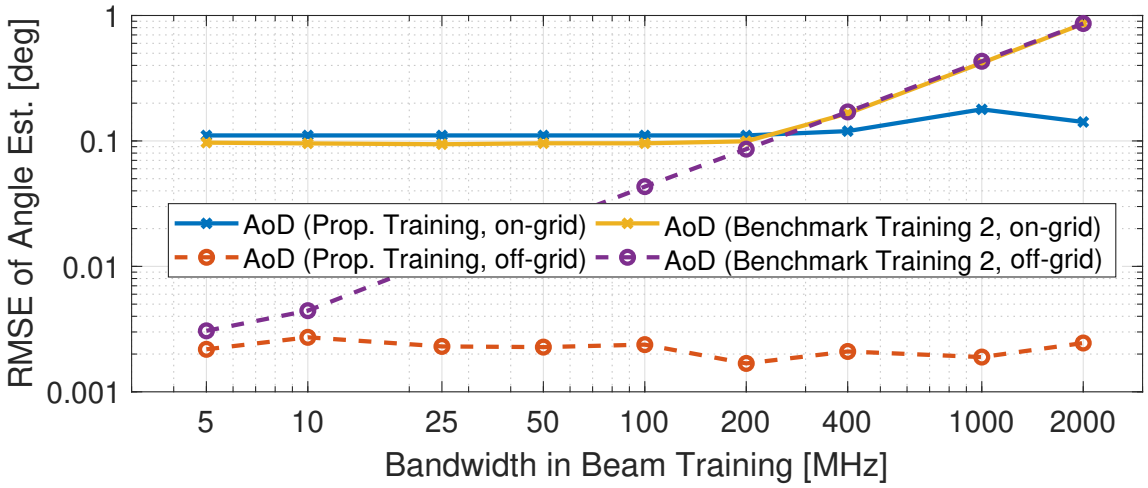


Figure 2.8: Angle-of-departure (AoD) estimation accuracy of the proposed squint robust training and the benchmark beam training method 2 for different training bandwidths.

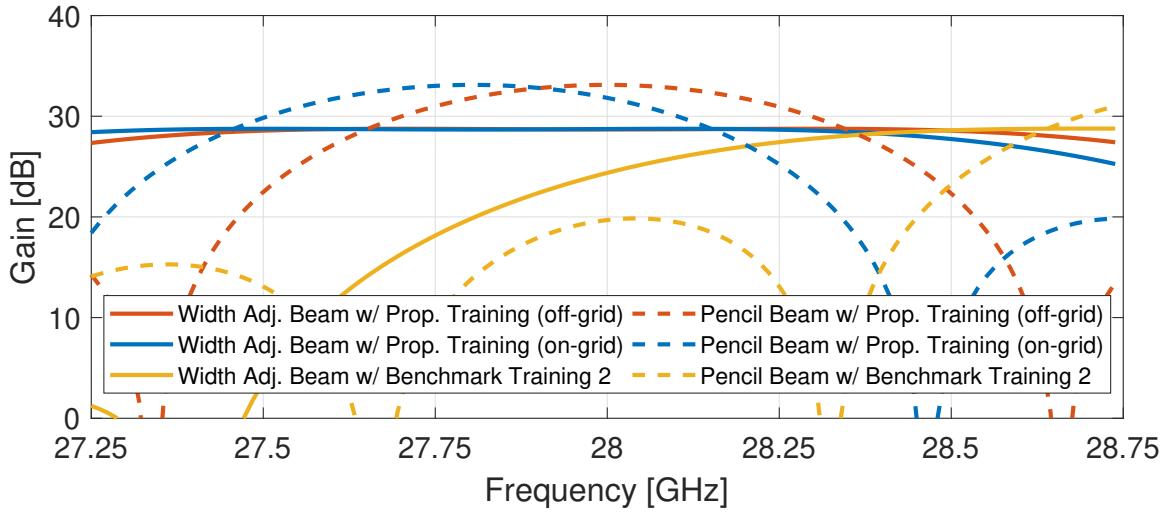


Figure 2.9: Beamforming gain after beam training across a wide frequency range. Two types of beams are compared: width adjusted beam that utilizes approach in Section 2.6 and pencil beam that uses steering vector. Beams are pointed to the center direction based on beam training.

This filter is required by IA regardless of sounding beam design. Secondly, the proposed approach sequentially estimates parameters and avoids multi-dimensional grid search. This feature is particularly appealing in wideband operation since the savings as compared to the benchmark training method 2 whose complexity scales with P, number of subcarriers.

2.9 Discussion on open issues

In this section, we discuss relevant issues in practical implementation of compressive IA and beam training.

Required a priori knowledge: Firstly, this work assumes coarse timing is available. It would be also important to study the case when timing is completely unknown, i.e., there is no a priori information about the range of ϵ_{T} in (2.1), which could cause SSB index misalignment to occur. Secondly, the compressive approach requires precise information about the sounding beam pattern $\tilde{\mathbf{a}}_k$ in (2.21). As a results, array geometry and sounding codebooks of both BS and UE need to be known a priori. This raises new challenges in communication protocol design to effectively incorporate this information. It also requires an increase in baseband operations if all dictionaries need to be computed on-the-fly. Further, mmW testbed experiments in [150] showed that the measured beam patterns commonly have mismatches with patterns predicted by codebook and array geometry model. Future research should address these impairments.

Channel sparsity: The efficacy of compressive approach is affected by the sparsity level in AoAs, AoDs, and multipath delays. Sparsity is endorsed by various mmW channel measurement campaigns, and urban NLOS, which is known with unfavorable sparsity, is tested in this work. However, severely rich scattering situation are modeled from standard perspective [6]. It is important for system that utilize CS-based approach to flexibly handle situation when channel sparsity disappears.

Array architecture: This work focuses on the scenario where UE uses a single RF-chain to process a single stream of IA signals. This allows other RF-chains, if available at BS

or UE, to operate in the band of data communication during IA. Since [30] shows that the hybrid analog/digital array and fully digital array are advantageous for DIA, it would be interesting to investigate benefits of compressive IA and beam training algorithm when they are adapted to utilize multiple RF-chain.

MIMO Multiplexing: The proposed beam training is compatible with multi-user multiplexing. In fact, multiplexing designs [16, 198] rely on each RF–chain and corresponding analog beamformer to provide adequate post-BF SNR, and use the digital baseband processing to handle multi-beam interference. However, as mentioned in *Remark 1*, the comparison with channel estimation based approaches, i.e., estimation of the entire wideband channel or its covariance during CSI-RS for optimal MIMO processing, is rarely investigated.

Phase coherency: To date, there is no coherent CS-based beam training prototype reported in mmW band. The only notable prototype [31] operates at 8 GHz with two phased arrays synchronized by cabled reference clock. Prototype [26] utilizes channel emulator to avoid issue of phase coherency. In addition to CFO, as emphasized in this work, the phase noise can also severely degrade coherency among channel observations. The phase noise detrimental impact becomes more severe with increased carrier frequency and shall be properly modeled and incorporated in signal processing techniques for mmW [190]. Proper phase noise compensation as well as non-coherent CS-based beam training [81, 141, 142] are naturally immune to phase error and are worth investigation.

2.10 Conclusions

In this work, quasi-omni pseudorandom sounding beam is proposed for the mmW initial access, synchronization, and beam training. We design associated signal processing algorithm based on the proposed sounding beam structure that is compatible with 5G-NR frame format. We provide theoretical analysis of cell discovery rate and CRLB of beam training performance, and evaluate them via simulations using the mmW hardware and urban channel models from the literature that are supported by measurements. The results show that

the proposed approach provides comparable performance to the state-of-the-art directional cell search for initial discovery, but achieves significantly more accurate angle estimation during initial beam training. This advantage holds true across different propagation condition (LOS/NLOS) and UE-BS distance at 28 GHz band. Due to the saving of additional radio resource, i.e., CSI-RS, for beam refinement, the proposed approach reduces up to two order of magnitude access latency compared to the directional initial access when the same signaling overhead and post-training beam steering SNR are targeted. An enhanced beam training algorithm that is robust to beam squint is proposed for future mobile network evolution when increased IA channel bandwidth is considered. The proposed squint robust CS based beam training algorithm is able to retain non-compromised AoD estimation accuracy and beamforming gain across a wide range of beam training bandwidths.

All numerical results are reproducible with scripts in [178].

CHAPTER 3

Implementation of Machine Learning Assisted Compressive Beam Alignment

3.1 Introduction

The mmW communication is a promising technology for future wireless networks, including 5G-NR and 60 GHz Wi-Fi. Due to abundant spectrum, mmW networks are expected to support ultra-fast data rates. As shown in both theory and prototypes, mmW systems requires BF with large antenna arrays and narrow beams at both the transmitter (Tx) and receiver (Rx) to combat severe propagation loss. Before data communication, directional beams must probe the channel to select a beam pair with adequate beamforming gain. This procedure is referred as beam alignment[1]. Existing mmW systems use analog phased arrays with beam sweeping, an exhaustive search approach, for beam alignment. However, this method introduces high communication overhead. Further, the required number of channel measurements linearly scales with number of antenna elements, a design parameter expected to increase with the evolution of mmW networks. In this work, we present the design and implementation of compressive beam alignment, a framework that accelerates beam alignment by allowing channel probing to logarithmicly scale with antenna array size. A neural network assisted design addresses the challenge of non-trivial model mismatch due to array hardware impairment. Our implementation achieves near perfect beam alignment in our experiment with a 60 GHz testbed and reduces 92.2% of the overhead as compared to exhaustive beam sweeps. To the authors' best knowledge, this is the first work to demon-

[1]It is also referred as beam training. However, we reserve training as a machine learning terminology in this chapter.

strate machine learning based beam alignment using a 60 GHz phased array testbed and real measurement data.

The rest of the chapter is organized as follows. Section 3.2 surveys mmW fast beam alignment designs and proofs-of-concept. In Section 3.3, we present the problem statement and the reasoning behind using machine learning to solve it. The proposed design is presented in Section 3.4 followed by the implementation details with our 60 GHz testbed in Section 3.5. The experimental results are presented in Section 3.6. Finally, Section 3.7 concludes the chapter.

Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. The (i, j)-th element of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. Similarly, The i-th element of a set $\mathcal{A}$ is denoted by $[\mathcal{A}]_i$. Conjugate, transpose, and Hermitian transpose are denoted by $(.)^*$, $(.)^{\mathrm{T}}$, and $(.)^{\mathrm{H}}$ respectively. Inner product between vector $\mathbf{a}$ and $\mathbf{b}$ is denoted as $\langle \mathbf{a}, \mathbf{b} \rangle$. $|\mathbf{a}|$ returns vector with magnitude of each element of $\mathbf{a}$.

3.2 Related works

Beam alignment for mmW is an active research area. While one direction is to focus on hardware innovations, e.g., fully-digital array and novel antenna design [69], others rely on signal processing.

Model-based signal processing algorithms for beam alignment mainly rely on the sparsity of mmW channels and a phased array model. From this class of algorithms, hierarchical beam alignment and compressive sensing based beam alignment each have overheads that logarithmically scale with array size. The former uses sounding beams that adapt with previous measurement, bisecting the beam width to reduce the search space [53]. The latter is based on either CS, i.e., with coherent complex sample measurements, or compressive sensing phase retrieval (CPR), i.e., with noncoherent RSS measurements, that exploit sparsity of mmW propagation [83, 141, 183]. However, the model mismatch due to channel and radio hardware impairments introduces non-trivial challenges.

Data-driven signal processing learns from extensive data to infer the best beam from various low-overhead, in-band measurements and/or out-of-band information. In-band measurements include channel impulse response estimated by omni-directionally received pilots [14] and a proportion of exhaustive beam search results [40, 107, 166]. Out-of-band information can be the terminal's geo-location [84,166]. To date, these works either use a statistic channel model [40, 107] or ray tracing simulations [84, 166] to generate data.

With the increased availability of mmW testbeds, there are many proofs-of-concept. Work [26] reports a chip-level demonstration of CS based beam alignment with a channel emulator. Works [141,196] showcase fast alignment by solving CPR problems. Works [24,81] design and demonstrate fast beam alignment using multi-lobe sounding beams and combinatorics inspired algorithms. Work [67] reports experimental work that effectively reduces overhead when more than one spatial streams is used in a hybrid array. Finally, some prototypes also rely on the side information, e.g., sub-6 GHz [124, 162] and visible light [76] measurements, for mmW beam alignment.

3.3 Noncoherent compressive beam alignment

In this section, we start with the mathematical model and problem of non-coherent fast beam alignment. Next, we describe the state-of-the-art model-based solutions and their limitations.

3.3.1 System model and problem statement

We consider mmW communication between an BS Tx and a UE Rx. The BS and UE are each equipped with an analog linear array with N_{T} and N_{R} elements. The channel between them follows L-path geometric model $\mathbf{H} = \sum_{l=1}^{L} g_l \mathbf{a}_{\mathrm{R}}(\phi_l)\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\phi_l)$, where $\mathbf{a}_{\mathrm{T}}(\theta) \in \mathbb{C}^{N_{\mathrm{T}}}$, $\mathbf{a}_{\mathrm{R}}(\phi) \in \mathbb{C}^{N_{\mathrm{R}}}$, and $g_l \in \mathbb{C}$ are the array responses in BS and mobile terminal (MS), and gain of the l-th path, respectively. Array responses are defined by their n-th element, i.e., $[\mathbf{a}_{\mathrm{R}}(\phi)]_n = \exp(j2\pi(n-1)d/\lambda \sin(\phi))$ and $[\mathbf{a}_{\mathrm{T}}(\theta)]_n = \exp(j2\pi(n-1)d/\lambda \sin(\theta))$, where d, λ, ϕ

and θ are the element spacing, carrier wavelength, AoA and angle of departure, respectively. We focus on the UE Rx side by assuming the the BS Tx AWV $\mathbf{v} = \mathbf{a}_{\mathrm{T}}(\theta_1)/\sqrt{N_{\mathrm{R}}}$ is pre-designed. Thus, the channel model in the rest of the chapter is

$$\mathbf{h} = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{R}}(\phi_l) \approx \alpha_1 \mathbf{a}_{\mathrm{R}}(\phi_1) \triangleq \alpha \mathbf{a}_{\mathrm{R}}(\phi^\star), \tag{3.1}$$

where $\alpha_l = \alpha_l \mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l)\mathbf{v}$ is the post-Tx-beam channel gain. The selection of Tx beam leads to $|\alpha_1| \gg |\alpha_l|, l > 1$ and it results in the approximation sign in (3.1). We defined $\phi^\star$ as the true AoA of the channel. When the Rx uses AWV $\mathbf{w}$, the received symbol is

$$y = \mathbf{w}^{\mathrm{H}}\mathbf{h}s + n, \tag{3.2}$$

where s is the Tx symbol and n is the post-combining thermal noise which is modeled as additive white Gaussian noise with variance σ_{n}^2. Without loss of generality, we let $s = 1$ and define $\mathrm{SNR} = |\alpha|^2/\sigma_{\mathrm{n}}^2$.

We consider a codebook based communication protocol which consists of two phases: a beam alignment phase and a data communication phase. The channel is unknown to UE Rx but can be assumed invariant between these two phases. During beam alignment, the UE Rx uses a sounding codebook, $\mathcal{W}_{\mathrm{S}}$ (with $|\mathcal{W}_{\mathrm{S}}| = M$ codewords), to probe the channel. The associated measurements are processed to select the best beam from a fixed directional codebook $\mathcal{W}_{\mathrm{D}}$ (with $|\mathcal{W}_{\mathrm{D}}| = K$ codewords), which is then used in the data communication. Each codeword of directional codebook is a steering vector, i.e., $[\mathcal{W}_{\mathrm{D}}]_k = \mathbf{a}_{\mathrm{R}}(\theta_k)/\sqrt{N_{\mathrm{R}}}$, and these directions $\{\theta_k\}_{k=1}^{K}$ cover an angular region of interest.

There are three additional assumptions from our implementation perspective. Firstly, the codebook $\mathcal{W}_{\mathrm{S}}$ is loaded into hardware in advance, and each codeword is called in a sequential manner. Adaption that uses on-the-fly measurements to change either the codebook or the codeword selection order, e.g., a hierarchy search, is not desired. In fact, we focus on pseudo-random sounding codebooks $\mathcal{W}_{\mathrm{S}}$, a well adopted design from compressive sensing literature when the Rx does not have prior knowledge of the channel [83, 141, 142, 183]. Specifically, the magnitude of each AWV in $\mathcal{W}_{\mathrm{S}}$ is $1/\sqrt{N_{\mathrm{R}}}$ and the phase is randomly picked from a set of $\{0, \pi/2, \pi, 2\pi/3\}$. These are referred to as pseudorandom noise (PN) beams in the remainder

of the chapter. Secondly, the received symbols in (3.2) are not directly observable. Instead, each channel measurement is taken from the preamble of a package, a sequence of pilot symbols, in the form of RSS. Lastly, the phased array experiences hardware impairment. An optimistic assumption is to model them as phase offset, i.e., multiplicative error $\mathbf{e} \in \mathbb{C}^{N_{\mathrm{R}}}$ irrelevant to codeword selection [196]. With the above assumptions, the M channel probings give RSS $\mathbf{p} = [p_1, \cdots, p_M]^{\mathrm{T}}$ where the m-th probing is

$$p_m = |\tilde{\mathbf{w}}_m^{\mathrm{H}} \mathbf{h}| + n_m. \tag{3.3}$$

In the equation, $\tilde{\mathbf{w}}_m = \mathrm{diag}(\mathbf{e})[\mathcal{W}_{\mathrm{S}}]_m$ is the receiver combiner with hardware impairment, and n_m is the error in RSS measurement. To that end, the compressive noncoherent beam alignment problem is:

Problem Design a signal processing algorithm that uses the noncoherent measurements $\mathbf{p}$ from (3.3) to infer the best directional beam for data communication, i.e., $\hat{\mathbf{w}} \in \mathcal{W}_{\mathrm{D}}$.

The performance metrics are the required number of measurements M and the post-alignment gain, i.e., normalized BF gain in data communication phase $G = |\mathbf{h}^{\mathrm{H}} \mathrm{diag}(\mathbf{e}) \hat{\mathbf{w}}|^2 / \|\mathbf{h}\|^2$. Note that an exhaustive search uses the same codebook for both alignment and communication, i.e., $\mathcal{W}_{\mathrm{S}} = \mathcal{W}_{\mathrm{D}}$. It is straightforward to find the optimal codeword $\mathbf{w}^\star = \max_{\mathbf{w} \in \mathcal{W}_{\mathrm{D}}} G$ with overhead cost $M = K$. The goal is to reduce M while introducing marginal impact to post-alignment gain as compared to an exhaustive search[2].

3.3.2 Model based solution and its limitation

The beam alignment with measurement (3.3) can be formulated as a CPR problem when the error $\mathbf{e}$ is assumed to be negligible, i.e.,

$$\mathbf{p} = |\mathbf{W}^{\mathrm{H}} \mathbf{h}| + \mathbf{n} = |\mathbf{W}^{\mathrm{H}} \mathbf{A}_{\mathrm{R}} \mathbf{g}| + \mathbf{n} \triangleq |\mathbf{\Psi} \mathbf{g}| + \mathbf{n}. \tag{3.4}$$

In the above equation, $[\mathbf{A}_{\mathrm{R}}]_k = \mathbf{a}_{\mathrm{R}}(\theta_k)$, θ_k is the steering directions in DFT codebook (the AoA hypothesis), and $\mathbf{g} \in \mathbb{C}^K$ is a sparse vector with all-zero elements except the $k^\star$-th being

[2]With increased DFT codebook size K, beam steering with AWV $\mathbf{w}^\star$ is asymptotically the same as the steering towards ground truth AoA $\phi_\star$. Thus, we do not directly compare with the latter in this work.

α (i.e. associated with true AoA $k^\star = \{k|\theta_k = \phi^\star\}$). The error vector is $\mathbf{n} = [n_1, \cdots, n_M]^{\mathrm{T}}$. The sensing matrix $\mathbf{W}_m$ is from error-free sounding AWVs $[\mathbf{W}]_m = \mathbf{w}_m$. The solution to a general CPR model is guaranteed with an adequate number of measurements M, which linearly scales with the sparsity level, i.e., non-zero elements in $\mathbf{g}$, and logarithmic scales with K [153]. Solutions to a general CPR can use convex optimizations [141,196] or approximate massage passing [153]. Solving CPR in this work, or finding the sparsity level 1 vector $\mathbf{g}$ from $\mathbf{p}$, directly leads to a solution of beam alignment since the non-zero element, say $\hat{k}$-th, can simply be used to select the best beam from DFT codebook $\hat{\mathbf{w}} = [\mathcal{W}_{\mathrm{D}}]_{\hat{k}}$. In this special case, the heuristic approach of received signal strength matching pursuit (RSS-MP) [142] also applies, where $\hat{k} = \arg\max_k \langle \mathbf{p}, |[\mathbf{\Psi}]_k| \rangle / \|[\mathbf{\Psi}]_k\|$.

The key concern of existing CPR solutions is the required knowledge of $\mathbf{\Psi}$ in (3.4). With hardware impairment, the sensing matrix $\mathbf{W}_m$ in (3.4) is composed of distorted sounding AWVs $[\mathbf{W}]_m = \mathrm{diag}(\mathbf{e})\mathbf{w}_m$. This can be problematic in a practical radio for three reasons. Firstly, in the production of radio hardware, the phase offsets $\mathbf{e}$ are due to a common offset among all devices and an additional device dependent offset. The former comes from phase rotation in the printed circuit board (PCB) routing between phase shifters and antenna ports. The latter is due to PCB fabrication variation. It is cost-effective to only calibrate and compensate the common offset and leave device-dependent ones untreated. Secondly, the mainlobe of DFT pencil beam is not sensitive to phase offset. Although sidelobes are more vulnerable to distortion, they are not directly used in the mmW system with beam sweep and beam steering. Thus, leaving a device-dependent array phase offset can be acceptable. Lastly, the beam patterns of PN AWVs are sensitive to phase offset. Hence, using a CPR model to directly design beam alignment is likely to experience model mismatch and degraded performance.

The above argument is quantitatively studied in Appendix B.1. In actual radios, modeling $\mathbf{e}$ is more challenging, as errors can also be associated with other array-inherent defects. An empirical evaluation of this concern is shown in Section 3.6.

3.4 Neural network assisted compressive beam alignment

To address the model mismatch in solving CPR due to hardware impairment, we propose a data driven approach for beam alignment.

3.4.1 Main idea

The key insight of our approach is that, although analytically solving the noncoherent beam alignment problem using model (3.4) is difficult, its solution can be easily found by exhaustive beam sweep. Therefore, we can resort to a data driven approach to learn how to solve the CPR problem (3.4) with unknown deterministic offset. The proposed system contains two stages, each covering a much longer time scale than the beam alignment or communication phases. We refer the first stage as the learning stage, where the radio uses a concatenated codebook $\mathcal{W} = \mathcal{W}_{\mathrm{D}} \cup \mathcal{W}_{\mathrm{S}}$ for multiple beam alignment phases. Specifically, the sounding results from exhaustive search ($\mathcal{W}_{\mathrm{D}}$) provides the solution to the beam alignment problem; the *labels* in machine learning terminology. The sounding results from $\mathcal{W}_{\mathrm{S}}$ are treated as the *features*, whose statistical relationship with the *labels* can be extracted by machine learning tools, e.g., neural network (NN) or support vector machine. Admittedly, the beam alignment overhead increases by M/K as compared to exhaustive search in this stage. The beam alignment *features* and *labels* must be collected in various environments to reliably generalize their relationship. In a practical system, this would arise from randomness in physical position and orientation of the MS, e.g., a phone held by a human with different posture in different places. In fact, the learning stage can be completely ambient and does not require dedicated interaction from the user [14]. We refer the second stage as the operation stage, where UE only uses codebook $\mathcal{W}_{\mathrm{S}}$ for beam alignment, compressing the overhead by M/K. The algorithm then only uses the *feature* to predict the *label*, i.e., the best beam $\mathbf{w}^{\star}$.

3.4.2 Neural network design

In this work, we designed a dense NN to classify the optimal gain DFT beam for a given antenna position, based on PN beam RSS measurements. The network used 3 fully connected (FC) layers, each using rectified linear unit (ReLU) activation functions. For all values of M tested, we used the same network architecture, with 64, 128, and K units in the first through third FC layers respectively.

Input data RSS data was expressed in linear scale and normalized by the maximum value. These feature transformations limit the data to the range $[0, 1]$, prevent activation function saturation, and improve the learning performance. Batch normalization layers were also used just before the ReLU activations in first and second FC layers as feature regularization to improve training efficiency.

Our design used a sparse categorical cross entropy loss function to produce our classification results over the K possible DFT beam physical angle labels. For training, we used the RMSprop optimizer with this loss function. The network architecture was implemented and trained in in Keras/Tensorflow. The total number of trainable parameters in this network depends on the input feature dimension (M) and the label dimension (K): $64M + 129K + 8768$.

3.5 Implementation with millimeter-wave testbed

In this section, we start with a brief introduction to the mmW testbed, including the capability of the hardware and software. Next, we describe our implementation of NN based beam alignment.

3.5.1 Millimeter-wave testbed

Our testbed is the Facebook Terragraph (TG) channel sounder, a pair of TG nodes customized for measurements of 60 GHz channels. Although the testbed is designed for model-

Table 3.1: Summary of millimeter-wave testbed

Parameters	**Values**
RF channel	IEEE802.11ad (58.32/60.48/62.64 GHz)
Array geometry	Planar array (36 by 8 elements)
Array control (az)	Individual phase (6 bits) & on-off control
Array control (el)	Fixed beam (bore-sight)
Scan region (az)	$-45°$ to $45°$ (due to element gain pattern)
Beam-width (az)	$2.8°$ (with directional codebook)
Max EIRP	45 dBm

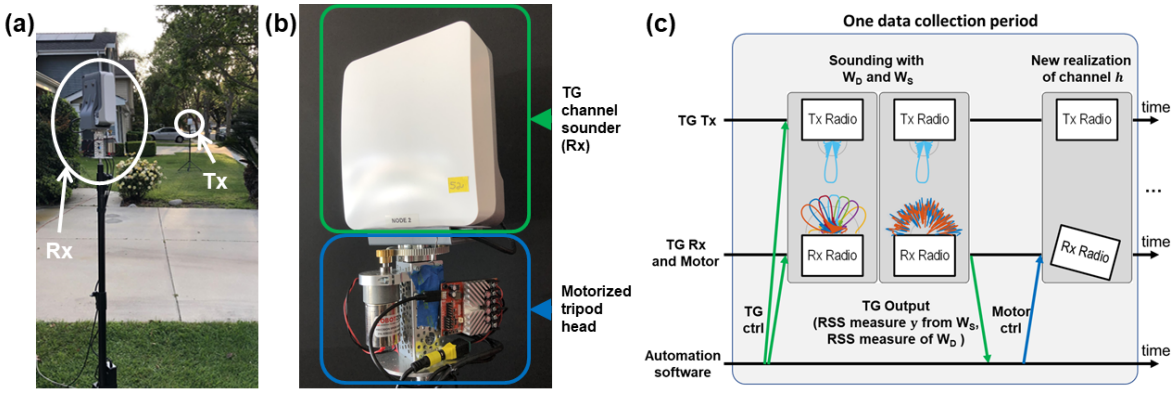


Figure 3.1: Overview of the experimental testbed. (a) testbed environment. (b) receiver setup. (c) testbed data capture procedure.

ing the channel characteristics such as the path loss, angular-power profile, and delay-power profile, e.g., [32], its capability adequately supports the compressive beam alignment experiments. The key hardware specifications of the testbed are listed in Table 3.1.

The testbed has an application programmable interface (API) that allow a host computer to customize AWVs when transmitting or receiving packets. The API also provides measurements from the received IEEE 802.11ad preamble, e.g., received signal strength indicator, short-training field specified SNR, and long-training field specific channel impulse response estimation. Note that although multiple automatic gain control amplifiers are involved in

preamble measurements, a look-up table is used to make the impact of automatic gain controller amplifier transparent, which indicates the fidelity of model (3.3). The array phase offset was calibrated and compensated using a golden design instead of in a device-by-device manner.

3.5.2 Data capture automation

As mentioned in Section 3.4, it is desired for the NN to learn from beam alignment data collected in various of locations and physical orientations of radios. To automate this procedure, we used a programmable motor underneath the testbed receiver. We designed a 3D printed bracket to attach the receiver to a motorized turntable kit (Servocity Gear Drive Pan Kit) controlled by a USB motor controller (Basicmicro Motion Control Roboclaw 2x30A), as shown in Figure 3.1b. While this design is capable of full rotational positioning, we steered the receiver between $-45°$ and $45°$ from the transmitter boresight. Figure 3.1c demonstrates how the receiver collects different physical AoAs using the automated turntable. We coarsely programmed the motor positions for data collection, achieving pseudo random beam alignment positions. Note that the motor is only used to emulate physical random movement and holding posture of UE used by human users. No such motor is required when generalizing this approach to actual scenarios.

3.6 Experiment and results

In this section, we describe the experiment details, including the environment and data collection. Then, we present results from the proposed method and compare with state-of-the-art methods.

3.6.1 Experiment details

We conducted the experiment in a LOS, suburban outdoor environment, shown in Figure 3.1a. The radios were mounted on tripods separated by approximately 14 m (91 dB

pathloss). Although the motor changed the azimuth AoA, the elevation angle was perfectly aligned and never changed. Similarly, the Tx directional beam was pointed towards the Rx at all time.

During each collection period, we collected 2,000 points. Each point consisted of 100 RSS measurements using the sounding codebook $\mathcal{W}$, i.e., $K = 64$ DFT beams such that adjacent pencil beams overlap by a half beamwidth and $M_0 = 36$ PN beams. Each data capture spent[3] 7 s when the Rx was static, and adjacent data capture had another 2 s gap for the motor movement to create a new LOS propagation direction. Although we collected data for $M_0 = 36$ PN beams, only the first M beams were used for training and testing with compressive beam alignment algorithms. A total of 3 collection periods from three different days were included in this chapter's results, each with a different SNR. We changed the SNR by modifying the transmit equivalent isotropically radiated power (EIRP) to be 31, 34, and 40 dBm, translating to median PN beam STF-SNRs of 10, 12, and 10 dB, whose physical meaning is consistent with SNR defined in (3.2). Note that the highest EIRP data had a low STF-SNR likely due to transmitter beam misalignment during the experiment. After data collection, AoAs with insufficient training data (at least 20 points per SNR) were eliminated, leaving $K = 51$ remaining labels. Of the data associated with the $K = 51$ DFT beam labels, a total of 3,060 data points were used for training and the rest were used for evaluation. During evaluation, the DFT sounding results were used as the ground truth to measure beam prediction performance from compressive PN probing.

For fair comparison, the training data was also available to the solution that analytically solves CPR. Note that when using Ψ_k in (3.4) as collected labels of training data, the system can estimate $|\tilde{\mathbf{w}}_m^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\theta_k)|$, i.e., the magnitude of dictionary $\mathbf{\Psi}_k$. Although such estimates cannot be directly used in CPR as phase information of $\mathbf{\Psi}$ are missing, they help the RSS-MP algorithm. Hence, we refer to the vanilla RSS-MP as one that uses only the model predicted dictionary, and dictionary refined RSS-MP as the one enhanced by training data.

[3]Since the latency of testbed API is not optimized, our goal is to achieve beam alignment with a compressed number of channel probings instead of high speed.

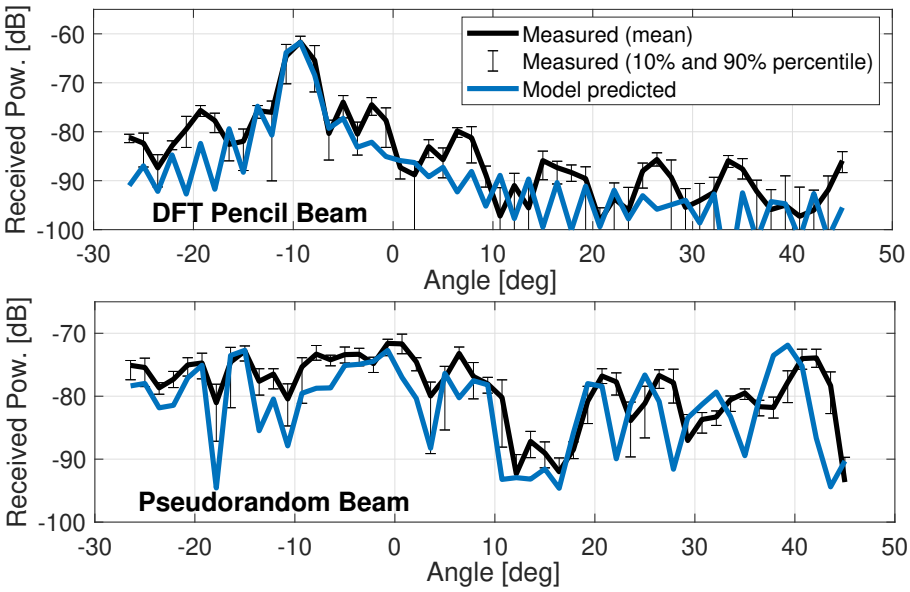


Figure 3.2: Measured and model predicted beam pattern of a discrete Fourier transform pencil beam and a pseudorandom sounding beam.

Also, note that our best efforts in applying the convex optimization based CPR solution led to unsatisfactory performance, likely due to the imperfect knowledge of the sensing matrix $\mathbf{\Psi}$ when the PN sounding codebook is used. [196] reported a similar finding.

3.6.2 Experiment results

The captured data allowed us to coarsely evaluate the beam pattern of the testbed. The beam pattern $|\tilde{\mathbf{w}}^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\theta)|^2$ at angle θ is measured by training data whose DFT beam exhaustive search gives θ. The pattern at each θ is statistically evaluated with 20 data points. A comparison between the measured pattern and model predicted pattern $|\mathbf{w}^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\theta)|^2$ is presented in Figure 3.2, showing an example DFT pencil beam and an example PN beam. The results verify the arguments in Section 3.3.2. Although the hardware impairment causes little distortion in the mainlobe of DFT pencil beam, DFT sidelobes and lobes of the PN beam are susceptible to larger distortion.

Using the neural network described in Section 3.4.2 with the experimental data, we

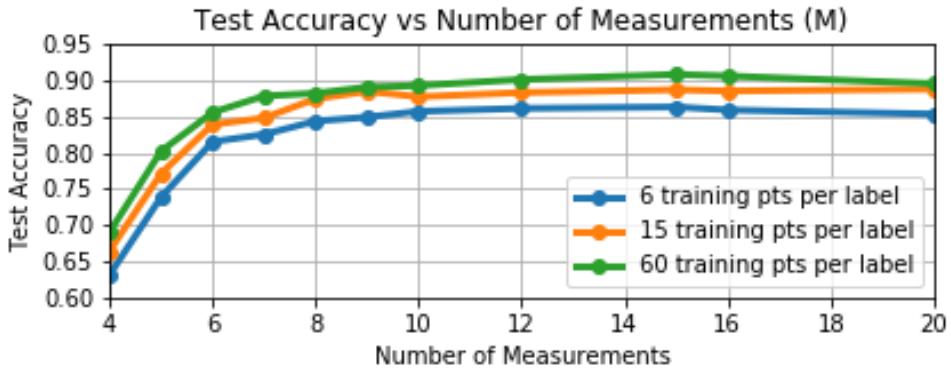


Figure 3.3: Test accuracy as a function of the number of pseudorandom beam measurements M for 3 training set sizes.

achieved very good accuracy for most desired numbers of measurements M. Figure 3.3 shows the classification accuracy of the $K = 51$ DFT beams used for $4 \leq M \leq 6$. When using $M \geq 6$, the test accuracy saturates around 89% for the full training set. Figure 3.3 also shows the network performance with three training set sizes. Performance does improve with more training data, but the benefit saturates with increasing training set size.

The post-alignment BF gain loss is presented in Figure 3.4 with a comparison of three algorithms using PN sounding results to predict best DFT beam for data communication. Vanilla RSS-MP suffered from the mismatched dictionary information and thus had the poor performance. Even with $M = 20$ channel probings, vanilla RSS-MP had more than 8 dB gain loss at the 90th percentile. Dictionary refinement improved RSS-MP. With $M = 15$ measurements (76.6% overhead savings), it sacrificed less than 2 dB BF gain in 90 percent of test cases. The proposed NN based approach provided further savings, requiring only $M = 5$ measurements (92.2% overhead saving) for comparable post-alignment gain. In fact, we observed 0 dB gain loss for 50 percent of test cases because the NN classification accuracy is above 50 percent for $M \geq 4$.

In Figure 3.5, we compare the required number of measurements as a function of array size for the dictionary refinement improved RSS-MP and proposed algorithms using exper-

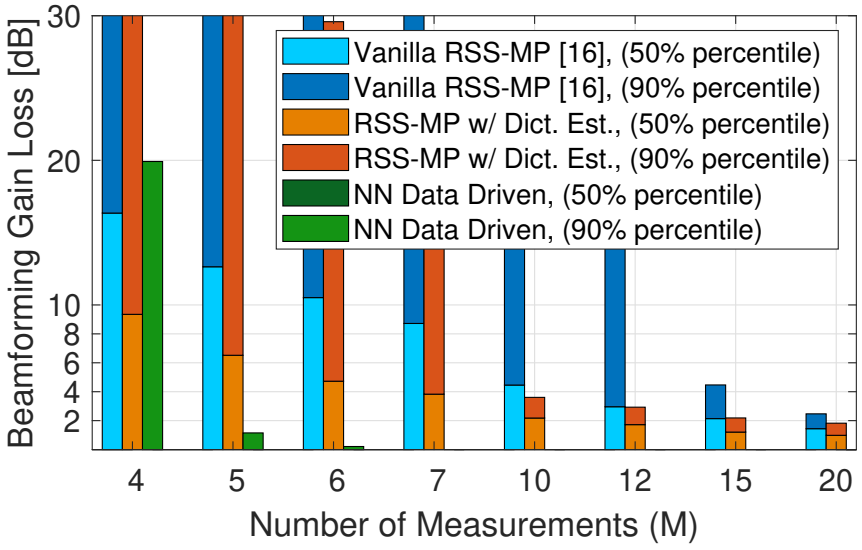


Figure 3.4: The post-alignment beamforming gain loss (as compared to a full exhaustive search with $K = 64$ measurements) as function the number of pseudorandom beam probings M. The best 50th percentile and 90th percentile are shown.

imental and simulation data. The simulation used the same PN sounding AWV realization as in the experiment, but did not include array hardware impairment. By comparing these two scenarios, we found the following three conclusions. Firstly, the required overhead of compressive beam alignment scaled logarithmically with array size, an appealing property for future mmW systems. Secondly, the proposed method effectively learned how to solve CPR and provided more accurate beam alignment than RSS-MP even in the simulations without model mismatch. This indicates RSS-MP is a sub-optimal solution of CPR, unsurprising as it is a heuristic algorithm. Lastly, the NN had no performance loss in the experimental implementation, unlike RSS-MP, because the data driven approach is immune to model mismatch due to array imperfection.

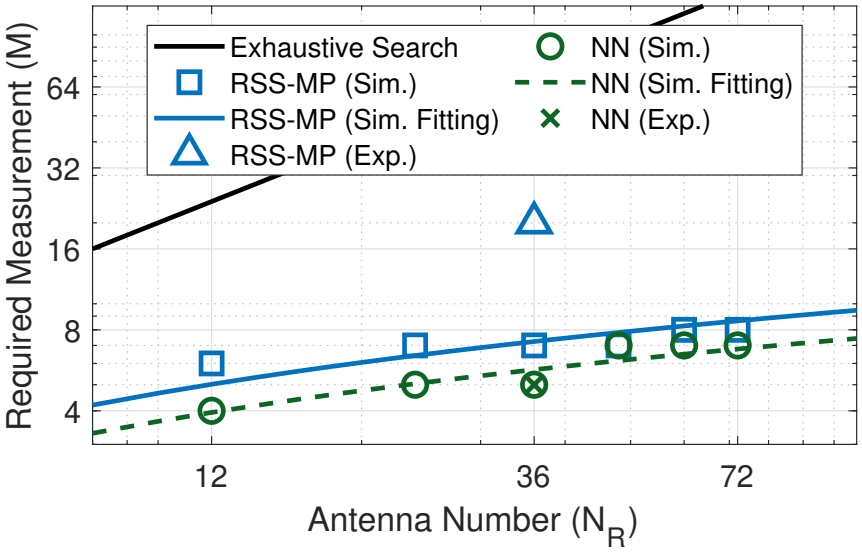


Figure 3.5: The required number of measurements M for beam alignment to reach less than 2dB gain loss in the 90-th percentile as function of receiver array size.

3.7 Conclusion and future works

In this chapter, we present a compressive beam alignment scheme that utilizes machine learning to address implementation challenges due to hardware impairment. The results demonstrate that compressive beam alignment has potential to significantly reduce the required number of channel probings. Our implementation on a 60 GHz testbed demonstrates an order of magnitude overhead savings with marginal post-alignment beamforming gain loss, as compared to exhaustive beam sweeps. It also outperforms purely model-based compressive beam alignment in the experiment.

There are open questions in this area. The approach and results of this chapter have yet to be generalized to more sophisticated mmW channels, e.g., non-line-of-sight. Further, the use of compressed channel probing to predict more information like an exhaustive search, e.g., multiple steering directions in multipath environments, has yet to be studied. Finally, a comparative study of different sounding codebooks, e.g., multi-lobe beams [24, 81], with consideration to array impairment and joint design of the codebook and beam alignment

algorithm with other machine learning tools are of interest.

CHAPTER 4

Rainbow Beam based Single-Shot Beam Training using True-Time-Delay Array

4.1 Introduction

The mmW communication is a promising technology for the future cellular network including 5G-NR. Due to abundant spectrum, it is expected that the mmW network will support ultra-fast data rates. As shown in both theory and prototypes, mmW systems requires BF with large antenna arrays at both BS and UE to combat severe propagation loss. The directional beam requires angular channel information for steering direction of analog beams. Such information is typically acquired by beam training in standardized mmW systems, including IEEE 802.11ad/ay [65] and 5G-NR [48]. However, with increased array size and reduced beam width, the training overhead increases. The challenge of overhead becomes more severe in the future mmW and sub-terahertz systems where the array size is expected to further increase due to carrier frequency increase [140].

The mmW beam training is an active research area. In the exhaustive sounding scheme, transceivers use one pair of pencil beams at a time. This approach has prohibitive overhead with increased array size. In the multi-stage sounding based scheme, transceivers adapt their beamwidth and pointing angle based on the training results of the previous stage [160]. However, its performance degrades with larger path number or user number. In the pseudorandom sounding based scheme, transceivers use pseudorandom beams and compressive sensing to exploit the sparse scattering nature of mmW propagation. Although promising results are reported, this approach faces implementation challenges [83]. The above

schemes use a single analog beam at a time. Steering multiple beams simultaneously can accelerate beam training. Existing approaches either require to use multiple RF-chains [51], fully digital architecture [29], or utilize leaky wave antenna [93], which introduce additional hardware complexity. Meanwhile, the TTD array architecture is an emerging array design for wireless communication systems, particularly in mmW and sub-terahertz bands [104]. Recent research of TTD arrays focus on mitigating squint distortion in beam steering of ultra-wideband signals [80, 147]. The frequency-controlled beam steering capability of this architecture [123] is overlooked and the beam training tailored for TTD array is rarely investigated.

In this work, we present the theory and signal processing algorithm of an one-shot beam training scheme using TTD array. Firstly, we derive the TTD array based mmW wideband system model with CP based OFDM waveforms. We show that TTD arrays can apply unique AWV to different subcarriers with a single RF-chain. Such beam is referred as rainbow beam. Secondly, we analyze the TTD array based rainbow beam steering and derive the relationship between subcarrier frequencies and the corresponding steering directions in the closed form. The required delay tap spacing of TTD arrays and number of subcarriers to simultaneously scan the entire angular space are also analytically determined. Thirdly, we design a novel low-complexity signal processing algorithm to exploit the TTD array for beam training. We conduct numerical evaluation using mmMAGIC channel model at 28 GHz for both LOS and NLOS [9]. The results show that the proposed approach can achieve accurate beam training using a single OFDM training symbol.

The rest of the chapter is organized as follows. In Section 4.2, we present the TTD array architecture and the mathematical system model for wideband beam training. Section 4.3.1 includes the analysis and design of TTD array sounding beams for simultaneous angle scan, followed by beam training algorithm in Section 4.3.2. The numerical results are presented in Section 4.4. Finally, Section 4.5 concludes the chapter and highlights open research questions.

Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. The (i, j)-th element of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. Conjugate, trans-

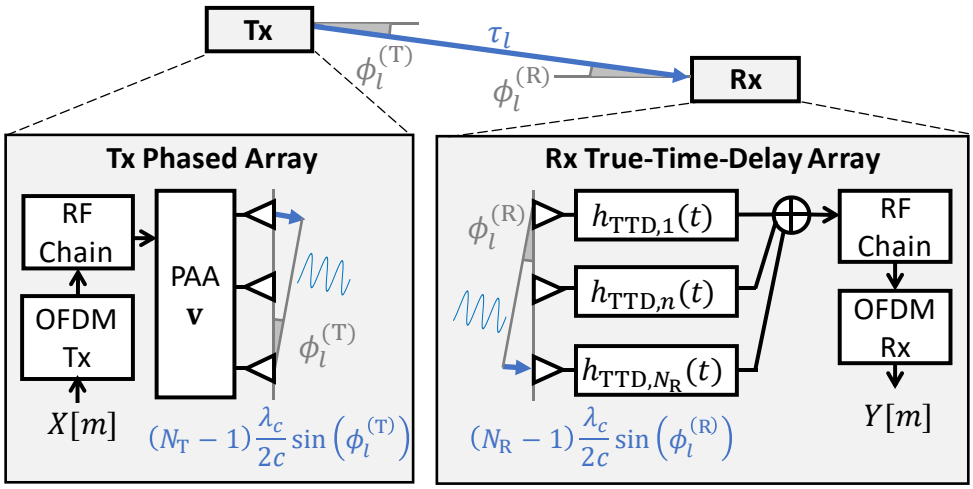


Figure 4.1: Illustration of the transceiver and channel model.

pose, Hermitian transpose, and pseudoinverse are denoted by $(.)^*$, $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$, and $(.)^\dagger$ respectively.

4.2 System model

This section introduces the system model of wideband beam training and problem formulation.

Consider a 5G-like single cell system with a BS equipped with N_T antennas. The system utilizes CP-OFDM with sampling duration T_s (bandwidth $\mathrm{BW} = 1/T_\mathrm{s}$, M subcarriers, and cyclic prefix length N_cp). The beam training pilots are denoted as $X[m]$, where m is the index of subcarrier. The beam training occurs in the downlink where BS transmits beamformed signal over mmW sparse multipath channel to UEs. We focus on the design where UE is equipped with TTD array with single RF-chain and N_R antennas. We assume a single stream of the training signal transmitted by the BS, that can use any array architecture[1]. During sounding, the BS transmitter use AWV $\mathbf{v} \in \mathbb{C}^{N_\mathrm{T}}$.

We consider the geometric multipath channel. The time domain channel between the

[1]Typically BS is equipped with multiple RF-chains. We assume they are used for other purposes, e.g., serving other users.

q-th transmitter antenna and the n-th receiver antenna is denoted as

$$h_{q,n}(t) = \rho \sum_{l=1}^{L} g_l p_c \left(t - \Gamma_{l,q,n}\right) \tag{4.1}$$

where $\rho = \sqrt{N_{\mathrm{T}} N_{\mathrm{R}}/L}$ is the normalization factor, L is the number of multipath component (MPC)[2]. Function $p_{\mathrm{c}}(t)$ is the time domain response filter due to limited hardware bandwidth. Scalars g_l and $\Gamma_{l,q,n}$ are the complex gain of the l-th MPC and propagation delay from the q-th transmitter element to the n-th receiver element via the l-th MPC, respectively. In this work, we focus on system where antenna array is linearly arranged with half wavelength spacing (corresponding to center frequency f_{c}). Thus, the delay $\Gamma_{l,q,n}$ is modeled as

$$\Gamma_{l,q,n} = \tau_l - \frac{(q-1)\lambda_c \sin\left(\phi_l^{(\mathrm{T})}\right)}{2c} + \frac{(n-1)\lambda_c \sin\left(\phi_l^{(\mathrm{R})}\right)}{2c}.$$

As illustrated in Figure 4.1, τ_l is delay between the first transmitter and the first receiver array element due to the l-th MPC. $\phi_l^{(\mathrm{T})}$, $\phi_l^{(\mathrm{R})}$, $\lambda_{\mathrm{c}} = c/f_{\mathrm{c}}$ and c are the AoD and AoA of the l-th MPC, wavelength of carrier, and the speed of light, respectively. At the UE receiver, each element introduces time domain thermal noise which is modeled as additive white Gaussian noise $n(t)$ with zero mean and spectral density $N_0/2$.

In principle, the TTD circuit block introduces constant group delay[3] to the received signal. Denoting this delay as $\tau_{\mathrm{TTD},n}$ in the n-th array element, such module has impulse response

$$h_{\mathrm{TTD},n}(t) = \delta(t - \tau_{\mathrm{TTD},n}). \tag{4.2}$$

Based on the above model, the received OFDM symbol after analog TTD array is determined by the following proposition.

---

[2]In the realistic mmW channel model, e.g., mmMAGIC [9], the number of MPC L is in the order of dozens to hundreds. Although some existing works exploit the clustering nature of MPC and deliberately reduce L, this work does not require assumption of sparsity.

[3]Delays can be introduced by transmission line [123]. But a more promising approach is to use digitally controlled analog circuits to apply delay either in RF [43] or baseband [64].

Proposition 3. *The received OFDM symbol of the m-th subcarrier is*

$$Y[m] = \mathbf{w}_{TTD}^{H}[m]\mathbf{H}[m]\mathbf{v}X[m] + N[m], \tag{4.3}$$

The combiner specified by TTD arrays $\mathbf{w}_{TTD}[m] \in \mathbb{C}^{N_R}$ *is frequency dependent, i.e., its n-th element as*

$$[\mathbf{w}_{TTD}[m]]_n = \exp\left(j2\pi f_m \tau_{TTD,n}\right). \tag{4.4}$$

where the RF frequency of the m-th subcarrier is denoted as f_m*, i.e.,*

$$f_m = \begin{cases} f_c + \frac{m}{M}BW, & 0 \le m < \frac{M}{2} \\ f_c - \frac{BW}{2} + \frac{\left(m-\frac{M}{2}\right)}{M}BW, & \frac{M}{2} \le m < M \end{cases}. \tag{4.5}$$

The channel at the m-th subcarrier $\mathbf{H}[m] \in \mathbb{C}^{N_R \times N_T}$ *is*

$$\begin{aligned} \mathbf{H}[m] = \rho \Bigg[\sum_{l=1}^{L} \tilde{g}_l \left(\sum_{i=0}^{M-1} e^{-j\frac{2\pi i m}{M}} p_c(iT_s - \tau_l) \right) \\ \cdot \mathbf{a}_R\left(\theta_l^{(R)}, f_m\right) \mathbf{a}_T^H\left(\theta_l^{(T)}, f_m\right) \Bigg]. \end{aligned} \tag{4.6}$$

In the above equation, $\tilde{g}_l = g_l \exp(-j2\pi f_c \tau_l)$ *is the channel gain. The array responses* $\mathbf{a}_R(\phi_l^{(R)}, f_m)$ *and* $\mathbf{a}_T(\phi_l^{(T)}, f_m)$ *are defined by their n-th and q-th element as*

$$\begin{aligned} \left[\mathbf{a}_R\left(\phi_l^{(R)}, f_m\right)\right]_n &= \exp\left[-j\pi(n-1)(f_m/f_c)\sin\left(\phi_l^{(R)}\right)\right], \\ \left[\mathbf{a}_T\left(\phi_l^{(T)}, f_m\right)\right]_q &= \exp\left[-j\pi(q-1)(f_m/f_c)\sin\left(\phi_l^{(T)}\right)\right]. \end{aligned}$$

This model holds true so long as the CP is longer than the cumulative delay of both MPC and TTD circuits, i.e.,

$$N_{cp}T_s > \max_{l,q,n} \Gamma_{l,q,n} + \max_{n} \tau_{TTD,n}, \tag{4.7}$$

The frequency domain noise $N[m]$ *is Gaussian distributed with zero mean and variance* $\mathbb{E}|N[m]|^2 = N_0 BW/(2M)$.

Proof. See Appendix. C.1 □

Remark1: Following the derivation in Appendix C.1, it is straightforward to generalize *Proposition 1* to system where transmitter also uses TTD array.

In this work, we focus on the system design where TTD array introduces uniformly spaced delay taps during beam training, i.e., $\tau_{\mathrm{TTD},n} = (n-1)\Delta\tau$ where $\Delta\tau$ denotes the delay spacing and it meets condition $\Delta\tau > 1/(2f_{\mathrm{c}})$. With straightforward mathematical manipulation, the receiver beamforming gain at direction α for the subcarrier with frequency f_m can be expressed $G(\alpha, f_m) = N_{\mathrm{R}}^{-1}|\mathbf{w}_{\mathrm{TTD}}^{\mathrm{H}}[m]\mathbf{a}_{\mathrm{R}}(\alpha, f_m)|^2$

$$G(\theta, f_m) = \frac{1}{N_{\mathrm{R}}}\left|\frac{\sin\left[\frac{N_{\mathrm{R}}\pi}{2}\left(2f_m\Delta\tau + (f_m/f_{\mathrm{c}})\sin(\theta)\right)\right]}{\sin\left[\frac{\pi}{2}\left(2f_m\Delta\tau + (f_m/f_{\mathrm{c}})\sin(\theta)\right)\right]}\right|^2. \tag{4.8}$$

In this work, we assume that pilots subcarrier have non-zero power when subcarrier indices are in set $\mathcal{M}$, i.e., $X[m] \neq 0, m \in \mathcal{M}$. Further, the transmit pilots have unit power constraint as $\sum_{m=0}^{M-1}|X[m]|^2 = M$.

We address two problems in TTD array based beam training.

Problem 1 (TTD sounding beams design): The objective is to design delay tap spacing $\Delta\tau$ and number of subcarriers M such that the sounding beams of subcarriers scan the entire angular region. In other words, for arbitrary AoA θ, there is at least one subcarrier which has sufficient beamforming gain in its direction. Mathematically, our goal is to find the feasible set $\mathcal{S}$ of design parameters such that sounding beamforming gain does not drop below $(1-\epsilon)N_{\mathrm{R}}$ for the least favorable AoA θ:

$$\mathcal{S} = \left\{(\Delta\tau, M) \,\middle|\, \min_{\theta}\max_{m} G(\theta, f_m) \geq (1-\epsilon)N_{\mathrm{R}}\right\}. \tag{4.9}$$

Problem 2 (TTD array based beam training): Using the design parameters from $\mathcal{S}$, the objective is to design beam training pilots $\mathcal{M}$ and signal processing algorithm to estimate dominant propagation directions from a single received symbol $Y[m]$. In this work, we focus on the receiver beam training and assume transmitter beamformer $\mathbf{v}$ has been aligned with AoD.

4.3 True-time-delay array based beam training design

In this section, we present the TTD based rainbow sounding beam design and beam training algorithm.

4.3.1 Analysis and design of rainbow sounding beams

We first analyze the frequency dependent receiver gain function (4.8). Denote $\Psi = 2f_m\Delta\tau + (f_m/f_c)\sin(\theta)$, the following observations are made [164, Chapt 7.2.4].

- $G(\Psi)$ has a peak at $\Psi = 0$.
- $G(\Psi)$ is a periodic function of Ψ with period 2.

Based on the above properties, the center of the sounding beam corresponding to subcarrier f_m, i.e., $\theta_m^\star = \arg\max_\theta G(\theta, f_m)$, is given in the following proposition.

Proposition 4. *Given delay tap spacing $\Delta\tau$, an approximation of the pointing direction of sounding beam that corresponds to the m-th subcarrier is*

$$\theta_m^\star \approx \sin^{-1}\left(\mathrm{mod}(2f_m\Delta\tau + 1, 2) - 1\right), \tag{4.10}$$

where mod() *is the modulo operation.*

Proof. This is by solving $\Psi = 2z, z \in \mathbb{Z}$ with given $\Delta\tau$ and f_m. The closed form solution (4.10) is available by using the approximation $f_m/f_c \approx 1, \forall m$ in Ψ. □

Next, we discuss the design of delay tap spacing $\Delta\tau$ and number of subcarriers M for a given BW in *problem 1*. For this purpose, we define ϵ-beamwidth as $\Omega(\epsilon, N_R)$ such that

$$G(\Psi) \geq (1-\epsilon)N_R, \quad \Psi \in [-\Omega(\epsilon, N_R), \Omega(\epsilon, N_R)] \tag{4.11}$$

for a given required sounding gain factor $1-\epsilon$ and receiver array size N_R. Note that the specific value of $\Omega(\epsilon, N_R)$ can be found numerically, e.g., 3dB-beamwidth is $\Omega(0.5, N_R) = 0.886/N_R$ [125, Chapt 22.7]. Note that directly solving *problem 1* is challenging. In the following proposition, we show a subset of the solution of *problem 1*.

Proposition 5. *A subset of $\mathcal{S}$, denoted as $\mathcal{S}_s$, is*

$$\mathcal{S}_s = \left\{ (\Delta\tau, M) \middle| \Delta\tau \geq \frac{1}{BW} + \frac{1}{2f_c}, M \geq \frac{BW\Delta\tau + \frac{BW}{2f_c}}{\Omega(\epsilon, N_R)} \right\}. \tag{4.12}$$

Proof. See Appendix C.2. □

In the system where BW $\ll f_{\mathrm{c}}$, the condition on delay tap spacing $\Delta\tau$ can be relaxed and the critical value is $\Delta\tau = 1/\mathrm{BW}$. With such delay tap spacing, the critical value of required number of subcarriers is $M = \lceil 1/\Omega(\epsilon, N_{\mathrm{R}}) \rceil$ where ceiling operator $\lceil a \rceil$ gives the smallest integer that is greater than a.

4.3.2 Rainbow beam based beam training

Using the proposed TTD rainbow sounding beams, the beam training is greatly simplified. Effectively, UE receiver only needs to configure TTD array based on predefined delay tap spacing $\Delta\tau$ during the scheduled time slot for beam training. A lookup table (LUT) can be constructed based on (4.10), which contains the pointing direction of all training pilots $m \in \mathcal{M}$ and sounding directions. During training, the UE measures the reference signal received power (RSRP) of the pilots, i.e.,

$$m_{\mathrm{best}} = \arg\max_{m} |Y[m]|^2 \tag{4.13}$$

and use the index of subcarrier that has highest RSRP m_{best} and LUT (4.10) to estimate AoA, i.e.,

$$\hat{\phi}^{(\mathrm{R})} = \theta^{\star}_{m_{\mathrm{best}}}. \tag{4.14}$$

The proposed TTD array beam training approach has an interesting relationship with PAA receiver beam training, when TTD delay tap spacing $\Delta\tau = 1/\mathrm{BW}$ and $f_{\mathrm{c}}\Delta\tau \in \mathbb{Z}$. In fact, the proposed simultaneous multi-beam can exactly represent DFT and oversampled DFT sounding beams, which is commonly used by PAA receiver. In the DFT beam based training procedure, the PAA receiver uses AWV

$$[\mathbf{w}_{\mathrm{PAA},k}]_n = \exp[j2\pi(n-1)k/K] \tag{4.15}$$

for the k-th training symbol, and it requires a total K OFDM training symbols (usually K is greater than N_{R} and it is power of 2). Comparing (4.4) and (4.15), it is straightforward that the TTD receiver can mimic this procedure by utilizing $M = K$ subcarriers in a single OFDM training symbol.

It is worth noting that the total number of subcarriers M is in the order of magnitude of receiver array size N_{R} when system intends to obey minimum condition in *proposition 3* or mimic PAA beam training. However, the total number of subcarrier in practical mmW system needs also accommodate coherent bandwidth, and therefore it can be much larger than N_{R}. As such, the TTD array based system needs to use *subcarrier selection* scheme. Based on (4.4), the following two TTD systems are identical: 1) system has a total number of subcarrier M and training utilizes all subcarriers $m < M$; 2) system has a total number of subcarrier MR for some $R \in \mathbb{Z}$ and training utilizes subcarriers within set $\mathcal{M} = \{mR | 0 \leq m < M\}$. With this feature, the proposed system is directly extendable to existing mmW protocols.

4.4 Numerical results

This section presents the numerical results of the beam pattern and the performance of the proposed beam training.

Figure 4.2 shows an example of the TTD array based simultaneous multi-beam and design parameters that follow *Proposition 3*. The simulation utilizes carrier frequency f_{c} =28 GHz, bandwidth 400 MHz, delay tap spacing $\Delta\tau$ =2.5 ns, and $N_{\mathrm{R}} = 8$ receiver antennas. The number of subcarrier $M = 8$ is designed based on constraint $1 - \epsilon = 0.4$ (4 dB loss), the corresponding $\delta(\epsilon, N_{\mathrm{R}}) = 0.1266$, and $M = \lceil 1/\delta(\epsilon, N) \rceil$. The figure shows that the gain of sounding beams meet the constraint for the least favorable AoA. It also verifies the relationship between subcarrier frequencies and the center of sounding beams in (4.10).

The beam training performance of the proposed BF training algorithm in a pure LOS channel, i.e., number of MPC is $L = 1$, is presented in Figure 4.3. The simulation parameters

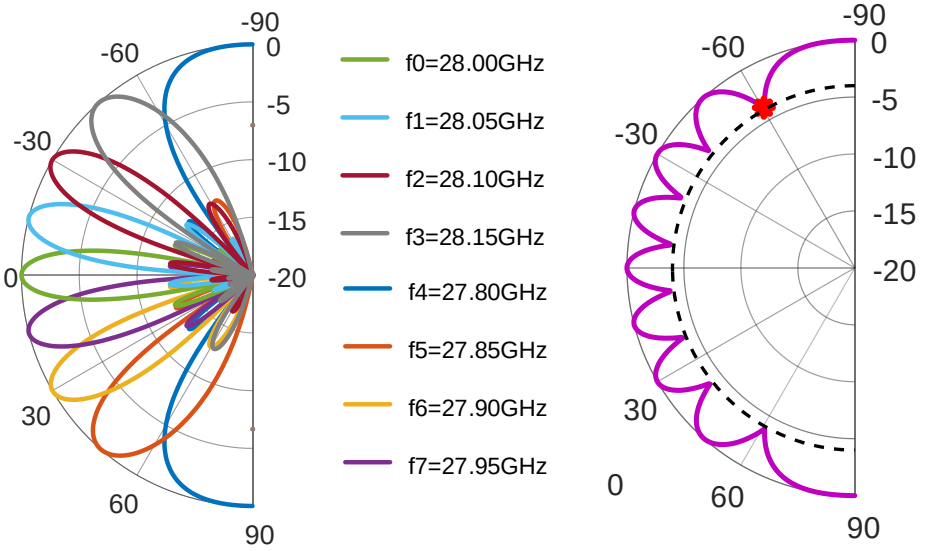


Figure 4.2: Example of true-time-delay (TTD) array based frequency dependent beam pattern in the logarithmic scale. In the left figure, different colored curves show beam patterns $G(\theta, f_m)$ for subcarriers f_m. In the right figure, the purple curve, the red circle, and the black dashed curve illustrate the function $\max_m G(\theta, f_m)$, $\min_\theta \max_m G(\theta, f_m)$, and gain constraint $1 - \epsilon$, respectively.

are identical to previous case, except the system has $N_{\mathrm{T}} = 64$ transmitter antennas and $N_{\mathrm{R}} = 16$ receiver antennas. Furthermore, a total of $M = 2048$ subcarriers are used. The SNR is defined as post-transmitter-beam SNR across all band, i.e.,

$$\mathrm{SNR} = \frac{\sum_m \|\mathbf{H}[m]\mathbf{v}\|^2}{\sum_m \mathbb{E}|N[m]|^2}.$$

The simulation follows the concept of frequency *resource block* in cellular systems [48]. Namely, when a subcarrier is used as the pilot, the 12 neighbor subcarriers are also used. The beam training utilizes only a single OFDM symbol. We use normalized post training beamforming gain as a metric, i.e., $N_{\mathrm{R}}^{-2}|\mathbf{a}_{\mathrm{R}}^{\mathrm{H}}(\hat{\phi}^{(\mathrm{R})})\mathbf{a}_{\mathrm{R}}(\phi_1^{(\mathrm{R})})|^2$. The figure shows that a smaller number of used subcarriers provides improvement in the low SNR regime, because more power is loaded in the corresponding pilots. However, it limits the training resolution of the proposed algorithm, and therefore the post training gain saturates in high SNR regime. Using an increased number of subcarriers improves performance in the high SNR regime, but the benefit saturates when the sounding beam number is more than $2N_{\mathrm{R}}$.

The post-beam-training spectral efficiency is shown in Figure 4.4. Particularly, the evaluation utilizes QuaDRiGa simulator [90] with mmMAGIC 28GHz channel model [9] in urban micro (UMi) LOS and NLOS environments. The number of MPC L in these two scenarios are 41 and 79, respectively. In the TTD array based system, 32 frequency *resource blocks* and 1 OFDM symbol are used for beam training. A benchmark system that uses a PAA receiver with same array geometry is used for comparison. The PAA receiver utilizes beam training AWV from sampled columns of a 32 by 32 DFT matrix when receiving the training symbols. The same frequency resources and power are used in the training pilots for both systems. In the LOS environment, the proposed system is equivalent to the benchmark using $2N_{\mathrm{R}}$ training symbols. Severe performance degradation is observed in the PAA with small number of training symbols, since the DFT beams fail to scan entire angular domain. In the NLOS environment, the proposed TTD array beam training method is outperformed by the PAA with the same angular coverage, i.e., using 32 symbols, by a small margin. This is because PAA provides robustness to the frequency selective channel gain when conducting wideband power measurement during beam training. In summary, the evaluation in realistic

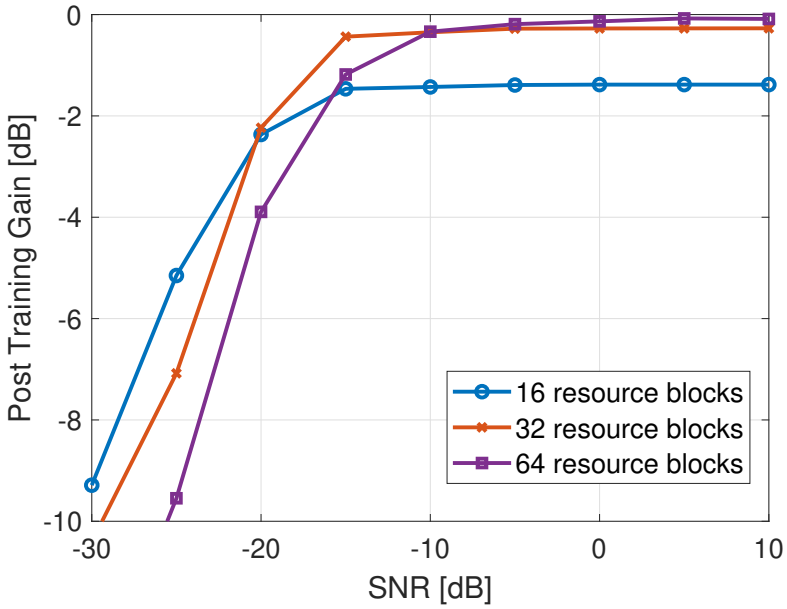


Figure 4.3: Post-beam-training gain of the proposed true-time-delay array rainbow beam training algorithm in pure line-of-sight environment.

mmW propagation channels reveals that the proposed method can use order of magnitude smaller number of training symbols as compared to PAA with DFT sounding beams.

4.5 Conclusions and future works

In this chapter, we present a fast beam training scheme that utilizes true-time-delay based millimeter-wave array. Exploiting the frequency dependence of this array architecture's antenna weight vector, the system can simultaneously scan multiple angles via frequency resources, and thus greatly accelerate the beam training procedure. We derive the condition for delay tap spacing and the required number of subcarriers. Based on these novel sounding beams, a one-shot, low-complex beam training algorithm is developed. The simulation results reveal that the proposed method utilizes a single training symbol to complete beam training. Such feature is appealing in the future mmW systems where conventional phased array based beam training meet increased training overhead due to increased array size.

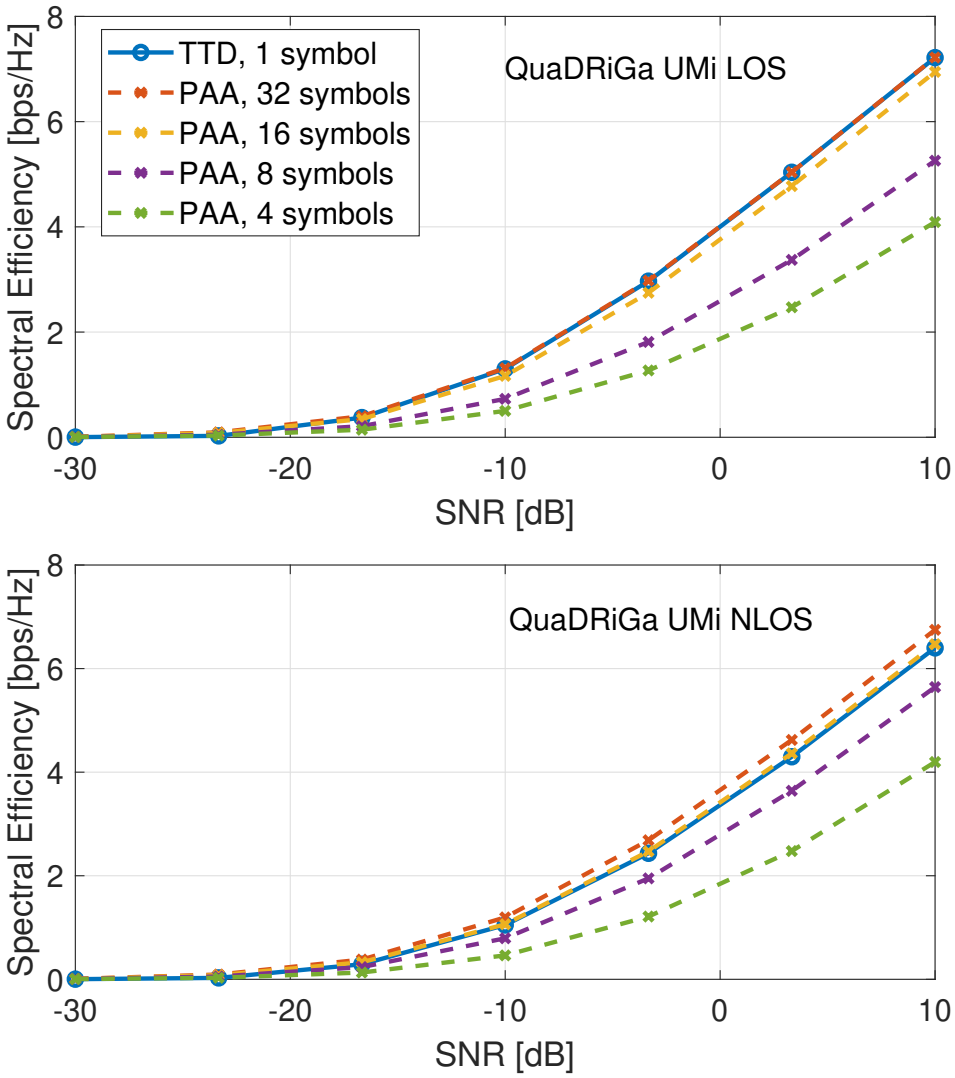


Figure 4.4: Post-training spectral efficiency comparison between the proposed true-time-delay array based system and system with the phased antenna array (PAA) and discrete Fourier transform (DFT) sounding beams.

There are many open questions in TTD array based signal processing and transceiver design. Firstly, the feasibility of TTD array based simultaneous multi-beam in 3D environment with planar array remains unknown. Secondly, it is of interest to exploit the frequency dependent AWV of TTD arrays to develop algorithms for super-resolution beam training, channel estimation, and covariance estimation in mmW wideband system. Lastly, it is of critical importance to understand the impact and required specification of hardware in RF and mixed signal domain for power efficient operation of TTD arrays.

CHAPTER 5

Comparative Study of Millimeter-wave Transmitter Array Architecture

5.1 Introduction

The mmW communications is a promising technology for the future 5G-NR cellular network [35, 131]. In the United States, the Federal Communications Commission has voted to adopt a new Upper Microwave Flexible Use service in the licensed bands, namely 28 GHz (27.5-28.35GHz band), 37 GHz (37-38.6GHz band), 39 GHz (38.6-40GHz) with a total 3.85 GHz bandwidth [57]. The abundant spectrum facilitates key performance indicators (KPI) of 5G, including 10Gbps peak rate, 1000 times higher traffic throughput than the current cellular system [21]. As shown in theory and measurements, mmW signals suffer higher free-space transmission loss [70], and is vulnerable to blockage [137]. As a consequence, radios require BF with large antenna arrays at both BS and UE to combat severe propagation loss [139]. This makes reliable communication range short and as a consequence, mmW BSs will be deployed in an ultra-dense manner with inter-site distance in the order of hundreds of meters [27, 33]. Due to these facts, performance, energy, and cost efficiency in the future mmW BS radios become more important than ever before.

Implementation and deployment of transceiver arrays in sub-6GHz have shown great success. In the 4G-LTE system, BS supports up to 8 antennas [1] and arrays with even larger size are being actively prototyped [156] and will be soon available in the LTE-A PRO (the pre-5G standard). Those systems exclusively have digital array architecture based on a dedicated radio-frequency transceiver chain, with data converter and up/down-conversion, per each

antenna, and rely on digital baseband for array processing. Many implementation challenges arise in scaling up array size [148] by an order of magnitude or more required for mmW bands. System designers are also concerned about the high cost and power consumption in digital array architecture with massive number of RF-chains and ultra-wide processing bandwidth [163].

Recently, an emerging concept of hybrid array has been proposed. A hybrid array uses two stage array processing. The analog beamforming implemented with variable phase shifter (PS) provides beamforming gain and the digital beamforming in the baseband provides flexibility for multiplexing multiple user streams [17, 78]. As a result, hybrid arrays support an RF transceiver count which is smaller than the array size. Such an architecture intends to reduce the power and cost penalty due to numerous transceivers. Based on the connectivity between RF-chain and antenna, there are two major variations, fully-connected hybrid array and partially connected hybrid array. Although both architectures were used for radar application [85] and were introduced for telecommunication application as early as a decade ago [195], they have recently gained much attention for mmW radios. Signal processing techniques, including channel estimation and beamforming, using hybrid architecture have been comprehensively studied [83]. Proposals for using hybrid architectures in mmW 5G have been considered in standardization organizations [176].

A handful comparative analyses exists for different mmW array architectures, with an emphasis on the signal process algorithms [11, 100, 116, 176]. Authors in [134] discussed circuits design challenges in implementing energy-efficient digital arrays. The relationship between spectral efficiency (SE) and energy efficiency in partially-connected hybrid architecture is studied in [62,78,82]. Works [116,165] provided comparison among array architectures and concluded that hybrid architecture can achieve higher energy efficiency than fully digital ones in the regime of point-to-point communication. Future 5G system, however, will certainly use multi-user multiplexing to provide higher network throughput. Moreover, existing works did not study trade-offs among array size, transmit power, and specifications of key circuit blocks in the three architectures. However, system designers need to understand these

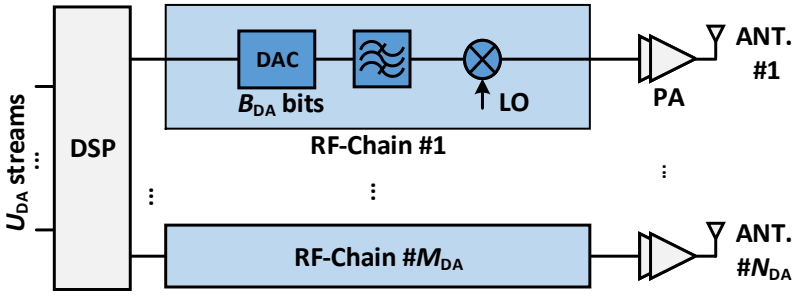
trade-offs and hardware implications to develop energy and cost efficient mmW systems [34].

This work aims to fulfill this gap. We intend to compare different array architectures in a comprehensive manner by considering trade-offs among capacity, energy and area efficiency. Specifically, we compare array architectures based on the criterion of achieving same capacity. All design trade-offs are carefully considered in reaching most efficient design in all architectures which meets the requirement of typical 5G use cases. Power consumption, including analog processing energy and digital computation energy, and integrated circuits (IC) area are then compared based on state-of-the-art circuits. We provide several design insights on scaling laws and the bottlenecks in each architecture which allow us to predict a trend for future wireless demands and technology scaling.

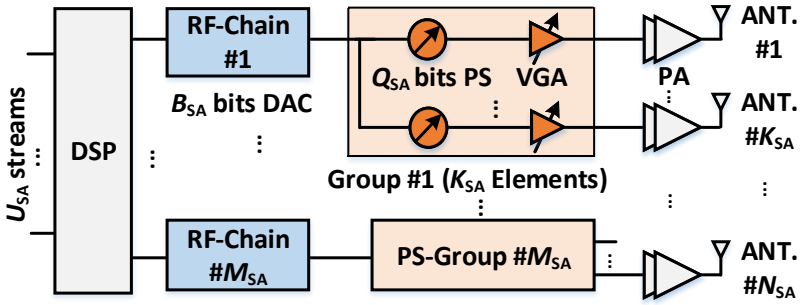
The chapter is organized as follows. In Section 5.2, we briefly introduce emerging mmW array architectures and typical 5G use cases. In Section 5.3, we discuss design trade-offs in all array architectures and the designs used for comparison. In Section 5.4, we study implementation issues in antenna arrays and their impact on different architectures. In Section 5.5, we present the state-of-the-art specifications of mmW beamforming circuits blocks and system level power consumption and IC area of the three architectures. Numerical results and open research questions are presented in Section 5.6 and Section 5.7, respectively. This leads us to the general conclusions in Section 5.8.

5.2 Comparative framework

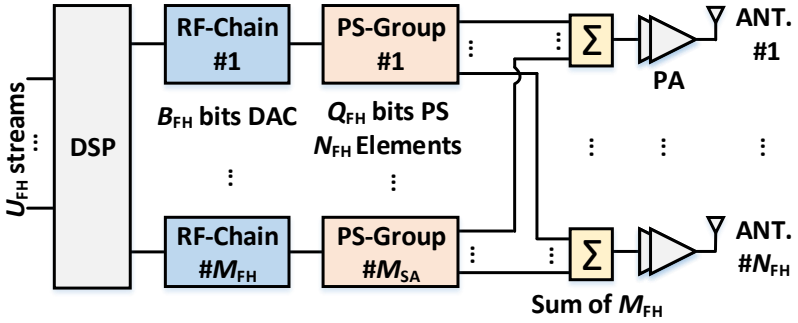
In this work, we focus on the comparison of transmitter antenna array architectures in a 5G mmW BS . We first introduce three commonly considered array architectures and summarize recent silicon implementations. Then, we describe the metrics used for comparison of the three architectures.



(a) Block diagram of digital array.



(b) Block diagram of sub-array. Each RF-chain has the same structure as (a).



(c) Block diagram of fully-connected hybrid array. Each RF-chain and PS group has the same structure as (b).

Figure 5.1: Three transmitter array architectures that are considered in this work.

5.2.1 Array architectures

There are three transmitter array architectures that are considered for adoption in 5G mmW system. Figure 5.1 depicts block diagrams of digital array and two variations of hybrid array, partially-connected hybrid array (we denote it as sub-array in this work), and fully-connected hybrid array. Key design parameters for each architecture are:

- Transmit power in all array elements: $P^{(\text{out})}$
- Number of antennas: N
- Number of RF-chains: M
- Number of simultaneous streams: U ($U \leq M$).
- Number of bits in DAC: B
- Number of bits in phase shifter: Q. This only applies to hybrid arrays.

In the rest of the chapter, we use digital array (DA), sub-array (SA) and fully-connected hybrid array (FH) when referring to digital array architecture, sub-array and fully-connected hybrid array architecture, respectively. Mathematical symbols with subscript indicate parameters associated with the specific architecture, e.g., N_{DA} represents number of antennas in digital array. The main differences among three array architecture are:

- *Digital Array:* As shown in Figure 5.1(a), N_{DA} antennas in DA are connected to M_{DA} RF-chains, i.e., $N_{\text{DA}} = M_{\text{DA}}$. The beamformer precoding occurs in the baseband (BB) digital signal processor (DSP).
- *Sub-Array:* The SA consists of multiple phased arrays. As shown in Figure 5.1(b), N_{SA} antennas are partitioned into M_{SA} group, each of which has one dedicated RF-chain, K_{SA} PS, variable gain amplifiers/attenuators (VGA), and power amplifier (PA)s. The array size, group number, and number of elements in a group follows relationship $N_{\text{SA}} = M_{\text{SA}} K_{\text{SA}}$. Using phase shifters, each group can transmit a beam towards specific

direction and SA is capable of transmitting/multiplexing up to M_{SA} simultaneous beams. When the required number of beams U_{SA} is smaller than M_{SA}, multiple array groups can form a virtual group. The increased array size for that specific beam provides better beamforming performance, e.g., higher gain and narrower beam-width. DSP facilitates precoding multiple beams in the baseband.

- *Fully-Connected Hybrid Array:* This architecture is also known as overlapped sub-array [85], multibeam active phased array [108], and high definition active antenna system [28]. Similar to SA, the FH architecture uses phase shifters for analog beamforming and DSP for digital beamforming. However, FH has different connecting structures between RF-chains and phase shifters. As shown in Figure 5.1(c), each of M_{FH} RF-chains connects with all N_{FH} antennas via N_{FH} phase shifters. Combiner networks are used to add M_{FH} RF signals before passing through the PAs. As a consequence, a total of $M_{\text{FH}}N_{\text{FH}}$ phase shifters are required in this architecture. FH is capable of transmitting up to M_{FH} simultaneous streams.

Recent IC implementations of all three architectures are summarized in Table 5.1, including complementary metal–oxide–semiconductor (CMOS) and other technology. Apart from array in 28 GHz band, Table 5.1 includes implementation in 60 GHz band for mmW indoor access, mmW backhaul and radar, because they share the same array architectures. Directly comparing array architectures from the table is difficult, because they use different silicon technology, and not all circuits components, e.g., local oscillator (LO) and associated up/down-conversion circuits, low-noise amplifier (LNA), and PA, are integrated. It is worth noting that SA and FH architectures in Table 5.1 implement phase shifters in the RF domain. A comprehensive survey of phase shifter implementations is covered in [133], including phase shifters in analog baseband, LO, and RF domain. Moreover, system level prototyping of 28 GHz arrays together with field test can be found in [176, 188].

There are other architectures that have been recently proposed, e.g., switch based antenna array [120] and lens antenna array [39]. Due to the lack of implementation details available in the literature, we do not include quantitative analysis of them in this work.

Table 5.1: Silicon implementations of millimeter-wave array architectures

Reference, Year	Architecture	Freq. (GHz)	Tx/Rx	Array Size	PA/ LNA	LO	Power Consumption per Array Element (mW)	Area per Array Element (mm^2)	Technology
[117], 2017	FH	25-30	Rx	8	-	✓	30 (Rx)	0.77	65 nm CMOS
[97], 2017	SA	28	TRx	2	-	-	0 (both Tx and Rx)	1.65	45 nm CMOS
[77], 2017	SA	28	TRx	4	✓	-	237.5 (Tx), 142.5 (Rx)	1.23	65 nm CMOS
[175], 2017	SA	57-64	Rx	4	-	✓	80 (Rx)	0.65	65 nm CMOS
[111], 2016	SA	57-64	TRx	4	✓	✓	167.5 (Tx), 107.8 (Rx)	1.97	28 nm CMOS
[36], 2014	SA	57-64	TRx	16	✓	✓	74.4 (Tx), 60 (Rx)	2.07	40 nm LP CMOS
[44], 2013	SA	57-64	TRx	32	✓	✓	37.5 (Tx), 26.6 (Rx)	0.89	90 nm CMOS
[149], 2017	SA	28	TRx	32	✓	✓	35.9 (Tx), 25.8 (Rx)	5.18	0.13 µm SiGe BiCMOS
[200], 2015	SA	57-64	Tx	256	✓	-	10.9 (Tx)	6.79	0.18 µm SiGe BiCMOS
[99], 2014	SA	76-85	TRx	8	✓	✓	118.74 (Tx), 143.8 (Rx)	3.26	0.13 µm SiGe BiCMOS
[167], 2013	SA	94	TRx	16	✓	✓	181.25 (Tx), 156.25 (Rx)	2.76	0.13 µm SiGe BiCMOS

5.2.2 Comparison metrics under 5G use cases

5G is characterized by a wide variety of use cases having different environments, communication distances, and performance requirements. Performance, in turn depends on connectivity density (defined as number of simultaneous connections for one wireless service operator in an given area), peak rate, and network traffic throughout. It is our vision that the mmW BS should be capable of using the same radio front-end arrays to handle various use cases and meet their demands.

We choose three representative use cases [10]: Dense Urban enhanced mobile broard band (eMBB), 50+Mbps Everywhere, and Self-Backhauling. They cover different MIMO processing schemes of transmitter array.

- *Dense Urban eMBB:* In dense urban area, large number of UEs require high-speed connections for applications like streaming, high-definition videos, and downloading files. According to 5G KPI requirement [10], the connection density is expected to be 150,000 connections per square kilometer, while the traffic throughput is up to 3.75Tbps/km^2 in such scenario. A typical 5G mmW BS deployment setting has inter-site distance of 200 m and each BS has 3 radio sectors [2]. With 850 MHz spectrum at 28 GHz band, the required SE in this use case is up to 58.8bps/Hz. Such a scenario often involves LOS environment and relatively good SNR is expected for each UEs so that SE greatly benefit from high multiplexing. We anticipate that at least 8 simultaneous streams are required[1].

- *50+Mbps Everywhere:* mmW electromagnetic waves are extremely vulnerable to blockage. Despite this, BS in the 5G mmW network need to sustain baseline performance (up to 100Mbps data rate [10]), even for those UEs under unfavorable propagation conditions. The 5G KPI requirement [10] also indicated that the connection density is up to 2,500 connections per square kilometer. With the same BS deployment assumption

[1]Till the time of writing, there is no specification for multiplexing in 5G mmW system. However, 8 streams are commonly used as assumption in the literature [173, 198]. Meanwhile, the next generation of 60 GHz indoor wireless system also targets to use 8 spatial streams [66].

as discussed in the previous use case, the required SE is 4.7bps/Hz. Due to a NLOS environment, severe propagation loss exists and more than 20 dB beamforming gain is required to close the link budget. Due to the requirement of high beamforming gain, we anticipated up to 8 simultaneous streams are adopted in this use case.

- *Self-Backhauling:* To facilitate ultra-dense mmW BS deployment, BSs are required to connect to core network through a backhaul link. Since the large array allows interference isolation in the spatial domain, it is expected that 5G BS is capable of using the same spectrum for both access and backhauling, which is refereed as self-backhauling. Self-backhauling using radio for 5G access significantly reduces cost of setting up high-speed fiber. We consider a scenario where mmW BS transmits uplink data of its local network to a macro-BS receiver which connects to core network. With assumption of one macro-BS deployed in every square kilometer, the self-backhauling link has up to 707m communication distance [130]. In this use case, LOS environment is assumed and 10Gbps rate is targeted by single data stream.

For fair comparison of power consumption and area among array architectures, each array architecture has to deliver the same target SE. In Table 5.2, the system parameters and link budgets are summarized, with a set of possible data streams number U and the corresponding signal-to-interference-plus-noise ratio (SINR) that reach SE objectives are also listed. In the Section 5.3, we study on the impact of design parameters on SE performance of different architectures and mainly focus on number of streams U, array size N and required transmit power $P^{(\text{out})}$. The power consumption and hardware resources comparison are then presented based on state-of-the-art device specifications.

5.3 Transmitter array design parameters

In this section, we discuss the impact of array design parameters on the SE performance of mmW multi-user MIMO system. We provide the design specification of components in array architectures to meet the SE requirement for each use case.

Table 5.2: Link Budget Estimation in Typical 5G Use Cases

Use Case	Dense Urban	50+Mbps Everywhere	Self-Backhauling
Channel Situation	Umi-LOS	Umi-NLOS	Uma-LOS
Freq. [GHz]	28	28	28
BW [MHz]	850	850	850
Distance [m]	100	100	707
Tx Power [dBm]	46.0	46.0	46.0
Tx Element Gain [dBi]	3.0	3.0	3.0
Pathloss[a] [dB]	104.4	125.1	118.3
Other Loss[b] [dB]	12.7	25.3	17.0
Rx Gain [dB]	12.0[c]	12.0[c]	27.1[d]
Rx Noise Figure [dB]	10.0	10.0	10.0
Rx Noise Power [dBm]	-74.7	-74.7	-74.7
SNR w/o Tx Array [dB]	18.7	-14.7	15.5
Target SE [bps/Hz]	58.8	4.7	11.8
Design Objective Examples			
Simultaneous Streams (U)	$\{8, 16, 32\}$	$\{2, 4, 8\}$	1
Per-UE SINR [dB]	$\{22.1, 10.7, 4.1\}$	$\{6.2, 1.0, -3.0\}$	35.5

a. Based on 3GPP model for above-6GHz band. [6].

b. Includes 3σ shadowing and 25mm/h rain absorption [25].

c. Based on 8 receiver antennas and 3dBi antenna gain.

d. Based on 256 receiver antennas and 3dBi antenna gain.

e. Based on equation $\text{SE} = U \log_2(1 + \text{SINR}_u)$.

5.3.1 System model of millimeter-wave multi-user MIMO

We consider a mmW system where a BS of interest transmits data to multiple UEs in mmW access or a hub in mmW self-backhauling. Both transmitter and receiver are equipped with antenna array. Linear precoding techniques over flat fading channel are considered. In case of frequency selective channel, the precoding can be extended using OFDM by considering per sub-carrier precoding. In the baseband equivalent model, the received symbol at the u^{th} UE is denoted as

$$y_u = \mathbf{w}_u^{\text{H}} \mathbf{H}_u \mathbf{R}(\mathbf{B}\mathbf{s} + \mathbf{z}_{\text{T}}) + \mathbf{w}_u^{\text{H}} \mathbf{z}_{\text{R}}. \tag{5.1}$$

In the above equation, vector $\mathbf{s} = [s_1 \cdots, s_U]$ contains the U symbols. Matrix $\mathbf{H}_u$ is the MIMO channel between transmitter and u^{th} UE receiver. Vector $\mathbf{w}_u$ represent the combining beamforming at the u^{th} receiver. $\mathbf{B}$ and $\mathbf{R}$ denote the precoding scheme in the baseband and RF domain on the transmitter side, respectively. The transmit noise due to DAC quantization error is denoted as $\mathbf{z}_{\text{T}}$ and thermal noise at the receiver is $\mathbf{z}_{\text{R}}$. Operation $\mathbf{a}^{\text{H}}$ is the Hermitian transpose of $\mathbf{a}$.

In DA architecture, the precoding occurs entirely in digital baseband and therefore there is no analog processing, i.e., $\mathbf{R}_{\text{DA}} = \mathbf{I}$. The digital precoder $\mathbf{B}_{\text{DA}}$ has dimension $N_{\text{DA}} \times U$.

In SA architecture, the digital precoder $\mathbf{B}_{\text{SA}}$ has dimension $M_{\text{SA}} \times U$ due to M_{SA} RFchains. The RF precoder $\mathbf{R}_{\text{SA}}$ has dimension $N_{\text{SA}} \times M_{\text{SA}}$. Due to the fact that every K_{SA} of phase shifters connect to one RF-chain, $\mathbf{R}_{\text{SA}}$ is a block diagonal matrix

$$\mathbf{R}_{\text{SA}} = \text{diag}\left(\mathbf{r}_{\text{SA},1}, \cdots, \mathbf{r}_{\text{SA},M}\right), \tag{5.2}$$

where column vector $\mathbf{r}_{\text{SA},m}$ with length K_{SA} represents K_{SA} phase shifters that connect to the m^{th} RF-chain. Each element of $\mathbf{r}_{\text{SA},m}$ has unit magnitude[2]. We define the set $\mathcal{S}_m = \{(m-1)K_{\text{SA}} + 1, \cdots, mK_{\text{SA}}\}$ that contains indices of array elements in the m^{th} group.

In FH architecture, the digital precoder $\mathbf{B}_{\text{FH}}$ has dimension $M_{\text{FH}} \times U$. The analog precoder matrix $\mathbf{R}_{\text{FH}}$ has dimension $N_{\text{FH}} \times M_{\text{FH}}$ and its m^{th} column $\mathbf{r}_{\text{FH},m}$ represents the

[2]In fact, analog precoding can be designed with both phase and magnitude tuning capability, which relaxes this constraint. The hardware aspect of phase shifter is discussed in Section V-C.

phase shifting from N_{FH} phase shifters connected to the m^{th} RF-chain, i.e.,

$$\mathbf{R}_{\mathrm{HF}} = \begin{bmatrix} \mathbf{r}_{\mathrm{HF},1} & \mathbf{r}_{\mathrm{HF},2} & \cdots & \mathbf{r}_{\mathrm{HF},M_{\mathrm{FH}}} \end{bmatrix}. \tag{5.3}$$

each element in $\mathbf{R}_{\mathrm{FH}}$ has unit magnitude.

We make the following assumptions. Firstly, the channel information $\mathbf{H}_u$ is known to both transmitter and receivers. A practical way of channel estimation can be found in [83]. Secondly, each UE receiver is equipped with a phased array with only one RF-chain. As a consequence, BS assigns one data stream to each UE receiver. Thirdly, all receivers have the same pre-beamforming SNR and BS assigns equal power among data streams. Fourthly, the combining vector of each receiver $\mathbf{w}_u$ is chosen as the primary left eigenvector of channel matrix $\mathbf{H}_u$ after magnitude normalization in each element.

The SINR at the u^{th} receiver array is denoted as

$$\mathrm{SINR}_u = \frac{\|g_u\|^2}{\sigma^2_{\mathrm{n,rx}} + \sigma^2_{\mathrm{n,tx}} + \sigma^2_{\mathrm{int}}} \tag{5.4}$$

where the signal power gain g_u is given by $g_u = \arg\min_g \mathbb{E}\|y_u - g s_u\|^2$. All signal, noise, and interference powers are relative powers, referenced to 46dBm transmit power based on Table 5.2. As a consequence, receiver thermal noise power $\mathbb{E}\|\mathbf{w}_u^{\mathrm{H}}\mathbf{z}_{\mathrm{R}}\|^2 = \sigma^2_{\mathrm{n,rx}}$ is treated as constant in each use case. The multi-user interference is $\sigma^2_{\mathrm{int}} = \mathbb{E}\|y_u - g_u s_u\|^2$.

In the remaining of the sections, we discuss how to design array parameters for each architecture to reach targeted SINR for three use cases.

5.3.2 Array size and transmit power gain

In principle, increased transmit power $P^{(\mathrm{out})}$ and array size N both improve signal power gain g_u in (5.4). Effectively, they provide higher EIRP and help achieve target SINR from Table 5.2.

In DA and FH, output power of each PA $P^{(\mathrm{out})}/N$ is split into U parts due to multiplexing and even power allocation. Thus each stream in each PA has output power $P^{(\mathrm{out})}/(NU)$. The coherent summation of N-elements via beamforming provides N^2 times increased power.

In SA, however, PAs are partitioned into groups to amplify different streams. For each stream, each PA element outputs $P_{\mathrm{SA}}^{(\mathrm{out})}/N_{\mathrm{SA}}$, while the beamforming gain is N_{SA}^2/U^2. As a consequence, maximum output signal power after beamforming in each architectures is

$$G_{\mathrm{DA}} = \frac{P_{\mathrm{DA}}^{(\mathrm{out})} N_{\mathrm{DA}}}{U}, G_{\mathrm{SA}} = \frac{P_{\mathrm{SA}}^{(\mathrm{out})} N_{\mathrm{SA}}}{U^2}, G_{\mathrm{FH}} = \frac{P_{\mathrm{FH}}^{(\mathrm{out})} N_{\mathrm{FH}}}{U}. \tag{5.5}$$

It is clear that SA is in an disadvantage in terms of signal power gain. SA requires to use more array elements, output power, or both for the comparable output power to DA and FH architectures.

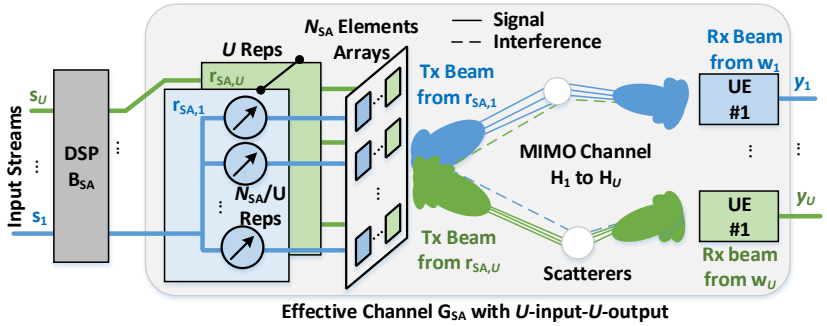
5.3.3 Precoder design

Given maximum signal output power G, the the precoder determines the actual signal power g_u and multi-user interference σ_{int}^2 in (5.4). In this subsection, we discuss precoding techniques for three architectures.

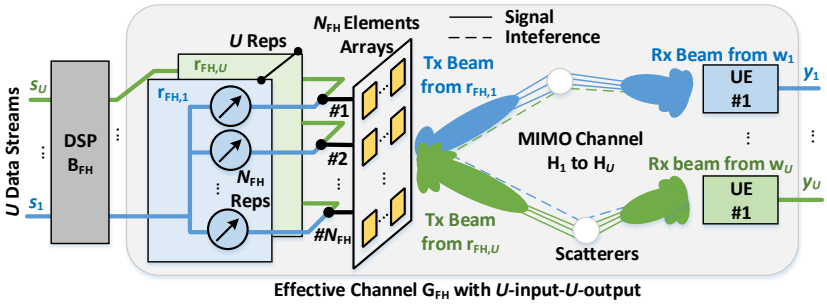
In DA architecture, maximum ratio transmission (MRT) and zero-forcing (ZF) are two commonly used linear precoding approaches. The former maximizes the signal strength at destination and approaches maximum gain discussed in Section 5.2.1, while the latter eliminates multi-user interference. It is commonly believed that because mmW signals suffer from severe propagation loss, the interference is generally less troublesome than sub-6GHz systems. However, the interference from transmitted sidelobes, if not properly handled, can still affect the achievable rate at receivers. In this work, we propose to use regularized ZF beamforming [129], where the introduced regularization coefficient α_{DA} facilitates controlling both signal strength and interference at the receiver.

$$\mathbf{B}_{\mathrm{DA}} = \kappa_{\mathrm{DA}} \mathbf{G}_{\mathrm{DA}}^{\mathrm{H}} (\mathbf{G}_{\mathrm{DA}} \mathbf{G}_{\mathrm{DA}}^{\mathrm{H}} + \alpha_{\mathrm{DA}} \mathbf{I})^{-1}, \tag{5.6}$$

In the above equation, $\mathbf{G}_{\mathrm{DA}}$ is the post-combining multi-user channel with the u^{th} row as $\{\mathbf{G}_{\mathrm{DA}}\}_u = \mathbf{w}_u^{\mathrm{H}} \mathbf{H}_u$. The regularization coefficient α_{DA} controls the behavior of the precoder, i.e., MRT when it approaches positive infinity and ZF when it approaches zero. One can expect SINR maximization when α_{DA} is selected to be the largest with constraint that $\sigma_{\mathrm{int}}^2 \ll$



(a) In sub-array architecture, each beam is steered through a group of N_{SA}/U antenna elements.



(b) In fully-connected hybrid array architecture, each beam is steered through all N_{FH} antenna elements.

Figure 5.2: Two-stage precoding in sub-array and fully-connected hybrid array architectures. The analog precoder steers spatial beams towards intended receivers. The digital precoder uses regularized zero-forcing over effective channel to handle interference.

$\sigma^2_{\mathrm{n,rx}}$. Power scaling parameter κ_{DA} is used to guarantee total transmit power constraint $\|\mathbf{B}_{\mathrm{SA}}\|^2 = P^{(\mathrm{out})}_{\mathrm{DA}}$.

Precoding approaches with SA and FH architectures are currently actively investigated by researchers and are mostly for systems where analog beamformer has phase-only tuning capability. The optimal hybrid precoding is a mixed integer programming problem and its optimal solution must be solved via potentially exhaustive search. Many sub-optimal methods have been proposed for near optimal performance, e.g., works in [16] for FH architecture. In [16], the analog precoder is selected to point beams towards directions of intended

receivers. The digital precoder is then used to handle associated interference among beams synthesized by phase shifters. In the following paragraphs regarding precoding algorithm for SA and FH, we adopt assumption of phase-only analog precoder.

In SA architecture, we propose to use the following approach as a modification of FH beamforming in [16] and the scheme is illustrated in Figure 5.2(a). We first merge adjacent M_{SA}/U phase shifter groups in SA into one virtual group. It leads to N_{SA}/U array elements within each virtual group in an ideal scenario[3]. The input signal of RF-chains within a virtual group are exactly the same. Let us denote set $\mathcal{V}_u$ as one that contains index of physical array groups within the u^{th} virtual group. The analog beamformer is chosen to synthesize beams towards primary propagation direction to U receivers

$$\mathbf{r}_{\mathrm{SA},m} = \exp\left[j\angle\left(\{\mathbf{H}_u^{\mathrm{H}}\mathbf{w}_u\}_{\mathcal{S}_m}\right)\right], m \in \mathcal{V}_u. \tag{5.7}$$

In the above equation, $\angle(\{\mathbf{a}\}_{\mathcal{S}_m})$ selects elements from vector $\mathbf{a}$ according to indices from set $\mathcal{S}_m$ and finds phases of selected elements. Let us denote the effective channel as $\mathbf{G}_{\mathrm{SA}}$ which contains the effect of receiver combiner and RF precoder in multi-user channel. The m^{th} row is defined as $\{\mathbf{G}_{\mathrm{SA}}\}_m = \mathbf{w}_m^{\mathrm{H}}\mathbf{H}_m\mathbf{R}_{\mathrm{SA}}$. Note the effective channel $\mathbf{G}_{\mathrm{SA}}$ is the channel between digitally precoded stream and UEs . As a consequence, the digital precoding problem in SA can be solved in the regularized-ZF framework

$$\mathbf{B}_{\mathrm{SA}} = \kappa_{\mathrm{SA}}\mathbf{G}_{\mathrm{SA}}^{\mathrm{H}}(\mathbf{G}_{\mathrm{SA}}\mathbf{G}_{\mathrm{SA}}^{\mathrm{H}} + \alpha_{\mathrm{SA}}\mathbf{I})^{-1} \tag{5.8}$$

The power scaling coefficient κ_{SA} is used to meet total output power constraint, i.e.,

$$\|\mathbf{R}_{\mathrm{SA}}\mathbf{B}_{\mathrm{SA}}\|^2 = P_{\mathrm{SA}}^{(\mathrm{out})}.$$

Similar to precoding in the digital array, the regularization coefficient α_{SA} is chosen to maximize SINR.

The precoding scheme in FH architecture is illustrated in Figure 5.2(b). Only U out of M_{FH} RF-chains are turned-on to provides U streams. Without loss of generality, the first U

[3]Ideal scenario is defined when the ratio M_{SA}/U is an integer. Using a reduced number of arrays can be used when it is not valid, but this scenario is not considered for simplicity.

RF-chains are active and the analog precoder is

$$\mathbf{r}_{\mathrm{FH},u} = \exp\left[j\angle(\mathbf{H}_u^{\mathrm{H}}\mathbf{w}_u)\right], u \leq U. \tag{5.9}$$

The digital precoder in FH is a regularized ZF over $\mathbf{G}_{\mathrm{FH}}$, the effective channel that contains the receiver combining and RF precoding in the multi-user channel

$$\mathbf{B}_{\mathrm{FH}} = \kappa_{\mathrm{FH}}\mathbf{G}_{\mathrm{FH}}^{\mathrm{H}}(\mathbf{G}_{\mathrm{FH}}\mathbf{G}_{\mathrm{FH}}^{\mathrm{H}} + \alpha_{\mathrm{FH}}\mathbf{I})^{-1}. \tag{5.10}$$

The u^{th} row is defined as $\{\mathbf{G}_{\mathrm{FH}}\}_u = \mathbf{w}_u^{\mathrm{H}}\mathbf{H}_u\mathbf{R}_{\mathrm{FH}}$. Similar to precoding in the SA architecture, κ_{FH} is the power scaling coefficient for $\|\mathbf{R}_{\mathrm{FH}}\mathbf{B}_{\mathrm{FH}}\|^2 = P_{\mathrm{FH}}^{(\mathrm{out})}$ and α_{FH} is the regularization coefficient.

5.3.4 Digital-to-analog converter precision

The transmit noise in (5.4) comes from the quantization error due to DACs with finite precision. A practical system design uses sufficient quantization precision such that the transmission noise level stays well below the receiver thermal noise. Different architectures require different values of effective number of bits (ENOB) for such goal. The required ENOB in three architectures are

$$\begin{aligned}
\tilde{B}_{\mathrm{DA}} &= \frac{\mathrm{PAPR} - 1.76 + D + 10\log_{10}\left(\frac{P_{\mathrm{DA}}^{(\mathrm{out})}}{\sigma_{\mathrm{n,rx}}^2}\right)}{6} \\
\tilde{B}_{\mathrm{SA}} &= \frac{\mathrm{PAPR} - 1.76 + D + 10\log_{10}\left(\frac{P_{\mathrm{SA}}^{(\mathrm{out})}}{\sigma_{\mathrm{n,rx}}^2}\frac{N_{\mathrm{SA}}}{U^2}\right)}{6} \\
\tilde{B}_{\mathrm{FH}} &= \frac{\mathrm{PAPR} - 1.76 + D + 10\log_{10}\left(\frac{P_{\mathrm{FH}}^{(\mathrm{out})}}{\sigma_{\mathrm{n,rx}}^2}\frac{N_{\mathrm{FH}}}{U}\right)}{6}
\end{aligned} \tag{5.11}$$

for transmit noise to be D dB lower than AWGN. In the above equation, peak-to-average power ratio (PAPR) represents the PAPR of the input signal of each DAC. Note that these expressions are accurate when DAC quantization errors are uncorrelated, which may not be valid with small number of bits, e.g., $B = 1$ bits. Derivations of (5.11) are provided in the Appendix D.1.

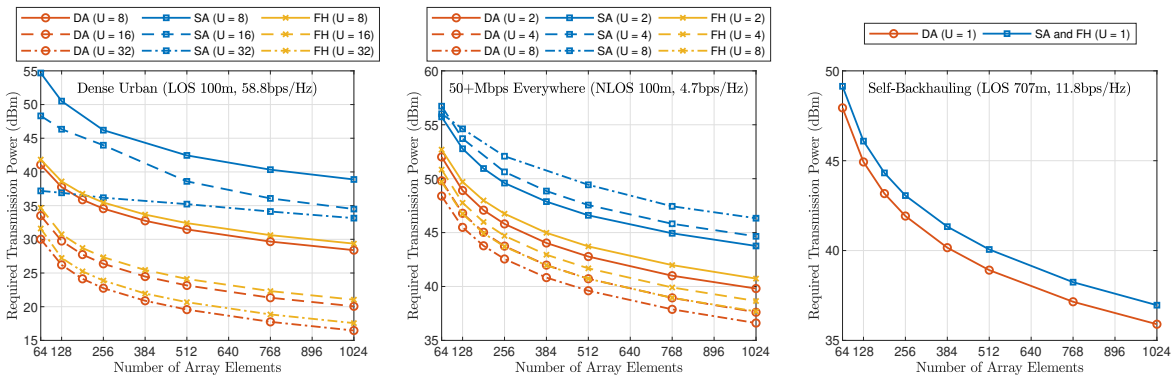


Figure 5.3: The required total transmit power $P^{(\text{out})}$ with different number of array elements (N) to reach spectral efficiency target in three typical 5G use cases.

Equation (5.11) together with (5.5) indicates following facts. Firstly, with fixed signal power gain G_{DA}, DACs precision in DA architecture can be reduced by increasing array size and decreasing transmit power. For SA and FH, however, the transmit noise remain constant regardless of the source of signal power gain. Secondly, with the same signal power gain and transmit power, DA architecture has lower requirement in DAC quantization as compared to SA and FH.

5.3.5 Phase shifter precision

In both SA and FH architectures, finite resolution of phase shifters leads to a changed power level of sidelobes and shifted location of nulls, as compared to system using ideal devices. More importantly, the locations of main lobe varies and associated signal gain drops. One might expect highly precise phase shifters are required to accurately control beams. In this subsection, we discuss the impact of finite resolution of phase shifters on SA and FH architectures.

The former issue regarding the distorted sidelobes is less troublesome in both SA or FH transmitter array architecture. Sidelobes lead to multi-user interference as seen from the off-diagonal elements in the effective channel $\mathbf{G}_{\text{SA}}$ and $\mathbf{G}_{\text{FH}}$. When system is aware of potential interference, digital precoding stage can be used to effectively suppress them. A practical way

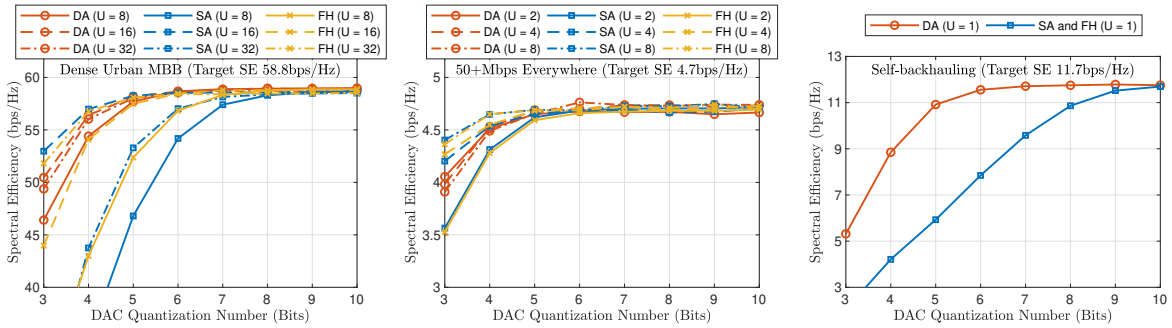


Figure 5.4: Spectral efficiency performance with quantization on the baseband precoding and digital-to-analog converter (DAC). Three architectures use 256 array elements and output power is adjusted according to Figure 5.3. The baseband precoding uses fixed point operation with precision 2 bits greater than associated DAC quantization which ensures negligible degradation as compared to baseband precoding with floating point operation.

to acquire the information of effective channel is via a training procedure where BS and UE use quantized analog beamformer to exchange pilot symbols and estimate effective channel $\mathbf{G}_{\mathrm{SA}}$ and $\mathbf{G}_{\mathrm{FH}}$. This training procedure is similar to the multi-beam scheme proposed for the next generation of mmW indoor system [66]. Meanwhile, the gain reduction due to finite phase shifter resolution is not severe either. In fact, the gain degradation is lower bounded by 0.68 dB, 0.16 dB and 0.04 dB with $Q = 3, 4, 5$ bits quantization of phase shifters and does not scale with the array size or multiplexing level. An analysis that supports these numbers is provided in the Appendix D.2. Equivalently, the gain degradation is bounded by 0.16 dB so long as angle error of phase shifters are no larger than 11.25°. Such specifications are not difficult to meet in state-of-the-art devices as it will be discussed in Section 5.5.3.

5.3.6 Simulation results

In this subsection, simulation results are presented to show the required design parameters to reach SE target in three array architectures.

In the simulation, 3D mmW MIMO channel between BS and U UEs are generated according to mmW sparse scattering model [16]. The channel between BS and each UE consists

of 20 multi-path rays in 3 multipath cluster and LOS cluster, if exists, is 10 dB stronger than the rest. AoA and AoD of clusters are uniform random variables within azimuth range $[-60°, 60°]$ and elevation range $[-30°, 30°]$. Azimuth and elevation AoA and AoD of rays within a cluster have random deviations from the cluster specific AoA and AoD, and they follow zero mean Laplacian distribution with 10° standard deviation. In dense urban eMBB, a scheduler is assumed such that the LOS paths of all target receivers are unique [68]. The mean SE is evaluated by taking average of SINR in (5.4) over U UEs and use Shannon capacity formula, i.e., $\text{SE} = \sum_{u=1}^{U} \log_2(1 + \text{SINR}_u)$. The data streams used in the simulation are Gaussian distributed and their magnitudes are truncated such that PAPR is 10 dB.

With ideal hardware, the required transmit power $P^{(\text{out})}$ to reach SE target with various antenna size N and number of data streams U in three architectures are shown in Figure 5.3.

We first focus on how transmit power changes with parameter N and U. Increasing array size N is effective in reducing transmit power in all scenarios since it helps improve both signal gain and interference control from narrow beams. When interference from multi-beam is negligible, the transmit power saving from increasing U depends on difference between the SINR target reduction in Table 5.2 and signal gain dropping in (5.5). For example, when the number of spatial streams U increases from 2 to 4 and 4 to 8 in eMBB, the SINR requirement reduces by 5.2 dB and 4 dB. Meanwhile, the signal gain changes by 3 dB, 6 dB and 3 dB in DA, SA, and FH, respectively. Therefore DA and FH save around 2.2 dB and 1 dB $P^{(\text{out})}$ and SA is forced to use around 0.8 dB and 2 dB higher $P^{(\text{out})}$. It is also true in high-N regime of DA and FH in the Dense Urban eMBB. When the number of spatial steams U increases from 8 to 16 and 16 to 32, the SINR requirement reduces by 11.4 dB and 6.6 dB. Therefore the power saving at $N = 1024$ is around 8.4 dB and 3.3 dB for both DA and FH. Power saving is more difficult to predict when system needs to trade power gain for interference control. Therefore the transmit power saving from increasing U with smaller antenna N and large multiplexing U is less accurately using the above analysis.

Then we focus on the comparison between array architectures. There is one universal conclusion that holds true for DA and FH in all scenarios. DA and FH have the same

maximum signal gain when $P^{(\text{out})}$ and N are the same according to (5.5). In simulation, FH actually requires near 1 dB higher $P^{(\text{out})}$ than DA in all scenarios. This gap is due to the loss from the two-stage precoding of FH. Further exploiting hardware capability, e.g., using phase-and-magnitude analog precoders, and designing better hybrid precoding algorithm in FH would reduce this gap.

Next, we compare array architectures in each use case. In self-backhauling where data stream number U is constraint by point-to-point environment, SA has the same performance as FH as both architectures become the same in model (5.1). They both require 1 dB higher transmit power than DA. Secondly, the difference of required transmit power between architecture can by analyzed by (5.5) in 50+Mbps Everywhere. Equation (5.5) reveals that SA has U times lower power gain than other architectures and it is shown in the figure that that SA requires U times higher $P^{(\text{out})}$ than FH for the same performance. Equation (5.5) predicts the gap between curves well in the since there is negligible interference with small number of beams. Thirdly, in eMBB use case the required transmit power gap between SA and FH in Dense Urban eMBB meet (5.5) when N is large, i.e., SA requires to use $9, 12, 15$dB higher $P^{(\text{out})}$ than FH when $U = 8, 16, 32$ beams are used. However, the transmit power gap between SA and FH deviates from what (5.5) predicts when N is small. This deviation is due to power gain and interference control trade-off. Dense Urban eMBB features a large number of simultaneous data streams and the mutual interference among streams becomes system bottleneck when beam-width is not small enough.

With $U = 8$, the transmit power gap between SA and FH increases from 9 dB to 13 dB when N reduces from 1024 to 64. The additional 4 dB gap is the cost of controlling interference in SA, because the SA uses nearly U times wider beam to carry each data stream as compared to FH. Further, the BB precoding of SA is forced to sacrifice more gain for interference control. With $U = 32$, the gap reduces from 9 dB to 6 dB when N reduces from 1024 to 64. One may expect each data stream in SA is carried by wide beams with $N/U = 2$ antennas and conclude the opposite results. However, with $U = 32$ data streams, each RF-chain is connected with at most $N/U = 2$ antennas and such architecture is effectively a

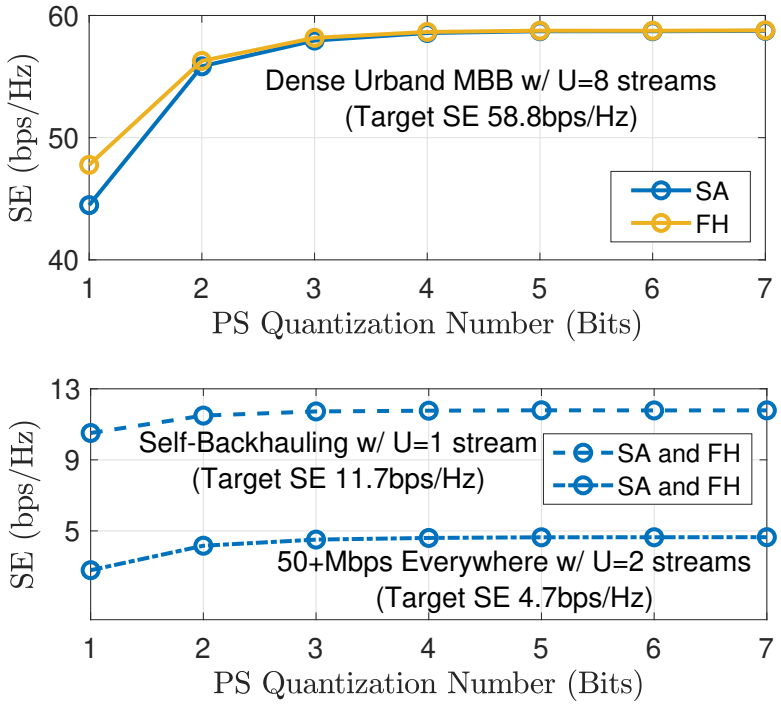


Figure 5.5: Spectral efficiency performance with quantization on radio frequency phase shifter of sub-array and fully-connected hybrid array. Both architectures use 256 array elements without baseband and digital-to-analog converter quantization error and output power is adjusted according to Figure 5.3. for target spectral efficiency.

digital array. In fact, the BB precoding stage in SA facilitates each stream to be transmitted by nearly all antenna elements and improves the signal gain. In fact, the intuition of hybrid precoding approach [16] may not be true and a better hybrid precoding scheme tailored for this regime would provide more additional power saving for SA.

With finite precision in the baseband precoding, DAC and phase shifters, the SE performance is shown in Figure 5.4 and Figure 5.5. For clarity, all array architectures use 256 antenna elements and the transmit power $P^{(\text{out})}$ in each architecture is chosen such that it delivers the same SE performance as in quantization free cases. Figure 5.4 shows the required quantization bits in BB precoding and DAC and it matches with the analysis. According to (5.11), the required ENOB for transmit noise to be $D = 15\,\text{dB}$ lower than AWGN in the

Dense Urban eMBB with $U = 8$ streams are 5.1, 8.0, and 7.7 in DA, SA, and FH architectures, respectively. The SE improvement in Figure. 5.4 is saturated once DAC quantization bits are beyond these values. Equation (5.11) also precisely matches with Self-backhauling use case where DA, SA, and FH requires 5.8, 10.0, and 10.0 ENOB, respectively. It is worth noting that the additive quantization error model becomes inaccuracy when the analytical ENOB from (5.11) is significantly small. For example, equation (5.11) estimates that system requires 1 to 4bits for the most scenarios in 50+Mbps Everywhere, while the required ENOB from simulation is close to 5bits. A rule-of-thumb is to use at least 5 bits. Note that this inaccuracy regime of (5.11) does not affect power consumption estimation of the system, because the direct current (DC) power of DAC does not effectively reduce by using less than 5 bits due to the fixed hardware overhead and it is discussed in details in Section. 5.5.1.

Moreover, the precision requirement in BB precoding and DAC of DA is in general lower than hybrid architectures throughout all scenarios and it suggests a system level power consumption saving. Last, Figure 5.5 shows that with the hybrid precoding approach in Section 5.3.3, the SE performance is negligibly affected by phase shifter quantization and it matches with our analysis in Section 5.3.5.

In summary, for the same target SE performance, DA requires a reduced transmit power or number of array elements as compared to SA and FH. Besides, the DAC quality of DA is relaxed as compared to the hybrid architecture. A fair comparison among architectures cannot overlook these factors by restricting architectures to use the same transmit power, number of array elements, or specification of hardware components. The design parameter trade-off analyzed in this section leads to a more practical comparison in Section 5.6.

5.4 Hardware design challenges of transmitter array

In this section, we discuss practical hardware design of mmW arrays with different architectures. We first introduce the distributed array processor module. Then, the necessary circuits blocks for BB signal and RF signals distribution are discussed.

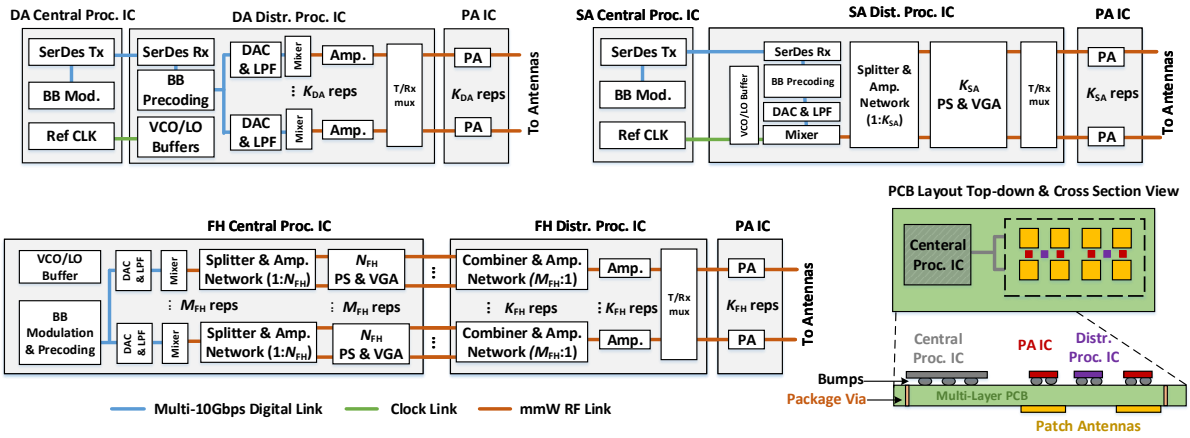


Figure 5.6: The proposed millimeter-wave antenna array circuits layout plan of three architectures. The functionality of transceivers arrays are divided into a centralized processing integrated circuits (Central Proc. IC), multiple distributed processing integrated circuits (Distr. Proc. IC) and power amplifiers which are close to patch antenna elements to reduce millimeter-wave signal Loss. Other abbreviations used in the figure include baseband modulation (BB Mod.), anti-aliasing low-pass filters (LPF), transmitter/receiver multiplexer (mux), reference clock (Ref CLK), amplifiers (Amp.), and repetitions of elements (reps).

5.4.1 Distributed array module

The conventional MIMO system integrates array processing module in an IC and delivers RF signal to antennas. Such centralized design may not be practical in mmW system with massive number of antennas. With a compact and centralized IC, mmW signals routed to hundreds of array elements suffer severe insertion loss[4]. Besides, the heat dissipation becomes a concerns for a centralized solution. Moreover, array size scalability becomes challenging since adding more elements requires completely new processing module.

A practical solution is to implemented processing hardware for antenna arrays in a dis-

[4]The wavelength at 28 GHz band is 10.7 mm, 256 antennas in a square alignment with half-wavelength require at least 7327mm^2.

tributed manner [134]. In DA and SA, each IC in a processing module integrates the processing circuits for K_{DA} and K_{SA} antennas and is located close to these antennas. Although a centralized digital processor is still necessary for some baseband functionality, e.g., symbol mapping and channel coding, the digital baseband precoding can be implemented in each distributed module. With such design, the system needs to deliver U digital signal streams rather than M digitally precoded signal streams to the processing modules [134]. It offers a significant saving of baseband signal distribution throughput given $M \gg U$ in DA and SA. The DAC, up-converter and RF signal processing are also included in the processing module. The digital signals from central processor are routed and recovered through Serializer/Deserializer (SerDes) sub-system in each of the processing modules. Note that the exact value of elements integrated in an IC affects system area and energy. But the discussion of that is beyond the scope of this work. The patch antenna is directly attached on the PCB.

The illustration of distributed DA hardware implementation is shown in Figure 5.6. In the remaining of the chapter, power consumption and cost estimation of DA system is based on design where each module contains $K_{\mathrm{DA}} = 8$ antenna elements and associated processing circuits. Each DA module contains SerDes, voltage-controlled oscillator (VCO) within a phase-locked loop (PLL), and RF-chains and T/Rx multiplexers. The power amplifiers for 5G mmW applications are expected to be built in non-silicon material, as shown in Section 5.5.4 and they are placed next of DA processing IC.

The illustration of SA implementation is illustrated in Figure 5.6. In SA, each module has processing circuits for K_{SA} antenna elements. Each of them contains SerDes, VCO and phase shifter networks.

There is no priori work on FH implementation with larger than 8 antennas. The RF signal routing is a challenging task in FH architecture, because the input signal for each antenna element is a combination of signals from all RF-chains. The most viable approach we could anticipate is illustrated in Figure 5.6. Opposite of DA and SA architectures, routing loss cannot be reduced by distributing RF-chains into a closer position, since their outputs are required to be delivered to entire PCB board. In the proposed design, each array module

integrates a combining network and delivers the combined signal to nearby antenna elements. It also contains RF amplifiers to compensate for insertion loss during the RF signal routing and combining.

In all array architectures, routing digital baseband signal and RF signals plays a critical roles. We discuss associated challenge and solutions in the next subsections.

5.4.2 Baseband signal distribution

The digitally precoded sample streams require to be routed into each processing module by serial-link transceivers in all array architectures. The state-of-the-art SerDes supports data rates over 50Gb/s using PAM-4 signaling in wireline chip-to-chip communication. The specific design of SerDes system is beyond the scope of this work. In Section 5.5, we use the specifications of ultra-high-speed transceivers.

5.4.3 Radio-frequency signal distribution

Multiple circuit components introduce non-negligible insertion losses that need to be carefully handled by system designers.

- *PCB and Inter-Connectors Loss:* RF signal suffers from interconnect loss between the silicon chip RF ports and the antenna elements. The low-loss PCB board, such as RO 3000 series and 4000 series, 28 GHz signal have 1.25dB/inch insertion. Besides, each IC chip require to be placed on organic or ceramic substrate (interposer) to distribute the chip ports to a ball-grid array and it has an additional 1 to 2 dB distribution loss. This implementation loss needs to be pre-compensated before the RF signal is fed into antenna.

- *Intra-Chip Transmission Lines Losses:* RF signal loss in silicon is significant at mmW band. According to [200], there is up to 0.6dB/mm transmission line loss at 28 GHz. The length of transmission line is proportional to the IC size but exact value is determined by actual IC design. According to a 60 GHz array design [199], phase shifter

and Wilkinson RF splitter take most of the IC area. The intra-chip routing loss can be roughly estimated by taking into account the required area of those components. With the practical components size in Section 5.5, the loss in an SA module with $K_{\mathrm{SA}} = 32$ phase shifters is less than 1 dB but up to 3 to 4 dB for FH since each RF-chain distributes signals into hundreds of phase shifters that require dozens of millimeters square area.

- *Power Splitters and Combiners Loss:* In the analog beamforming stage of SA and FH architectures, output signals of RF-chains need to be fed into phase shifter network for phase rotation. The Wilkinson power splitters are commonly used for such purpose [97, 117, 199]. Moreover, the fully-connected hybrid architecture uses same Wilkinson structure to combine multiple RF signals before power amplification. An ideal power splitter/combiner introduces 3 dB insertion loss in each of the one-to-two splitter (1:2) or two-to-one combiner (2:1) unit. Practical design often has an additional 1 dB implementation loss. It results in a $4\log_2(K_{\mathrm{SA}})$ dB power drop in the SA architecture. For FH architecture, the splitters and combiners introduce total $4\log_2(N_{\mathrm{FH}}M_{\mathrm{FH}})$ dB loss.

All the above RF insertion losses lead to an reduced EIRP at the antenna and therefore need to be properly compensated. The detailed distribution budget in all architectures is discussed in Section. 5.5.4.

5.5 Hardware power and cost modeling

In this section, we first provide the power and cost model of necessary circuits blocks based on a survey of the state-of-the-art circuits design and measurement. The power consumption contains DSP module for precoding, SerDes, mixed signal components, and RF components. Note that other hardware blocks such as power supply, active cooling may consume considerable power [22]. We omit them in this work since these are constant hardware overhead. Then, examples are provided for signal distribution budgets calculation in order to determine

necessary RF amplifiers to compensate insertion loss. Finally, we summarize the total power and cost calculating formula for all architectures operating with different design parameters.

5.5.1 Digital signal processing power

Due to large bandwidth, the array processing in the digital baseband needs to support such high throughput. The DSP for array processing mainly consumes power for digital precoding and digital signal routing. Note that tasks such as channel coding, higher layer processing in the communication standard stack are not included since they have equal power consumption for all architectures. Channel estimation and precoder computation are also omitted since they occur at time scale that is several orders of magnitude longer than symbol duration.

The DSP power estimation contains linear precoding and 4096 point inverse DFT[5]. The precoding requires multiplication of $M \times U$ complex matrix with $U \times 1$ complex vector. It has $6UM$ fixed points operations. Note that the number of operation does not change with different design choices of N_{FFT}, because the number of precoder slices in sub-carriers and symbol duration change. The latter consists of $\log_2(N_{\mathrm{FFT}}) = 12$ complex multiplication per sample per RF-chain, and it results in $6 \times 12M \times \mathrm{BW}$ operations per second. We use the figure-of-merit (FOM) of digital backend $\mathrm{FOM}_{\mathrm{DSP}} = 13\mathrm{GOPS/mW}$ in 40 nm CMOS as state-of-the-art fixed point digital computation efficiency [194]. As a consequence, the power consumption in the digital precoding is

$$P_{\mathrm{Precoding}} = \frac{(6UM + 72M) \times \mathrm{BW}}{\mathrm{FOM}_{\mathrm{DSP}}} \tag{5.12}$$

where BW is the signal bandwidth. The power consumption $P_{\mathrm{Precoding}}$ has unit Watt.

The power of SerDes system is modeled in the following equation

$$P_{\mathrm{SerDes}} = \mathrm{FOM}_{\mathrm{SerDes}} \times \mathrm{BW}_{\mathrm{OS}} \times \mathrm{ENOB} \times U \tag{5.13}$$

In the above, ENOB is the required precision in the digital precoding and DAC of mmW transmitter and its value is determined according to the analysis in (5.11) and Figure 5.4.

[5]We assume $N_{\mathrm{FFT}} = 4096$ point inverse-DFT for 850 MHz signal bandwidth to achieve subcarrier spacing 240 kHz [4]

P_{SerDes} scales with the number of independent data stream U due to the distributed digital precoding. The figure-of-merit of SerDes is adopted as $\text{FOM}_{\text{SerDes}} = 10\text{mW/(Gb/s)}$ [47] in this work. Note here we use BW_{OS} as the oversampled data rate after considering a factor of 2 oversampling ratio, i.e., $\text{BW}_{\text{OS}} = 1.7\text{GS/s}$.

5.5.2 Power model of mixed signal components

In Section 5.3.4, we analyze the impact of DAC quantization in different array architecture. The DAC power consumption is mainly determined by the sampling frequency and effective number of bits. The total power consumption in each DAC is computed using the following equation

$$P_{\text{DAC}} = \text{FOM}_{\text{DAC}} \times \left(2^{\text{ENOB}} \times \text{BW}_{\text{OS}}\right) + P_{\text{buffer}} \tag{5.14}$$

where P_{DAC} has unit. BW_{OS} and are similarly define in (5.13). The state-of-the-art specification of DAC is $\text{FOM}_{\text{DAC}} = 0.08\text{pJ/conversion}$ [121]. A constant hardware overhead for signal amplification is modeled as $P_{\text{buffer}} = 10\text{mW}$ for -14dBm output signal power. Therefore further reducing precision has limited power saving benefits when P_{buffer} dominates.

5.5.3 Power model of radio-frequency circuit components

In this section, we estimate the required power consumption in the RF IC, including the power for signal amplification and analog array processing for hybrid architecture. The components are phase shifter, LO using PLL, mixer, RF amplifier for gain compensation, and the power amplifier for transmission.

- *LO and mixer:* The phase noise of an oscillator is inversely proportional to the power dissipated [134]. The state-of-the-art VCO design [54, 55, 58, 59] facilitates phase noise lower than -110dBc/Hz at 1MHz by using less than 30mW DC power consumption, and system performance is not affected by such noise specification [98]. Considering the required buffer at the output, the power consumption of VCO block can be $P_{\text{VCO}} = 60\text{mW}$ for each element. Mixer can be made by active or passive devices. Practically,

passive mixers are easier to implement and have better linearity and noise. Mixers require enough LO signal power to be driven. In this work, we select the input LO power to be at least -5dBm and the power consumption of mixer is $P_{\text{Mixer}} = 10\text{mW}$. The total power consumption of LO is $P_{\text{LO}} = 70\text{mW}$

- *Phase shifter:* RF phase shifting can be implemented in various ways, see [133] for a comprehensive survey. The state-of-the-art work uses reflective-type phase shifter (RTPS) and switch-type phase shifter (STPS) as main approaches of passive PS [63,77, 97,113,157]. Such approaches use delay line with controllable length to generate desired phase shifting. Although nearly zero DC power consumption is required, passive PS often has high insertion loss and large IC area due to the delay line. The active approach uses vector modulator (VM), which consists of variable gain amplifier in both In-Phase and Quadratic RF path to generate a complex gain as magnitude adjustment and phase shifting coefficient. VM requires active devices and has higher power consumption than STPS or RTPS. Meanwhile, VM requires less IC area [117, 158, 199]. In this work, we use VM for building block of hybrid architecture and the power model is $P_{\text{PS}} = 10\text{mW}$ with 2 dB gain.

5.5.4 RF signal amplification power

The RF signals amplification has two categories: gain compensation amplifier and power amplifier.

- *RF amplifier:* Gain compensation amplifiers are used to compensate insertion loss in the analog beamforming for hybrid architectures. As discussed in Section 5.4, hybrid architectures require to distribute up-converted RF signal into phase shifter networks. During this procedure, insertion loss is introduced in power splitter, transmission line and power combiner. These losses need to be properly compensated in order to deliver sufficient radiated signal power at the antenna. From the cost perspective, it is better to provide the gain before power splitting occurs since it requires fewer number of

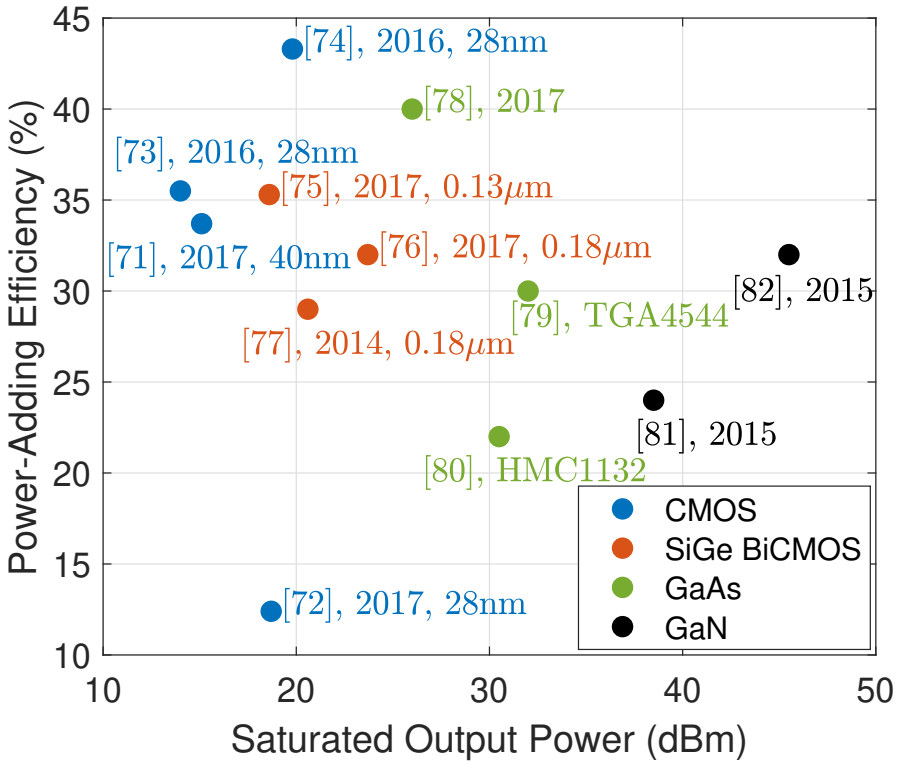
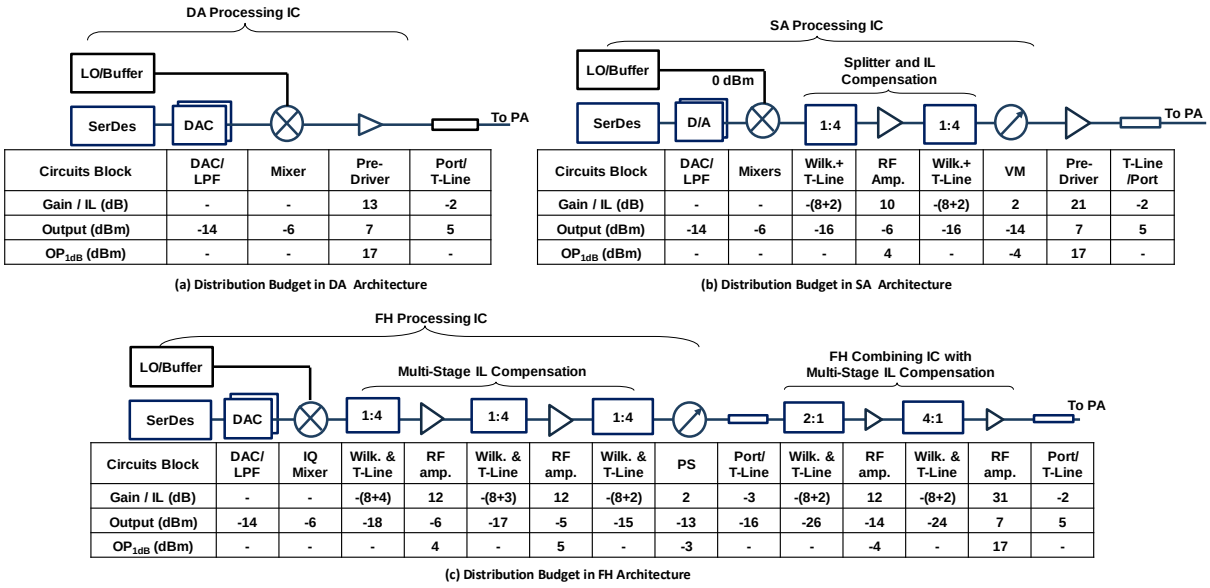


Figure 5.7: The saturated output power and power-added efficiency at that power level (PAE_{max}) of the state-of-the-art power amplifiers at 28 GHz. Data are from [20, 52, 61, 95, 118, 122, 127, 135, 151, 152, 154, 155] and labels include publication date and silicon technology.

amplifiers. However, it raises the linearity concern of CMOS amplifier. As it is shown in the next subsection, a large hybrid array has more than 20 dB insertion loss in the distribution route and in order to pre-compensate such loss immediately after up-conversion leads to a severe nonlinear distortion in RF signals. A practical design typically places amplifiers in a hierarchical manner along RF signal distribution route [199]. Besides, the gain compensation amplifiers need to be carefully designed and their power consumption cannot be overlooked. The power model adopted in this work considers gain compensation amplifier design from [200], where each amplifier has up to 15 dB gain with $P_{\text{Amp}} = 40\text{mW}$ power consumption. Note that active combining [117] is an alternative approach that combines RF signal in current mode using low-noise amplifiers. Although insertion loss can be avoided, there is power consumption in each combiner. We do not discuss this approach in details.



Circuits Block	DAC/ LPF	Mixer	Pre-Driver	Port/ T-Line
Gain / IL (dB)	-	-	13	-2
Output (dBm)	-14	-6	7	5
OP_{1dB} (dBm)	-	-	17	-

(a) Distribution Budget in DA Architecture

Circuits Block	DAC/ LPF	Mixers	Wilk.+ T-Line	RF Amp.	Wilk.+ T-Line	VM	Pre-Driver	T-Line /Port
Gain / IL (dB)	-	-	-(8+2)	10	-(8+2)	2	21	-2
Output (dBm)	-14	-6	-16	-6	-16	-14	7	5
OP_{1dB} (dBm)	-	-	-	4	-	-4	17	-

(b) Distribution Budget in SA Architecture

Circuits Block	DAC/ LPF	IQ Mixer	Wilk. & T-Line	RF amp.	Wilk. & T-Line	RF amp.	Wilk. & T-Line	PS	Port/ T-Line	Wilk. & T-Line	RF amp.	Wilk. & T-Line	RF amp.	Port/ T-Line
Gain / IL (dB)	-	-	-(8+4)	12	-(8+3)	12	-(8+2)	2	-3	-(8+2)	12	-(8+2)	31	-2
Output (dBm)	-14	-6	-18	-6	-17	-5	-15	-13	-16	-26	-14	-24	7	5
OP_{1dB} (dBm)	-	-	-	4	-	5	-	-3	-	-	-4	-	17	-

(c) Distribution Budget in FH Architecture

Figure 5.8: The signal distribution budget example of three array architectures.

- *Power amplifier:* Power amplifiers consume large amount of power in current base-stations operating in sub-6GHz band. In the mmW BS system design there are two conflicting scaling direction. On one hand, the transmit power of each PA is relaxed

due to the use of massive antenna array for similar total power. On the other hand, the power amplifier efficiency is lower than those designed for sub-6GHz band. In Figure 5.7, specifications of the state-of-the-art mmW power amplifier at 28 GHz are shown. Specifically, the power-added-efficiency (PAE) at saturated output power and associated saturated output power are presented. Different semiconductor technologies, e.g., CMOS, BiCMOS, Gallium Arsenide (GaAS), and Gallium Nitride (GaN) are included. The state-of-the-art CMOS or SiGe BiCMOS PAs are not suitable due to the low saturated output power. Assuming 10 dB PAPR margin, even with an extremely large array of 1024 elements, the 46dBm total transmitter power leads to 16 dBm output power for each element. Thus the PA is likely to require a saturation point of 26dBm and this is a challenging target for PAs suitable for deployment in arrays. GaAs PAs are generally cheaper than GaN PAs and are expected for 5G array applications without operating in strongly nonlinear region. In the proposed PA power consumption model, a PA efficiency is $\eta_{\text{PA}} = 0.185$ is adopted. Specifically, the calculation of PA efficiency is based on 0.3 peak PAE, 10 dB power back-off, and a Doherty PA architecture[6]. Accordingly, the power consumption in each PA element is

$$P_{\text{PA}} = \frac{P^{(\text{out})}}{N\eta_{\text{PA}}}, \tag{5.15}$$

where the number of array elements N and output power $P^{(\text{out})}$ are from Figure 5.3 in each architecture.

5.5.5 Summary of specifications of circuits blocks for transmitter array

In Figure 5.8, we present the signal distribution budget example of three array architectures with 64 elements. Specifically, we focus on the insertion loss in PCB, silicon, and RF devices

[6]In Doherty PA, the PAE remain constant when the instantaneous output magnitude a is no more than 3 dB weaker than the peak magnitude $a_{\max}$, i.e., $\text{PAE}(a) = \text{PAE}_{\max}, a \geq a_{\max}/2$. Otherwise, the PAE drops as a linear function of instantaneous output magnitude, i.e., $\text{PAE}(a) = \frac{2a}{a_{\max}}\text{PAE}_{\max}, a < a_{\max}/2$. Thus, the average efficiency is $\eta_{\text{PA}} = \int_a f_A(a)\text{PAE}(a)da$, where $f_A(a)$ is the probability distribution of signal magnitude. When $\text{PAE}_{\max} = 0.3$ and the signal magnitude is Rayleigh distributed with average power 10 dB below the peak, PA efficiency is $\eta_{\text{PA}} = 0.185$.

as modeled as in Section 5.4. There is more than 10 dB loss for every two stages of Wilkinson splitters/combiners plus associated transmission line. As a consequence, RF amplifiers can be placed to compensate such loss in SA as shown in Figure 5.8(b). For FH, multi-stage compensation is required to avoid saturation as shown in Figure 5.8(c). Such design is commonly adopted in implementation of phased array [199]. Moreover, a combining network in FH also needs similar design. For a splitting or combining network with N_{wilk} ports, we use an approximation number of $\sum_{n=1}^{\infty} N_{\text{wilk}}/4^n \approx N_{\text{wilk}}/3$ amplifiers for simplicity. Therefore, FH requires a total $UN_{\text{FH}}/3$ amplifiers in both splitting and combining network. Moreover, for all architectures, we assume a 5dBm signal strength is required at the input of PA [20,135]. The output of each mixer is -6dBm. For a Wilkinson splitter or combiner with N_{wilk} ports, a total $N_{\text{wilk}} - 1$ splitting (1:2) or combining (2:1) units are required. As a consequence, the required number of Wilkinson units are $(K_{\text{SA}} - 1)N_{\text{SA}}/K_{\text{SA}}$ and $U(N_{\text{FH}} - 1) + N_{\text{FH}}(U - 1)$ in the SA and FH architectures, respectively. A summary of specifications of circuits blocks, total number of blocks in each architectures, and required number of blocks per antenna element are summarized in Table 5.3.

5.6 Comparison results

In this section, we present the power and hardware cost comparison among three architectures. Then, we discuss the scalability of these architectures for future trends. Specifically, we focus on the impact of increased throughput requirements and improved energy efficiency in digital computation due to silicon scaling.

5.6.1 Power consumption of millimeter-wave array architectures

The required power consumption in three use cases is presented in Figure 5.9 to Figure 5.11. All designs meet the SE requirement and the quantizations in DSP, SerDes, DAC, and PS are optimized. We observe that the system power consumption is a concave function of array size except few exceptions that will be discussed in later paragraphs. The concavity

Table 5.3: Summary of circuits blocks in array architectures

Circuits Block	DSP	SerDes[h]	DAC	LO/Mixer	PS	Splitter/Combiner[b]	RF Amp. (IL Comp.)	RF Amp. (Pre-Driver)	PA
DC Power per Block	(5.12)	(5.13)	(5.14)	60+10mW	10mW	-	40mW	40mW	(5.15)
IC Area per Block (mm^2)	Varies[i]	1.21[c]	0.05[d]	0.18[e]	0.05[f]	0.04[g]	0.025[f]	0.025[f]	-
Blocks in DA[a]	1	N_{DA}/K_{DA}	N_{DA}		-	N_{DA}	-	N_{DA}	N_{DA}
Blocks in SA[a]	1	N_{SA}/K_{SA}	N_{SA}/K_{SA}		N_{SA}	$N_{SA} - N_{SA}/K_{SA}$	$N_{SA}/4$	N_{SA}	N_{SA}
Blocks in FH	1	-	U		UN_{FH}	$2N_{FH}U - N_{FH} - U$	$2UN_{FH}/3$	N_{FH}	N_{FH}
Blocks per DA Antenna	N/A	$1/N_{DA}$	1		-	1	-	1	1
Blocks per SA[a] Antenna	N/A	$1/K_{SA}$	$1/K_{SA}$		1	$1 - 1/K_{SA}$	1/4	1	1
Blocks per FH Antenna	N/A	-	U/N_{FH}		U	$2U - 1 - U/N_{FH}$	$2U/3$	1	1

a. We do not focus on varying the number of elements in module. $K_{DA} = 8$ and $K_{SA} = 16$ are treated as constants.

b. It refers to a 1:2 or 2:1 Wilkinson splitting or combining unit.

c. We use $0.89mm^2$ [47] and $0.32mm^2$ [60] for SerDes receiver and transmitter respectively. They are fabricated in 28 nm and 16 nm CMOS.

d. Specification is from [121] and the DAC has 8 bits precision and uses 28 nm fabrication.

e. Specification of 28 GHz LO and mixer are from [55] and 65 nm CMOS fabrication is used.

f. Specification is estimated from figure in [199]. 0.18μm BiCMOS is used for fabrication.

g. Specification is estimated from figure in [199] and scaled by wave-length due to its direct impact in Wilkison divider.

h. Assuming SerDes module is used for each module define in Section. 5.4.1.

i. $A_{Precoding} = \frac{(6UM+72M)\times BW}{FOM_{DSP,area}}$ with $FOM_{DSP,area} = 500GOPS/mm^2$, 10 times scaling from [193] due to potential advanced CMOS process.

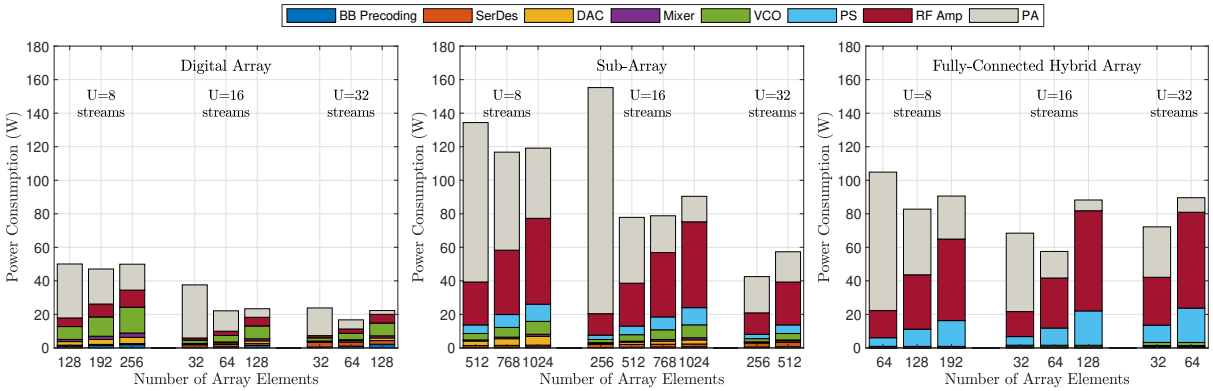


Figure 5.9: Total power consumption for three architectures operating in the Dense Urban enhanced mobile broadband use case. For each array architecture with varying array size, other design parameters are chosen according the analysis in Section 5.3 and 58.8bps/Hz spectral efficiency target demands are guaranteed in the corresponding line-of-sight environment listed in Table 5.2.

comes from the trade-off between PA power and processing power in other circuits blocks for different antenna array sizes. In the figures, the range of antenna element number N for all scenarios is chosen to be close to green point, one that minimizes system power consumption.

Taking a closer look at Dense Urban eMBB use case in Figure 5.9, we have the following conclusions. Firstly, DA and FH have similar green point of array size when the same number of streams U is used, while green point of SA is much larger. This is due to the inefficiency of array gain (5.5) when SA splits antenna with sub-groups. The exception occurs in SA with $U = 32$ streams. When SA uses small antenna number and high multiplexing level, it effectively becomes a digital array. In fact, the green point for SA with $U = 32$ streams occurs at $N = 32$. It requires RF-chain to be connected to one antenna which makes SA a fully digital array. In the rest of comparison discussion, we focus on regime where each RF-chain is connected to $K_{\text{SA}} = 8$ antennas and do not further consider regime for $N < 256$ with $U = 32$ streams. Secondly, increasing U reduces system power consumption in DA and SA. With the fixed N, increasing U reduces required transmit power and thus saves DC power of PA. Besides, increasing U does not require additional hardware resources except baseband

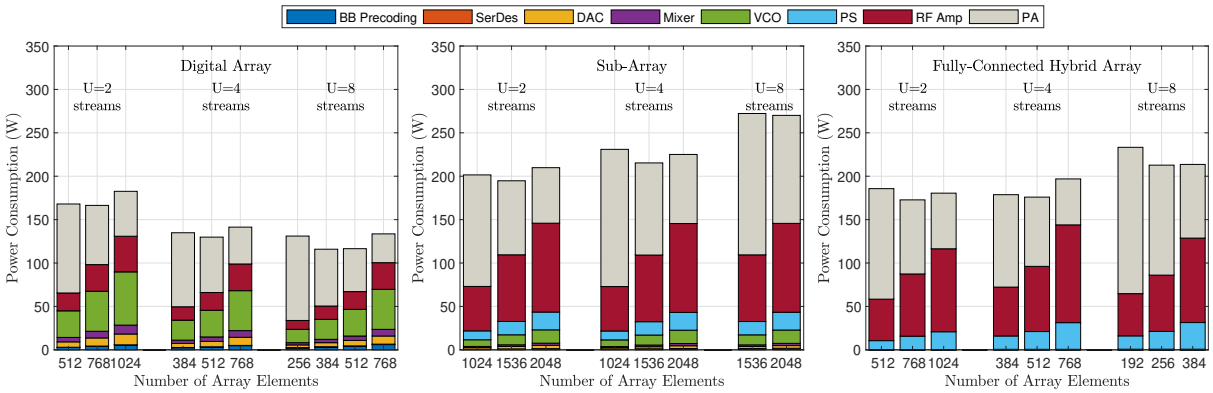


Figure 5.10: Total power consumption for three architectures operating in the 50+Mbps Everywhere use case. For each array architecture with varying array size, other design parameters are chosen according the analysis in Section 5.3 and 4.7bps/Hz spectral efficiency target demands are guaranteed in the corresponding non-line-of-sight environment listed in Table 5.2.

precoding and SerDes throughput. With the benefits of quantization requirement reduction from Figure 5.4 and high DSP efficiency, the negative impact of additional hardware resources is marginal. Thirdly, the transmit power and power consumption of PA reduces when FH uses higher U, but the system does not necessarily benefits. Part of the reason is that power in other circuits blocks linearly scales with stream number and they become system bottleneck in high-U regime. Another important fact is that a power efficient design tends to reduce N to save processing power when U is increased. It implies FH needs to deal with higher interference from the increased beam-width. In fact, FH with $N = 16$ antennas cannot meet SE requirement when using $U = 32$ beams. At last, comparing with the best designs of all architectures, we conclude that DA is the most power efficient architecture. The best design of SA becomes DA and the best design of FH still requires 240% more power than DA.

The system power consumption in 50+Mbps Everywhere is shown in Figure 5.10. We have the following findings. Firstly, the benefits of using higher multiplexing are not as prominent as in eMBB case. According to Figure 5.3 and corresponding analysis, it is mainly

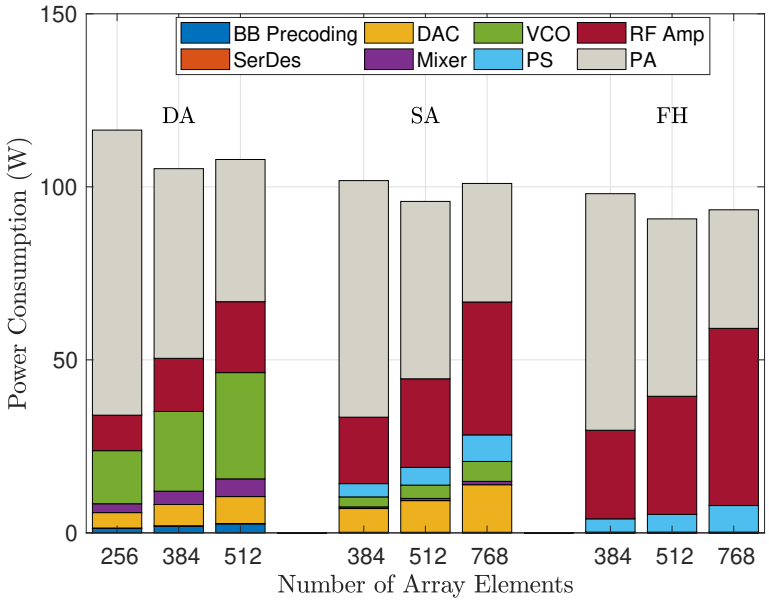


Figure 5.11: Total power consumption for three architectures operating in the Self-backhauling use case. For each array architecture with varying array size, other design parameters are chosen according the analysis in Section 5.3 and 11.8bps/Hz spectral efficiency target demands are guaranteed in the corresponding line-of-sight environment listed in Table 5.2.

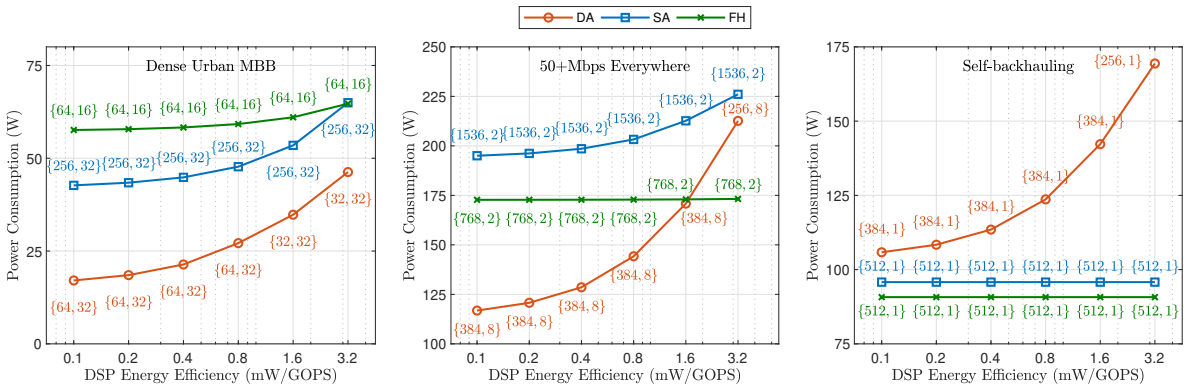


Figure 5.12: System power consumption with different digital signal processing energy efficiency in the unit of mW/GOPS. In all cases, optimal design parameters that reach spectral efficiency target with lowest power consumption are chosen. The optimal antenna number N of multiplexing level U are labeled insider brackets $\{N, U\}$ which are adjacent to corresponding data markers of system power consumption.

caused by smaller target SINR relaxation by reducing U. In fact, SA requires to use higher transmit power and thus DC power of PA. Secondly, large array size N is required for power efficient system. Overall, system requires more hardware and power consumption than in Dense Urban eMBB and it implies the intrinsic disadvantage of mmW to provide ubiquitous connection even in small cell size. At last, DA remains the most efficient architecture while the best design of hybrid architecture requires nearly 50% more power. This is a surprising result. One may expects that hybrid architecture outperforms DA when system is optimized for beamforming rather than multiplexing in this NLOS environment. With $U = 2$, we do observe comparable power consumption. However, DA further reduces its power by levering on increasing U with negligible additional processing power consumption. Hybrid architectures either require higher transmit power, e.g., SA, or excessive processing power, e.g., FH, to increase U.

The only use case in our survey that hybrid architectures outperform DA is Self-backhauling where multiplexing level is limited due to point-to-point communication environment of LOS channel. In Figure 5.10, the DA requires 18% more power as compared to hybrid architec-

tures. This small power margin is due to the fact that the DA requires nearly 4 bits smaller quantization than hybrid architectures according to Figure 5.4 and it prevents excessive power consumption in BB precoding, SerDes and DAC. Overall in this use case, the SA and FH have similar power consumption. In fact, SA and FH have the same the number of phase shifters when using same number of antenna elements. The difference between them lies in the power consumption of signal routing. The SA has more RF-chains than FH and therefore SA requires more power in high precision DAC and VCOs. The FH has only one RF-chain but it requires more power for RF signal distribution than SA.

In Figure 5.9 to 5.11, DAC and BB precoding power has small proportion in the DA system, even when high multiplexing or large array size is used. Part of the reason is the ENOB requirement relaxation according to Section 5.3. A more important factor is the DSP energy efficiency. Our study is based on the assumption that baseband processing is implemented on application-specific integrated circuits (ASIC). In deploying mmW DA, programmable DSP or Field-Programmable Gate Array (FPGA) based BB processor provide flexibility of reconfiguring BB precoding scheme, with the cost of order-of-magnitude more power consumption [193]. In Figure 5.12, the system power of all architectures are compared when different DSP efficiencies are used. Throughout all cases, all design parameters are optimized such that lowest power consumed in reaches SE target, and the required array size N and multiplexing level U is labeled in the figure. We have the following findings. Firstly, DA is most sensitive to the decreased DSP efficiency. An efficient design would use smaller array size when BB precoding becomes bottleneck since it effectively reduces DSP burden. SA is less sensitive due to a much smaller number of RF-chains except in Dense Urban eMBB where SA effective behaves as a digital array. FH is least sensitive to DSP efficiency. Secondly, with 3.2mW/GOPS, a FOM that can be reached by reconfigurable digital processor using 90 to 130 nm process [193], DA remains the best architecture in Dense Urban eMBB. In the rest use cases, DA becomes less competitive in terms of power consumption.

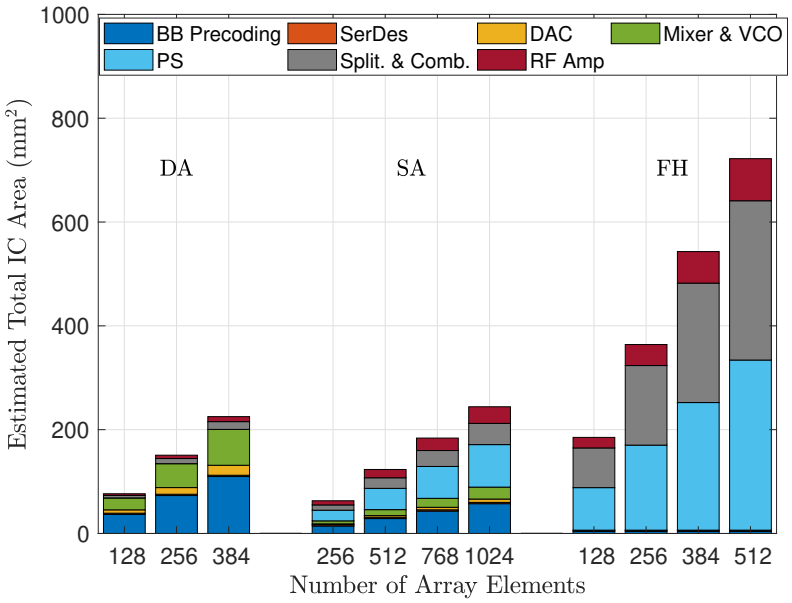


Figure 5.13: Integrated circuits area breakdown of three architectures.

5.6.2 Intergrated circuits areas of millimeter-wave array architectures

In Figure 5.13, the required IC area is presented as a function of array size. Note that increasing the multiplexing capability forces DA to have more powerful and larger DSP, and it also forces FH to have more RF-chain and complicated distribution network. Since maximum multiplexing of $U = 16$ does not significantly affect the optimal design for power consumption, we use $U = 16$ for DA and FH while $U = 32$ for SA. As shown in the figure, the largest contributor in DA area is the DSP, which is expected to be further reduced so long as Moore Law reduces silicon area. SA remains competitive in IC area with DA. However, the cost of PAs, which is likely to be fabricated with other material, is likely to require additional cost for SA due to the requirement of larger antenna number for power efficiency. FH requires the largest IC area due to the full connection nature between RF-chains and large number of antenna elements.

5.7 Discussions on open research challenges

Admittedly, the power and IC area analysis for three array architectures provided are preliminary estimates based on the surveyed literature. In particular, the effect of the extra digital processing on power consumption and area depends on actual design and is hard to analyze at this point. Besides, some open research questions remain and were not covered in this chapter. First one is the issues of synchronization among large number of array elements. In the centralized LO distribution architecture, each element re-generates clock from the same references but global LO distribution may not be area and energy efficient [86]. Under distributed LO scheme, independent LOs help reduce impact of phase noise [132] but system needs to be calibrated periodically to avoid loss of coherency across elements. Second issue is related to compensation of PA nonlinearity. Digital predistortion (DPD) is important in massive transmitter array design. Conventionally, DPD is designed for DA where DSP is implemented for each pair of transmitter chain and PA. Due to increased processing and power of DPD, the overall gains in power efficiency for large number antenna arrays need to be analyzed and optimized. DPD for SA [13, 106] and FH [182] are actively investigated by researchers. Thirdly, other design variations, including phase-and-magnitude analog precoder and active RF splitter and combiner [192] can help reduce the complexity and power consumption of the hybrid arrays. Lastly, our survey reveals the benefits of using high multiplexing level for power saving in the hardware. However, high multiplexing brings additional burden in higher layers of system, e.g., network layer faces more challenges to schedule users with non-overlapping propagation paths, and their impact needs to be incorporated in more comprehensive study.

In this work, we reveal that the conventional belief that hybrid array architecture is more cost and energy efficient than digital architecture is not necessarily true when comprehensive hardware block is modeled and system adopts optimized design parameters. Similar findings were reported for the receiver array during the period when this work is written [12, 146]. It is worth noting that these works, including ours, focus on the additive uniformly distributed quantization error model and linear MIMO processing model. However, such quantization

error model becomes less precise when data samples and quantization error are correlated, which occurs when data converters have significantly small number of bits. Besides, linear MIMO processing is not optimal. In fact, in the receiver array a variety of nonlinear combining and decoding algorithms are proposed, e.g., successive interference cancellation based combining [62], approximate message passing [174]. Besides, the precision requirement of DAC and analog-to-digital converter (ADC) devices are strongly dependent on processing algorithms, e.g., algorithm tailored for 1-bit ADC [115]. It remains open research question how to use advanced signal processing to further reduce power consumption and cost of mmW array.

The Matlab code for simulation and data for system level power comparison is released in [177] for readers that are interested in results with different design choices and hardware specifications.

5.8 Conclusion

Implementation of energy and cost efficient massive antenna array transmitters is one of the major challenges in deploying mmW networks in the 5G era. In this work, we study and compare three array architecture candidates, digital architecture and two variation of analog-digital hybrid architectures, and discuss various hardware design trade-offs in their implementation. The analysis and comparison are based on the modeling of required power and IC area of circuits blocks, derived from the state-of-the-art mmW circuits design and measurement results. We compare three array architectures when their design parameters are optimized to meet the spectral efficiency targets in three representative 5G-NR use cases. The results show that digital architecture is the most efficient in terms of power and IC area. The key intuition behind this finding is that digital array architecture can effectively save system power and area by levering high multiplexing gains due to digital precoding of multiple spatial streams, which effectively reduces requirements for array size, transmit power, and hardware specifications of the RF-chains. On the other hand, the hybrid architectures require additional power to support more simultaneous spatial beams.

We reveal that the bottleneck of hybrid architectures is the RF signal distribution network in RF beamforming stage. Furthermore, additional transmit power is required in sub-array architecture to compensate for the array splitting loss.

CHAPTER 6

Conclusion

6.1 Summary of contributions

In this dissertation, we focus on a key physical layer procedure of the millimeter-wave cellular system: initial access and beam alignment. We have designed and analyzed novel signal processing algorithm to enable fast and scalable initial access and beam alignment, and implemented one approach in the mmW testbed. Further, we have studied the performance, power, and cost trade-off in millimeter-wave array architecture for beamforming and MIMO multiplexing.

In Chapter 2, we focus on the initial cell discovery, synchronization, and beam alignment problem. We propose to use the novel quasi-omni pseudorandom sounding beam during this procedure to accelerate the initial access. We have designed signal processing algorithm based on the proposed sounding beam structure that is compatible with 5G-NR frame format. The cell detection and integer timing synchronization algorithm have been designed and thoroughly analyzed. The results showed that the pseudorandom sounding beam only marginally sacrifice detection sensitivity. Further, compressive sensing based signal processing algorithm has been designed for initial beam alignment by reusing the received pilots. The proposed approach precisely estimates the angular channel information of the sparse millimeter-wave channel, and is robust to frequency synchronization error during this procedure. The approach also achieves CRLB in high signal-to-noise radio regime. We have evaluated the approaches via simulations using the 5G-compliant frame structure practical synchronization offset in cellular system, and measurement endorsed urban channel models at 28 GHz. The results show that the proposed approach provides comparable performance

to the state-of-the-art directional cell search for initial discovery, but achieves significantly more accurate angle estimation during initial beam training. This advantage holds true across different propagation condition (LOS/NLOS) and UE-BS distance. Due to the saving of additional radio resource, i.e., CSI-RS, for beam refinement, the proposed approach reduces up to two order of magnitude access latency compared to the directional initial access when the same signaling overhead and post-training beam steering SNR are targeted.

When the bandwidth and array size increase as millimeter-wave cellular system evolves, the array response of different frequency components cannot be treated as the same. Such phenomenon is referred as beam squint. An enhanced beam training algorithm that is robust to beam squint. The proposed squint robust CS based beam training algorithm is able to retain non-compromised AoD estimation accuracy and beamforming gain across a wide range of beam training bandwidths.

In Chapter 3, we continue investigating the compressive beam alignment in mmW system using pseudorandom sounding beams. We have identified that practical mmW radios commonly suffers from hardware impairments in the form of array phase offset. Such imperfection can greatly degrades the performance of compressive beam alignment algorithm as precise knowledge of sensing dictionary is required. Further, radios may only provide signal strength of the pilots instead of providing complex sample outputs. We have designed a novel noncoherent compressive beam alignment scheme that utilizes machine learning to address implementation challenges. Further, we have implemented the algorithm in a 60 GHz as proof-of-concept. The experiments in a suburban line-of-sight environment showed that an order of magnitude overhead savings with marginal post-alignment beamforming gain loss is achievable, as compared to exhaustive beam sweeps. The proposed approach also outperforms purely model-based compressive beam alignment in the experiment.

In Chapter 4, we design a fast beam training approach based on the emerging true-time-delay based analog millimeter-wave array. We have theoretically proved that by designing proper delay taps, the array can steer different subcarrier of an OFDM waveform into different direction. Such scheme is referred as rainbow beam. Exploiting the frequency de-

pendence nature of rainbow beam, the system can scan multiple angles simultaneously, and thus greatly accelerate the beam training procedure. We derive the condition for delay tap spacing and the required number of subcarriers. Based on these novel sounding beams, a single-shot, low-complex, and noncoherent beam training algorithm is developed. The simulation results in realistic urban channel model at 28GHz reveal that the proposed method is capable to use a single training symbol to complete beam training in both LOS and NLOS environment. Such feature is appealing in the future mmW systems as the required beam alignment overhead and latency does not further increases with the increased array size.

In Chapter 5, we study the Implementation of energy and cost efficient massive antenna array transmitters is one of the major challenges in deploying mmW networks in the 5G era. In this chapter, we study and compare three array architecture candidates, digital architecture and two variation of analog-digital hybrid architectures, and discuss various hardware design trade-offs in their implementation. The analysis and comparison are based on the modeling of required power and IC area of circuits blocks, derived from the state-of-the-art mmW circuits design and measurement results. We compare three array architectures when their design parameters are optimized to meet the spectral efficiency targets in three representative 5G-NR use cases. The results show that digital architecture is the most efficient in terms of power and IC area. The key intuition behind this finding is that digital array architecture can effectively save system power and area by levering high multiplexing gains due to digital precoding of multiple spatial streams, which effectively reduces requirements for array size, transmit power, and hardware specifications of the RF-chains. On the other hand, the hybrid architectures require additional power to support more simultaneous spatial beams. We reveal that the bottleneck of hybrid architectures is the RF signal distribution network in RF beamforming stage. Furthermore, additional transmit power is required in sub-array architecture to compensate for the array splitting loss.

6.2 Future works

There are many open research problem in the topics that this dissertation covers.

In Chapter 2 and 3, our design intends to identify the best beam steering directions and our evaluation occurs in a LOS channel. The approach can be further generalized to identify one or more paths in mmW NLOS channels. Another feature of the proposed approach is that only 2 bits phase shifter resolution is required. It is of interest to design a system where both beam alignment and beam steering in data communication use 2 bits phase shifter. As such, the RF phase shifter design and associated insertion loss can be greatly reduced, which potentially provides additional power saving. The practical system design and proof-of-concept is left as future work. Lastly, the machine learning assisted beam alignment that utilize prior knowledge of beam steering direction in dynamic environment, i.e., beam tracking [37, 179, 184], is a promising direction.

In Chapter 4, we designed receiver beam alignment approach in a 2D environment. Extending this approach in 3D environment, i.e., by joint design delay taps of antenna element in the azimuth and elevation plane is an interesting direction. Further, joint transmitter and receiver beam alignment design using TTD array is also of interest. Further, this work did not discuss hardware aspects of TTD array. It is importance to understand the impact and required specification of hardware in RF and mixed signal domain for power efficient operation of TTD arrays [38].

In Chapter 5, we focus on the transmitter side. Similar methodology can be applied to receiver side of the base station. The advanced signal processing technique can be applied to further enhance the performance of fully-digital array with low resolution DACs/DACs. Further, the impact of hardware nonlinearity [185] and power saving scheme by reducing phase shifter quantization can be investigated.

APPENDIX A

Appendix for Chapter 2

A.1 Initial discovery performance

The noise after correlation

$$\tilde{z}[n] = \frac{1}{P}\sum_{k=0}^{P-1}(\mathbf{w}^{\mathrm{H}}[n+k]\mathbf{z}[n+k])s_{\mathrm{zc}}^{*}[k]$$

follows distribution $\mathcal{NC}(0, \sigma_{\mathrm{n}}^2/P)$. Thus $|\tilde{z}[n]|^2$ is Chi-Square distributed with degree-of-freedom 2, mean σ_{n}^2/P, and variance $\sigma_{\mathrm{n}}^4/P^2$. We denote detection statistic in PT and NT scenario under $\mathcal{H}_0$ and $\mathcal{H}_1$ as denoted as $\gamma_{\mathrm{PT},0}$, $\gamma_{\mathrm{PT},1}$, $\gamma_{\mathrm{NT},0}$, $\gamma_{\mathrm{NT},1}$, respectively, and find their distribution.

$\gamma_{\mathrm{PT},0}$ is the sum of squared $N_{\mathrm{c}}M$ realizations of $\tilde{z}[n]$ divided by M, thus central limit theory (CLT) applies. The distribution of $\gamma_{\mathrm{PT},0}$ is $\mathcal{N}(\mu_{\mathrm{PT},0}, \sigma_{\mathrm{PT},0})$, where $\mu_{\mathrm{PT},0} = N_{\mathrm{c}}\sigma_{\mathrm{n}}^2/P$ and $\sigma_{\mathrm{PT},0} = \sqrt{N_{\mathrm{c}}\sigma_{\mathrm{n}}^4/(P^2M)}$, respectively. As a result, the optimal detection threshold that reaches target false alarm rate $P_{\mathrm{FA}}^{\star}$ is given by (2.7). Similarly, the detection statistic under $\mathcal{H}_0$ with TO is denoted as $\gamma_{\mathrm{NT},0}$. It is the maximum operation with degrees of freedom $\epsilon_{\mathrm{T,max}}$ of $\gamma_{\mathrm{PT},0}$. With large $\epsilon_{\mathrm{T,max}}$, $\gamma_{\mathrm{NT},0}$ follows extreme value distribution, *Gumbel Distribution*, where the mean and standard deviation are $\mu_{\mathrm{NT},0} = \mu_{\mathrm{PT},0} + \sigma_{\mathrm{PT},0}\mathrm{Q}^{-1}(1/\epsilon_{\mathrm{T,max}})$ and and $\sigma_{\mathrm{NT},0} = \sigma_{\mathrm{PT},0}/\mathrm{Q}^{-1}(1/\epsilon_{\mathrm{T,max}})$, respectively. Using its inverse cumulative distribution function, the optimal detection threshold is $\eta_{\mathrm{NT}}^{\star} = \mu_{\mathrm{NT},0} - (\sqrt{6}\pi)\sigma_{\mathrm{NT},0}\ln(-\ln(1-P_{\mathrm{FA}}^{\star}))$. It gives (2.7) using expressions of $\mu_{\mathrm{NT},0}$, $\sigma_{\mathrm{NT},0}$ and $\sqrt{6}/\pi \approx 0.78$.

Detection statistic $\gamma_{\mathrm{PT},1}$ is the sum of noise energy and signal energy, i.e.,

$$\gamma_{\mathrm{PT},1} = \gamma_{\mathrm{PT},0} + (\sum_{m=1}^{M}\sum_{l=0}^{L}|\tilde{g}_{m,l}\sum_{n=1}^{P}|s_{\mathrm{zc}}[n]|^2 e^{j\epsilon_{\mathrm{F}}n}|^2)/(PMN_{\mathrm{T}}N_{\mathrm{R}}),$$

where $\tilde{g}_{m,l}$ is defined in Section 2.5.1. Using the fact $|s_{\rm zc}[n]| = 1$, definition $\kappa(0,\epsilon_{\rm F}) \triangleq |\sum_{n=1}^{P} e^{j\epsilon_{\rm F} n}|^2$ in (2.10), and approximation that different multipaths are resolvable, i.e., $p_{\rm c}(dT_{\rm s} - \tau_l) = 1, d \in \mathcal{S}_{\rm d}$ where $\mathcal{S}_{\rm d}$ has L integers in range $[0, N_{\rm c}-1]$, the above equation becomes $\gamma_{\rm PT,1} = \kappa(0,\epsilon_{\rm F}) \sum_{m=1}^{M} \zeta_m/M + \gamma_{\rm PT,0}$ where $\zeta_m = \sum_{l=0}^{L} |g_l \mathbf{w}_m^{\rm H} \mathbf{a}_{\rm R}(\phi_l) \mathbf{a}_{\rm T}^{\rm H}(\theta_l) \mathbf{v}_m|^2/(N_{\rm T}N_{\rm R})$. Using the fact that ζ_m are mutually independent due to independent $\mathbf{v}_m$ and $\mathbf{w}_m$, the mean and variance of ζ_m are

$$\mathbb{E}(\zeta_m) = \sum_{l=1}^{L} |g_l|^2 \, \mathbb{E}\left|\mathbf{w}_m^{\rm H} \mathbf{a}_{\rm R}(\phi_l)\right|^2 \mathbb{E}\left|\mathbf{a}_{\rm T}^{\rm H}(\theta_l)\mathbf{v}_m\right|^2/(N_{\rm T}N_{\rm R}) = \sigma_{\rm g}^2$$

and

$$\begin{aligned}\mathrm{var}(\zeta_m) =& (N_{\rm T}N_{\rm R})^{-2} \sum_{l=1}^{L} |g_l|^4 \, \mathbb{E}\left|\mathbf{w}_m^{\rm H} \mathbf{a}_{\rm R}(\phi_l)\right|^4 \mathbb{E}\left|\mathbf{a}_{\rm T}^{\rm H}(\theta_l)\mathbf{v}_m\right|^4 - \sigma_{\rm g}^4 \\ =& \sigma_{\rm g}^4 \left(2 - \frac{1}{N_{\rm T}}\right)\left(2 - \frac{1}{N_{\rm R}}\right) - \sigma_{\rm g}^4 \approx 3\sigma_{\rm g}^4,\end{aligned}$$

respectively. The above approximation holds true with typical antenna array sizes $N_{\rm R}$ and $N_{\rm T}$ in mmW. Therefore, according to CLT $\gamma_{\rm PT,1} \sim \mathcal{CN}(\kappa(0,\epsilon_{\rm F})\sigma_{\rm g}^2 + \mu_{\rm PT,0}, 3\kappa^2(0,\epsilon_{\rm F})\sigma_{\rm g}^4/M + \sigma_{\rm PT,0}^2)$, which gives the miss detection probability $P_{\rm MD,PT} = \mathrm{Q}[(\mathbb{E}(\gamma_{\rm PT,1}) - \eta_{\rm PT}^{\star})/\sqrt{\mathrm{var}(\gamma_{\rm PT,1})}]$, and it equals to (2.9).

In NT scenario, we make the following approximations: 1) the detection statistic $\gamma_{\rm NT,1}$ corresponds to the correlation peaks for the correct timing $\epsilon_{\rm T}$; 2) the abrupt beamformer changes during m-th PSS reception, when present, result in an independent realization of sounding beam $\tilde{\mathbf{w}}_m$. Although the former is not valid with low SNR, the MD rate with typical threshold in such SNR regime already approaches 1. Therefore, impact of such loose approximation is negligible. Based on these assumptions, we evaluate distribution of $\gamma_{\rm NT,1}$ as $\gamma_{\rm NT,1} = \gamma_{\rm PT,0} + \frac{1}{PMLN_{\rm T}N_{\rm R}}(\sum_{m=1}^{M}\sum_{l=0}^{L} |\tilde{g}_{m,l}^{(1)} \sum_{n_1=1}^{K-1} |s_{\rm zc}[n_1]|^2 e^{j\epsilon_{\rm F} n_1} + \tilde{g}_{m,l}^{(2)} \sum_{n_2=K}^{P} |s_{\rm zc}[n_2]|^2 e^{j\epsilon_{\rm F} n_2}|^2)$ where $\tilde{g}_{m,l}^{(1)} = g_l \mathbf{w}_m^{\rm H} \mathbf{a}_{\rm R}(\phi_l)\mathbf{a}_{\rm T}^{\rm H}(\theta_l)\mathbf{v}_m$ and $\tilde{g}_{m,l}^{(2)} = g_l \tilde{\mathbf{w}}_m^{\rm H} \mathbf{a}_{\rm R}(\phi_l)\mathbf{a}_{\rm T}^{\rm H}(\theta_l)\mathbf{v}_m$ are the post-BF channel gain due to partially overlapped burst window in BS and UE. In other words, K follows (2.11) and $n_1 \in [1, K-1]$ and $n_2 \in [K, P]$ are the sample window where K represents the abrupt change in BF. The independent $\mathbf{w}_m$ and $\tilde{\mathbf{w}}_m$ lead to uncorrelated $\tilde{g}_{m,l}^{(1)}$ and $\tilde{g}_{m,l}^{(2)}$. For notational convenience of finding statistic of $\gamma_{\rm NT,1}$, we define $\zeta_{m,l}$ as $\zeta_{m,l} \triangleq (|\tilde{g}_{m,l}^{(1)} \frac{1-e^{jK\epsilon_{\rm F}}}{1-e^{j\epsilon_{\rm F}}} +$

$\tilde{g}_{m,l}^{(2)}\frac{1-e^{j(P-K)\epsilon_{\mathrm{F}}}}{1-e^{j\epsilon_{\mathrm{F}}}}|^2)/(N_{\mathrm{T}}N_{\mathrm{R}})$ in $\gamma_{\mathrm{NT},1}$ after simplification with the fact $|s_{\mathrm{zc}}[n]|^2=1, \forall n\in\mathcal{S}$ as well as $\sum_{n=1}^{K}e^{j\epsilon_{\mathrm{T}}n}=(1-e^{jK\epsilon_{\mathrm{F}}})/(1-e^{j\epsilon_{\mathrm{F}}})$. The mean and variance of $\zeta_{m,l}$ are $\mathbb{E}\left(\zeta_{m,l}\right)=\kappa(\epsilon_{\mathrm{F}},\epsilon_{\mathrm{T}})\sigma_{\mathrm{g}}^2$, and $\mathrm{var}\left(\zeta_{m,l}\right)\approx 3\sigma_{\mathrm{g}}^4\zeta^2(\epsilon_{\mathrm{F}},\epsilon_{\mathrm{T}})$ after plugging in definition of $\kappa(\epsilon_{\mathrm{F}},\epsilon_{\mathrm{T}})$ from (2.10). Using CLT and statistic of $\zeta_{m,l}$, $\gamma_{\mathrm{NT},1}\sim\mathcal{CN}(\mu_{\mathrm{PT},0}+\kappa(\epsilon_{\mathrm{F}},\epsilon_{\mathrm{T}})\sigma_{\mathrm{g}}^2,\sigma_{\mathrm{PT},0}^2+3\sigma_{\mathrm{g}}^4\kappa^2(\epsilon_{\mathrm{F}},\epsilon_{\mathrm{T}})/M$. The MD rate $P_{\mathrm{MD,NT}}=\mathrm{Q}[(\mathbb{E}(\gamma_{\mathrm{NT},1})-\eta_{\mathrm{NT}}^{\star})/\sqrt{\mathrm{var}(\gamma_{\mathrm{NT},1})}]$ reduces to (2.9).

A.2 Cramer-Rao lower bound of joint estimation problem

The FIM has the following form

$$\mathbf{J}=\frac{1}{\sigma_{\mathrm{n}}^2}\begin{bmatrix}\Phi_{\epsilon_{\mathrm{F}},\epsilon_{\mathrm{F}}} & \Phi_{\epsilon_{\mathrm{F}},\theta} & \Phi_{\epsilon_{\mathrm{F}},\phi} & \Phi_{\epsilon_{\mathrm{F}},\tau} & \Phi_{\epsilon_{\mathrm{F}},\alpha} & \Phi_{\epsilon_{\mathrm{F}},\beta}\\ \Phi_{\theta,\epsilon_{\mathrm{F}}} & \Phi_{\theta,\theta} & \Phi_{\theta,\phi} & \Phi_{\theta,\tau} & \Phi_{\theta,\alpha} & \Phi_{\theta,\beta}\\ \Phi_{\phi,\epsilon_{\mathrm{F}}} & \Phi_{\phi,\theta} & \Phi_{\phi,\phi} & \Phi_{\phi,\tau} & \Phi_{\phi,\alpha} & \Phi_{\phi,\beta}\\ \Phi_{\tau,\epsilon_{\mathrm{F}}} & \Phi_{\tau,\theta} & \Phi_{\tau,\phi} & \Phi_{\tau,\tau} & \Phi_{\tau,\alpha} & \Phi_{\tau,\beta}\\ \Phi_{\alpha,\epsilon_{\mathrm{F}}} & \Phi_{\alpha,\theta} & \Phi_{\alpha,\phi} & \Phi_{\alpha,\tau} & \Phi_{\alpha,\alpha} & 0\\ \Phi_{\beta,\epsilon_{\mathrm{F}}} & \Phi_{\beta,\theta} & \Phi_{\beta,\phi} & \Phi_{\beta,\tau} & 0 & \Phi_{\beta,\beta}\end{bmatrix}$$

where $\Phi_{x,x}$ denotes for $\Phi_{x,x}=\partial^2 L(\mathbf{y};\boldsymbol{\xi})/\partial x\partial y=(\partial L(\mathbf{x}(\boldsymbol{\xi})/\partial x)^{\mathrm{H}}(\partial L(\mathbf{x}(\boldsymbol{\xi}))/\partial y)$. The exact expressions of each elements in FIM are summarized in Table A.1, where for notational convenience the following matrices are defined. The derivative over CFO matrix is a diagonal matrix whose p-th diagonal element is $[\dot{\mathbf{Q}}_m]_{p,p}=j[(m-1)N_{\mathrm{B}}+(p-1)]e^{j\epsilon_{\mathrm{F}}[(m-1)N_{\mathrm{B}}+(p-1)]}$. The vector $\dot{\mathbf{f}}=\partial\mathbf{f}(\tau)/\partial\tau$ whose p-th element is $[\dot{\mathbf{f}}]_p=j2\pi(p-1)T_{\mathrm{s}}e^{j2\pi(p-1)\epsilon_{\mathrm{F}}T_{\mathrm{s}}}$ Other expression in Table A.1 include $\mathbf{f}^{\mathrm{H}}(\tau)\mathbf{F}^{\mathrm{H}}\mathbf{Q}_m^{\mathrm{H}}\mathbf{Q}_m\mathbf{F}\mathbf{f}(\tau)=P,\forall m$, $C_{\mathrm{df}}=\sum_{p=0}^{P-1}2\pi pT_{\mathrm{s}}=(P-2)(P-1)\pi T_{\mathrm{s}}$, $C_{\mathrm{dq},m}\triangleq\mathbf{f}^{\mathrm{H}}(\tau)\mathbf{F}^{\mathrm{H}}\dot{\mathbf{Q}}_m^{\mathrm{H}}\mathbf{Q}_m\mathbf{F}\mathbf{f}(\tau)=(m-1)T_{\mathrm{B}}+\frac{(P-2)(P-1)T_{\mathrm{s}}}{2}$, , and $C_{\mathrm{d2q},m}=\sum_{p=0}^{P-1}\left[(m-1)T_{\mathrm{B}}+pT_{\mathrm{s}}\right]^2$.

Table A.1: Elements of fisher information matrix

Symb.	Expressions
$\Phi_{\epsilon_{\mathrm{F}},\epsilon_{\mathrm{F}}}$	$\sum_{m=1}^{M}(C_{\mathrm{d2q}}g)\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2$
$\Phi_{\epsilon_{\mathrm{F}},\theta}$	$\sum_{m=1}^{M}(C_m g)\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\Re\left\{\left[\mathbf{v}_m^{\mathrm{H}}\dot{\mathbf{a}_{\mathrm{T}}}(\theta)\right]\left[\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right]\right\}$
$\Phi_{\epsilon_{\mathrm{F}},\tau}$	$\Re\left\{\sum_{m=1}^{M}g\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2\mathbf{f}^{\mathrm{H}}(\tau)\mathbf{F}\dot{\mathbf{Q}}_m^{\mathrm{H}}\mathbf{Q}_m\mathbf{F}^{\mathrm{H}}\dot{\mathbf{f}}(\tau)\right\}$
$\Phi_{\epsilon_{\mathrm{F}},\alpha}$	$\sum_{m=1}^{M}\left[C_{\mathrm{dq},m}\Re(g)\right]\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2$
$\Phi_{\epsilon_{\mathrm{F}},\beta}$	$\sum_{m=1}^{M}\left[C_{\mathrm{dq},m}\Im(g)\right]\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2$
$\Phi_{\theta,\theta}$	$\sum_{m=1}^{M}(P\lvert g\rvert^2)\left\|\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\dot{\mathbf{a}_{\mathrm{T}}}(\theta)\right\|^2$
$\Phi_{\phi,\phi}$	$\sum_{m=1}^{M}(P\lvert g\rvert^2)\left\|\mathbf{w}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{R}}(\phi)\right\|^2\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2$
$\Phi_{\phi,\theta}$	$\Re\left\{\sum_{m=1}^{M}P\lvert g\rvert^2[\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)][\mathbf{w}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{R}}(\phi)][\mathbf{v}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{T}}(\theta)][\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)]\right\}$
$\Phi_{\phi,\tau}$	$\sum_{m=1}^{M}C_{\mathrm{df},m}\lvert g\rvert^2\Re\left\{[\mathbf{w}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{R}}(\phi)][\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)]\right\}\left\|\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\right\|^2$
$\Phi_{\phi,\alpha}$	$\Re\left\{\sum_{m=1}^{M}Pg[\mathbf{w}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{R}}(\phi)][\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)]\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2\right\}$
$\Phi_{\theta,\alpha}$	$\Re\left\{\sum_{m=1}^{M}Pg\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2[\mathbf{v}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{T}}(\theta)][\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)]\right\}$
$\Phi_{\phi,\beta}$	$\Re\left\{\sum_{m=1}^{M}jgP[\mathbf{w}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{R}}(\phi)][\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)]\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2\right\}$
$\Phi_{\theta,\beta}$	$\Re\left\{\sum_{m=1}^{M}jPg\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2[\mathbf{v}_m^{\mathrm{H}}\dot{\mathbf{a}}_{\mathrm{T}}(\theta)][\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)]\right\}$
$\Phi_{\tau,\tau}$	$\Re\left\{\sum_{m=1}^{M}\lvert g\rvert^2\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2\left[\dot{\mathbf{f}}^{\mathrm{H}}(\tau)\mathbf{Q}_m^{\mathrm{H}}\mathbf{Q}_m\dot{\mathbf{f}}(\tau)\right]\right\}$
$\Phi_{\tau,\alpha}$	$\Re\left\{\sum_{m=1}^{M}g\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2\left[\dot{\mathbf{f}}^{\mathrm{H}}(\tau)\mathbf{Q}_m^{\mathrm{H}}\mathbf{Q}_m\mathbf{f}(\tau)\right]\right\}$
$\Phi_{\tau,\beta}$	$\Re\left\{\sum_{m=1}^{M}jg\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2\left[\dot{\mathbf{f}}^{\mathrm{H}}(\tau)\mathbf{Q}_m^{\mathrm{H}}\mathbf{Q}_m\mathbf{f}(\tau)\right]\right\}$
$\Phi_{\alpha,\alpha}$	$\sum_{m=1}^{M}P\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2$
$\Phi_{\beta,\beta}$	$-\sum_{m=1}^{M}P\lvert\mathbf{w}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{R}}(\phi)\rvert^2\lvert\mathbf{v}_m^{\mathrm{H}}\mathbf{a}_{\mathrm{T}}(\theta)\rvert^2$

APPENDIX B

Appendix for Chapter 3

B.1 Pattern and dictionary mismatch

We evaluate the pattern distortion due to array phase offset. We assume each element of $\mathbf{e}$ has unit modulus and its phase is uniform randomly distributed with boundary $|\angle[\mathbf{e}]_n| \leq \psi_{\max}$. Further, we define angular region of interest $\mathcal{B}(\theta_0, \delta) = \{\epsilon | \theta_0 - \delta \leq \epsilon \leq \theta_0 + \delta\}$, i.e., centered at θ_0 with span 2δ. Hence, the beam pattern distortions of AWV $\mathbf{w}$ in such region is

$$E(\mathbf{w}, \mathcal{B}(\theta_0, \delta), \psi_{\max}) = \frac{\int_{\mathcal{B}(\theta_0,\delta)} |G[\mathrm{diag}(\mathbf{e})\mathbf{w}, \epsilon] - G(\mathbf{w}, \epsilon)|^2 \, d\epsilon}{\int_{\mathcal{B}(\theta_0,\delta)} |G(\mathbf{w}, \epsilon)|^2 \, d\epsilon}$$

where the beam pattern of AWV $\mathbf{w}$ at angle ϵ is defined as $G(\mathbf{w}, \epsilon) = |\mathbf{a}_{\mathrm{R}}^{\mathrm{H}}(\epsilon)\mathbf{w}|$. We further define $\mathcal{B}(0, \pi/2)$ as the entire angular region of linear array. Hence, the distortion in the mainlobe and sidelobes of a DFT pencil beam towards θ_0 are $E_0 = E(\mathbf{a}_{\mathrm{R}}(\theta_0), \mathcal{B}(\theta_0, \delta), \Psi_{\max})$ and $E_1 = E(\mathbf{a}_{\mathrm{R}}(\theta_0), \mathcal{B} \backslash \mathcal{B}(\theta_0, \delta), \Psi_{\max})$, respectively. Further, the PN beam distortion across all angular region is defined as $E_2 = E(\mathbf{w}_{\mathrm{PN}}, \mathcal{B}, \Psi_{\max})$, where $\mathbf{w}_{\mathrm{PN}}$ is a random AWV. In Figure B.1, all the error metrics are numerically evaluated as function of $\psi_{\max}$ using a setup with $\theta_0 = 0$, $\delta = 0.89/N_{\mathrm{R}}$ rad, i.e., half 3dB beamwidth, $N_{\mathrm{R}} = 36$, and a single realization of random AWV. The distortion in DFT mainlobe is least sensitive to phase offset $\psi_{\max}$, which implies a $\pm 20^\circ$ offset marginally affects a system using beam sweep and steering. On the other hand, the sidelobes of DFT beam and lobes of PN beams are more vulnerable to the phase error in the array.

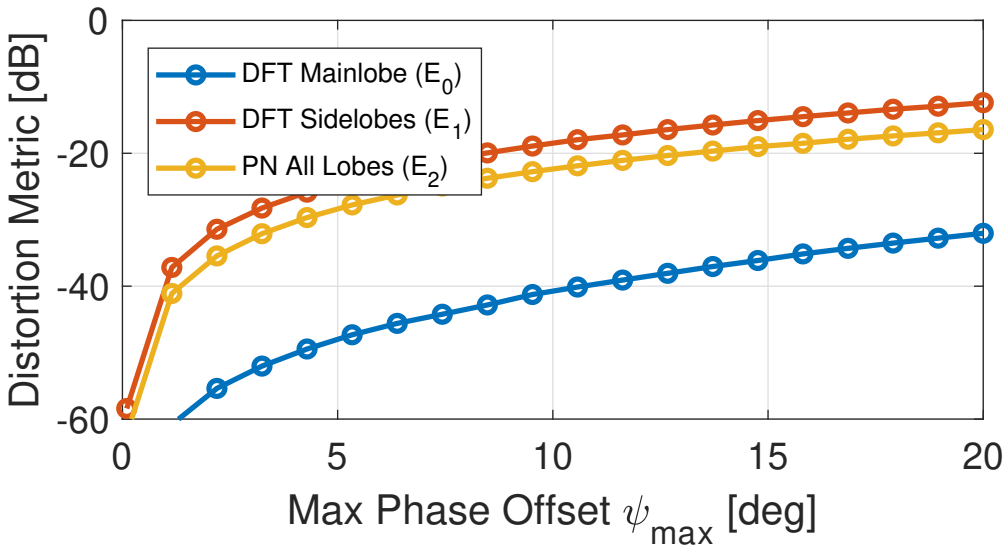


Figure B.1: Simulated beam pattern distortion metrics of different beam type as function of phase offset.

APPENDIX C

Appendix for Chapter 4

C.1 Proof of received signal mode with true-time-delay array

Consider the continuous time channel between the q-th transmitter and n-th receiver element due to the l-th MPC *before* and *after* the TTD circuits as $\tilde{h}^{(\mathrm{RF})}_{l,q,n}(t)$ and $h^{(\mathrm{RF})}_{l,q,n}(t)$, respectively. These two systems are modeled as $\tilde{h}^{(\mathrm{RF})}_{l,q,n}(t) = g_l p_c\left(t - \Gamma_{l,q,n}\right)$ and $h^{(\mathrm{RF})}_{l,q,n}(t) = \tilde{h}_{l,q,n}(t) * h_{\mathrm{TTD},n}(t) = g_l p_c(t - \tau_{l,q,n})$, where $*$ is the convolution operator. For notation clarity, we denote $\tau_{l,q,n} = \Gamma_{l,q,n} + \tilde{\tau}_n$ and $\tilde{\tau}_n \triangleq \tau_{\mathrm{TTD},n}$.

For a real pass-band signal $x(t) = \Re[s(t)e^{j2\pi ft}]$, the complex baseband equivalent model of the impulse response is written as

$$h_{l,q,n}(t) = g_l p_c(t - \tau_{l,q,n}) e^{j\theta_{l,q,n}}, \tag{C.1}$$

where $\theta_{l,q,n} \triangleq -2\pi f_c \tau_{l,q,n} + \theta_q$ and $\theta_q = \angle[\mathbf{v}]_q$ is the phase shift value introduced by the q-th transmitter circuit.

The discrete time channel with sampling duration T_s is denoted as

$$h_{l,q,n}[i] = h_{l,q,n}(iT_\mathrm{s}) = g_l p_c(iT_s - \tau_{l,q,n}) e^{j\theta_{l,q,n}}. \tag{C.2}$$

The overall channel response is then written in vector form by $\mathbf{h}_{l,n} \in \mathbb{C}^{N_\mathrm{cip}}$ where $[\mathbf{h}_{l,n}]_i = h_{l,n}[i]$. Note that $N_\mathrm{cip} T_\mathrm{s} \geq \max_{l,q,n} \Gamma_{l,q,n} + \max_n \tau_{\mathrm{TTD},n}$.

The time domain OFDM sample sequence after sampling of combined N_R antenna signals can be written as matrix form

$$\mathbf{y} = \sum_{n=1}^{N_\mathrm{R}} \sum_{l=1}^{L} \mathbf{H}_{l,n} \tilde{\mathbf{x}} \tag{C.3}$$

where $\mathbf{y} = [y[M-1], y[M-2], \cdots, y[0]]^{\mathrm{T}} \in \mathbb{C}^{M}$ is the received signal in the time domain that corresponds to one OFDM symbol. $\tilde{\mathbf{x}} = [x[M-1], \cdots, x[0], x[-1], \cdots, x[-N_{\mathrm{cp}}]]^{\mathrm{T}} \in \mathbb{C}^{M+N_{\mathrm{cp}}}$ is the transmits time domain signal after both CP and signal. Channel $\mathbf{H}_{l,n} \in \mathbb{C}^{M\times(M+N_{\mathrm{cp}})}$ is a cyclic matrix for $1 \leq n \leq N_{\mathrm{R}}$ according to the linear convolution of channel [73, Chapt 12.22]. The specific expression of this channel matrix for the n-th antenna element ($1 \leq n \leq N_{\mathrm{R}}$) by its k-th row ($1 \leq k \leq M$) as

$$[\mathbf{H}_{l,n}]_{k,:} = [\mathbf{0}_{k-1}^{\mathrm{T}}, \mathbf{h}_{l,n}^{\mathrm{T}}, \mathbf{0}_{M+N_{\mathrm{cp}}-N_{\mathrm{cip}}-k+1}^{\mathrm{T}}]. \tag{C.4}$$

Note that we generalized the definition of all-zero vector $\mathbf{0}_n \in \mathbb{C}^n$ such that $\mathbf{0}_0$ is empty. Further, for valid dimension to exist in above equation, the CP length needs to follow condition $N_{\mathrm{cp}} \geq N_{\mathrm{cip}}$, i.e., (4.7).

Next, we discuss the the effective channel after CP removal. Due to the fact that both TTD analog combining and CP removal are linear operations, the cyclic matrix that corresponds digital baseband after CP-removal is denoted as $\check{\mathbf{H}} \in \mathbb{C}^{M\times M}$. This cyclic matrix is defined by its first row $\check{\mathbf{h}}^{\mathrm{T}}$ as

$$\check{\mathbf{h}}^{\mathrm{T}} = [\check{\mathbf{H}}]_{1,:} = \left[\sum_{n=1}^{N_{\mathrm{R}}}\sum_{l=1}^{L} \mathbf{h}_{l,n}, \mathbf{0}_{M-N_{\mathrm{cip}}}^{\mathrm{T}}\right] \tag{C.5}$$

Note that the frequency domain channel that takes account analog precoder and combiner is the eigenvalues of $\check{\mathbf{H}}$. Due to the property of cyclic matrix, it can be achieved by taking discrete Fourier transform $\mathcal{F}()$ of the row vector $\check{\mathbf{h}}^{\mathrm{T}}$. Particularly, its m-th component is denoted as in (C.6), where the second equality sign is due to the *sampling theorem* and relationship between *discrete time Fourier transform* and *discrete Fourier transform*. Therefore (C.6) shows the post-beamformer channel is $\mathbf{w}_{\mathrm{TTD}}^{\mathrm{H}}[m]\mathbf{H}[m]\mathbf{v}$ as in (4.3).

Further, the TTD operation $h_{\mathrm{TTD},n}(t)$ does not change power spectral density of thermal noise in each element, and therefore the noise term is $N[m]$ defined in (4.3).

C.2 Requried delay tap and subcarrier number

Let define set $\mathcal{P} = \{\Psi | G(\Psi) > (1-\epsilon)N_{\mathrm{R}}\}$. From properties of function (4.8), this set is equivalent to $\mathcal{P} = \{\Psi | \min_{z\in\mathbb{Z}} |\Psi - 2z| \leq \delta(\epsilon, N_{\mathrm{R}})\}$. Thus, the set in terms of f and θ such that $\mathcal{P} = \{(f_m, \theta) | G(\theta, f_m) > N_{\mathrm{R}}(1-\epsilon)\}$ can be written as

$$\mathcal{P} = \left\{ (f_m, \theta) \,\middle|\, \min_{z\in\mathbb{Z}} \left| 2z + 2f_m \Delta\tau + \frac{f_m \sin(\theta)}{f_{\mathrm{c}}} \right| \leq \delta(\epsilon, N_{\mathrm{R}}) \right\}.$$

Therefore, *problem 1* becomes to find set $\mathcal{S}$ such that

$$\mathcal{S} = \left\{ (\Delta\tau, M) \,\middle|\, \max_{\theta} \min_{z\in\mathbb{Z}, m<M} \left| z + \frac{m\xi}{M} - C_\theta \right| \leq \frac{\delta(\epsilon, N_{\mathrm{R}})}{2} \right\},$$

where we define parameter $C_\theta = -f_{\mathrm{c}}\Delta\tau - \sin(\theta)/2$ and $\xi = \mathrm{BW}[\Delta\tau + \sin(\theta)/(2f_{\mathrm{c}})]$ for clarity.

Due to the degree of freedom of integer z, the condition function in the previous set is equivalent to

$$\max_{\theta} \min_{m<M} \left| \frac{m}{M}\xi - \tilde{C}_\theta \right| \leq \frac{\delta(\epsilon, N_{\mathrm{R}})}{2} \tag{C.7}$$

where $\tilde{C}_\theta \in [0, 1)$. Denote $f_{\mathrm{quan}}(m) = m\xi/M$ as a quantifier with range ξ and granularity ξ/M. A sufficient condition for (C.7) to be valid is that $f_{\mathrm{quan}}(m)$ has its range greater than 1 and its granularity less than $\delta(\epsilon, N_{\mathrm{R}})$, i.e.,

$$\xi \geq 1, \forall\theta \quad \text{and} \quad \xi/M \leq \delta(\epsilon, N_{\mathrm{R}}), \forall\theta \tag{C.8}$$

which leads to $\mathcal{S}_s$ of (4.12).

$$
\begin{aligned}
[\mathcal{F}(\check{\mathbf{h}})]_m &= \sum_{l=1}^{L}\sum_{n=1}^{N_\mathrm{R}}\sum_{q=1}^{N_\mathrm{T}}\sum_{i=0}^{M-1} e^{-j\frac{2\pi mi}{M}} \left[g_l p\left(iT_\mathrm{s} - \tau_l - \frac{(n-1)\lambda_\mathrm{c}\sin\left(\phi_l^{(\mathrm{R})}\right)}{c} + \frac{(q-1)\lambda_\mathrm{c}\sin\left(\phi_l^{(\mathrm{T})}\right)}{c} - \tilde{\tau}_n \right) e^{j\theta_{l,q,n}} \right] \\
&= \sum_{l=1}^{L}\sum_{n=1}^{N_\mathrm{R}}\sum_{q=1}^{N_\mathrm{T}} \left[\sum_{i=0}^{M-1} e^{-j\frac{2\pi mi}{M}} g_l p\left(iT_\mathrm{s} - \tau_l\right) \right] e^{-j\frac{2\pi m(n-1)\lambda_\mathrm{c}\sin(\phi_l)}{T_\mathrm{s}Mc}} e^{j\frac{2\pi m(q-1)\lambda_\mathrm{c}\sin(\phi_l)}{T_\mathrm{s}Mc}} e^{-j\frac{2\pi m\tilde{\tau}_n}{T_\mathrm{s}M}} e^{j\theta_{l,q,n}} \\
&= \sum_{l=1}^{L} \Bigg\{ \underbrace{g_l e^{-j2\pi f_\mathrm{c}\tau_l}}_{\tilde{g}_l} \left[\sum_{i=0}^{M-1} e^{-j\frac{2\pi mi}{M}} p\left(iT_\mathrm{s} - \tau_l\right) \right] \sum_{n=1}^{N_\mathrm{R}} \underbrace{\left(e^{-j\frac{2\pi m\tilde{\tau}_n}{T_\mathrm{s}M}} e^{-j2\pi f_\mathrm{c}\tilde{\tau}_n} \right)}_{[\mathbf{w}^*[m]]_n} \underbrace{\left(e^{-j\frac{2\pi m(n-1)\lambda_\mathrm{c}\sin(\phi_l)}{T_\mathrm{s}cM}} e^{-j\frac{2\pi f_\mathrm{c}(n-1)\lambda_\mathrm{c}\sin(\phi_l)}{2c}} \right)}_{[\mathbf{a}_\mathrm{R}(\phi_l, f_m)]_n} \\
&\quad \cdot \sum_{q=1}^{N_\mathrm{T}} \underbrace{\left(e^{j\frac{2\pi m(q-1)\lambda_\mathrm{c}\sin(\phi_l)}{T_\mathrm{s}cM}} e^{j\frac{2\pi f_\mathrm{c}(q-1)\lambda_\mathrm{c}\sin(\phi_l)}{2c}} \right)}_{[\mathbf{a}_\mathrm{T}^*(\phi_l, f_m)]_q} \underbrace{e^{j\theta_q}}_{[\mathbf{v}]_q} \Bigg\}
\end{aligned}
\tag{C.6}
$$

APPENDIX D

Appendix for Chapter 5

D.1 Required digital-to-analog converter quantization bits

In this subsection, we provide analysis of transmit noise $\sigma^2_{\text{n,tx}}$ in each architecture.

For each DAC, the quantization error is uniformly distributed in $[-A/2^B, A/2^B]$ where A is the largest quantization level. Without signal cropping, A depends on the peak-to-average-power-ratio (PAPR), i.e., PAPR $= A^2$ with unit signal power. The power of DAC quantization noise is

$$\epsilon_{\text{DAC}}(B) = 10\log_{10}\left[\frac{(2A)^2}{12(2^B)^2}\right] = 10\log_{10}(A^2/3) - 6B \text{ [dB]}. \tag{D.1}$$

Note that the above power is normalized with the input signal power of each DAC.

In DA architecture, the input signal power of DAC is amplified to $P^{(\text{out})}_{\text{DA}}/N_{\text{DA}}$. As a consequence, the transmitter noise power at output of each PA is $P^{(\text{out})}_{\text{DA}}\epsilon_{\text{DAC}}(B_{\text{DA}})/N_{\text{SA}}$. With the uncorrelated[1] quantization errors in each DAC, transmit noise is $\sigma^2_{\text{n,tx}} = P^{(\text{out})}_{\text{DA}}\epsilon_{\text{DAC}}(B_{\text{DA}})$.

In SA architecture, due to the identical input signal of DACs in a virtual group, quantization noise remains the same as well. The quantization noises are coherent at the outputs of N_{SA}/U PAs within a virtual group and each has power $P^{(\text{out})}_{\text{SA}}\epsilon_{\text{DAC}}(B_{\text{SA}})/N_{\text{SA}}$. As a consequence, the transmit noise is $\sigma^2_{\text{n,tx}} = P^{(\text{out})}_{\text{SA}}N_{\text{SA}}\epsilon_{\text{DAC}}(B_{\text{DA}})/U^2$.

In FH architecture, the quantization noise from each DAC is amplified to $P^{(\text{out})}_{\text{FH}}/(NU)$ in each PA. As a result, the total transmitter noise power is $\sigma^2_{\text{n,tx}} = P^{(\text{out})}_{\text{FH}}N_{\text{FH}}\sigma^2_{\text{DAC}}(B_{\text{FH}})/U$.

[1]Correlation among quantization errors of DACs become non-negligible when quantization level is significantly small, e.g., one bit. Dithering is a technique to de-correlate them but is beyond the scope of this work.

D.2 Impact of phase shifter quantization and random error on beamforming gain

Consider a linear phased array system with N antenna elements that steers a beam towards direction γ in a 2D plane. Beamforming vector is given by $[e^{j\phi_1}, \cdots, e^{j\phi_N}]$, where $\phi_n = (n-1)\pi \sin(\gamma)$. In the next, we derive beamforming gain at the main lobe for system with ideal and non-ideal phase shifters.

Let us denote the signal at the n^{th} elements as w_n with $|w_n| = 1/\sqrt{N}, \forall n$ when all phase shifters are ideal. Clearly, the phase shifter needs to be set such that signals are constructively added in the intended direction, i.e., $w_n e^{\phi_n} = 1/\sqrt{N}$, and the beamforming gain is

$$G = \left| \sum_{n=1}^{N} w_n e^{j\phi_n} \right|^2 = N$$

When all phase shifters are non-ideal, the signal at the n^{th} element is denoted as $w'_n = w_n \exp(j\psi_n)$, where ψ_n is the phase error due to quantization and random implementation impairment. With Q bits quantization, the phase error ψ_n is bounded as $|\psi_n| \leq \epsilon$ where $\epsilon = \pi/2^Q$. The corresponding beamforming gain is

$$\begin{aligned}
G' =& \left| \sum_{n=1}^{N} (w'_n e^{j\phi_n}) \right|^2 = \left| \sum_{n=1}^{N} (w_n e^{j\psi_n}) e^{j\phi_n} \right|^2 \\
=& \frac{1}{N} \left| \sum_{n=1}^{N} e^{j\psi_n} \right|^2 = \frac{1}{N} \left| \sum_{n=1}^{N} \cos(\psi_n) + j \sum_{n=1}^{N} \sin(\psi_n) \right|^2 \\
=& \frac{1}{N} \left[\sum_{n=1}^{N} \cos(\psi_n) \right]^2 + \frac{1}{N} \left[\sum_{n=1}^{N} \sin(\psi_n) \right]^2 \\
\geq& \frac{1}{N} \left[\sum_{n=1}^{N} \cos(\psi_n) \right]^2 \\
\geq& N \cos^2(\epsilon)
\end{aligned}$$

where the second inequality is valid so long as $Q \geq 1$, i.e., $|\psi_n| \leq \pi/2, \forall n$.

Therefore the gain reduction is bounded by

$$10\log_{10}\left[\frac{G}{G'}\right] \leq -20\log_{10}\left[\cos\left(\frac{\pi}{2^Q}\right)\right] \text{[dB]}$$

The above derivation shows that the gain drop in the main lobe is less than 0.68dB, 0.16dB and 0.04dB with $Q = 3$ to 5 bits quantization. Besides, these values are independent from the antenna size N. Equivalently, when phase shifter implementation error is less than $\epsilon = 22.5°, 11.25°$, and $5.625°$, gain drop is also bounded by 0.68dB, 0.16dB and 0.04dB, respectively.

Bibliography and Acronym

[1] 3GPP. LTE-Advanced. June 2013.

[2] 3GPP. TR 38.913 Study on scenarios and requirements for next generation access technologies. September 2016.

[3] 3GPP. Finalization of the NR-SSS. *Huawei, HiSilicon - Tdoc R1-1709901*, June 2017.

[4] 3GPP. NR; Overall description; Stage-2," 3rd generation partnership project (3GPP), TS 38.300. December 2017.

[5] 3GPP. TR38.802 study on new radio access technology physical layer aspects. *Tech. Rep.*, 2017.

[6] 3GPP. TR38.901 study on channel model for frequencies from 0.5 to 100 GHz. *Tech. Rep.*, 2017.

[7] 3GPP. TS38.211 NR physical channels and modulation (release 15). 2018.

[8] 3GPP. RP 190831 key directions for release 17. June 2019.

[9] 5GPPP. D2.2 measurement results and final mmMAGIC channel models. *Tech. Rep.*

[10] 5GPPP. D1.1 use case characterization, KPIs and preferred suitable frequency ranges for future 5G systems between 6 GHz and 100 GHz. November 2017.

[11] 5GPPP. D5.2 Final multi-node and multi-antenna transmitter and receiver architectures and schemes. June 2017.

[12] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi. Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs. *IEEE Trans. Wireless Commun.*, 16(12):8131–8146, December 2017.

[13] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama. Digital predistortion for hybrid MIMO transmitters. *IEEE J. Sel. Topics Signal Process.*, 12(3):445–454, June 2018.

[14] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic. Deep learning coordinated beamforming for highly-mobile millimeter wave systems. *IEEE Access*, 6:37328–37348, 2018.

[15] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Topics Signal Process.*, 8(5):831–846, October 2014.

[16] A. Alkhateeb, G. Leus, and R. W. Heath. Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE Trans. Wireless Commun.*, 14(11):6481–6494, November 2015.

[17] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath. MIMO precoding and combining solutions for millimeter-wave systems. *IEEE Commun. Mag.*, 52(12):122–131, December 2014.

[18] A. Alkhateeb, Y. H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath. Initial beam association in millimeter wave cellular systems: Analysis and design insights. *IEEE Trans. Wireless Commun.*, 16(5):2807–2821, May 2017.

[19] A. Alkhateeby, G. Leusz, and R. W. Heath. Compressed sensing based multi-user millimeter wave systems: How many measurements are needed? In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2909–2913, April 2015.

[20] Analog-Devices. Datasheet of HMC1132.

[21] 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 J. Sel. Areas Commun.*, 32(6):1065–1082, June 2014.

[22] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske. How much energy is needed to run a wireless network? *IEEE Wireless Commun. Mag.*, 18(5):40–49, October 2011.

[23] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath. Spatially sparse precoding in millimeter wave mimo systems. *IEEE Transactions on Wireless Communications*, 13(3):1499–1513, March 2014.

[24] I. Aykin, B. Akgun, and M. Krunz. Multi-beam transmissions for blockage resilience and reliability in millimeter-wave systems. *IEEE Journal on Selected Areas in Communications*, 37(12):2772–2785, 2019.

[25] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez, D. Hwang, and T. S. Rappaport. 28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York city. In *2013 IEEE International Conference on Communications (ICC)*, pages 5143–5147, June 2013.

[26] M. Bajor, T. Haque, G. Han, C. Zhang, J. Wright, and P. R. Kinget. A flexible phased-array architecture for reception and rapid direction-of-arrival finding utilizing pseudo-random antenna weight modulation and compressive sampling. *IEEE Journal of Solid-State Circuits*, 54(5):1315–1328, May 2019.

[27] R. Baldemair, T. Irnich, K. Balachandran, E. Dahlman, G. Mildh, Y. Selén, S. Parkvall, M. Meyer, and A. Osseiran. Ultra-dense networks in millimeter-wave frequencies. *IEEE Commun. Mag.*, 53(1):202–208, January 2015.

[28] Mihai Banu. HDAAS: An efficient massive-MIMO technology. In *Brooklyn 5G Summit 2017*, 2017.

[29] 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 Trans. Wireless Commun.*, 15(12):7926–7940, Dec 2016.

[30] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport. Directional cell discovery in millimeter wave cellular networks. *IEEE Trans. Wireless Commun.*, 14(12):6664–6678, December 2015.

[31] D. E. Berraki, T. H. Barratt, M. A. Beach, S. M. D. Armour, and A. R. Nix. Practical demonstration of limited feedback beamforming for mmWave systems. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2015.

[32] Lorenzo Bertizzolo, Michele Polese, Leonardo Bonati, Abhimanyu Gosain, Michele Zorzi, and Tommaso Melodia. Mmbac: Location-aided mmwave backhaul management for uav-based aerial cells. In *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, mmNets'19, page 7–12, New York, NY, USA, 2019. Association for Computing Machinery.

[33] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer. Network densification: the dominant theme for wireless evolution into 5G. *IEEE Commun. Mag.*, 52(2):82–89, February 2014.

[34] E. Björnson, M. Matthaiou, and M. Debbah. Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design. *IEEE Trans. Wireless Commun.*, 14(8):4353–4368, August 2015.

[35] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski. Five disruptive technology directions for 5G. *IEEE Commun. Mag.*, 52(2):74–80, February 2014.

[36] M. Boers, B. Afshar, I. Vassiliou, S. Sarkar, S. T. Nicolson, E. Adabi, B. G. Perumana, T. Chalvatzis, S. Kavvadias, P. Sen, W. L. Chan, A. H. T. Yu, A. Parsa, M. Nariman, S. Yoon, A. G. Besoli, C. A. Kyriazidou, G. Zochios, J. A. Castaneda, T. Sowlati, M. Rofougaran, and A. Rofougaran. A 16TX/16RX 60 GHz 802.11ad chipset with single coaxial interface and polarization diversity. *IEEE J. Solid-State Circuits*, 49(12):3031–3045, December 2014.

[37] 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*, pages 1274–1279, 2018.

[38] Veljko Boljanovic, Han Yan, Erfan Ghaderi, Deukhyoun Heo, Subhanshu Gupta, and Danijela Cabric. Design of millimeter-wave single-shot beam training for true-time-delay array, 2020.

[39] J. Brady, N. Behdad, and A. M. Sayeed. Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements. *IEEE Trans. Antennas Propag.*, 61(7):3814–3827, July 2013.

[40] D. Burghal, N. A. Abbasi, and A. F. Molisch. A machine learning solution for beam tracking in mmwave systems. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 173–177, 2019.

[41] Mingming Cai. *Modeling and mitigating beam squint in millimeter wave wireless communication*. PhD thesis, University of Notre Dame, 2018.

[42] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato. Channel-reconstruction-based hybrid precoding for millimeter-wave multiuser MIMO systems. *IEEE J. Sel. Topics Signal Process.*, 12(2):383–398, May 2018.

[43] M. Cho, I. Song, and J. D. Cressler. A true time delay-based SiGe bi-directional T/R chipset for large-scale wideband timed array antennas. In *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 272–275, June 2018.

[44] E. Cohen, M. Ruberto, M. Cohen, O. Degani, S. Ravid, and D. Ritter. A CMOS bidirectional 32-element phased-array transceiver at 60 GHz with LTCC antenna. *IEEE Trans. Microw. Theory Tech.*, 61(3):1359–1375, March 2013.

[45] Computerworld. Self-driving cars could create 1gb of data a second. 2016.

[46] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura. Experimental mm wave 5g cellular system. In *2014 IEEE Globecom Workshops (GC Wkshps)*, pages 377–381, Dec 2014.

[47] D. Cui, H. Zhang, N. Huang, A. Nazemi, B. Catli, H. G. Rhew, B. Zhang, A. Momtaz, and J. Cao. 3.2 A 320mW 32Gb/s 8b ADC-based PAM-4 Analog Front-End with Programmable Gain Control and Analog Peaking in 28nm CMOS. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, January 2016.

[48] E. Dahlman, S. Parkvall, and J. Skold. *5G NR: The Next Generation Wireless Access Technology*. Elsevier Science, 2018.

[49] C. Dehos, J. L. González, A. D. Domenico, D. Kténas, and L. Dussopt. Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5g mobile communications systems? *IEEE Communications Magazine*, 52(9):88–95, September 2014.

[50] A. Demir, A. Mehrotra, and J. Roychowdhury. Phase noise in oscillators: a unifying theory and numerical methods for characterization. *IEEE Trans. Circuits Syst. I*, 47(5):655–674, May 2000.

[51] 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*, pages 1926–1930, Nov 2014.

[52] S. Din, M. Wojtowicz, and M. Siddiqui. High power and high efficiency Ka band power amplifier. In *2015 IEEE MTT-S International Microwave Symposium*, pages 1–4, May 2015.

[53] D. De Donno, J. Palacios, and J. Widmer. Millimeter-wave beam training acceleration through low-complexity hybrid transceivers. *IEEE Trans. Wireless Commun.*, 16(6):3646–3660, June 2017.

[54] S. Ek, T. Påhlsson, A. Carlsson, A. Axholt, A. K. Stenman, and H. Sjöland. A 16-20 GHz LO system with 115 fs jitter for 24-30 GHz 5G in 28 nm FD-SOI CMOS. In *ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference*, pages 251–254, September 2017.

[55] W. El-Halwagy, A. Nag, P. Hisayasu, F. Aryanfar, P. Mousavi, and M. Hossain. A 28GHz quadrature fractional-N synthesizer for 5G mobile communication with less than 100fs jitter in 65nm CMOS. In *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 118–121, May 2016.

[56] Ericsson. Ericsson mobility report. 2019.

[57] 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.

[58] M. Ferriss, A. Rylyakov, J. A. Tierno, H. Ainspan, and D. J. Friedman. A 28 GHz hybrid PLL in 32 nm SOI CMOS. *IEEE J. Solid-State Circuits*, 49(4):1027–1035, April 2014.

[59] M. Ferriss, B. Sadhu, A. Rylyakov, H. Ainspan, and D. Friedman. A 13.1-to-28GHz fractional-N PLL in 32nm SOI CMOS with a Delta-Sigma noise-cancellation scheme. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, February 2015.

[60] Y. Frans, S. McLeod, H. Hedayati, M. Elzeftawi, J. Namkoong, W. Lin, J. Im, P. Upadhyaya, and K. Chang. 3.7 A 40-to-64Gb/s NRZ transmitter with supply-regulated front-end in 16nm FinFET. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 68–70, January 2016.

[61] K. Fujii. Low cost Ka-band 7W GaAs PHEMT based HPA with GaN PHEMT equivalent performance. In *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 207–210, May 2015.

[62] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath. Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.*, 34(4):998–1009, April 2016.

[63] R. Garg and A. S. Natarajan. A 28-GHz low-power phased-array receiver front-end with 360° RTPS phase shift range. *IEEE Trans. Microw. Theory Tech.*, 65(11):4703–4714, November 2017.

[64] E. Ghaderi, A. Sivadhasan Ramani, A. A. Rahimi, D. Heo, S. Shekhar, and S. Gupta. An integrated discrete-time delay-compensating technique for large-array beamformers. *IEEE Trans. Circuits Syst. I*, 66(9):3296–3306, September 2019.

[65] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly. IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi. *IEEE Commun. Mag.*, 55(12):186–192, Dec 2017.

[66] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly. IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi. *IEEE Commun. Mag.*, 55(12):186–192, December 2017.

[67] Y. Ghasempour, M. K. Haider, C. Cordeiro, and E. W. Knightly. Multi-user multi-stream mmwave wlans with efficient path discovery and beam steering. *IEEE Journal on Selected Areas in Communications*, 37(12):2744–2758, Dec 2019.

[68] Y. Ghasempour, N. Prasad, M. Khojastepour, and S. Rangarajan. Novel combinatorial results on downlink MU-MIMO scheduling with applications. In *2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 152–159, February 2017.

[69] Yasaman Ghasempour, Rabi Shrestha, Aaron Charous, Edward Knightly, and Daniel M Mittleman. Single-shot link discovery for terahertz wireless networks. *Nature Communications*, 11(1):1–6, 2020.

[70] 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 J. Sel. Areas Commun.*, 32(6):1152–1163, June 2014.

[71] M. Giordani, M. Mezzavilla, and M. Zorzi. Initial access in 5G mmWave cellular networks. *IEEE Commun. Mag.*, 54(11):40–47, November 2016.

[72] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi. A tutorial on beam management for 3GPP NR at mmWave frequencies. *IEEE Commun. Surveys Tuts.*, 21(1):173–196, Firstquarter 2019.

[73] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[74] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath. Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems. *IEEE J. Sel. Topics Signal Process.*, 12(2):353–367, May 2018.

[75] J. P. González-Coma, W. Utschick, and L. Castedo. Hybrid LISA for wideband multiuser millimeter-wave communication systems under beam squint. *IEEE Trans. Wireless Commun.*, 18(2):1277–1288, February 2019.

[76] Muhammad Kumail Haider, Yasaman Ghasempour, Dimitrios Koutsonikolas, and Edward W. Knightly. Listeer: Mmwave beam acquisition and steering by tracking indicator leds on wireless aps. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, page 273–288, New York, NY, USA, 2018. Association for Computing Machinery.

[77] J. Han, J. Kim, J. Park, and J. Kim. A Ka-band 4-ch bi-directional CMOS T/R chipset for 5G beamforming system. In *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 41–44, June 2017.

[78] S. Han, C. l. I, Z. Xu, and C. Rowell. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun. Mag.*, 53(1):186–194, January 2015.

[79] K. Haneda, Lei Tian, H. Asplund, Jian Li, Yi Wang, D. Steer, C. Li, T. Balercia, Sunguk Lee, YoungSuk Kim, A. Ghosh, T. Thomas, T. Nakamurai, Y. Kakishima, T. Imai, H. Papadopoulas, T. S. Rappaport, G. R. MacCartney, M. K. Samimi, Shu Sun, O. Koymen, Sooyoung Hur, Jeongho Park, Jianzhong Zhang, E. Mellios, A. F. Molisch, S. S. Ghassamzadeh, and A. Ghosh. Indoor 5g 3gpp-like channel models for office and shopping mall environments. In *2016 IEEE International Conference on Communications Workshops (ICC)*, pages 694–699, May 2016.

[80] H. Hashemi, T. Chu, and J. Roderick. Integrated true-time-delay-based ultra-wideband array processing. *IEEE Commun. Mag.*, 46(9):162–172, Sep. 2008.

[81] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. Fast millimeter wave beam alignment. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '18, pages 432–445, New York, NY, USA, 2018. ACM.

[82] S. He, C. Qi, Y. Wu, and Y. Huang. Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems. *IEEE Access*, 4:9895–9905, 2016.

[83] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Topics Signal Process.*, 10(3):436–453, April 2016.

[84] Y. Heng and J. G. Andrews. Machine learning-assisted beam alignment for mmwave systems. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2019.

[85] J. S. Herd and M. D. Conway. The evolution to modern phased array architectures. *Proc. IEEE*, 104(3):519–529, March 2016.

[86] R. Ho, K. W. Mai, and M. A. Horowitz. The Future of Wires. *Proc. IEEE*, 89(4):490–504, April 2001.

[87] Y. Huang, Y. Li, H. Ren, J. Lu, and W. Zhang. Multi-panel mimo in 5g. *IEEE Communications Magazine*, 56(3):56–61, March 2018.

[88] IEEE. 802.15.3c-2009 - ieee standard for information technology– local and metropolitan area networks– specific requirements– part 15.3: Amendment 2: Millimeter-wave-based alternative physical layer extension. 2009.

[89] IEEE. 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 amendment 3: Enhancements for very high throughput in the 60 ghz band. 2012.

[90] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele. QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials. *IEEE Trans. Antennas Propag.*, 62(6):3242–3256, June 2014.

[91] J. Jeon. NR wide bandwidth operations. *IEEE Commun. Mag.*, 56(3):42–46, March 2018.

[92] M. Jian, F. Gao, Z. Tian, S. Jin, and S. Ma. Angle-domain aided UL/DL channel estimation for wideband mmWave massive MIMO systems with beam squint. *IEEE Trans. Wireless Commun.*, pages 1–1, 2019.

[93] Nicholas J Karl, Robert W McKinney, Yasuaki Monnai, Rajind Mendis, and Daniel M Mittleman. Frequency-division multiplexing in the terahertz range using a leaky-wave antenna. *Nature Photonics*, 9(11):717, 2015.

[94] S. Kay. A fast and accurate single frequency estimator. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(12):1987–1990, December 1989.

[95] K. Kim and C. Nguyen. A 16.5 to 28 GHz 0.18-um BiCMOS power amplifier with flat 19.4 pm 1.2 dBm output power. *IEEE Microwave and Wireless Components Letters*, 24(2):108–110, February 2014.

[96] Yungsoo Kim, Hyun-Yong Lee, Philyeong Hwang, Ranjeet Kumar Patro, Jaekon Lee, Wonil Roh, and Kyungwhoon Cheun. Feasibility of Mobile Cellular Communications at Millimeter Wave Frequency. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):589–599, April 2016.

[97] U. Kodak and G. M. Rebeiz. Bi-directional flip-chip 28 GHz phased-array core-chip in 45nm CMOS SOI for high-efficiency high-linearity 5G systems. In *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 61–64, June 2017.

[98] R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober. Linear massive MIMO precoders in the presence of phase noise: A large-scale analysis. *IEEE Trans. Veh. Technol.*, 65(5):3057–3071, May 2016.

[99] B. H. Ku, O. Inac, M. Chang, H. H. Yang, and G. M. Rebeiz. A high-linearity 76 to 85-GHz 16-element 8-transmit/8-receive phased-array chip with high isolation and flip-chip packaging. *IEEE Trans. Microw. Theory Tech.*, 62(10):2337–2356, October 2014.

[100] S. Kutty and D. Sen. Beamforming for millimeter wave communications: An inclusive survey. *IEEE Commun. Surveys Tuts.*, 18(2):949–973, Secondquarter 2016.

[101] C. Li and W. Huang. A constructive representation for the Fourier dual of the Zadoff-Chu sequences. *IEEE Trans. Inf. Theory*, 53(11):4221–4224, November 2007.

[102] Y. Li, F. Baccelli, J. G. Andrews, and J. C. Zhang. Directional cell search delay analysis for cellular networks with static users. *IEEE Trans. Commun.*, 66(9):4318–4332, Sep. 2018.

[103] Y. Li, J. Luo, M. H. Castañeda Garcia, R. Böhnke, R. A. Stirling-Gallacher, W. Xu, and G. Caire. On the beamformed broadcasting for millimeter wave cell discovery: Performance analysis and design insight. *IEEE Trans. Wireless Commun.*, 17(11):7620–7634, November 2018.

[104] C. Lin and G. Y. L. Li. Terahertz communications: An array-of-subarrays solution. *IEEE Commun. Mag.*, 54(12):124–131, December 2016.

[105] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting. Millimeter wave beam alignment: Large deviations analysis and design insights. *IEEE J. Sel. Areas Commun.*, 35(7):1619–1631, July 2017.

[106] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng. Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters. *IEEE Trans. Microw. Theory Tech.*, 66(7):3419–3432, July 2018.

[107] W. Ma, C. Qi, and G. Y. Li. Machine learning for beam alignment in millimeter wave massive mimo. *IEEE Wireless Communications Letters*, pages 1–1, 2020.

[108] K. J. Maalouf and E. Lier. Theoretical and experimental study of interference in multibeam active phased array transmit antenna for satellite communications. *IEEE Trans. Antennas Propag.*, 52(2):587–592, February 2004.

[109] D. Malioutov, M. Cetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.*, 53(8):3010–3022, August 2005.

[110] B. Mamandipoor, D. Ramasamy, and U. Madhow. Newtonized orthogonal matching pursuit: Frequency estimation over the continuum. *IEEE Trans. Signal Process.*, 64(19):5066–5081, October 2016.

[111] G. Mangraviti, K. Khalaf, Q. Shi, K. Vaesen, D. Guermandi, V. Giannini, S. Brebels, F. Frazzica, A. Bourdoux, C. Soens, W. Van Thillo, and P. Wambacq. A 4-antenna-path beamforming transceiver for 60GHz multi-Gb/s communication in 28nm CMOS. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 246–247, January 2016.

[112] Z. Marzi, D. Ramasamy, and U. Madhow. Compressive channel estimation and tracking for large arrays in mm-Wave picocells. *IEEE J. Sel. Topics Signal Process.*, PP(99):1–1, April 2016.

[113] F. Meng, K. Ma, K. S. Yeo, and S. Xu. A 57-to-64-GHz 0.094-mm^2 5-bit passive phase shifter in 65-nm CMOS. *IEEE Trans. VLSI Syst.*, 24(5):1917–1925, May 2016.

[114] X. Meng, X. Gao, and X. G. Xia. Omnidirectional precoding and combining based synchronization for millimeter wave massive MIMO systems. *IEEE Trans. Commun.*, 66(3):1013–1026, March 2018.

[115] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath. Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs. *IEEE Trans. Wireless Commun.*, 16(4):2274–2287, April 2017.

[116] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda. Hybrid beamforming for massive MIMO: A survey. *IEEE Commun. Mag.*, 55(9):134–141, September 2017.

[117] S. Mondal, R. Singh, A. I. Hussein, and J. Paramesh. A 25-30 GHz 8-antenna 2-stream hybrid beamforming receiver for MIMO communication. In *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 112–115, June 2017.

[118] B. Moret, V. Knopik, and E. Kerherve. A 28GHz self-contained power amplifier for 5G applications in 28nm FD-SOI CMOS. In *2017 IEEE 8th Latin American Symposium on Circuits Systems (LASCAS)*, pages 1–4, February 2017.

[119] N. J. Myers and R. 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)*, pages 1–5, July 2017.

[120] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath. Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches? *IEEE Access*, 4:247–267, 2016.

[121] A. Nazemi, K. Hu, B. Catli, D. Cui, U. Singh, T. He, Z. Huang, B. Zhang, A. Momtaz, and J. Cao. A 36Gb/s PAM4 transmitter using an 8b 18Gs/s DAC in 28nm CMOS. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, February 2015.

[122] D. P. Nguyen, B. L. Pham, and A. V. Pham. A compact 29% PAE at 6 dB power back-off E-mode GaAs pHEMT MMIC doherty power amplifier at Ka-band. In *2017 IEEE MTT-S International Microwave Symposium (IMS)*, pages 1683–1686, June 2017.

[123] T. Nishio, Y. Wang, Y. Qian, and T. Itoh. A novel K-band frequency-controlled beam-steering quasi-Yagi array with mixing frequency compensation. In *2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No.02CH37278)*, volume 2, pages 1345–1348 vol.2, June 2002.

[124] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer. Steering with eyes closed: Mm-wave beam steering without in-band measurement. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2416–2424, 2015.

[125] S. J. Orfinidas. *Electromagnetic Waves and Antennas*, (accessed in Dec 1, 2019).

[126] M. Pajovic, P. Wang, T. Koike-Akino, and P. Orlik. Estimation of frequency unsynchronized millimeter-wave channels. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1205–1209, November 2017.

[127] B. Park, S. Jin, D. Jeong, J. Kim, Y. Cho, K. Moon, and B. Kim. Highly linear mm-Wave CMOS power amplifier. *IEEE Trans. Microw. Theory Tech.*, 64(12):4535–4544, December 2016.

[128] S. Park and R. W. Heath. Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach. *IEEE Trans. Wireless Commun.*, 17(12):8047–8062, December 2018.

[129] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst. A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. *IEEE Trans. Commun.*, 53(1):195–202, January 2005.

[130] Z. Pi, J. Choi, and R. Heath. Millimeter-wave gigabit broadband evolution toward 5G: fixed access and backhaul. *IEEE Commun. Mag.*, 54(4):138–144, April 2016.

[131] Z. Pi and F. Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE Commun. Mag.*, 49(6):101–107, June 2011.

[132] A. Pitarokoilis, E. Björnson, and E. G. Larsson. Performance of the massive MIMO uplink with OFDM and phase noise. *IEEE Commun. Lett.*, 20(8):1595–1598, August 2016.

[133] A. S. Y. Poon and M. Taghivand. Supporting and enabling circuits for antenna arrays in wireless communications. *Proc. IEEE*, 100(7):2207–2218, July 2012.

[134] A. Puglielli, A. Townley, G. LaCaille, V. Milovanović, P. Lu, K. Trotskovsky, A. Whitcombe, N. Narevsky, G. Wright, T. Courtade, E. Alon, B. Nikolić, and A. M. Niknejad. Design of energy- and cost-efficient massive MIMO arrays. *Proc. IEEE*, 104(3):586–606, March 2016.

[135] Qorvo. Datasheet of TGA4544-SM.

[136] V. Raghavan, A. Partyka, A. Sampath, S. Subramanian, O. H. Koymen, K. Ravid, J. Cezanne, K. Mukkavilli, and J. Li. Millimeter-wave mimo prototype: Measurements and experimental results. *IEEE Communications Magazine*, 56(1):202–209, Jan 2018.

[137] S. Rangan, T. S. Rappaport, and E. Erkip. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proc. IEEE*, 102(3):366–385, March 2014.

[138] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE Trans. Commun.*, 63(9):3029–3056, September 2015.

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

[140] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos. Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond. *IEEE Access*, 7:78729–78757, 2019.

[141] Maryam Eslami Rasekh and Upamanyu Madhow. Noncoherent compressive channel estimation for mm-wave massive mimo. *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Oct 2018.

[142] 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, pages 13–18, New York, NY, USA, 2017. ACM.

[143] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic. Channel estimation for frequency-selective mmWave MIMO systems with beam-squint. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, December 2018.

[144] J. Rodríguez-Fernández and N. González-Prelcic. Joint synchronization and compressive estimation for frequency-selective mmWave MIMO systems. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 1280–1286, October 2018.

[145] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath. Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems. *IEEE Trans. Wireless Commun.*, 17(5):2946–2960, May 2018.

[146] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek. A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI. *IEEE J. Sel. Topics Signal Process.*, 12(3):484–498, June 2018.

[147] R. Rotman, M. Tur, and L. Yaron. True time delay in phased arrays. *Proc. IEEE*, 104(3):504–518, March 2016.

[148] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Process. Mag.*, 30(1):40–60, January 2013.

[149] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, Ö. Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J. E. Thillberg, L. Rexberg, M. Yeck, X. Gu, D. Friedman, and A. Valdes-Garcia. A 28GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4° beam-steering resolution for 5G communication. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 128–129, February 2017.

[150] Swetank Kumar Saha, Yasaman Ghasempour, Muhammad Kumail Haider, Tariq Siddiqui, Paulo De Melo, Neerad Somanchi, Luke Zakrajsek, Arjun Singh, Roshan Shyamsunder, Owen Torres, Daniel Uvaydov, Josep Miquel Jornet, Edward Knightly, Dimitrios Koutsonikolas, Dimitris Pados, Zhi Sun, and Ngwe Thawdar. X60: A programmable testbed for wideband 60GHz WLANs with phased arrays. *Computer Communications*, 133:77 – 88, 2019.

[151] A. Sarkar, F. Aryanfar, and B. A. Floyd. A 28-GHz SiGe BiCMOS PA with 32% efficiency and 23-dBm output power. *IEEE J. Solid-State Circuits*, 52(6):1680–1686, June 2017.

[152] A. Sarkar and B. A. Floyd. A 28-GHz harmonic-tuned power amplifier in 130-nm SiGe BiCMOS. *IEEE Trans. Microw. Theory Tech.*, 65(2):522–535, February 2017.

[153] P. Schniter and S. Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

[154] S. Shakib, M. Elkholy, J. Dunworth, V. Aparin, and K. Entesari. A wideband 28GHz power amplifier supporting 8 by 100MHz carrier aggregation for 5G in 40nm CMOS. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 44–45, February 2017.

[155] S. Shakib, H. C. Park, J. Dunworth, V. Aparin, and K. Entesari. A highly efficient and linear power amplifier for 28-GHz 5G phased array radios in 28-nm CMOS. *IEEE J. Solid-State Circuits*, 51(12):3020–3036, December 2016.

[156] Clayton Shepard, Hang Yu, Narendra Anand, Erran Li, Thomas Marzetta, Richard Yang, and Lin Zhong. Argos: Practical many-antenna base stations. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, Mobicom '12, pages 53–64, 2012.

[157] G. S. Shin, J. S. Kim, H. M. Oh, S. Choi, C. W. Byeon, J. H. Son, J. H. Lee, and C. Y. Kim. Low insertion loss, compact 4-bit phase shifter in 65 nm CMOS for 5G applications. *IEEE Microwave and Wireless Components Letters*, 26(1):37–39, January 2016.

[158] W. Shin and G. M. Rebeiz. 60 GHz active phase shifter using an optimized quadrature all-pass network in 45nm CMOS. In *2012 IEEE/MTT-S International Microwave Symposium Digest*, pages 1–3, June 2012.

[159] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi. Millimeter wave cellular networks: A MAC layer perspective. *IEEE Transactions on Communications*, 63(10):3437–3458, Oct 2015.

[160] J. Song, J. Choi, and D. J. Love. Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays. *IEEE Trans. Commun.*, 65(4):1859–1872, April 2017.

[161] X. Song, S. Haghighatshoar, and G. Caire. A scalable and statistically robust beam alignment technique for millimeter-wave systems. *IEEE Trans. Wireless Commun.*, 17(7):4792–4805, July 2018.

[162] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. Wifi-assisted 60 ghz wireless networks. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, page 28–41, New York, NY, USA, 2017. Association for Computing Machinery.

[163] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino. Millimeter-wave massive MIMO: the next wireless revolution? *IEEE Commun. Mag.*, 52(9):56–62, September 2014.

[164] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[165] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten. On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems. *IEEE J. Sel. Areas Commun.*, 35(9):1980–1995, September 2017.

[166] V. Va, T. Shimizu, G. Bansal, and R. W. Heath. Online learning for position-aided millimeter wave beam training. *IEEE Access*, 7:30507–30526, 2019.

[167] A. Valdes-Garcia, A. Natarajan, D. Liu, M. Sanduleanu, X. Gu, M. Ferriss, B. Parker, C. Baks, J. O. Plouchart, H. Ainspan, B. Sadhu, M. Islam, and S. Reynolds. A fully-integrated dual-polarization 16-element W-band phased-array transceiver in SiGe BiCMOS. In *2013 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 375–378, June 2013.

[168] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath. Channel estimation for hybrid architecture-based wideband millimeter wave systems. *IEEE J. Sel. Areas Commun.*, 35(9):1996–2009, September 2017.

[169] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li. Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems. *IEEE Trans. Signal Process.*, 66(13):3393–3406, July 2018.

[170] B. Wang, F. Gao, S. Jin, H. Lin, G. Y. Li, S. Sun, and T. S. Rappaport. Spatial-wideband effect in massive MIMO with application in mmWave systems. *IEEE Commun. Mag.*, 56(12):134–141, December 2018.

[171] Bolei Wang, Mengnan Jian, Feifei Gao, Geoffrey Ye Li, Shi Jin, and Hai Lin. Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems. *CoRR*, 2019.

[172] Y. Wang, Z. Tian, S. Feng, and P. Zhang. Efficient channel statistics estimation for millimeter-wave MIMO systems. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3411–3415, March 2016.

[173] Z. Wang, M. Li, X. Tian, and Q. Liu. Iterative hybrid precoder and combiner design for mmwave multiuser mimo systems. *IEEE Commun. Lett.*, 21(7):1581–1584, July 2017.

[174] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting. Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs. *IEEE Trans. Signal Process.*, 64(10):2541–2556, May 2016.

[175] L. Wu, H. F. Leung, A. Li, and H. C. Luong. A 4-element 60-GHz CMOS phased-array receiver with beamforming calibration. *IEEE Trans. Circuits Syst. I*, 64(3):642–652, March 2017.

[176] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. L. I, and A. Ghosh. Millimeter wave communications for future mobile networks. *IEEE J. Sel. Areas Commun.*, 35(9):1909–1935, September 2017.

[177] H. Yan. Matlab simulation and data sheet for millimeterwave transmitter array architecture comparison, 2018.

[178] H. Yan. Matlab simulation for millimeter-wave compressive initial access and beam training, 2019.

[179] H. Yan, V. Boljanovic, and D. Cabric. Tracking sparse mmwave channel: Performance analysis under intra-cluster angular spread. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, June 2018.

[180] H. Yan, V. Boljanovic, and D. Cabric. Wideband millimeter-wave beam training with true-time-delay array architecture. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1447–1452, 2019.

[181] H. Yan and D. Cabric. Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 620–624, December 2016.

[182] H. Yan and D. Cabric. Digital predistortion for hybrid precoding architecture in millimeter-wave massive MIMO systems. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3479–3483, March 2017.

[183] 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, Sep. 2019.

[184] H. Yan, S. Chaudhari, and D. Cabric. Wideband channel tracking for mmwave mimo system with hybrid beamforming architecture: (invited paper). In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, Dec 2017.

[185] Han Yan and Danijela Cabric. DSP linearization for Millimeter-Wave All-Digital receiver array with Low-Resolution ADCs. In *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (IEEE SPAWC 2019)*, Cannes, France, July 2019.

[186] Han Yan, Benjamin Domae, and Danijela Cabric. mmRAPID: Machine Learning assisted Noncoherent Compressive Millimeter-Wave Beam Alignment. submitted to the 4th ACM mmNets Workshop, 2020.

[187] Han Yan, Sridhar Ramesh, Timothy Gallagher, Curtis Ling, and Danijela Cabric. Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures. *arXiv preprint arXiv:1807.07201*, 2018.

[188] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong. Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications. *IEEE Trans. Microw. Theory Tech.*, 66(7):3403–3418, July 2018.

[189] P. Yang, Y. Xiao, M. Xiao, and S. Li. 6g wireless communications: Vision and potential techniques. *IEEE Network*, 33(4):70–75, July 2019.

[190] X. Yang, M. Matthaiou, J. Yang, C. Wen, F. Gao, and S. Jin. Hardware-constrained millimeter-wave systems for 5G: Challenges, opportunities, and solutions. *IEEE Commun. Mag.*, 57(1):44–50, January 2019.

[191] Y. Yang, H. S. Ghadikolaei, C. Fischione, M. Petrova, and K. W. Sung. Reducing initial cell-search latency in mmwave networks. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 686–691, April 2018.

[192] Y. Yu, P. G. M. Baltus, A. de Graauw, E. van der Heijden, C. S. Vaucher, and A. H. M. van Roermund. A 60 GHz phase shifter integrated with LNA and PA in 65 nm CMOS for phased array systems. *IEEE J. Solid-State Circuits*, 45(9):1697–1709, September 2010.

[193] F. Yuan, C. C. Wang, T. Yu, and D. Marković. A multi-granularity FPGA with hierarchical interconnects for efficient and flexible mobile computing. *IEEE J. Solid-State Circuits*, 50(1):137–149, January 2015.

[194] Fang-Li Yuan and D. Marković. A 13.1GOPS/mW 16-core processor for software-defined radios in 40nm cmos. In *2014 Symposium on VLSI Circuits Digest of Technical Papers*, pages 1–2, June 2014.

[195] Xinying Zhang, A. F. Molisch, and Sun-Yuan Kung. Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection. *IEEE Trans. Signal Process.*, 53(11):4091–4103, November 2005.

[196] Yi Zhang, Kartik Patel, Sanjay Shakkottai, and Robert W. Heath Jr. Side-information-aided noncoherent beam alignment design for millimeter wave systems. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '19, page 341–350, New York, NY, USA, 2019. Association for Computing Machinery.

[197] 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, Sep. 2019.

[198] L. Zhao, D. W. K. Ng, and J. Yuan. Multi-user precoding and channel estimation for hybrid millimeter wave systems. *IEEE J. Sel. Areas Commun.*, 35(7):1576–1590, July 2017.

[199] S. Zihir, O. D. Gurbuz, A. Kar-Roy, S. Raman, and G. M. Rebeiz. 60-GHz 64- and 256-elements wafer-scale phased-array transmitters using full-reticle and subreticle stitching techniques. *IEEE Trans. Microw. Theory Tech.*, 64(12):4701–4719, December 2016.

[200] S. Zihir, O. D. Gurbuz, A. Karroy, S. Raman, and G. M. Rebeiz. A 60 GHz single-chip 256-element wafer-scale phased array with EIRP of 45 dBm using sub-reticle stitching. In *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pages 23–26, May 2015.

5G-NR the fifth generation mobile network new radio. 1–4, 8, 10–14, 19, 30, 37, 49, 51, 67, 81, 122, 124, 126

ADC analog-to-digital converter. 122

AoA angle-of-arrival. 3, 15, 17, 18, 22, 24–29, 32, 37, 48, 54–56, 61, 70, 72, 74, 75, 99

AoD angle-of-departure. 3, 15, 17, 18, 22, 24–29, 31–34, 37, 38, 46, 48, 50, 70, 72, 99, 125

AWGN additive white Gaussian noise. 31, 37, 96, 101

AWV antenna weight vector. 4, 5, 54, 56, 64, 68, 69, 74, 77, 80, 132

BB baseband. 85, 100–102, 119

BF beamforming. 8, 9, 14, 18, 22, 28, 29, 33, 35, 36, 41, 46, 49, 51, 55, 63, 67, 75, 81

BS base station. 2, 3, 8, 9, 12–17, 29, 30, 32, 37, 38, 48, 50, 53, 54, 67, 69, 81, 83, 88, 89, 91, 92, 98, 111, 125

CDF cumulative distribution function. 44

CFO carrier frequency offset. 4, 11, 15, 16, 18, 21, 23–28, 36, 38, 44, 49

CMOS complementary metal–oxide–semiconductor. 86, 87, 107, 111, 112, 114

CP cyclic-prefix. 5, 16, 23, 39, 68, 71

CPR compressive sensing phase retrieval. 52, 53, 55–57, 61, 62, 64

CRLB Cramér–Rao lower bound. 11, 22, 27, 29, 41, 49, 124

CS compressive sensing. 9–11, 22, 25, 27, 46, 48–50, 52, 53, 125

CSI-RS channel state information reference signal. 12, 14, 15, 18, 19, 29, 34, 35, 38, 44, 49, 50, 125

DA digital array. 85, 91–93, 97, 99, 100, 102, 104, 114–116, 118–121

DAC digital-to-analog converter. 6, 85, 91, 96, 97, 101, 102, 104, 107, 108, 113, 114, 119, 122

DFT discrete Fourier transform. 4, 15, 23, 55, 56, 58, 61–63, 74, 77, 78, 107, 132

DIA directional initial access. 9, 11, 13, 14, 18, 22, 29, 35, 37, 38, 41, 44, 46, 49

DPD digital predistortion. 121

DSP digital signal processor. 85, 86, 106, 107, 114, 116, 119, 121

EIRP equivalent isotropically radiated power. 61, 92, 106

eMBB enhanced mobile broard band. 88, 99, 100, 102, 115, 116, 118, 119

ENOB effective number of bits. 96, 101, 102, 119

FC fully connected. 58

FH fully-connected hybrid array. 85–87, 91–97, 99, 100, 102, 104, 106, 113–116, 118–121

FOM figure-of-merit. 107, 119

IA initial access. 2–4, 8–16, 18, 19, 21, 23, 24, 28, 33–38, 46, 48–50

IC integrated circuits. 83, 86, 103–106, 108, 109, 114, 120–122, 126

LNA low-noise amplifier. 86, 87

LO local oscillator. 86, 108, 109, 114, 121

LOS line-of-sight. 11, 22, 27, 29, 41, 44, 50, 60, 61, 68, 75, 77, 88, 89, 99, 118, 125–127

LUT lookup table. 74

MIMO multiple-input multiple-output. 3, 7, 9, 15, 17, 18, 49, 88, 89, 91, 98, 103, 121, 122, 124

MLE maximum-likelihood estimator. 24, 26

mmW millimeter-wave. 1–10, 12, 14, 33, 37, 44, 48, 49, 51–53, 56, 58, 64, 65, 67, 68, 70, 75, 78, 80–83, 85, 86, 88, 89, 91, 93, 98, 102–105, 107, 111, 112, 118, 119, 122, 126, 127

MPC multipath component. 70, 71

MRT maximum ratio transmission. 93

MS mobile terminal. 53, 57

NLOS non-line-of-sight. 11, 27, 44, 48, 50, 68, 77, 89, 118, 125–127

NN neural network. 57, 58, 60, 63, 64

OFDM orthogonal frequency-division multiplexing. 5, 6, 12, 13, 15, 16, 24, 35, 68, 71, 75, 77, 91, 125

OH overhead. 35

PA power amplifier. 85, 86, 92, 93, 111–116, 118, 121

PAA phased antenna array. 2–5, 74, 75, 77, 78

PAPR peak-to-average power ratio. 96, 99, 112

PBCH physical broadcast channel. 12, 13, 19

PCB printed circuit board. 56, 104, 105, 112

PLL phase-locked loop. 104, 108

PN pseudorandom noise. 54, 56, 58, 61–64, 132

PS phase shifter. 82, 85, 109, 113, 114

PSS primary synchronization signal. 12, 13, 15, 16, 20–23, 25, 28, 39, 46

ReLU rectified linear unit. 58

RF radio-frequency. 2, 5, 6, 15, 48, 49, 68, 69, 82, 85, 86, 91, 92, 95, 96, 100, 102–109, 111–115, 119–123, 126, 127

RMSE root mean square error. 41, 46

RSRP reference signal received power. 74

RSS received signal strength. 10, 29, 52, 55, 58, 61

RSS-MP received signal strength matching pursuit. 56, 61, 63, 64

Rx receiver. 51, 53, 54, 61

S-V Saleh-Valenzuela. 17, 37

SA sub-array. 85–87, 91, 93–97, 99–102, 104, 106, 113, 115, 116, 118–121

SE spectral efficiency. 82, 88, 89, 98, 99, 101, 102, 113, 116, 119

SerDes Serializer/Deserializer. 104–108, 113, 114, 116, 119

SINR signal-to-interference-plus-noise ratio. 89, 92, 93, 95, 99, 118

SNR signal-to-noise ratio. 4, 6, 11, 18, 21, 35, 38, 41, 44, 49, 50, 61, 77, 88, 92, 125

SSB synchronization signal burst. 12–16, 18, 19, 22–25, 29, 34, 39, 41, 48

SSS secondary synchronization signal. 12

STO sample timing offser. 4, 11, 14, 15, 18, 20–23, 38

TTD true-time-delay. 5, 68–75, 80, 127

Tx transmitter. 51, 53, 54, 61

UE user equipment. 2, 3, 8, 9, 11–19, 21–23, 29, 30, 35, 37–39, 41, 44, 48–50, 53, 54, 57, 60, 67, 69, 81, 88, 91, 92, 95, 98, 99, 125

ULA uniform linear array. 17

UPA uniform planar array. 17, 37, 38

VCO voltage-controlled oscillator. 104, 108, 119

ZF zero-forcing. 93, 96