

Compressive Initial Access and Beamforming Training for Millimeter-Wave Cellular Systems

Han Yan , *Student Member, IEEE*, and Danijela Cabric, *Senior Member, IEEE*

Abstract—Initial access (IA) is a fundamental procedure in cellular systems where user equipment (UE) detects base station (BS) and acquires synchronization. Due to the necessity of using antenna arrays for IA in millimeter-wave (mmW) systems, BS simultaneously performs beam training to acquire angular channel state information. The state-of-the-art directional IA (DIA) uses a set of narrow sounding beams in IA, where different beam pairs are sequentially measured, and the best candidate is determined. However, the directional beam training accuracy depends on scanning beam angular resolution, and consequently its improvement requires additional dedicated radio resources, access latency, and overhead. To remedy the problem of access latency and overhead in DIA, this paper proposes to use quasi-omni pseudorandom sounding beams for IA, and develops an algorithm for joint initial access and fine resolution initial beam training without requiring additional radio resources. It comprehensively models realistic timing and frequency synchronization errors encountered in IA. We provide the analysis of the proposed algorithm’s miss detection rate under timing synchronization errors, and we further derive 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 design the beam squint robust algorithm. For realistic performance evaluation under mmW channels, we use QuaDRiGa simulator with mmMAGIC model at 28 GHz to show that the proposed approach is advantageous to DIA. The proposed algorithm offers orders of magnitude access latency saving compared to DIA, when the same discovery, post training SNR, and overhead performance are targeted. This conclusion holds true in various propagation environments and three-dimensional locations of a mmW pico-cell with up to 140 m radius. Furthermore, our results demonstrate that the proposed beam squint robust algorithm is able to retain unaffected performance with increased beam training bandwidth.

Index Terms—Millimeter wave radio, 5G, cell discovery, initial access, beam management, synchronization, beamforming, compressive sensing.

I. INTRODUCTION

THE millimeter-wave (mmW) communication is a promising technology for the future cellular network including

Manuscript received January 31, 2019; revised June 18, 2019; accepted July 15, 2019. Date of publication July 26, 2019; date of current version September 20, 2019. This work was supported in part by NSF under Grant 1718742 and in part by the ComSenTer and CONIX Research Centers, which are two of the six centers in JUMP, a Semiconductor Research Corporation program sponsored by DARPA. The guest editor coordinating the review of this paper and approving it for publication was Prof. Feifei Gao. This paper was presented in part at the IEEE GlobalSIP, Washington, DC, USA, December 2016. (*Corresponding author: Han Yan.*)

The authors are with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095 USA (e-mail: yhaddint@ucla.edu; danijela@ee.ucla.edu).

Digital Object Identifier 10.1109/JSTSP.2019.2931206

5G New Radio (5G-NR) [2]. 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 base station (BS) and user equipment (UE) to combat severe propagation loss [3]. Significant differences in propagation characteristics and hardware architectures for mmW band compared to microwave band require novel signal processing techniques [4] and physical layer procedures [5].

Initial access (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-6 GHz 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 [6], [7]. 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.

A. 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 [6]–[15]. Directional IA (DIA) is first studied in [6] where a Generalized Likelihood Ratio Test (GLRT) is proposed to solve the cell discovery problem under unknown multiple-input multiple-output (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 [8] where overhead and access latency are analyzed. Works [9] and [10] study DIA and its access latency in large networks using stochastic geometry. Impact of beam-width of sounding beams in DIA is researched in [11]. The comparison between omni-directional and DIA is also discussed in [12]. IA using out-of-band information, e.g., location, sub-6GHz measurement, are discussed in [7], [13]. 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 [8]. It is common that DIA is paired with directional beam training [14], [15] 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 [16]–[24]. 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 [16], [17] proposed a CS-based narrowband BF training with pseudorandom sounding beamformers in the downlink, and [18], [19] extended this approach for a wideband channel. Other related works include channel covariance estimation [20]–[22] which requires periodic channel observations, and UE centric uplink training [23], [24]. It is worth nothing that all recent works focus on channel estimation alone while assuming perfect cell discovery and synchronization. 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 [1], [25]–[27]. 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 [28], Hash table [29], and sparse phase retrieval [30]. 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 [31], [32]. Recent works reveal that a beam squint unaware array processing results in a compromised channel estimation performance in both sub-6GHz and mmW systems [33]–[36]. 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.

B. 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., carrier frequency offset (CFO) and timing offset (TO).

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-LOS (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 28GHz 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.

C. Organizations and Notations

The rest of the paper is organized as follows. We start with a brief introduction of 5G-NR frame structure, IA and beam training in Section II. In Section III, we present the system model and problem statement. Section IV includes the proposed algorithm for cell discovery and timing acquisition followed by associated performance analysis. In Section V we present the algorithm and analysis for initial beam training under CFO. The squint robust beam training is presented in Section VI. The access latency, overhead, and complexity analysis is included in Section VII. The numerical results are presented in Section VIII. Open research issues are summarized in Section IX. Finally, Section X concludes the paper.

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 $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^\dagger$ respectively. The inner product is $\langle \mathbf{a}, \mathbf{b} \rangle \triangleq \mathbf{a}^H \mathbf{b}$. The l_2 -norm of \mathbf{h} is denoted by $\|\mathbf{h}\|$. $\text{diag}(\mathbf{a})$ 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 .

II. 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

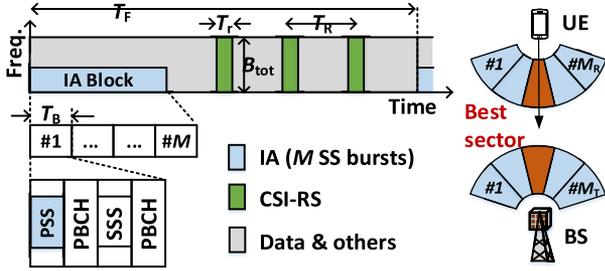


Fig. 1. The 5G-NR mmW frame structure with emphasis in beam management function and the illustration for directional initial access.

IA scheme as well as beam training. The reader is referred to work [7] for a more detailed survey.

Frame Structure: Fig. 1 shows the frame structure of 5G-NR. We focus on two functional blocks, namely synchronization signal (SS) burst and channel state information reference signal (CSI-RS). 5G-NR uses orthogonal frequency division multiplexing (OFDM), and the subcarrier spacing is either 120 or 240 KHz for mmW band. The SS signal is transmitted by a BS with period T_F , typically 20 ms. The SS 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 Fig. 1, each SS burst has 4 OFDM symbols, i.e., primary synchronization signal (PSS), physical broadcast channel (PBCH), and secondary 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 [37], while in 5G-NR PSS is replaced by Maximum Length Sequences (M-sequences) [38]. There are $N_{\text{PPS}} = 3$ and 336 unique sequences of PSS and SSS, respectively, and these 1008 combination define the cell identifier (ID) 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 SS bursts are intended to facilitate multi-antenna processing in BS and UE when no a priori channel information is available. Referring to Fig. 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 T/Rx beam is used at a time, for all $M = M_T M_R$ SS bursts.

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 [39]. Such design covers the angular space of BS and UE within M bursts, but achieves coarse propagation directions estimation [8]. Thus DIA relies on directional beam training to refine angular resolution where BS and UE steer

TABLE I
NOMENCLATURE

Symbol	Explanations
p, P	Index and total number of subcarriers
m, M	Index and total number of SS bursts
l, L	Index and total number of multipaths
N_T, N_R	Number of antenna in BS and UE
T_s	Sample duration of IA signal
T_B, N_B	Duration and sample number in each SS burst
T_F	Period of SS bursts
T_R, T_r	Period and duration of CSI-RS
N_c, N_{cp}	Max. excess delay taps and length of CP
N_{train}, N_U	Required CSI-RS and UE number
$\Delta f, \epsilon_F$	CFO in [Hz] and normalized in [rad/samp.]
ϵ_T	Initial TO in UE (number of sample)
$\mathbf{H}[d]$	MIMO channel at d -th delay sample
$\mathbf{a}_T(\theta), \mathbf{a}_R(\phi)$	Spatial responses of BS and UE
$\phi_l, \theta_l, g_l, \tau_l$	Gain/AoA/AoD/delay of l -th multipath
α_l, β_l	Real and imaginary parts of g_l
$\mathbf{s}, \bar{\mathbf{s}}, \bar{\mathbf{s}}[n]$	F/T domain PSS vector and sequence
$\mathbf{v}_m, \mathbf{w}_m$	RF precoder/combiner of the m -th burst
$\mathbf{z}[n], \mathbf{z}_m, \sigma_n^2$	AWGN sequence, vector, and power
Initial discovery (detection)	
P_{FA}^*	Target FA prob. in initial discovery
$P_{\text{MD,PT}}, P_{\text{MD,NT}}$	MD prob. w/ and w/o perfect timing
$\gamma_{\text{PT}}, \eta_{\text{PT}}$	Detection stat. and TH w/ perfect timing
$\gamma_{\text{NT}}, \eta_{\text{NT}}$	Detection stat. and TH w/ unknown timing
Initial beamforming training (estimation)	
ξ	Unknown parameters in BF training
\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.
$\mathbf{Q}(\epsilon_F), \mathbf{F}$	Phase error matrix and DFT matrix

narrow sounding beams within the sectors of interest during CSI-RS periods.

III. 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 I.

A. Asynchronous Received Signal Model in IA

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 SS burst structure and uses predefined receiver beamformers to capture the signal according to [7]. As illustrated in Fig. 2, when the signal is present, the received samples, sampled at T_s , is denoted as

$$y[n] = \sum_{d=0}^{N_c-1} e^{j(\epsilon_F n + \psi[n])} \mathbf{w}^H[n] \mathbf{H}[d] \mathbf{v}[n-d-\epsilon_T] s[n-d-\epsilon_T] + \mathbf{w}^H[n] \mathbf{z}[n], \quad n \in [0, N_F - 1]. \quad (1)$$

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

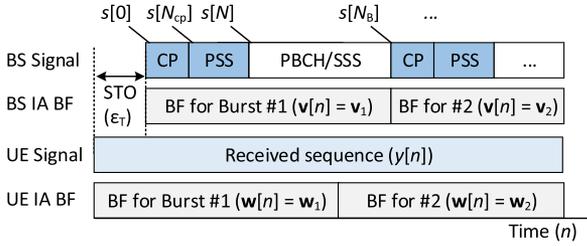


Fig. 2. Illustration of the received signal model as time sequence.

In the above equation, ϵ_T is the unknown integer sample TO within range² $0 \leq \epsilon_T \leq \epsilon_{T,\max} \leq N_B$, where $\epsilon_{T,\max}$ is the largest offset known to the system and N_B is the number of samples in one SS burst, i.e., $N_B = T_B/T_s$. The phase measurement error $e^{j(\epsilon_F n + \psi[n])}$ comes from two sources. ϵ_F is the normalized initial CFO, i.e., $\epsilon_F = 2\pi T_s \Delta f$ 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 cyclic prefix (CP) $N_{CP} > N_c$ for OFDM symbols is designed. $s[n]$ is the time domain signals of SS bursts. Referring to Fig. 2, we focus on the PSS and treat other symbols as zero [6], i.e.,

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

where $\mathcal{S}_{CP,m} \triangleq [(m-1)N_B, (m-1)N_B + N_{CP} - 1]$, $\mathcal{S}_{PSS,m} \triangleq [(m-1)N_B + N_{CP}, (m-1)N_B + N - 1]$ are the sets with sample index corresponding to CP and PSS in the m -th burst, respectively. $|s_{zc}[n]| = 1, n \in [0, P-1]$ is the Fourier dual of a known PSS sequence, and $N = P + N_{CP}$ is the number of samples in PSS including CP. $\mathbf{z}[n]$ is the Additive White Gaussian noise (AWGN) and $\mathbf{z}[n] \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_R})$. 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., $\mathbf{w}[n] \in \mathcal{W} \triangleq \{\mathbf{w}_1, \dots, \mathbf{w}_M\}$ and $\mathbf{v}[n] \in \mathcal{V} \triangleq \{\mathbf{v}_1, \dots, \mathbf{v}_M\}$. 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 $\lfloor n/N_B \rfloor = m$ and $\mathbf{v}[n] = \mathbf{v}_m$, if $\lfloor n/N_B \rfloor = m$. 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 $\mathcal{S}_T = \{\pm 1/\sqrt{N_T}, \pm j/\sqrt{N_T}\}$, and $\mathcal{S}_R = \{\pm 1/\sqrt{N_R}, \pm j/\sqrt{N_R}\}$. 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 ($d < N_c$) is denoted as $\mathbf{H}[d] \in \mathbb{C}^{N_R \times N_T}$. Following the extended Saleh Valenzuela (S-V) model in [4], we express $\mathbf{H}[d]$ as $\mathbf{H}[d] = \frac{1}{\sqrt{N_T N_R}} \sum_{l=1}^L \sum_{r=1}^R [g_{l,r} p_c(dT_s - \tau_{l,r}) \cdot \mathbf{a}_R(\phi_{l,r}^{(az)}, \phi_{l,r}^{(el)}) \mathbf{a}_T^H(\theta_{l,r}^{(az)}, \theta_{l,r}^{(el)})]$, where L and R are the number of multipath clusters (typically small, $L \leq 4$ [40]) and sub-paths

(rays), respectively. Scalar $g_{l,r}, \tau_{l,r}, \theta_{l,r}^{(az)}, \theta_{l,r}^{(el)}$ and $\phi_{l,r}^{(az)}, \phi_{l,r}^{(el)}$ are the complex gain, excessive delay, angle of departure (AoD) in azimuth and elevation plane, and angle of arrival (AoA) in two planes of the r -th sub-path within the l -th cluster, respectively. Function $p_c(t)$ 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 $\mathbf{a}_T(\theta) \in \mathbb{C}^{N_T}$ and $\mathbf{a}_R(\phi) \in \mathbb{C}^{N_R}$. In uniform planar array (UPA) with $N_R^{(az)}$ by $N_R^{(el)}$ element ($N_R^{(az)} N_R^{(el)} = N_R$), the receiver array response is defined as $[\mathbf{a}_R(\phi^{(az)}, \phi^{(el)})]_{(k_v-1)N_R^{(az)}+k_h} = \exp[j\pi(k_h-1)\sin(\phi^{(az)})\sin(\phi^{(el)}) + (k_v-1)\cos(\phi^{(el)})]$. The transmitter array response is similarly defined.

Note that the above model aligns with measurement-endorsed mmMAGIC channel model [41] and is used for the system performance evaluation in Section VIII. However, for the sake of tractable algorithm design and analysis, the following assumptions and definitions are made.

Assumption 1: Assuming BS and UE use ULA with omnidirectional element pattern in the 2D environment, i.e., array response reduces to $[\mathbf{a}_R(\phi^{(az)})]_k = \exp[j\pi(k-1)\sin(\phi^{(az)})]$ and $[\mathbf{a}_R(\phi^{(az)})]_k = \exp[j\pi(k-1)\sin(\phi^{(az)})]$. 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 paper for clarity. The phase error process is solely from CFO i.e., phase noise process is $\psi[n] = 0, \forall n$ in (1). The complex path gain g_l is deterministic complex value, i.e., $\sum_{l=1}^L |g_l|^2 = \sigma_g^2$.

Definition 1: The pre-BF signal to noise ratio (SNR) is defined as $\text{SNR} \triangleq \sigma_g^2 / \sigma_n^2$.

B. 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 SS burst from in-band received samples (1). This problem is a binary hypothesis testing with unknown channel $\mathbf{H}[d]$ and synchronization errors ϵ_T and ϵ_F .

$$\mathcal{H}_0 : y[n] = \mathbf{w}^H[n] \mathbf{z}[n],$$

$$\mathcal{H}_1 : y[n] = \sum_{d=0}^{N_c-1} \left(e^{j\epsilon_F n} \mathbf{w}^H[n] \mathbf{H}[d] \mathbf{v}[n-d-\epsilon_T] \cdot s[n-d-\epsilon_T] \right) + \mathbf{w}^H[n] \mathbf{z}[n]. \quad (2)$$

In addition, the TO 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 (1) to estimate the AoD and AoA of a path with significant power, say θ^* and ϕ^* , and they are then used in designing beamformers, $\mathbf{v}^* = \mathbf{a}_T(\theta^*)$ and $\mathbf{w}^* = \mathbf{a}_R(\phi^*)$, 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,

²We assume coarse timing synchronization is available with 10 μ s level accuracy that corresponds to current LTE-A. Practically it is achievable via GPS clock or non-standalone mmW network [7].

although *Problem 2* has overlap with parametric channel estimation, approaches from this class are not directly comparable. In fact, [19], [23] 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 ID 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.

IV. INITIAL DISCOVERY AND TIMING SYNCHRONIZATION

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

A. Initial Discovery and Timing Synchronization Algorithm

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

$$\tilde{y}[n] = \frac{1}{P} \sum_{k=0}^{P-1} y[n+k] s_{zc}^*[k]. \quad (3)$$

Intuitively, there are M correlation peaks across M SS bursts. 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 SS bursts and compares it with the threshold. In contrast to previous works [6], [9], [10] 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_{FA}^* .

To understand the impact of timing synchronization error, we first consider a *Genie* scenario where the UE has perfect timing (PT) information, i.e., $\epsilon_T = 0$. 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

$$\gamma_{PT} \triangleq \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_c-1} |\tilde{y}[k+mN_b]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta_{PT}, \quad (4)$$

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

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

$$\gamma_{NT} \triangleq \max_{0 \leq n < \epsilon_{T,\max}} \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_c-1} |\tilde{y}[n+k+mN_b]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta_{NT} \quad (5)$$

that searches all possible instances within TO window $\epsilon_T \in [0, \epsilon_{T,\max}]$ and uses the highest energy collected for the hypothesis test. The sample index corresponding to the highest energy

in (5) is the estimate of TO, namely

$$\hat{\epsilon}_T = \arg \max_{0 \leq n < \epsilon_{T,\max}} \frac{1}{M} \sum_{m=0}^{M-1} \sum_{k=0}^{N_c-1} |\tilde{y}[n+k+mN_b]|^2. \quad (6)$$

B. 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 $E \in \{\text{NT}, \text{PT}\}$.

Proposition 1: The optimal threshold of (5) that reaches target FA rate $\Pr(\gamma_E \geq \eta_E^* | \mathcal{H}_0) = P_{FA}^*$ is approximately³

$$\eta_E^* = \sigma_n^2 \left[\frac{N_c}{P} + \sqrt{\frac{N_c}{MP^2} \xi_E(\epsilon_{T,\max}, P_{FA}^*)} \right], \quad (7)$$

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

$$\xi_E = \begin{cases} Q^{-1}(P_{FA}^*), & E = \text{PT} \\ Q^{-1}\left(\frac{1}{\epsilon_{T,\max}}\right) - \frac{0.78 \ln(-\ln(1-P_{FA}^*))}{Q^{-1}\left(\frac{1}{\epsilon_{T,\max}}\right)}, & E = \text{NT} \end{cases}, \quad (8)$$

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

$$P_{MD,E} = Q\left(\frac{\kappa(\epsilon_T, \epsilon_F) \text{SNR} - \sqrt{\frac{N_c}{MP^2} \xi_E(\epsilon_{T,\max}, P_{FA}^*)}}{\sqrt{\frac{2\kappa^2(\epsilon_T, \epsilon_F) \text{SNR}^2}{M} + \frac{N_c}{P^2 M}}}\right), \quad (9)$$

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

$$\kappa(\epsilon_T, \epsilon_F) = \frac{2 - \Re(e^{jK(\epsilon_T)\epsilon_F}) - \Re(e^{j[P-K(\epsilon_T)]\epsilon_F})}{P^2 [1 - \Re(e^{j\epsilon_F})]}, \quad (10)$$

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

$$K(\epsilon_T) = \begin{cases} N_B - \epsilon_T, & \text{if } N_B - P \leq \epsilon_T < N_B \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

Proof: See Appendix A. ■

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

We gain two main insights from MD expressions (9) corresponding to threshold adjustment factor $\xi_E(\epsilon_{T,\max}, P_{FA}^*)$ and SNR degradation factor $\kappa(\epsilon_F, \epsilon_T)$. Firstly, the CFO affects MD performance by effectively reducing SNR via term $\kappa(\epsilon_F, \epsilon_T)$. Under maximum CFO at UE of $\pm 5\text{ppm}$ and typical frame parameters P, M, N_c specified in Section VIII, the SNR degradation is bounded by 4 dB, i.e., $10 \log_{10}[\kappa(\epsilon_F, \epsilon_T)] \geq -4 \text{ dB}, \forall \epsilon_T$. Secondly, the TO has impact on both factors. As seen in (10), the SNR in the detection problem degrades when severe TO exists. In fact, $K(\epsilon_T)$ in $\kappa(\epsilon_F, \epsilon_T)$ models phenomenon that receiver sounding beam switches during the reception of PSS,

³ Approximation is tight when TO search window size $\epsilon_{T,\max} \geq 100$.

i.e., $K(\epsilon_T) \neq 0$. In addition, the presence of TO forces system to use peak detection scheme (5) where system searches peak location over a sample window with length $\epsilon_{T,\max}$, i.e., the worst case in (5). Under \mathcal{H}_0 , the algorithm picks strongest noise realization over $\epsilon_{T,\max}$ samples and thus system needs to use higher threshold than in PT scenario, as seen in (7) and (8). Note that such degradation does not depend on the value of ϵ_T , and the degradation in (9) is not critical with practical maximum TO uncertainty $\epsilon_{T,\max} \leq N_B$. In summary, synchronization offset does not severely affect discovery performance of the proposed scheme.

C. Benchmark Approach: Directional Initial Discovery

For completeness, we briefly introduce the benchmark approach using directional sounding beam in initial discovery [6]. The system model of DIA is similar to Section III, except that sounding beamformers \mathcal{W} and \mathcal{V} are codebooks that steer directional sector beams, e.g., [14], [42]. Adapting the approach in [6] for the wideband channel and known PSS in SS burst, the cell discovery in DIA uses the following detector

$$\gamma_{\text{DIA}} \triangleq \max_n |\tilde{y}_{\text{DIA}}[n]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta_{\text{DIA}} \quad (12)$$

where γ_{DIA} and η_{DIA} are the detection statistic and threshold in DIA. Sequence $\tilde{y}_{\text{DIA}}[n]$ is the correlation output in (3) that corresponds to directional sounding beams. Refer to Fig. 1, the UE detects the burst with maximum power and denotes the index as m_{DIA}^* which is used in directional beam training.

V. 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.

A. 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 (1) is correctly detected and TO 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

$$\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_m^T, \dots, \mathbf{y}_M^T]^T, \\ \{\mathbf{y}_m\}_p = y[\hat{\epsilon}_T + N_{\text{CP}} + (p-1) + (m-1)N_B], p \leq P. \quad (13)$$

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

$$\mathbf{y}_m = \underbrace{\sum_{l=1}^L \tilde{g}_{m,l} \mathbf{Q}(\epsilon_{\text{F}}) \mathbf{F}^H [\mathbf{f}(\tau_l) \circ \mathbf{s}] + \mathbf{z}_m}_{\mathbf{x}_m(\boldsymbol{\xi})}, \quad (14)$$

In the above equation, deterministic vector $\mathbf{x}_m(\boldsymbol{\xi}) \in \mathbb{C}^P$ is observations model of unknown parameters

$\boldsymbol{\xi} \triangleq [\epsilon_{\text{F}}, \dots, \theta_l, \phi_l, \tau_l, \alpha_l, \beta_l, \dots]^T$, 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 $\mathbf{x}(\boldsymbol{\xi}) = [\mathbf{x}_1^T(\boldsymbol{\xi}), \dots, \mathbf{x}_M^T(\boldsymbol{\xi})]^T$. Specifically, in (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

$$[\mathbf{f}(\tau_l)]_p = \exp[-j2\pi(p-1)\tau_l/(PT_s)]. \quad (15)$$

Matrix $\mathbf{F} \in \mathbb{C}^{P \times P}$ is discrete Fourier transform (DFT) matrix.⁴ The effective channel gain is defined as $\tilde{g}_{m,l} = e^{j\epsilon_{\text{F}} N_B (m-1)} g_l \mathbf{w}_m^H \mathbf{a}_{\text{R}}(\phi_l) \mathbf{a}_{\text{T}}^H(\theta_l) \mathbf{v}_m$, and it includes the contribution of phase rotation across SS bursts due to CFO and IA beamformers \mathbf{v}_m and \mathbf{w}_m . Matrix $\mathbf{Q}(\epsilon_{\text{F}}) = \text{diag}([1, e^{j\epsilon_{\text{F}}}, \dots, e^{j(P-1)\epsilon_{\text{F}}}]^T)$ contains phase rotations within an OFDM symbol.

B. Baseline CS Formulation

Directly estimating $\boldsymbol{\xi}$ from (14) via maximum likelihood (ML) 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 [4, Sec. V], the vector $[\tilde{\mathbf{g}}_l]_m = \tilde{g}_{m,l}$ in (14) can be re-formulated as

$$\tilde{\mathbf{g}}_l = \tilde{\mathbf{Q}}(\epsilon_{\text{F}}) \tilde{\mathbf{A}}^H \text{vec}(\tilde{\mathbf{H}}_l), \quad (16)$$

where $\tilde{\mathbf{A}} \in \mathbb{C}^{G_T \times G_R \times M}$ is defined by the Hermitian conjugate of its m -th column as $([\tilde{\mathbf{A}}]_m)^H = (\mathbf{v}_m^T \otimes \mathbf{w}_m^H) (\mathbf{A}_{\text{T}}^* \otimes \mathbf{A}_{\text{R}})$. Note that the above equation is different from [4, Sec. V] which requires M^2 sounding beam pairs. The matrix $\tilde{\mathbf{Q}}(\epsilon_{\text{F}}) \in \mathbb{C}^{M \times M}$ contains the phase rotation in each SS burst due to CFO.

$$\tilde{\mathbf{Q}}(\epsilon_{\text{F}}) = \text{diag}([1, e^{jN_B \epsilon_{\text{F}}}, \dots, e^{jN_B (M-1)\epsilon_{\text{F}}}]^T). \quad (17)$$

In fact, matrices $\mathbf{A}_{\text{T}} \in \mathbb{C}^{N_T \times G_T}$ and $\mathbf{A}_{\text{R}} \in \mathbb{C}^{N_R \times G_R}$ 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}_{T} and \mathbf{A}_{R} are $[\mathbf{A}_{\text{R}}]_k = \mathbf{a}_{\text{R}}([\mathbf{r}]_k)$ and $[\mathbf{A}_{\text{T}}]_k = \mathbf{a}_{\text{T}}([\mathbf{t}]_k)$, respectively, where $[\mathbf{r}]_k$ and $[\mathbf{t}]_k$ are the vectors that contain angle candidates.

$$[\mathbf{r}]_k = -\frac{\pi}{2} + (k-1)\Delta\phi, \quad [\mathbf{t}]_k = -\frac{\pi}{2} + (k-1)\Delta\theta. \quad (18)$$

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 $\Delta\theta = 2\pi/G_T$ and $\Delta\phi = 2\pi/G_R$. Matrix $\tilde{\mathbf{H}}_l \in \mathbb{C}^{G_R \times G_T}$ 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, (16) reduces to the baseline problem in [4, Sec. V]. However, (14) implies that the former assumption is non-trivial unless $\mathbf{s} = \mathbf{1}$, $\tau_l = 0$, $\forall l$, e.g., [17]. Moreover, algorithm designed with latter assumption is sensitive to CFO [1]. Finally, the AoA/AoD estimators are commonly confined in \mathbf{r} and \mathbf{t} [17]. We address these challenges in the following three subsections.

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

C. Effective Gain Estimation

To address the challenge discussed in *Remark 3*, we propose the following approach. We treat $\mathbf{Q}(\epsilon_F)$ in (14) as identity matrix and estimate delay of dominant path and gain, say τ_l and $\tilde{g}_{m,l}$, by ML approach. Actually, the proposed algorithm uses sparse impulse support $[\mathbf{d}]_q = q\Delta\tau$ to construct a dictionary, where $\Delta\tau = N_c T_s / G_D$ is the step-size of delay candidates. Based on the knowledge of the model (15) and PSS signal \mathbf{s} , the delay estimation is implemented as

$$\hat{q} = \arg \max_{1 \leq q \leq G_D} \langle \mathbf{p}_q, \bar{\mathbf{y}} \rangle / \|\mathbf{p}_q\|^2 \text{ and } \hat{\tau} = [\mathbf{d}]_{\hat{q}}, \quad (19)$$

where $\bar{\mathbf{y}} = \sum_{m=1}^M \mathbf{y}_m / M$ is the received PSS samples averaged over M SS bursts. The vector \mathbf{p}_q contains PSS samples when the true delay of dominant path is $[\mathbf{d}]_q$, i.e., $\mathbf{p}_q \triangleq \mathbf{F}^H [\mathbf{f}([\mathbf{d}]_q) \circ \mathbf{s}]$, where $\mathbf{f}([\mathbf{d}]_q)$ is by plugging in $[\mathbf{d}]_q$ into (15). The estimated delay tap $\hat{\tau}$ enables estimating effective gain of a significant path by

$$\hat{\mathbf{g}} = (\mathbf{p}_{\hat{q}}^H \otimes \mathbf{I}_M) \mathbf{y}, \quad (20)$$

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

D. On-Grid Joint AoA and AoD Estimation Robust to CFO

The second step uses a modified matching pursuit to solve CS problem (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} [16]. The proposed heuristic treats AoA and AoD as known in the k -th step, and uses the ML estimator of CFO $\hat{\epsilon}_{F,k}$ which is available in closed form. The modified matching pursuit is expressed as

$$\hat{k} = \arg \max_{1 \leq k \leq G_R G_T} \langle \tilde{\mathbf{Q}}(\hat{\epsilon}_{F,k}) \tilde{\mathbf{a}}_k, \hat{\mathbf{g}} \rangle / \|\tilde{\mathbf{a}}_k\|^2, \quad (21)$$

where $\tilde{\mathbf{a}}_k \triangleq [\tilde{\mathbf{A}}]_k$ from (16). The matrix $\tilde{\mathbf{Q}}(\hat{\epsilon}_{F,k})$ has structure as (17). The input $\hat{\epsilon}_{F,k}$ is the ML CFO estimator when treating AoA/AoD as they correspond to ones in $[\tilde{\mathbf{A}}]_k$. Specifically, the CFO estimator relies on the estimator in [43] by treating $\bar{\mathbf{y}}_k = \tilde{\mathbf{a}}_k^* \circ \hat{\mathbf{g}}$ as a tone with frequency ϵ_F .

$$\hat{\epsilon}_{F,k} = \frac{1}{N_B} \angle \left(\frac{1}{M-1} \sum_{m=1}^{M-1} [\bar{\mathbf{y}}_k]_m^* [\bar{\mathbf{y}}_k]_{m+1} \right). \quad (22)$$

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 (18) after the following adjustment $\hat{k}_R = \lfloor (\hat{k} - 1) / G_T \rfloor + 1$ and $\hat{k}_T = \hat{k} - (\hat{k}_R - 1)G_T$,

$$\hat{\phi} = [\mathbf{r}]_{\hat{k}_R}, \quad \hat{\theta} = [\mathbf{t}]_{\hat{k}_T}, \quad \hat{\epsilon}_F = \hat{\epsilon}_{F,\hat{k}}. \quad (23)$$

E. Off-Grid Refinement

The aforementioned 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 [44] and the Newtonized gradient refinement [45]. In this work, we

Algorithm 1: Compressive Initial Access and BF Training.

Input: Received IA signal sequence $y[n]$
Output: Discovery decision; Beam pair \mathbf{v}^* , \mathbf{w}^*
 % Initial Discovery
 1: PSS correlation (3).
 2: Energy detection (5) and timing acquisition (6).
 3: **if** PositiveDecision **then**
 % Initial BF Training (Coarse)
 4: Arrange sequence $y[n]$ into vector \mathbf{y} as (13).
 5: Estimate excessive delay as (19).
 6: Estimate effective channel gain as (20).
 7: Matching pursuit (21) with CFO estimation (22).
 8: Get AoA, AoD, and CFO estimators in (23).
 % Initial BF Training (Fine)
 9: **while** $\|\mathbf{e}^{(k)}\| > \epsilon_0$ in (24) **do**
 10: Use refinement steps (25) and (26); $k = k + 1$.
 11: **end while**
 12: Report beam pair $\mathbf{w}^* = \mathbf{a}_R(\phi_l^{(k)})$, $\mathbf{v}^* = \mathbf{a}_T(\theta_l^{(k)})$.
 13: **end if**

propose to use first order descent approach. As initialization of refinement, the estimator from previous steps is saved into $\hat{\xi}^{(k)}$ for $k = 1$. In the k -th iteration, the error vector is evaluated

$$\mathbf{e}^{(k)} = \mathbf{y} - \mathbf{x}(\hat{\xi}^{(k)}), \quad (24)$$

where \mathbf{y} is the received signal after rearrangement as (13), $\mathbf{x}(\hat{\xi}^{(k)})$ is obtained by plugging in estimated parameters into parametric model (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

$$\hat{g}^{(k+1)} = (\nabla_{\mathbf{x}_g})^\dagger \mathbf{y}, \quad (25)$$

where $\nabla_{\mathbf{x}_g} = (\partial \mathbf{x}(\xi) / \partial g)|_{\xi=\hat{\xi}^{(k)}}$ is the partial derivative of $\mathbf{x}(\xi)$ over parameter g in (14) evaluated at $\hat{\xi}^{(k)}$. 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., $x = \{\tau, \epsilon_F, \theta, \phi\}$. The refinement steps are

$$\hat{x}^{(k+1)} = \hat{x}^{(k)} + \mu_x \Re \left[(\nabla_{\mathbf{x}_x})^\dagger \mathbf{e}^{(k)} \right], \quad x = \{\tau, \epsilon_F, \theta, \phi\}, \quad (26)$$

where μ_x is the step-size, vector $\nabla_{\mathbf{x}_x} = (\partial \mathbf{x}(\xi) / \partial x)|_{\xi=\hat{\xi}^{(k)}}$ is the the partial derivative of $\mathbf{x}(\xi)$ in (14) over parameter of interest. The above approach iteratively runs by appending updated parameter into $\mathbf{x}(\hat{\xi}^{(k+1)})$ 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 [16], [19], [21]–[23], [25], [26]. 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 [6]–[13], namely single path training, is evaluated.

The algorithm is summarized in Algorithm 1.

F. Performance Bound of Initial BF Training in LOS

In this subsection, we provide lower bound of AoA/AoD estimation variance in pure LOS⁵ scenario, namely CRLB in joint estimating $\boldsymbol{\xi} = [\epsilon_F, \theta_1, \phi_1, \tau_1, \alpha_1, \beta_1]^T$. Based on (14), the likelihood function is $\Pr(\mathbf{y}; \boldsymbol{\xi}) = (2\pi\sigma_n^{2MP})^{-1} \exp(-(\|\mathbf{y} - \mathbf{x}(\boldsymbol{\xi})\|^2)/(\sigma_n^2))$. 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

$$\text{var}(\hat{\phi}_1) \geq [\mathbf{J}^{-1}]_{2,2}, \quad \text{var}(\hat{\theta}_1) \geq [\mathbf{J}^{-1}]_{3,3} \quad (27)$$

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

Proof: See Appendix B. \blacksquare

G. Benchmark BF Training 1: Hierarchical Directional Search

The directional beams in SS burst allow BS and UE to coarsely estimate the propagation directions [46]. Although approach in [46] 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 SS burst 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 Fig. 1, the estimated θ^* and ϕ^* are the centers of the \hat{m}_T -th and \hat{m}_R -th sounding beams in BS and UE [46], respectively. Note that the estimated angle sector indices \hat{m}_T and \hat{m}_R are computed from the SS burst index m_{DIA}^* in (12), i.e., $\hat{m}_R = \lfloor (m_{\text{DIA}}^* - 1) / M_T \rfloor + 1$, and $\hat{m}_T = m_{\text{DIA}}^* - (\hat{m}_R - 1)M_T$. 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.

VI. SQUINT ROBUST BEAMFORMING 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 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.

A. Received Signal With Spatial Wideband Effect

By extending the channel model of [34] to multi-antenna UEs scenario, the channel in the p -th subcarrier is denoted as

$$\mathbf{H}^{(f)}[p] = \frac{1}{\sqrt{N_T N_R}} \sum_{l=1}^L g_l e^{-j2\pi\tau_l \frac{(p-1)}{PT_s}} \mathbf{a}_R(\phi_l) \tilde{\mathbf{a}}_T^H(\theta_l, p). \quad (28)$$

Due to the fact that antenna array at BS is commonly larger than in mobile terminal, the beam squint is modeled at the

⁵In 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.

transmitter end. Thus the p -th subcarrier experiences a unique transmitter array response, which is defined by its n -th element $[\tilde{\mathbf{a}}_T(\theta, p)]_n = e^{j\pi(n-1)\frac{f_p}{2f_c} \sin(\theta)} = e^{j\pi(n-1)(1+\frac{\tilde{p}(p)}{PT_s f_c}) \sin(\theta)}$. 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 $f_p = f_c + \tilde{p}(p)/(PT_s)$ [47]. For notational convenience, we use \tilde{p} instead of $\tilde{p}(p)$ in the rest of the paper. Note that when the bandwidth in beam training is negligible as compared to the carrier frequency,⁶ i.e., $1/(T_s f_c) \ll 1$, the transmitter array response vector reduces to the one in Section III, i.e., $\tilde{\mathbf{a}}_T(\theta, p) \approx \mathbf{a}_T(\theta), \forall p$. 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 [31].

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 V-A, the received signal in the frequency domain is

$$y_m^{(f)}[p] = \mathbf{w}_m^H \mathbf{H}^{(f)}[p] \mathbf{v}_m + z_m^{(f)}[p], \quad (29)$$

where $z_m^{(f)}[p]$ is the post combining AWGN in the m -th SS-burst and the p -th subcarrier. Using the similar approach as in V-B and ignoring noise for notational clarity, the above received signal can be reformulated as $y_m^{(f)}[p] = (\mathbf{v}_m^T \otimes \mathbf{w}_m^H) (\tilde{\mathbf{A}}_T^*[p] \otimes \mathbf{A}_R \otimes \mathbf{a}_d^T[p]) \mathbf{g}$. In this expression, $\tilde{\mathbf{A}}_T[p] \in \mathbb{C}^{N_T \times G_T}$ is defined by its k -th column as $\tilde{\mathbf{a}}_T([\mathbf{t}]_k, p)$ where the on-grid AoD candidates are from (18), and \mathbf{A}_R is defined in Section V. The vector $\mathbf{a}_d[p] \in \mathbb{C}^{G_D}$ is defined as by its k -th element as $e^{-j2\pi(p-1)[d]_k/(PT_s)}$. $\mathbf{g} \in \mathbb{C}^{G_T G_R G_D}$ is L -sparse vector that contains path gain that corresponds to a tuple of AoA, AoD, and delay on the grid.

B. Squint Robust Beam Training

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

$$[\tilde{\mathbf{A}}_T]_{n,k} = e^{j\pi(n-1) \sin([\mathbf{t}]_k)} \cdot \frac{e^{-j\frac{\pi(n-1) \sin([\mathbf{t}]_k)}{2f_c T_s}} - e^{j\frac{\pi(n-1) \sin([\mathbf{t}]_k)}{2f_c T_s}}}{P \left(1 - e^{j\frac{\pi(n-1) \sin([\mathbf{t}]_k)}{P f_c T_s}}\right)}, \quad n \neq 1. \quad (30)$$

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

$$\bar{\mathbf{y}} = \frac{1}{P} \sum_{p=1}^P \mathbf{y}^{(f)}[p] e^{j2\pi \frac{\hat{\tau}_l (p-1)}{T_s P}} \quad (31)$$

With an accurate excessive delay estimate $\hat{\tau}_l$, the compressive sensing problem is formulated for the signal $\bar{\mathbf{y}}$ as

$$\bar{\mathbf{y}} = \underbrace{\begin{bmatrix} (\mathbf{v}_1^T \otimes \mathbf{w}_1^H) (\tilde{\mathbf{A}}_T^* \otimes \mathbf{A}_R) \\ \vdots \\ (\mathbf{v}_M^T \otimes \mathbf{w}_M^H) (\tilde{\mathbf{A}}_T^* \otimes \mathbf{A}_R) \end{bmatrix}}_{\Psi} \bar{\mathbf{g}} \quad (32)$$

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

where $\bar{\mathbf{g}} = (\mathbf{1}_{G_D}^T \otimes \mathbf{I}_{G_T G_R}) \mathbf{g} \in \mathbb{C}^{G_T G_R}$ 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 (31) significantly reduce contribution of other paths in (32) when operating in wideband. The beam squint aware dictionary $\Psi \in \mathbb{C}^{M \times G_T G_R}$ utilize the knowledge of $\hat{\mathbf{A}}_T$ in (30), \mathbf{A}_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 (32). Further, the off-grid accuracy can be achieved in a similar manner as Section V-E. 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 (29). The residual error in the k -th iteration is defined by $e^{(k)} = \sum_{m,p} |y_m^{(f)}[p] - \mathbf{w}_m \hat{\mathbf{H}}^{(f)}[p] \mathbf{v}_m|^2$ where $\hat{\mathbf{H}}^{(f)}[p]$ in the k -th iteration is computed by plugging the estimated channel parameters into (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 flat⁷ beamformer \mathbf{v}^* are design such that it provides constant gain for all frequency range of data bandwidth, i.e., $[f_c - B_{\text{tot}}/2, f_c + B_{\text{tot}}/2]$. Given estimated AoD θ^* , the squint aware beam \mathbf{v}^* 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 $\theta_{\min} = \sin^{-1}[(1 - \frac{B_{\text{tot}}}{2f_c}) \sin(\theta^*)]$ and $\theta_{\max} = \sin^{-1}[(1 + \frac{B_{\text{tot}}}{2f_c}) \sin(\theta^*)]$.

C. Benchmark BF Training 2: Squint Non-Aware CS

There are various wideband mmW channel parameter estimation approaches that use compressive sensing [18], [19], [23], [24], [34], [36]. To compare the performance of proposed squint robust beam training, we use [18] as the second benchmark method. In this method, the channel parameters are estimated for each subcarrier separately, i.e., the problem is decoupled into multiple parallel parameter estimations, which are then solved via compressive sensing algorithm. Although [18] 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 with our wideband AoD estimates.

VII. ACCESS LATENCY, OVERHEAD AND DSP 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.

⁷Frequency flat beamformers can be implemented by analog array architecture. Readers are referred to [34], [36] for approaches using frequency dependent beamformer design in handling squint after beam training, where digital or hybrid array architecture is used.

Algorithm 2: Squint Robust BF Training.

Input: Received IA signal sequence $y[n]$
Output: AoA/AoD pair ϕ^*, θ^*
 % ——— Beam Training (Coarse) ———
 1: Estimate excessive delay $\hat{\tau}_l$ as (19).
 2: Convert $y[n]$ to frequency symbols $\mathbf{y}^{(f)}[p]$
 3: Filter frequency domain measurements as (31).
 4: On-grid angle estimates via matching pursuit in (32) that utilize squint aware dictionary (30).
 % ——— BF Training (Fine) ———
 5: **while** $e^{(k)} > \epsilon_0$ **do**
 6: Use gradient refinement based on (29); $k = k + 1$.
 7: **end while**
 8: Report AoA/AoD pair $\phi^* = \phi_l^{(k)}, \theta^* = \theta_l^{(k)}$.

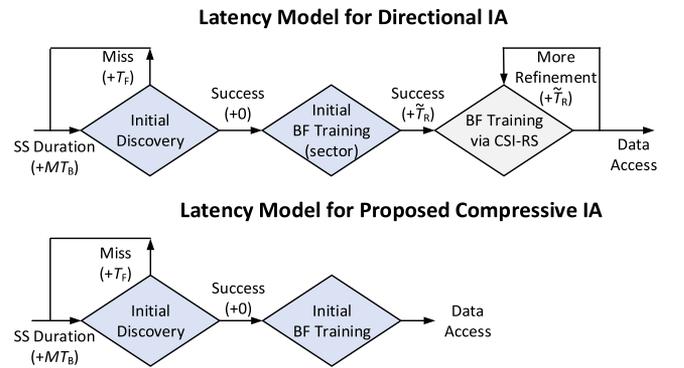


Fig. 3. Initial access latency model for directional IA and the proposed compressive IA. The associated latency in each step is shown under bracket.

Based on [7], we propose to use the latency model⁸ for both SS burst and CSI-RS as shown in Fig. 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 CSI-RS according to the required number N_{train} . Thus the access latency is

$$T_{\text{latency}} = MT_B + \sum_{k=0}^{+\infty} P_{\text{MD}}^k (1 - P_{\text{MD}}) k T_F + \tilde{T}_R N_{\text{Train}} \quad (33)$$

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

$$\tilde{T}_R = \frac{1}{N_U} \left[\sum_{k=0}^{K_F} \sum_{q=1}^{K_R} ((k-1)T_F + qT_R) + \sum_{q=1}^{K_{\text{res}}} (K_F T_F + qT_R) \right]$$

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 $K_R = \lfloor (T_F - MT_B) / T_R \rfloor$ within one IA period, more than one frame duration is required to meet scheduling of large number of UE N_U . Therefore, in (33) $K_F = \lfloor (N_U - 1) / K_R \rfloor$ is the number of frames required to assign all CSI-RS to UEs

⁸Note 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.

TABLE II
DIGITAL BASEBAND OPERATIONS (COMPLEX MULTIPLICATIONS)

Function Block	Equation	Operations
Initial Discovery		
PSS FIR corr.	(3)	PN_B
Detection and time sync.	(4) or (5)	N_B
Initial BF training (on-grid stage)		
Excess. delay est.	(19) and (20)	$PG_D + PM$
AoA/AoD est.	(21)	$MG_T G_R$
CFO est.	(22)	$2MG_T G_R$
Initial BF training (off-grid stage)		
Alternative updates	(25) and (26)	$\mathcal{O}(K_{ite} N_T N_R MP)$
Error norm evaluation	-	$K_{ite} MP$

and $K_{res} = N_U - K_{cyc} K_R$ 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 [7], the overhead (OH) ratio is modeled by counting the time-frequency resource in IA and CSI-RS

$$OH = \frac{MB_{IA}T_B + K_R B_{tot} T_T}{B_{tot} T_F} \times 100\% \quad (34)$$

where $B_{IA} = 1/T_s$ 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 [6], 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 (33). Furthermore, the overhead remains the same since $B_{IA}T_B$ in (34) remains constant.

The required baseband operations of the proposed approach are summarized in Table II, 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 $\mathcal{O}(PMG_T G_R)$ when all P subcarriers are used [18], or $\mathcal{O}(P_{sel}MG_T G_R)$ where a selective number of P_{sel} subcarriers are used [19]. The proposed approach reduces this most computationally demanding steps into $\mathcal{O}(MG_T G_R)$. Admittedly, the refinement stage involves higher complexity, since each iteration contains the computation of gradient and gradient based updates (26) and (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 VI 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 (19), $\tilde{\mathbf{a}}_k$ in (21). Lastly, directional IA involves the computation of (4) or (5) and its computational complexity is given in Table II.

VIII. RESULTS

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

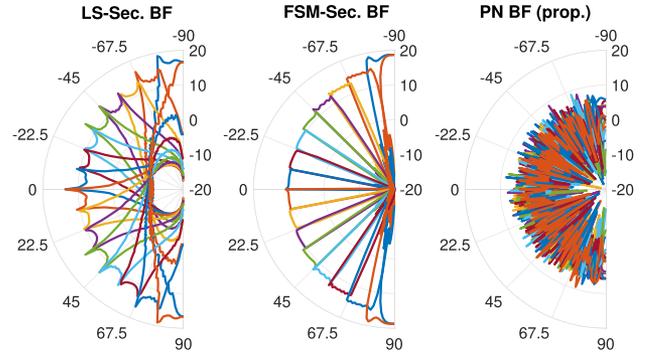


Fig. 4. Beam patterns of two sector beam designs [14], [42] with $M_T = 16$ transmit sectors and one realization of 16 pseudorandom beams. In the polar plot, the r -axis refers to the gain in dB and the angular axis refers to steering angle in degrees. All patterns correspond to $N_T = 128$ ULA.

A. 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 (LS-Sec) codebook [14] and frequency sampling method based sector beamforming (FSM-Sec) codebook [42, C23.4]. Examples of beam patterns⁹ are shown in Fig. 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 VIII-C we simulate the system with QuaDRiGa simulator [48] based on mmMAGIC model [41] in 28 GHz urban-micro (UMi). We remove Assumptions 1, 2 from Sections III-A and V-A. Uniform planar arrays (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 (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 $-170 + 10 \log_{10}(BW)$ dBm, where the noise bandwidth is $1/T_s$ and $B_{tot} = 400$ MHz [49] for IA and data stage, respectively. Moreover, the UE phase noise, $\psi[n]$ in (1), is modeled as Weiner process [50] that corresponds to oscillator with phase noise spectrum -114 dBc/Hz at 1 MHz offset [51]. The other detailed simulations setting in QuaDRiGa can be found in the supplementary material [52]. The DIA and beam training are also extended for UPA and 3D channel, i.e., FSM-Sec 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

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

TABLE III
 SUMMARY OF SIMULATION SETTINGS

Parameters	Values in Simulations
Frame Structure	
SS Signal BW	$1/T_s = 57.6$ MHz [7]
Carrier and PSS Length	$P = 128$ [7]
Max Exces. Delay and CP	$N_c = \{4, 32\}$ and $N_{cp} = \{8, 32\}$
SS Burst Duration	$N_B = 1024$ ($T_B = 17.84\mu\text{s}$) [7]
SS Burst Num.	$M = 64$ [7]; $M_T = 16$, $M_R = 4$
SS Signal Period	$T_F = 20$ ms [7]
Initial Synchronization Offset	
Freq. Offset at UE	Up to ± 5 ppm [53]
Timing Offset at UE	$\epsilon_T = \{170, 960\}$, ($\Delta\tau = 3, 17\mu\text{s}$)
STO Search Window	$\epsilon_{T,\max} = 1024$
Algorithm Design	
Target False Alarm	$P_{FA}^* = 0.01$
Dictionary Size	$G_D = 500$, $G_T = 2N_T$, $G_R = 2N_R$

bisect previous scanned azimuth and elevation angular regions. We use post-training SNR as performance indicator, which is evaluated by dividing channel gain $P_{\text{out}} \sum_{d=0}^{N_c-1} |(\mathbf{w}^*)^H \mathbf{H}[d] \mathbf{v}^*|^2$ 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 $N_T = 128$ antenna elements, and the AoD is chosen to be far from bore-sight, i.e., randomly drawn from $[35^\circ, 45^\circ]$. The UE uses $N_R = 16$ 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. III.

B. Performance in Simplified S-V Channel Model

The miss detection rate¹⁰ of the proposed approach for initial discovery is shown in Fig. 5, and it is verified against the theoretical expressions (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 ($\leq 10 \mu\text{s}$) 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., $17 \mu\text{s}$ STO which corresponds to large $K(\epsilon_T)$ in (11), severe sensitivity loss is introduced as shown in grey dotted curves. In summary, these simulations verified the findings from Section IV that practical initial synchronization error introduces up to few dB sensitivity loss as compared to perfect synchronization scenario.

The comparison among proposed approach and benchmark DIA based discovery approaches is also presented in Fig. 5. Although common sense may doubt the efficacy of the proposed approach since there is no significant angular gain for any beam

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

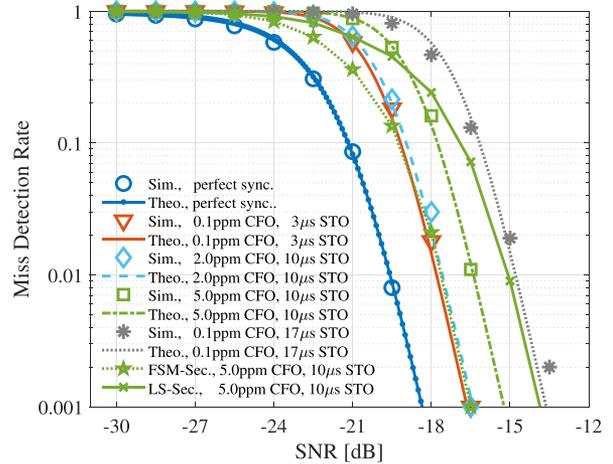


Fig. 5. Simulated (Sim.) and theoretical (Theo.) results of the miss detection rate of the proposed initial discovery with various synchronization errors. The discovery rate of the directional initial access is also included as benchmark and both LS-Sec. and FSM-Sec. are used as sector beams. The BS and UE have $N_T = 128$ and $N_R = 32$ ULA and SV channel has $L = 2$ multipaths.

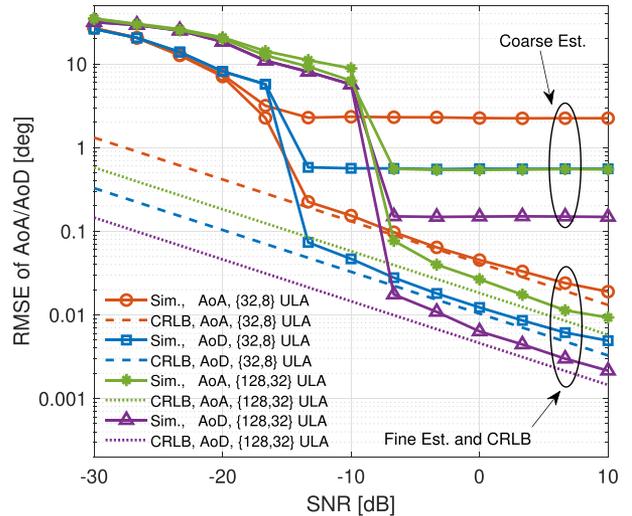
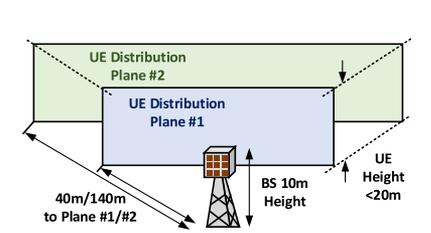


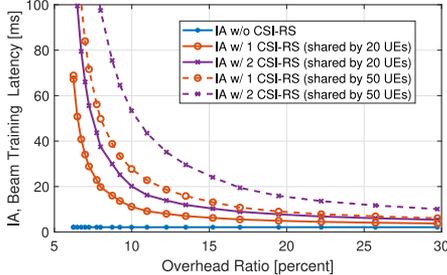
Fig. 6. Simulated results of the proposed algorithm, with and without refinement steps, and theoretical bound of RMSE of AoA/AoD estimation in LOS. Both array geometry setting, $\{N_T, N_R\} = \{32, 8\}$ ULA and $\{N_T, N_R\} = \{128, 32\}$ ULA are evaluated. System has 5ppm CFO.

pattern, as illustrated in Fig. 4, the results show that there is only a couple of dB 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 SS bursts which in fact gives equivalent energy measurement as directional approach where energy collection occurs only when a sector beam aligns with true propagation direction.

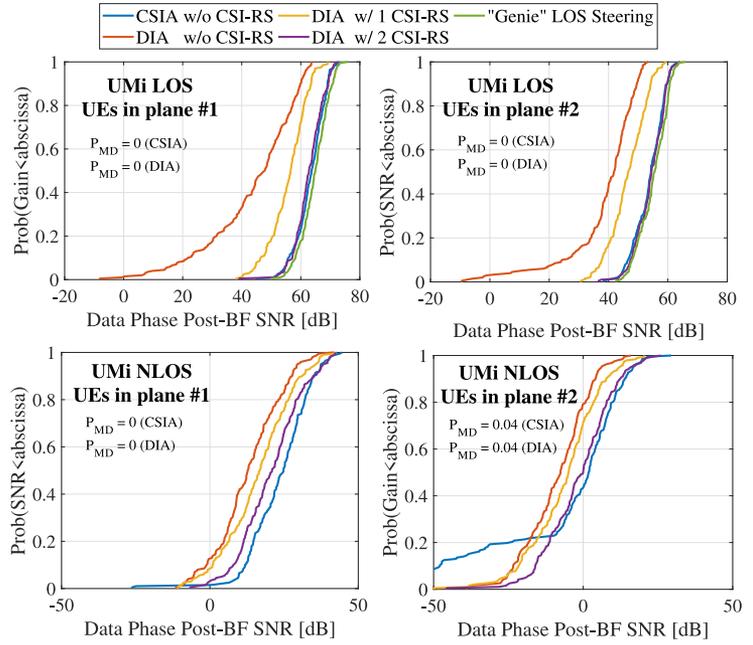
The beam training performance of the proposed BF training algorithm in LOS is presented in Fig. 6. The performance metrics are the residual mean square error defined by $\text{RMSE}_{\text{AoA}} = \sqrt{\mathbb{E}|\hat{\phi}_1 - \phi_1|^2}$ and $\text{RMSE}_{\text{AoD}} = \sqrt{\mathbb{E}|\hat{\theta}_1 - \theta_1|^2}$. The simulations are conducted with Assumption 2. The same pseudorandom setting is used in both simulation and theoretical CRLB evaluation. The



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



(c) Access latency versus overhead of both IA scheme (33) and (34) with different number of UEs that share the scheduled CSI-RS. P_{MD} is 0.04.



(b) The CDF of post-training beam steering SNR in the data phase. For the DIA, different number of CSI-RS N_{train} are considered. The SNR distribution corresponding to beam steering towards true LOS path (when existing) is also included as benchmark. The miss detection rates are included in each plots.

Fig. 7. Initial access and beam training of proposed and directional beam training evaluated in 3D outdoor UMi network using 28GHz mmMAGIC channel model [41]. The trade off between post-training SNR in the data phase, required overhead, and access latency are also studied.

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., $102^\circ/N_T$ and $102^\circ/N_R$. Finally, Fig. 5 and 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.

C. Performance in QuaDRiGa Channel Simulator

Fig. 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 Fig. 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 $N_{train} = 2$ 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 $N_{train} = 2$ 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 Fig. 7(c). As explained in Section VII, 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.

D. Performance Under Beam Squint Phenomenon

Fig. 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

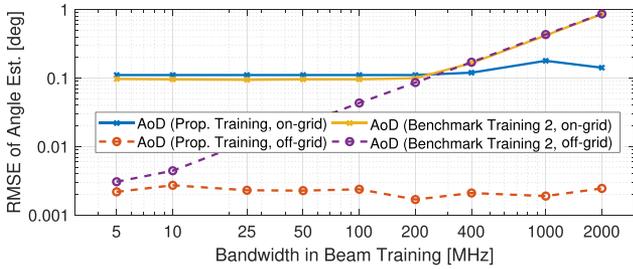


Fig. 8. AoD estimation accuracy of the proposed squint robust training and the benchmark BF training method 2 for different training bandwidths.

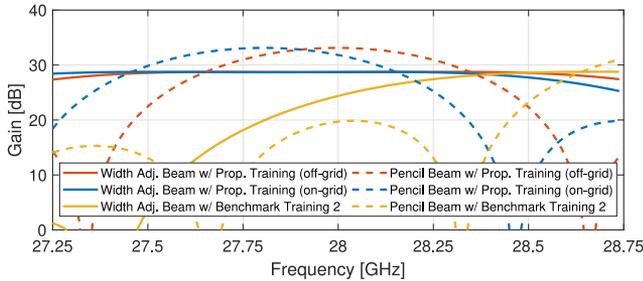


Fig. 9. Beamforming gain after beam training across a wide frequency range. Two types of beams are compared: width adjusted beam that utilizes approach in Section VI and pencil beam that uses steering vector. Beams are pointed to the center direction based on beam training.

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 Fig. 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 VI, with vector coefficients as in [42, C23.4]. The pencil beam uses conventional steering vector. The beamforming gain across a wideband range is presented in Fig. 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.

E. Baseband Processing Requirements

Using the simulation parameters in Table III to evaluate required operations in Table II, the baseband resource of the proposed method are in the same order of magnitude with DIA, i.e., $(PN_B + PG_d + 3MG_T G_R)/(PN_B) \approx 7.2$. There are two reasons for this finding. Firstly, exhaustive PSS correlation filter

(3) is extreme computational demanding in IA. 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.

IX. 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 (1), which could cause SS burst index misalignment to occur. Secondly, the compressive approach requires precise information about the sounding beam pattern $\tilde{\mathbf{a}}_k$ in (21). As a result, 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 [54] 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 [55]. 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 [6] 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 multiuser multiplexing. In fact, multiplexing designs [24], [56] 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 [57] operates at 8GHz with two phased arrays synchronized by cabled reference clock. Prototype [58] 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

TABLE IV
ELEMENTS OF FISHER INFORMATION MATRIX

Symb.	Expressions	Symb.	Expressions
$\Phi_{\epsilon_F, \epsilon_F}$	$\sum_{m=1}^M (C_{dq}^2 g) \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$	$\Phi_{\epsilon_F, \theta}$	$\sum_{m=1}^M (C_m g) \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \Re \{ [\mathbf{v}_m^H \dot{\mathbf{a}}_T(\theta)] [\mathbf{v}_m^H \mathbf{a}_T(\theta)] \}$
$\Phi_{\epsilon_F, \tau}$	$\Re \left\{ \sum_{m=1}^M g \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 \mathbf{f}^H(\tau) \mathbf{F} \mathbf{Q}_m^H \mathbf{Q}_m \mathbf{F} \mathbf{f}(\tau) \right\}$	$\Phi_{\epsilon_F, \alpha}$	$\sum_{m=1}^M [C_{dq, m} \Re(g)] \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$
$\Phi_{\epsilon_F, \beta}$	$\sum_{m=1}^M [C_{dq, m} \Im(g)] \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$	$\Phi_{\theta, \theta}$	$\sum_{m=1}^M (P g ^2) \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \dot{\mathbf{a}}_T(\theta) ^2$
$\Phi_{\phi, \phi}$	$\sum_{m=1}^M (P g ^2) \mathbf{w}_m^H \dot{\mathbf{a}}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$	$\Phi_{\phi, \theta}$	$\Re \left\{ \sum_{m=1}^M P g ^2 [\mathbf{w}_m^H \dot{\mathbf{a}}_R(\phi)] [\mathbf{w}_m^H \mathbf{a}_R(\phi)] [\mathbf{v}_m^H \dot{\mathbf{a}}_T(\theta)] [\mathbf{v}_m^H \mathbf{a}_T(\theta)] \right\}$
$\Phi_{\phi, \tau}$	$\sum_{m=1}^M C_{df, m} g ^2 \Re \{ [\mathbf{w}_m^H \dot{\mathbf{a}}_R(\phi)] [\mathbf{w}_m^H \mathbf{a}_R(\phi)] \} \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$	$\Phi_{\phi, \alpha}$	$\Re \left\{ \sum_{m=1}^M P g [\mathbf{w}_m^H \dot{\mathbf{a}}_R(\phi)] [\mathbf{w}_m^H \mathbf{a}_R(\phi)] \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 \right\}$
$\Phi_{\theta, \alpha}$	$\Re \left\{ \sum_{m=1}^M P g \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 [\mathbf{v}_m^H \dot{\mathbf{a}}_T(\theta)] [\mathbf{v}_m^H \mathbf{a}_T(\theta)] \right\}$	$\Phi_{\phi, \beta}$	$\Re \left\{ \sum_{m=1}^M j g P [\mathbf{w}_m^H \dot{\mathbf{a}}_R(\phi)] [\mathbf{w}_m^H \mathbf{a}_R(\phi)] \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 \right\}$
$\Phi_{\theta, \beta}$	$\Re \left\{ \sum_{m=1}^M j P g \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 [\mathbf{v}_m^H \dot{\mathbf{a}}_T(\theta)] [\mathbf{v}_m^H \mathbf{a}_T(\theta)] \right\}$	$\Phi_{\tau, \tau}$	$\Re \left\{ \sum_{m=1}^M g ^2 \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 [\mathbf{f}^H(\tau) \mathbf{Q}_m^H \mathbf{Q}_m \mathbf{f}(\tau)] \right\}$
$\Phi_{\tau, \alpha}$	$\Re \left\{ \sum_{m=1}^M g \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 [\mathbf{f}^H(\tau) \mathbf{Q}_m^H \mathbf{Q}_m \mathbf{f}(\tau)] \right\}$	$\Phi_{\tau, \beta}$	$\Re \left\{ \sum_{m=1}^M j g \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2 [\mathbf{f}^H(\tau) \mathbf{Q}_m^H \mathbf{Q}_m \mathbf{f}(\tau)] \right\}$
$\Phi_{\alpha, \alpha}$	$\sum_{m=1}^M P \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$	$\Phi_{\beta, \beta}$	$-\sum_{m=1}^M P \mathbf{w}_m^H \mathbf{a}_R(\phi) ^2 \mathbf{v}_m^H \mathbf{a}_T(\theta) ^2$

detrimental impact becomes more severe with increased carrier frequency and shall be properly modeled and incorporated in signal processing techniques for mmW [59]. Proper phase noise compensation as well as non-coherent CS-based beam training [28]–[30] are naturally immune to phase error and are worth investigation.

X. 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 (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 5G 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 [52].

APPENDIX

A. Initial Discovery Performance

The noise after correlation $\tilde{z}[n] = \frac{1}{P} \sum_{k=0}^{P-1} (\mathbf{w}^H[n+k] \mathbf{z}[n+k]) s_{zc}^*[k]$ is $\mathcal{NC}(0, \sigma_n^2/P)$. Thus $|\tilde{z}[n]|^2$ is Chi-Square distributed with degree-of-freedom 2, mean σ_n^2/P , and variance σ_n^4/P^2 . We

denote detection statistic in PT and NT scenario under \mathcal{H}_0 and \mathcal{H}_1 as denoted as $\gamma_{PT,0}$, $\gamma_{PT,1}$, $\gamma_{NT,0}$, $\gamma_{NT,1}$, respectively, and find their distribution.

$\gamma_{PT,0}$ is the sum of squared $N_c M$ realizations of $\tilde{z}[n]$ divided by M , thus central limit theory (CLT) applies. The distribution of $\gamma_{PT,0}$ is $\mathcal{N}(\mu_{PT,0}, \sigma_{PT,0})$, where $\mu_{PT,0} = N_c \sigma_n^2/P$ and $\sigma_{PT,0} = \sqrt{N_c \sigma_n^4/(P^2 M)}$, respectively. As a result, the optimal detection threshold that reaches target false alarm rate P_{FA}^* is given by (7). Similarly, the detection statistic under \mathcal{H}_0 with TO is denoted as $\gamma_{NT,0}$. It is the maximum operation with degrees of freedom $\epsilon_{T, \max}$ of $\gamma_{PT,0}$. With large $\epsilon_{T, \max}$, $\gamma_{NT,0}$ follows extreme value distribution, *Gumbel Distribution*, where the mean and standard deviation are $\mu_{NT,0} = \mu_{PT,0} + \sigma_{PT,0} Q^{-1}(1/\epsilon_{T, \max})$ and $\sigma_{NT,0} = \sigma_{PT,0}/Q^{-1}(1/\epsilon_{T, \max})$, respectively. Using its inverse cumulative distribution function, the optimal detection threshold is $\eta_{NT}^* = \mu_{NT,0} - (\sqrt{6}\pi)\sigma_{NT,0} \ln(-\ln(1 - P_{FA}^*))$. It gives (7) using expressions of $\mu_{NT,0}$, $\sigma_{NT,0}$ and $\sqrt{6}/\pi \approx 0.78$.

Detection statistic $\gamma_{PT,1}$ is the sum of noise energy and signal energy, i.e., $\gamma_{PT,1} = \gamma_{PT,0} + (\sum_{m=1}^M \sum_{l=0}^L |\tilde{g}_{m,l} \sum_{n=1}^P |s_{zc}[n]|^2 e^{j\epsilon_F n}|^2)/(PMN_T N_R)$, where $\tilde{g}_{m,l}$ is defined in Section V-A. Using the fact $|s_{zc}[n]| = 1$, definition $\kappa(0, \epsilon_F) \triangleq |\sum_{n=1}^P e^{j\epsilon_F n}|^2$ in (10), and approximation that different multipaths are resolvable, i.e., $p_c(dT_s - \tau_l) = 1, d \in \mathcal{S}_d$ where \mathcal{S}_d has L integers in range $[0, N_c - 1]$, the above equation becomes $\gamma_{PT,1} = \kappa(0, \epsilon_F) \sum_{m=1}^M \zeta_m/M + \gamma_{PT,0}$ where $\zeta_m = \sum_{l=0}^L |g_l \mathbf{w}_m^H \mathbf{a}_R(\phi_l) \mathbf{a}_T^H(\theta_l) \mathbf{v}_m|^2/(N_T N_R)$. 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 $\mathbb{E}(\zeta_m) = \sum_{l=0}^L |g_l|^2 \mathbb{E} |\mathbf{w}_m^H \mathbf{a}_R(\phi_l)|^2 \mathbb{E} |\mathbf{a}_T^H(\theta_l) \mathbf{v}_m|^2/(N_T N_R) = \sigma_g^2$ and $\text{var}(\zeta_m) = (N_T N_R)^{-2} \sum_{l=0}^L |g_l|^4 \mathbb{E} |\mathbf{w}_m^H \mathbf{a}_R(\phi_l)|^4 \mathbb{E} |\mathbf{a}_T^H(\theta_l) \mathbf{v}_m|^4 - \sigma_g^4 = \sigma_g^4 (2 - \frac{1}{N_T}) (2 - \frac{1}{N_R}) - \sigma_g^4 \approx 3\sigma_g^4$, respectively. The above approximation holds true with typical antenna array sizes N_R and N_T in mmW. Therefore, according to CLT $\gamma_{PT,1} \sim \mathcal{CN}(\kappa(0, \epsilon_F) \sigma_g^2 + \mu_{PT,0}, 3\kappa^2(0, \epsilon_F) \sigma_g^4/M + \sigma_{PT,0}^2)$, which gives the miss detection probability $P_{MD,PT} = Q[(\mathbb{E}(\gamma_{PT,1}) - \eta_{PT}^*)/\sqrt{\text{var}(\gamma_{PT,1})}]$, and it equals to (9).

In NT scenario, we make the following approximations: 1) the detection statistic $\gamma_{NT,1}$ corresponds to the correlation peaks for the correct timing ϵ_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 $\gamma_{\text{NT},1}$ as $\gamma_{\text{NT},1} = \gamma_{\text{PT},0} + \frac{1}{PMLN_T N_R} (\sum_{m=1}^M \sum_{l=0}^L |\tilde{g}_{m,l}^{(1)} \sum_{n_1=1}^{K-1} |s_{zc}[n_1]|^2 e^{j\epsilon_F n_1} + \tilde{g}_{m,l}^{(2)} \sum_{n_2=K}^P |s_{zc}[n_2]|^2 e^{j\epsilon_F n_2}|^2)$ where $\tilde{g}_{m,l}^{(1)} = g_l \mathbf{w}_m^H \mathbf{a}_R(\phi_l) \mathbf{a}_T^H(\theta_l) \mathbf{v}_m$ and $\tilde{g}_{m,l}^{(2)} = g_l \tilde{\mathbf{w}}_m^H \mathbf{a}_R(\phi_l) \mathbf{a}_T^H(\theta_l) \mathbf{v}_m$ are the post-BF channel gain due to partially overlapped burst window in BS and UE. In other words, K follows (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 $\gamma_{\text{NT},1}$, we define $\zeta_{m,l}$ as $\zeta_{m,l} \triangleq (|\tilde{g}_{m,l}^{(1)} \frac{1-e^{jK\epsilon_F}}{1-e^{j\epsilon_F}} + \tilde{g}_{m,l}^{(2)} \frac{1-e^{j(P-K)\epsilon_F}}{1-e^{j\epsilon_F}}|^2) / (N_T N_R)$ in $\gamma_{\text{NT},1}$ after simplification with the fact $|s_{zc}[n]|^2 = 1, \forall n \in \mathcal{S}$ as well as $\sum_{n=1}^K e^{j\epsilon_F n} = (1 - e^{jK\epsilon_F}) / (1 - e^{j\epsilon_F})$. The mean and variance of $\zeta_{m,l}$ are $\mathbb{E}(\zeta_{m,l}) = \kappa(\epsilon_F, \epsilon_T) \sigma_g^2$, and $\text{var}(\zeta_{m,l}) \approx 3\sigma_g^4 \zeta^2(\epsilon_F, \epsilon_T)$ after plugging in definition of $\kappa(\epsilon_F, \epsilon_T)$ from (10). Using CLT and statistic of $\zeta_{m,l}$, $\gamma_{\text{NT},1} \sim \mathcal{CN}(\mu_{\text{PT},0} + \kappa(\epsilon_F, \epsilon_T) \sigma_g^2, \sigma_{\text{PT},0}^2 + 3\sigma_g^4 \kappa^2(\epsilon_F, \epsilon_T) / M)$. The MD rate $P_{\text{MD,NT}} = \mathcal{Q}[(\mathbb{E}(\gamma_{\text{NT},1}) - \eta_{\text{NT}}^*) / \sqrt{\text{var}(\gamma_{\text{NT},1})}]$ reduces to (9).

B. CRLB of Joint Estimation Problem

The FIM has the following form

$$\mathbf{J} = \frac{1}{\sigma_n^2} \begin{bmatrix} \Phi_{\epsilon_F, \epsilon_F} & \Phi_{\epsilon_F, \theta} & \Phi_{\epsilon_F, \phi} & \Phi_{\epsilon_F, \tau} & \Phi_{\epsilon_F, \alpha} & \Phi_{\epsilon_F, \beta} \\ \Phi_{\theta, \epsilon_F} & \Phi_{\theta, \theta} & \Phi_{\theta, \phi} & \Phi_{\theta, \tau} & \Phi_{\theta, \alpha} & \Phi_{\theta, \beta} \\ \Phi_{\phi, \epsilon_F} & \Phi_{\phi, \theta} & \Phi_{\phi, \phi} & \Phi_{\phi, \tau} & \Phi_{\phi, \alpha} & \Phi_{\phi, \beta} \\ \Phi_{\tau, \epsilon_F} & \Phi_{\tau, \theta} & \Phi_{\tau, \phi} & \Phi_{\tau, \tau} & \Phi_{\tau, \alpha} & \Phi_{\tau, \beta} \\ \Phi_{\alpha, \epsilon_F} & \Phi_{\alpha, \theta} & \Phi_{\alpha, \phi} & \Phi_{\alpha, \tau} & \Phi_{\alpha, \alpha} & 0 \\ \Phi_{\beta, \epsilon_F} & \Phi_{\beta, \theta} & \Phi_{\beta, \phi} & \Phi_{\beta, \tau} & 0 & \Phi_{\beta, \beta} \end{bmatrix}$$

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

REFERENCES

- [1] H. Yan and D. Cabric, "Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2016, pp. 620–624.
- [2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [3] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [4] R. W. Heath, N. Gonzalez-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.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [5] "Study on new radio access technology physical layer aspects, v14.2.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.802, 2017.
- [6] C. N. Barati *et al.*, "Directional cell discovery in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.
- [7] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surv. Tuts.*, vol. 21, no. 1, pp. 173–196, Jan.–Mar. 2019.
- [8] Y. Li *et al.*, "On the beamformed broadcasting for millimeter wave cell discovery: Performance analysis and design insight," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7620–7634, Nov. 2018.
- [9] 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.*, vol. 66, no. 9, pp. 4318–4332, Sep. 2018.
- [10] Y. Yang, H. S. Ghadikolaie, C. Fischione, M. Petrova, and K. W. Sung, "Reducing initial cell-search latency in mmwave networks," in *Proc. IEEE INFOCOM*, Apr. 2018, pp. 686–691.
- [11] 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.*, vol. 16, no. 5, pp. 2807–2821, May 2017.
- [12] X. Meng, X. Gao, and X. G. Xia, "Omnidirectional precoding and combining based synchronization for millimeter wave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1013–1026, Mar. 2018.
- [13] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.
- [14] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [15] 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.*, vol. 35, no. 7, pp. 1619–1631, Jul. 2017.
- [16] Z. Marzi, D. Ramasamy, and U. Madhoo, "Compressive channel estimation and tracking for large arrays in mm-Wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.
- [17] A. Alkhateeb, G. Leusz, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 2909–2913.
- [18] K. Venugopal, A. Alkhateeb, N. Gonzalez Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [19] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [20] Y. Wang, Z. Tian, S. Feng, and P. Zhang, "Efficient channel statistics estimation for millimeter-wave MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 3411–3415.
- [21] X. Song, S. Haghghatshoar, and G. Caire, "A scalable and statistically robust beam alignment technique for millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4792–4805, Jul. 2018.
- [22] S. Park and R. W. Heath, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8047–8062, Dec. 2018.
- [23] J. P. Gonzalez-Coma, J. Rodriguez-Fernandez, N. Gonzalez-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.*, vol. 12, no. 2, pp. 353–367, May 2018.
- [24] 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.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
- [25] N. J. Myers and R. W. Heath, "A compressive channel estimation technique robust to synchronization impairments," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2017, pp. 1–5.
- [26] M. Pajovic, P. Wang, T. Koike-Akino, and P. Orlik, "Estimation of frequency unsynchronized millimeter-wave channels," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Nov. 2017, pp. 1205–1209.

- [27] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, "Joint synchronization and compressive estimation for frequency-selective mmWave MIMO systems," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1280–1286.
- [28] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. 18th Int. Workshop Mobile Comput. Syst. Appl.*, ser. HotMobile '17, New York, NY, USA: ACM, 2017, pp. 13–18.
- [29] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Proc. Conf. ACM Special Interest Group Data Commun.*, ser. SIGCOMM '18, New York, NY, USA: ACM, 2018, pp. 432–445.
- [30] M. E. Rasekh and U. Madhow, "Noncoherent compressive channel estimation for mm-wave massive MIMO," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 889–894.
- [31] B. Wang *et al.*, "Spatial-wideband effect in massive MIMO with application in mmWave systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 134–141, Dec. 2018.
- [32] M. Cai, "Modeling and mitigating beam squint in millimeter wave wireless communication," Ph.D. dissertation, University of Notre Dame, Notre Dame, IN, USA, 2018.
- [33] 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.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
- [34] B. Wang, M. Jian, F. Gao, G. Y. Li, S. Jin, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *CoRR*, 2019. [Online]. Available: <http://arxiv.org/abs/1903.01340>
- [35] 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.*, vol. 18, no. 7, pp. 3515–3527, Jul. 2019.
- [36] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, "Channel estimation for frequency-selective mmWave MIMO systems with beam-squint," in *Proc. IEEE Global Commun. Conf.*, Dec. 2018, pp. 1–6.
- [37] C. Li and W. Huang, "A constructive representation for the Fourier dual of the Zadoff-Chu sequences," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4221–4224, Nov. 2007.
- [38] 3GPP, "TS38.211 NR physical channels and modulation (release 15)," 2018. [Online]. Available: <https://portal.3gpp.org>
- [39] 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.*, vol. 65, no. 4, pp. 1859–1872, Apr. 2017.
- [40] 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.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [41] M. Peter *et al.*, "Measurement Results and Final mmMAGIC Channel Models." [Online]. Available: https://bscw.5g-mmmagic.eu/pub/bscw.cgi/d202656/mmMAGIC_D2-2.pdf, Accessed: Dec. 2018.
- [42] S. J. Orfanidas, *Electromagnetic Waves and Antennas*. [Online]. Available: <http://ecceweb1.rutgers.edu/~orfanidi/ewa/>, Accessed: Dec. 2018.
- [43] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1987–1990, Dec. 1989.
- [44] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [45] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.
- [46] M. R. Castellanos *et al.*, "Channel-reconstruction-based hybrid precoding for millimeter-wave multi-user MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 383–398, May 2018.
- [47] J. P. Gonzalez-Coma, W. Utschick, and L. Castedo, "Hybrid LISA for wideband multiuser millimeter-wave communication systems under beam squint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1277–1288, Feb. 2019.
- [48] S. Jaeckel, L. Raschkowski, K. Brner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
- [49] J. Jeon, "NR wide bandwidth operations," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 42–46, Mar. 2018.
- [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*, vol. 47, no. 5, pp. 655–674, May 2000.
- [51] W. El-Halwagy, A. Nag, P. Hisayasu, F. Aryanfar, P. Mousavi, and M. Hossain, "A 28 GHz quadrature fractional-N synthesizer for 5G mobile communication with less than 100fs jitter in 65 nm CMOS," in *Proc. IEEE Radio Frequency Integr. Circuits Symp.*, May 2016, pp. 118–121.
- [52] H. Yan, "Matlab simulation for millimeter-wave compressive initial access and beam training," 2019. [Online]. Available: <https://github.com/yhaddint/MillimeterWaveCSIA>
- [53] 3GPP, "Finalization of the NR-SSS," *Huawei, HiSilicon, Shenzhen, China, Tech. Rep. Tdoc R1-1709901*, Jun. 2017.
- [54] S. K. Saha *et al.*, "X60: A programmable testbed for wideband 60GHz WLANs with phased arrays," *Comput. Commun.*, vol. 133, pp. 77–88, 2019.
- [55] "Study on channel model for frequencies from 0.5 to 100 GHz, v14.2.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, 2017.
- [56] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [57] 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 *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [58] 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 J. Solid-State Circuits*, vol. 54, no. 5, pp. 1315–1328, May 2019.
- [59] 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.*, vol. 57, no. 1, pp. 44–50, Jan. 2019.



Han Yan received the B.E. degree from Zhejiang University, Hangzhou, China, in 2013, and the M.S. degree in electrical engineering from the University of California, Los Angeles, Los Angeles, CA, USA, in 2015. He is currently working toward the Ph.D. degree at the University of California, Los Angeles, Los Angeles, CA, USA. His research interests include signal processing for millimeter-wave communications, cognitive radio, and coordinated wireless communications. He was a recipient of the UCLA Dissertation Year Fellowship in 2018 and Qualcomm

Innovation Fellowship in 2019.



Danijela Cabric received the M.S. degree in electrical engineering from the University of California, Los Angeles, Los Angeles, CA, USA, in 2001, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, Berkeley, CA, USA, in 2007. He is a Professor in electrical and computer engineering with the University of California, Los Angeles. Her research interests include novel radio architecture, signal processing, and networking techniques for cognitive radio, 5G, and massive MIMO systems. She was the recipient of the Samueli

Fellowship in 2008, the Okawa Foundation Research Grant in 2009, Hellman Fellowship in 2012, and the National Science Foundation Faculty Early Career Development (CAREER) Award in 2012. She is now an Associated Editor for the IEEE TRANSACTIONS OF COGNITIVE COMMUNICATIONS AND NETWORKING and the IEEE TRANSACTIONS OF WIRELESS COMMUNICATIONS. She is an IEEE ComSoc Distinguished Lecturer.