Fast Beam Search and Refinement for Millimeter-Wave Massive MIMO Based on Two-Level Phased Arrays

Song Noh, *Member, IEEE*, Jiho Song, *Member, IEEE*, and Youngchul Sung, *Senior Member, IEEE*

***Abstract*— In this paper, a new method of fast beam search and refinement is proposed for millimeter-wave hybrid beamforming systems. The proposed method is based on a two-level phased array approach in which the first-level phased array is formed by actual analog-domain subarrays, and the second-level phased array is virtually formed by the outputs of the first-level subarray phased combiners. Exploiting the fact that the overall beam pattern of the two-level array is the product of the beam patterns of the two levels, the proposed method searches angle-of-arrivals by sweeping coarse-resolution analog training beams at the first-level phased array and matching the first-level subarray outputs with fine-resolution digital training beams by parallel fast Fourier transform (FFT) filtering. In this way, the overall beam search resolution of the proposed method is given by the fine resolution of the second-level virtual array only with the overhead of coarse beam sweeping at the first-level subarrays. Hence, the proposed method provides a very efficient way of beam training and refinement. The performance of the proposed method including the directional ambiguity, beam search latency, and computational complexity is analyzed. Numerical results show the effectiveness of the proposed method.**

***Index Terms*— Millimeter wave, massive MIMO, hybrid beamforming, beam search, beam refinement, grating lobes.**

I. Introduction

MILLIMETER-WAVE (mmWave) communication is one of the key technologies to support extremely high data rates in 5G-and-beyond wireless systems, but radio channels in the mmWave band are not so favorable as compared to those in the previously-used lower cellular frequency bands. Radio propagation in the mmWave band is highly-directional with large path loss and very few multi-paths. In order to overcome such large path loss and to realize reliable communication in the mmWave band, highly-directional beamforming is necessary for mmWave communication. Fortunately, highly-directional beamforming is feasible with large-scale antenna arrays which can be packed into small areas due to the short wavelength in the mmWave band. However, such highly-directional beamforming requires accurate knowledge of angle-of-arrivals (AoAs) and angle-of-departures (AoDs) of the mmWave propagation channel paths to achieve beam alignment between the transmitter and the receiver. Typically, AoA and/or AoD search for beam alignment is performed during the beam training before data transmission, and the beam training overhead is a major burden that reduces the spectral efficiency in mmWave systems [1]. For example, the beam management procedure of 3GPP includes beam sweeping that covers the entire angular space for initial link setup or link failure recovery [1]. In general, the beam training overhead scales with the level of spatial resolution and the dynamics of the propagation environment [2], and thus the beam training overhead increases with high beamforming gain with narrow beamwidth. Hence, the invention of fast and efficient beam training methods for high-directional beamforming is one of the key challenges to realize spectrally-efficient mmWave communication systems. In this paper, among several beamforming architectures [3], [4], we consider the hybrid beamforming architecture, which is considered to be a practical trade-off solution to mmWave beamforming from the perspective of hardware complexity and performance in-between full digital beamforming and full analog beamforming [5], [6], and tackle the problem of fast beam search and refinement in uplink beam training for the hybrid beamforming architecture.

A. Contributions of the Paper

Most prior work on hybrid beamforming in mmWave systems focused on achieving spectral-efficient transmission compared to full digital beamforming. On the other hand, this work mainly studies the impact of the number of RF chains in a hybrid beamforming framework on beam search to present highly-efficient beam training design. The contributions of the paper are summarized as follows.

Manuscript received July 11, 2019; revised December 1, 2019 and April 22, 2020; accepted June 16, 2020. Date of publication July 7, 2020; date of current version October 9, 2020. This work was supported in part by Electronics and Telecommunications Research Institute (ETRI) grant funded by ICT R&D program of IITP (No. 2018-0-00218, Speciality Laboratory for Wireless Backhaul Communications based on Very High Frequency) and in part by Incheon National University Research Grant (No. 2020-0265). The associate editor coordinating the review of this article and approving it for publication was N. Gonzalez-Prelcic. *(Corresponding author: Youngchul Sung.)*

Song Noh is with the Department of Information and Telecommunication Engineering, Incheon National University, Incheon 22012, South Korea (e-mail: songnoh@inu.ac.kr).

Jiho Song is with the School of Electrical Engineering, University of Ulsan, Ulsan 44610, South Korea (e-mail: jihosong@ulsan.ac.kr).

Youngchul Sung is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: ysung@ee.kaist.ac.kr).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2020.3004916

- We propose a new beam search and refinement framework for uplink beam training based on a two-level phased array approach. In the proposed approach, the first-level phased array is formed by actual analog-domain subarrays and the second-level phased array is virtually formed by the outputs of the first-level subarray phased combiners. The proposed method achieves high-resolution low-complexity AoA search by sweeping coarse-resolution analog training beams at the first-level phased array and matching the first-level subarray outputs with fine-resolution pipelined FFT filtering.
- We show that the overall beam pattern of the two-level array is the product of the beam patterns of the first-level array and the second-level virtual array, and the overall beam search resolution of the proposed method is given by the fine resolution of the second-level virtual array. Since only coarse analog training beams are swept while the second-level digital beam training for each analog beam is done in a parallel manner by digital processing, the overall beam training overhead in terms of signalling in space is the same as that of the coarse analog beam sweeping at the first-level subarrays.
- We provide error analysis on the proposed beam search method in terms of the directional ambiguity. We present a comparison of beam search latency and computational complexity of several beam search methods.
- We present extensions of the proposed method to an interleaved linear subarrays, multi-user case, and joint transmit and receive beam search.

B. Related Works

We here cover only the most relevant references to our work from the perspective of beam training overhead reduction due to space limitation, although there exist many works on hybrid beamforming from other perspectives. One line of research to reduce the antenna training time and the associated overhead is hierarchical beamforming. In this line of research, much of the prior work investigates the design of variable beamwidth codebooks [7]–[10]. However, one drawback of the hierarchical beam search is asymmetric coverage of wide and narrow beams. This makes hierarchical beam search complicated because multiple transmissions of wide beams increase the beam alignment delay, and allocating more power to transmit wide beams and less power to transmit narrow beams requires a wide dynamic range for power amplifiers. The most related work to our approach is previous subarray-based methods. In the previous subarray-based beamforming methods, in order to reduce the beam training overhead, the time delay from different phased subarrays is exploited for beam search per overlapped subarray [11] and per non-overlapped subarray [12]. In [12]–[18], the authors present cross-correlation-based beam search algorithms based on subarrays [12]–[14] and subspace projection-based AoA estimation algorithms [15], [16], under the assumption that the same phase-shift is applied to each analog subarray. The inherent phase ambiguity in AoA detection associated with subarray-based methods is resolved by sounding additional training symbols [12], [16], post-processing in the frequency domain [13], noise subspace projection [15], and iterative update [17], [18]. In [19]–[21], the authors consider using different phase-shift to each analog subarray in order to reduce the beam search latency, and the gains of consecutive subarrays are coherently combined by calibrating the signs of cross correlations among the outputs of subarrays.

The main difference between the proposed method and the previous subarray-based methods is that in the proposed method we form a second-level phased array based on the analog subarray outputs, and this enables us to use the narrow grating lobes of the undersampled second-level virtual phased array and to exploit the combined beam pattern of the first- and second-level phased arrays for efficient unambiguous beam search without additional training symbols or complicated processing such as serial interference cancellation used in the previous methods.

C. Notations and Organization

Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^\dagger$ indicate the transpose, Hermitian transpose, and Moore-Penrose inverse of $\mathbf{A}$, respectively. $\mathrm{diag}(\cdot)$ denotes a diagonal matrix. For a vector $\mathbf{a}$, $\|\mathbf{a}\|$ represents vector ℓ_2-norm. $\mathbf{0}_n$ and $\mathbf{1}_n$ are the zero vector of length n and the one vector of length n, respectively. $\mathbf{e}_n(i)$ denotes the i-th column vector of the identity matrix of size $n \times n$. $\mathbf{I}_n$ is the identity matrix of size $n \times n$. $\mathbb{E}\{\cdot\}$ denotes statistical expectation. $\mathbf{a} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that random vector $\mathbf{a}$ is complex circularly-symmetric Gaussian distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbb{N}$ and $\mathbb{C}$ denote the sets of natural numbers and complex numbers, respectively. The symbol $\setminus$ and $\otimes$ denote the set minus operation and the Kronecker product, respectively. $j := \sqrt{-1}$.

This paper is organized as follows: The system model is described in Section II. Section III describes the proposed beam search framework. Analysis on the proposed method is provided in Section IV. Extensions of the proposed method to an interleaved linear arrays and the cases of multiples users and joint transmit and receive beam search are discussed in Section V. Numerical results are provided in Section VI, followed by conclusions in Section VII.

II. System Model

A. System Setup

In this paper, we consider a wideband hybrid beamforming system composed of a base station (BS) equipped with a uniform linear array (ULA) of N elements and a single-antenna user, and focus on the beam search at the BS in uplink training.[1] We assume that the ULA at the BS has half-wavelength inter-element spacing $d = \lambda/2$, where λ is the carrier wavelength, and has a multi-subarray architecture. That is, the N antenna elements are grouped into M analog phased subarrays so that each analog phased subarray has

[1]The proposed beam search algorithm can be extended to two-dimensional uniform planar subarrays, and the trained beams in the uplink can be used directly in the downlink in the TDD case.

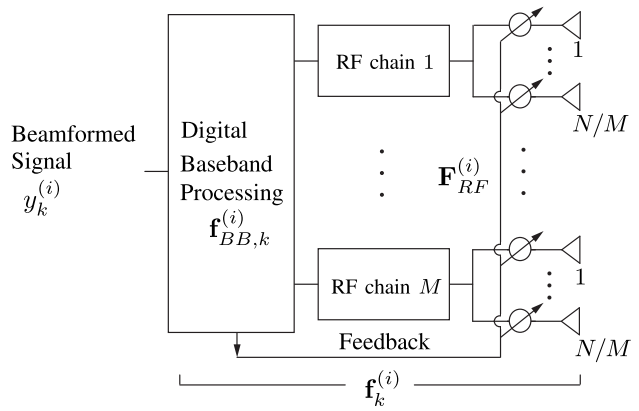


Fig. 1. The considered subarray-based hybrid beamforming architecture.

N/M antennas elements and is attached to a radio-frequency (RF) chain after phased combining.[2] Then, the overall M RF chains are connected to a digital processing unit, as shown in Fig. 1. We assume that M is a divisor of N with $M \geq 2$, i.e., $N/M \in \mathbb{N}$. Thus, the spacing between two adjacent subarray centers is $d\frac{N}{M} = \frac{\lambda}{2}\frac{N}{M}$.

We assume orthogonal frequency-division multiplexing (OFDM) for wideband processing. Under the assumption of OFDM with the discrete Fourier transform (DFT) size K, the received signal at the k-th subcarrier of the i-th OFDM symbol is processed by the hybrid combiner $\mathbf{f}_k^{(i)} = \mathbf{F}_{RF}^{(i)}\mathbf{f}_{BB,k}^{(i)}$ with unit power constraint $\|\mathbf{f}_k^{(i)}\|^2 = 1$, where $\mathbf{F}_{RF}^{(i)} \in \mathbb{C}^{N\times M}$ is the wideband analog beamformer shared by all subcarriers and $\mathbf{f}_{BB,k}^{(i)} \in \mathbb{C}^M$ is the digital beamformer in the baseband for the k-th subcarrier. We assume block-wise transmission in which each transmission block of N_B OFDM symbols consists of the initial N_P pilot symbols and the following $N_B - N_P$ data symbols. Then, the received signal at the k-th subcarrier of the i-th symbol after hybrid beamforming can be expressed as

$$y_k^{(i)} = \mathbf{f}_{BB,k}^{(i)H}\left(\mathbf{F}_{RF}^{(i)H}\left(\mathbf{h}_k x_k^{(i)} + \mathbf{v}_k^{(i)}\right)\right) \quad (1)$$

$$= \mathbf{f}_{BB,k}^{(i)H}\left(\mathbf{F}_{RF}^{(i)H}\mathbf{h}_k x_k^{(i)} + \mathbf{n}_k^{(i)}\right) \quad (2)$$

$$= \mathbf{f}_k^{(i)H}\mathbf{h}_k x_k^{(i)} + z_k^{(i)}, \quad (3)$$

where $k \in \{1,\cdots,K\}$, $i \in \{1,\cdots,N_B\}$, $x_k^{(i)}$ is the transmit symbol with power constraint $\mathbb{E}\{|x_k^{(i)}|^2\} = \rho$, $\mathbf{h}_k \in \mathbb{C}^N$ is the uplink channel vector, $\mathbf{v}_k^{(i)} \in \mathbb{C}^N$ is independent and identically distributed (i.i.d.) zero-mean Gaussian noise with $\mathbf{v}_k^{(i)} \sim \mathcal{CN}(\mathbf{0}_N, \sigma_n^2\mathbf{I}_N)$, $\mathbf{n}_k^{(i)} = \mathbf{F}_{RF}^{(i)H}\mathbf{v}_k^{(i)}$ with $\mathbf{F}_{RF}^{(i)H}\mathbf{F}_{RF}^{(i)} = \mathbf{I}_M$, and $z_k^{(i)} := \mathbf{f}_{BB,k}^{(i)H}\mathbf{n}_k^{(i)} \sim \mathcal{CN}(0,\sigma_n^2)$ with $\|\mathbf{f}_{BB,k}^{(i)}\|^2 = 1$. During the training period (i.e., $1 \leq i \leq N_P$), $x_k^{(i)}$ is the uplink pilot symbol and $\mathbf{F}_{RF}^{(i)}$ is the receive antenna weight matrix to scan different spatial directions for receiver beam training. Once the best $\mathbf{F}_{RF}^{(i)\star}$ is found, the digital precoder is designed to optimize the desired performance criterion, e.g., as in [22]

$$\mathbf{f}_{BB,k}^{(i)\star} = \tilde{\mathbf{h}}_k^H\left(\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^H + (\sigma_n^2/\rho)\|\mathbf{f}_{BB,k}^{(i)}\|^2\right)^{-1}, \quad (4)$$

where $\tilde{\mathbf{h}}_k = \mathbf{F}_{RF}^{(i)H}\mathbf{h}_k$. During the data transmission period (i.e., $N_P < i \leq N_B$), the receiver-side processing of the data symbol $x_k^{(i)}$ is based on the hybrid beamformer $\mathbf{F}_{RF}^{(i)\star}\mathbf{f}_{BB,k}^{(i)\star}$ obtained during the training period. From (3), the received signal-to-noise ratio (SNR) with the obtained hybrid combiner is expressed as

$$SNR_k = \frac{\rho}{\sigma_n^2}\left|\mathbf{f}_k^{(i)H}\mathbf{h}_k\right|^2. \quad (5)$$

From here on, the superscript (i) representing the symbol index will be omitted if unnecessary.

Under the assumed phased subarray architecture for analog beamforming, the m-th column vector $\mathbf{f}_{RF,m}$ of $\mathbf{F}_{RF}$ is given by

$$\mathbf{f}_{RF,m}(\psi_m) = \left[\mathbf{0}_{(m-1)\frac{N}{M}}^T, \sqrt{\frac{M}{N}}\mathbf{u}_{\frac{N}{M},d}^T(\psi_m), \mathbf{0}_{(M-m)\frac{N}{M}}^T\right]^T, \quad (6)$$

where $1 \leq m \leq M$, ψ_m is the normalized look-angle of the m-th phased subarray, and $\mathbf{u}_{L,d}(\cdot) \in \mathbb{C}^L$ is the array response vector given by

$$\mathbf{u}_{L,d}(\varphi) = \left[1, e^{j\frac{2\pi d}{\lambda}\varphi}, \cdots, e^{j\frac{2\pi d}{\lambda}\varphi(L-1)}\right]^T. \quad (7)$$

Here, d is the inter-element spacing assumed to be $d = \lambda/2$, and the normalized angle φ is defined as $\varphi = \cos(\phi) \in [-1,1)$ for the one-sided physical azimuth angle $\phi \in (0,\pi]$. There is a one-to-one mapping between φ and ϕ in the visible region $-1 \leq \varphi < 1$ [23, Ch. 22]. Note that due to the subarray structure, $\mathbf{f}_{RF,m}(\psi_m)$ has N/M non-zero elements corresponding to the m-th subarray.

B. Channel Model

We adopt the sparse geometry-based channel model to incorporate directional beams in mmWave communication [24], [25]. Under this model, the multi-tap channel vector $\boldsymbol{\eta}(\nu)$ for ULA in time domain is given by

$$\boldsymbol{\eta}(\nu) = \sum_{\ell=1}^{N_{cp}}\alpha_\ell p(\nu T_s - \tau_{\ell,\nu})\mathrm{diag}(\mathbf{r}(\theta_\ell))\mathbf{u}_{N,d}(\theta_\ell), 0 \leq \nu < D, \quad (8)$$

for the ν-th delay tap, where α_ℓ, $\tau_{\ell,\nu}$, and θ_ℓ denote the gain, delay, and AoA of the ℓ-th path, respectively; N_{cp} is the number of paths; $p(\cdot)$ represents the combination of transmit and receive pulse-shaping filters; and $\mathbf{r}(\cdot) \in \mathbb{C}^N$ denotes the radiation pattern assumed to be omnidirectional, i.e., $\mathbf{r}(\cdot) = \mathbf{1}_N$. Then, the channel vector $\mathbf{h}_k$ in frequency domain at the k-th subcarrier after DFT is given by

$$\mathbf{h}_k = \sum_{\ell=1}^{N_{cp}}\alpha_\ell g_{\ell,k}\mathbf{u}_{N,d}(\theta_\ell), \quad (9)$$

[2]In the fully-connected hybrid beamforming case, we turn on only the corresponding phase shifters to form a set of M subarrays in the beam training period. After AoA identification, all antenna elements per each RF chain can be used for directional beamforming to the identified AoA.

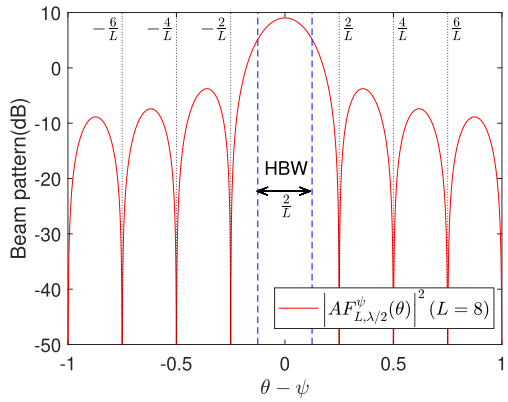


Fig. 2. $|AF^{\psi}_{L,\delta}(\theta)|^2$ with critical sampling $d = \lambda/2$: $L = 8$ in this example.

where $g_{\ell,k} = \sum_{\nu=0}^{D-1} p(\nu T_s - \tau_{\ell,\nu}) e^{-j2\pi \frac{k\nu}{K}}$. Eq. (9) can be rewritten in matrix form as

$$\mathbf{h}_k = \mathbf{U}\text{diag}(\alpha_1, \cdots, \alpha_{N_{cp}})\mathbf{g}_k, \tag{10}$$

where $\mathbf{U} = \left[\mathbf{u}_{N,d}(\theta_1), \cdots, \mathbf{u}_{N,d}(\theta_{N_{cp}})\right] \in \mathbb{C}^{N \times N_{cp}}$ and $\mathbf{g}_k = [g_{1,k}, \cdots, g_{N_{cp},k}]^T = \left(\mathbf{D}(k,:)[\mathbf{p}_1, \cdots, \mathbf{p}_{N_{cp}}]\right)^T$. Here, $\mathbf{D}(k,:)$ denotes the k-th row of the K-point DFT matrix $\mathbf{D}$, and $\mathbf{p}_\ell = [p(-\tau_{\ell,0}), p(T_s - \tau_{\ell,1}), \cdots, p((D-1)T_s - \tau_{\ell,D-1})]^T \in \mathbb{C}^D$.

Definition 1 (Array Factor): The array factor (AF) for ULA is defined by

$$\begin{aligned} AF^{\psi}_{L,\delta}(\theta) :&= \frac{1}{\sqrt{L}} \mathbf{u}^H_{L,\delta}(\psi)\mathbf{u}_{L,\delta}(\theta) \\ &= e^{j\frac{\pi\delta}{\lambda}(\theta-\psi)(L-1)} \frac{\sin\left(L\frac{\pi\delta}{\lambda}(\theta-\psi)\right)}{\sqrt{L}\sin\left(\frac{\pi\delta}{\lambda}(\theta-\psi)\right)}. \end{aligned} \tag{11}$$

where θ is the AoA and ψ is the look-angle. Here, the second part of the right-hand side (RHS) of (11) is known as the Dirichlet kernel [26].

Then, the received signal output at the m-th subarray after analog phased combining (excluding receive noise) can be expressed as

$$\begin{aligned} \tilde{h}^{\psi_m}_k(m) :&= \left(\mathbf{f}_{RF,m}(\psi_m)\right)^H \mathbf{h}_k \\ &= \sum_{\ell=1}^{N_{cp}} \alpha_\ell g_{\ell,k} e^{j\pi\theta_\ell \frac{N}{M}(m-1)} AF^{\psi_m}_{N/M,\lambda/2}(\theta_\ell). \end{aligned} \tag{12}$$

As seen in (12), the AF represents the strength of the received signal as a function of AoA and look-angle and hence it determines the beam pattern. An example of $d = \lambda/2$ and $L = 8$ is shown in Fig. 2. In the case of critical sampling ($d = \lambda/2$) with L elements, we have a single main lobe, and the half of the base of the main beamwidth (HBW) is defined the angle width from the main lobe center to the main lobe null point, given by $2/L$, as shown in Fig. 2 [23]. The HBW is not equal to but close to the half-power beamwidth. Since it is analytically simple, this definition of HBW will be used in this paper.

III. The Proposed Beam Search Framework

In this section, we present our beam search method based on a virtual two-level phased array for the subarray-based hybrid beamforming framework.

A. Beam Search and Refinement Based on Two-Level Phased Arrays

In the proposed beam search and refinement method, we first set the look-angle values $\psi_1, \cdots, \psi_M$ for the M analog phased subarrays in (6) to be equal, i.e., $\psi_1 = \cdots = \psi_M = \psi$.[3] Next, the phase-combined output signals of the M analog subarrays at the k-th subcarrier are further combined by the digital combiner, as seen in (2). Regarding the design of digital combiner for the given phased analog subarrays, we have the following proposition:

Proposition 1: Suppose that we aim to detect each single AoA expressed in terms of the channel vector as $\mathbf{h}_k = \alpha_1 \ g_{1,k}\mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1)$. Then, an optimal hybrid combiner that maximizes the received SNR (5) is given by spatial matched filter, i.e., $\mathbf{f}_{RF,m} = \mathbf{e}_M(m) \otimes \sqrt{\frac{M}{N}}\mathbf{u}_{\frac{N}{M},\frac{\lambda}{2}}(\theta_1)$ and $\mathbf{f}_{BB,k} = \frac{1}{\sqrt{M}}\mathbf{u}_{M,\frac{\lambda N}{2M}}(\theta_1)$.

Proof: From (5), the received SNR is given by

$$SNR_k = \frac{\rho}{\sigma_n^2}\left|\mathbf{f}_k^H\mathbf{h}_k\right|^2 = \frac{\rho}{\sigma_n^2}|\alpha_1 g_{1,k}|^2 \left|\mathbf{f}_k^H\mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1)\right|^2 \tag{13}$$

$$\overset{(a)}{\leq} \frac{\rho}{\sigma_n^2}|\alpha_1 g_{1,k}|^2 \left\|\mathbf{f}_k\right\|^2 \left\|\mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1)\right\|^2 = \frac{\rho}{\sigma_n^2}|\alpha_1 g_{1,k}|^2 N \tag{14}$$

where the inequality (a) holds by the Cauchy-Schwartz inequality, and the SNR upper bound is achieved when $\mathbf{f}_k = c\mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1)$ for some constant c due to the properties of the Cauchy-Schwartz inequality. On the other hand, when the RF beamformer $\mathbf{f}_{RF,m} = \mathbf{e}_M(m) \otimes \sqrt{\frac{M}{N}}\mathbf{u}_{\frac{N}{M},\frac{\lambda}{2}}(\theta_1)$ is combined with digital beamformer $\mathbf{f}_{BB,k} = \frac{1}{\sqrt{M}}\mathbf{u}_{M,\frac{\lambda N}{2M}}(\theta_1)$, the combined beamformer becomes

$$\begin{aligned} \mathbf{f}_k = \mathbf{F}_{RF}\mathbf{f}_{BB,k} &= \left(\mathbf{I}_M \otimes \sqrt{\frac{M}{N}}\mathbf{u}_{\frac{N}{M},\frac{\lambda}{2}}(\theta_1)\right)\frac{1}{\sqrt{M}}\mathbf{u}_{M,\frac{\lambda N}{2M}}(\theta_1) \\ &= \left(\frac{1}{\sqrt{M}}\mathbf{u}_{M,\frac{\lambda N}{2M}}(\theta_1) \otimes \sqrt{\frac{M}{N}}\mathbf{u}_{\frac{N}{M},\frac{\lambda}{2}}(\theta_1)\right) \\ &= \frac{1}{\sqrt{N}}\mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1). \end{aligned} \tag{15}$$

Note that the combined beamformer is the beamformer achieving the SNR upper bound (14). ■

Based on Proposition 1, in order to maximize the received SNR for AoA detection, we set the digital combiner as another phased array combiner with look-angle ξ:

$$\mathbf{f}_{BB,k}(\xi) := \frac{1}{\sqrt{M}}\left[1, e^{j\pi\xi\frac{N}{M}}, \cdots, e^{j\pi\xi\frac{N}{M}(M-1)}\right]^T \in \mathbb{C}^M. \tag{16}$$

[3]The proposed algorithm can be extended to the case in which distinct analog beam look-angle values are used at the first-level subarrays. However, this case leads to a biased estimate of AoA.

The second-level combiner $\mathbf{f}_{BB,k}(\xi)$ looks at the normalized angle ξ with a *virtual* array of M elements with spacing $\frac{\lambda}{2}\frac{N}{M}$, which is the distance between two adjacent analog subarray centers. Note that for quantized values of ξ, the digital combiner can be implemented simply by using DFT. In the remainder of the paper, we will use the term "the first-level subarray" to refer to each of M subarrays of N/M elements with inter-spacing $\lambda/2$ and the term "the second-level virtual array" to refer to the M outputs of the M first-level subarrays with inter-spacing $\frac{\lambda}{2}\frac{N}{M}$.

The output of the second-level virtual phased array combiner is written as

$$y_k^{\psi,\xi} := \mathbf{f}_{BB,k}(\xi)^H \left(\tilde{\mathbf{h}}_k^{\psi} x_k + \mathbf{n}_k\right) = \tilde{h}_k^{\psi,\xi} x_k + z_k, \quad (17)$$

where $\tilde{\mathbf{h}}_k^{\psi} = \mathbf{F}_{RF}^H \mathbf{h}_k = [\tilde{h}_k^{\psi}(1), \cdots, \tilde{h}_k^{\psi}(M)]^T \in \mathbb{C}^M$ with $\tilde{h}_k^{\psi}(m)$ defined in (12), and $z_k \sim \mathcal{CN}(0, \sigma_n^2)$. In (17), $\tilde{h}_k^{\psi,\xi} := \mathbf{f}_{BB,k}(\xi)^H \tilde{\mathbf{h}}_k^{\psi}$ represents the effective channel gain at the k-th subcarrier obtained jointly by the first-level phased subarray with look-angle ψ and the second-level virtual array with look-angle ξ, and is expressed as

$$\tilde{h}_k^{\psi,\xi} = \textstyle\sum_{\ell=1}^{N_{cp}} \alpha_\ell g_{\ell,k} AF_{N/M,\lambda/2}^{\psi}(\theta_\ell) AF_{M,\frac{\lambda N}{2M}}^{\xi}(\theta_\ell), \quad (18)$$

where $AF_{N/M,\lambda/2}^{\psi}(\theta_\ell)$ and $AF_{M,\frac{\lambda N}{2M}}^{\xi}(\theta_\ell)$ represent the AFs of the first-level phased subarray and the second-level virtual phased array, respectively.

Proposition 2: When the channel has a single AoA, i.e., $\mathbf{h}_k = \alpha_1 g_{1,k} \mathbf{u}_{N,\frac{\lambda}{2}}(\theta_1)$ in (9), the following holds for the two-level phased array: $\left|\tilde{h}_k^{\psi,\xi}\right|^2 \le \left|\tilde{h}_k^{\psi,\theta_1}\right|^2 \le \left|\tilde{h}_k^{\theta_1,\theta_1}\right|^2 = N\left|\alpha_1 g_{1,k}\right|^2$.

Proof: From (18), we have

$$\begin{aligned}
&\left|\tilde{h}_k^{\psi,\xi}\right|^2 / |\alpha_1 g_{1,k}|^2 \\
&= \left|AF_{N/M,\lambda/2}^{\psi}(\theta_1) AF_{M,\frac{\lambda N}{2M}}^{\xi}(\theta_1)\right|^2 \\
&= \left(\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}(\theta_1-\psi)\right)}{\sqrt{\frac{N}{M}}\sin\left(\frac{\pi}{2}(\theta_1-\psi)\right)} \frac{\sin\left(M\frac{\pi N}{2M}(\theta_1-\xi)\right)}{\sqrt{M}\sin\left(\frac{N}{M}\frac{\pi}{2}(\theta_1-\xi)\right)}\right)^2 \quad (19) \\
&\overset{(a)}{\le} \left(\frac{M}{\sqrt{N}} \frac{\sin\left(\frac{N}{M}\frac{\pi}{2}(\theta_1-\psi)\right)}{\sin\left(\frac{\pi}{2}(\theta_1-\psi)\right)}\right)^2 \\
&= \left|AF_{N/M,\lambda/2}^{\psi}(\theta_1) AF_{M,\frac{\lambda N}{2M}}^{\theta_1}(\theta_1)\right|^2 = \left|\tilde{h}_k^{\psi,\theta_1}\right|^2 / |\alpha_1 g_{1,k}|^2 \\
&\overset{(b)}{\le} \left|AF_{N/M,\lambda/2}^{\theta_1}(\theta_1) AF_{M,\frac{\lambda N}{2M}}^{\theta_1}(\theta_1)\right|^2 = \left|\tilde{h}_k^{\theta_1,\theta_1}\right|^2 / |\alpha_1 g_{1,k}|^2 \\
&= N = \left|AF_{N,\lambda/2}^{\theta_1}(\theta_1)\right|^2. \quad (20)
\end{aligned}$$

In step (a), we used the fact that $\frac{\sin\left(M\frac{\pi N}{2M}(\theta_1-\psi)\right)}{\sqrt{M}\sin\left(\frac{N}{M}\frac{\pi}{2}(\theta_1-\psi)\right)} \le \sqrt{M} = |AF_{M,\frac{\lambda N}{2M}}^{\theta_1}(\theta_1)|$ for the second-level virtual array, and in step (b), we used the fact that $\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}(\theta_1-\psi)\right)}{\sqrt{\frac{N}{M}}\sin\left(\frac{\pi}{2}(\theta_1-\psi)\right)} \le \sqrt{N/M} = |AF_{N/M,\frac{\lambda}{2}}^{\theta_1}(\theta_1)|$ for the first-level subarrays. ■

Proposition 2 states that the two-level phased array architecture achieves the full array combining gain N, when both the look-angles of the first-level subarrays and the second-level virtual array hit the true AoA, i.e., $\psi = \xi = \theta_1$. Thus, the remaining task for beam training is to determine the look-angles ψ and ξ of the two-level arrays during the training period. For beam training and search, we here consider quantized beam search. Note that *the search for the look-angle ψ of the first-level subarrays should be done in analog domain with actual beam sweeping, but the search for the look-angle ξ of the second-level virtual array for a given ψ is done in digital domain. Hence, the search for multiple values of ξ in a parallel manner is possible.* Considering these facts, we assume a coarse spatial resolution for the first-level actual analog beam sweeping and a fine spatial resolution for the second-level parallel beam search. Thus, we design the training beam (or look-angle) sets B_{I_1} for ψ and B_{I_1,I_2} for ξ as follows. First, we design B_{I_1} as

$$B_{I_1} = \left\{ \psi(i_1) = -1 + \frac{2}{I_1}\left(i_1 - \frac{1}{2}\right) \middle| \, i_1 = 1, \cdots, I_1 \right\}, \quad (21)$$

for some integer I_1. The I_1 quantized angles in B_{I_1} are evenly spaced over the entire normalized angle value range $[-1, 1)$ with spacing $2/I_1$. Next, we design B_{I_1,I_2} as

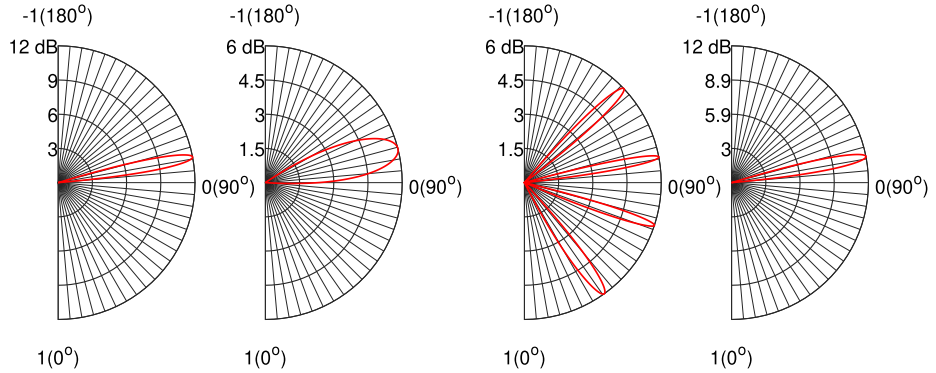
$$B_{I_1,I_2} = \left\{ \xi(i_1,i_2) = \psi(i_1) + \frac{2\left(i_2 - 1 - (I_2-1)/2\right)}{I_1\ I_2} \middle| \, \psi(i_1) \in B_{I_1},\ \begin{matrix} i_1 = 1, \cdots, I_1 \\ i_2 = 1, \cdots, I_2 \end{matrix} \right\}, \quad (22)$$

where $\xi(i_1 - 1, I_2) + \frac{2}{I_1 I_2} = \xi(i_1, 1) \le \xi(i_1, i_2) \le \xi(i_1, I_2) = \xi(i_1+1, 1) - \frac{2}{I_1 I_2}$. As seen in (22), the quantized angles in B_{I_1,I_2} are evenly spaced over $[-1, 1)$ with spacing $2/(I_1\ I_2)$. Note that B_{I_1,I_2} has a finer resolution than B_{I_1}, and $\{\xi(i_1, 1), \cdots, \xi(i_1, I_2)\}$ is a set of I_2 finely quantized steering angles around $\psi(i_1)$ for given $i_1 \in \{1, \cdots, I_1\}$. Here, we set I_1 and I_2 as follows:

$$I_1 \ge N/M \quad \text{and} \quad I_2 \ge M. \quad (23)$$

$I_1 \ge N/M$ implies that the number of analog training beams is at least as many as the number of orthogonal beams of each analog subarray with N/M elements, and $I_2 \ge M$ ensures that the fine training beam spacing $2/(I_1 I_2)$ is less than or equal to the HBW $2/N$ of one of the dominant lobes of the second-level array. (This will become clear shortly.)

The main advantage of the considered virtual two-level beam search approach is that although the actually-swept analog training beams have a coarse resolution, the overall resolution of the two-level array approach follows the fine resolution of the second-level virtual array. This is because the overall training beam pattern of the two-level phased array approach is given by the product of the array factors $AF_{N/M,\lambda/2}^{\psi}(\theta)$ and $AF_{M,\frac{\lambda N}{2M}}^{\xi}(\theta)$ of the first-level and second-level phased arrays, as seen in (17) and (18). Let us explain this with an example shown in Fig. 3. In this example, we consider the case of total $N = 16$ receive antenna elements at the BS. First, we consider the beam pattern of a single-level ULA with $N = 16$ and critical spatial sampling $d = \lambda/2$, which can be obtained by analog beamforming with $N = 16$. Fig. 3(a) shows the beam pattern of a single-level ULA with $N = 16$ and $d = \lambda/2$ looking at near $105°$.



(a) A beam pattern at the N-antenna ULA with spacing $d = \frac{\lambda}{2}$

(b) A beam pattern at the N/M-antenna first-level subarray $(I_1 = 4, i_1 = 2)$

(c) A beam pattern at the M-antenna second-level array $(i_1 = 2, i_2 = 3)$

(d) A beam pattern at the N-antenna two-level phased array $(i_1 = 2, i_2 = 3)$

Fig. 3. Beam patterns for $N = 16$, $M = 4$, $I_1 = 4$, $I_2 = 4$, $i_1 = 2$, and $i_2 = 3$: (a) $\left|AF^{\xi(i_1,i_2)}_{N,\lambda/2}(\theta)\right|^2$, (b) $\left|AF^{\psi(i_1)}_{N/M,\lambda/2}(\theta)\right|^2$, (c) $\left|AF^{\xi(i_1,i_2)}_{M,\lambda N/(2M)}(\theta)\right|^2$, and (d) $\left|AF^{\psi(i_1)}_{N/M,\lambda/2}(\theta)AF^{\xi(i_1,i_2)}_{M,\lambda N/(2M)}(\theta)\right|^2$. (Patterns below 0 dB are not drawn.)

The HBW of the beam pattern in this case is narrow and given by $2/N = 1/8$ [23]. Next, consider the two-level approach for the same $N = 16$ with $M = 4$, $I_1 = 4$, and $I_2 = 4$. Fig. 3(b) shows the beam pattern of each of the four $(M = 4)$ first-level subarrays (each having 4 antenna elements) with $\psi_1 = \cdots = \psi_4 = \psi(2) \in B_{I_1}$ and $d = \lambda/2$. In this case, $N/M = 4$ antenna elements in each subarray critically sample the space, and hence we see a single main lobe appearing in the visible region in Fig. 3(b). This single main lobe has the HBW $2M/N = 2 \cdot 4/16 = 1/2$ which is $M = 4$ times wider than the HBW in Fig. 3(a). Note that there exists 6 dB (i.e., factor of $1/M$) beamforming gain loss in Fig. 3(b) as compared to Fig. 3(a). Fig. 3(c) shows the beam pattern of the second-level virtual array beam pattern (i.e., the beam pattern of an ULA with four elements and spacing $\frac{\lambda N}{2M}$ $(> \frac{\lambda}{2})$). Note that this second-level virtual array undersamples the space with inter-element spacing $\frac{\lambda N}{2M}$ $(> \frac{\lambda}{2})$. Thus, the corresponding beam pattern has multiple $(N/M = 4)$ dominant peaks consisting of the main and *grating* lobes with inter-beam spacing $2M/N$. The key point is that the HBW of each dominant beam is $2/N$ [23], which is the same as that of the single-level full-array beam pattern in Fig. 3(a). Finally, Fig. 3(d) shows the combined beam pattern obtained by the first-level subarrays and the second-level virtual array for $\xi(i_1, 8)$. Note that the unwanted grating lobes of the second-level virtual array outside the single main lobe of the first-level subarray are suppressed. Hence, the combined beam pattern shows a single narrow main lobe, as shown in Fig. 3(d). Note also that the 6 dB beamforming gain loss in Fig. 3(b) is now restored in Fig. 3(d) due to 6 dB beamforming gain at the second-level virtual array.

Since the HBW of each grating lobe of the second-level beam pattern is narrow as $2/N$, the direction of the second-level beam in Fig. 3(c) can be adjusted with fine granularity within the wide main lobe of Fig. 3(b) by setting the look-angle of the second-level virtual array by digital processing. Exploiting this fact, we illustrate the proposed hybrid beam search framework in Fig. 4 under the setup of $N = 16$, $M = 4$, and $I_1 = I_2 = 4$. In the proposed method, as shown in Fig. 4, all the first-level subarrays each with 4 antenna elements probe the same-look angle $\psi(i_1)$ with critical spatial sampling $d = \lambda/2$ so that the beam pattern of the first-level subarray has only one single main lobe with HBW $2M/N = 2{\cdot}4/16 = 1/2$. In the figure, we used $\psi(2)$ for the first-level look-angle. The analog-combined output signal from each first-level subarray is followed by an RF chain and converted into a digital signal. Then, the $M = 4$ digital signals are fed to four parallel fast Fourier transform (FFT) filters $i_2 = 1, 2, 3, 4$ (implementing the second-level digital combiners) with look-angles $\xi(i_1 = 2, i_2 = 1), \ldots, \xi(i_1 = 2, i_2 = 4)$, respectively, so that we can simultaneously probe $I_2 = 4$ beam directions $\xi(i_1 = 2, i_2 = 1), \ldots, \xi(i_1 = 2, i_2 = 4)$ within the main-lobe beamwidth of the first-level beam pattern. (Note that the digital combining in (16) at the second-level virtual array is basically the DFT-based matched filtering.) As seen in Fig. 3(d), the combined beam pattern of the first-level subarray plus each digital FFT filter is a narrow beam with HBW $2/N = 2/16 = 1/8$. Thus, the proposed method simultaneously scans $I_2 = 4$ narrow beam directions within a wide main-lobe beamwidth of the first-level subarrays in the digital domain. Thus, we only need to scan the analog beam requiring pilot symbol transmission only for $N/M = 16/4 = 4$ times instead of $N = 16$ times.

Based on the above discussion, we present the proposed beam search and refinement method based on the two-level phased array approach in Algorithm 1. In the proposed method, we sweep the coarse analog beams from $\psi(1)$ to $\psi(I_1)$ at the first-level analog subarrays. For each $\psi(i_1) \in B_{I_1}$, we compute the sum received power across all subcarriers:

$$T^{i_1,i_2} := \frac{1}{K}\sum_{k=1}^{K}\left|y_k^{\psi(i_1),\xi(i_1,i_2)}\right|^2, \quad \forall i_2 = 1, \cdots, I_2 \quad (24)$$

to exploit all subcarriers for increased detection reliability, and select $\hat{i}_2(i_1)$ as

$$\hat{i}_2(i_1) = \ argmax_{1\le i_2\le I_2}\ T^{i_1,i_2}, \quad (25)$$

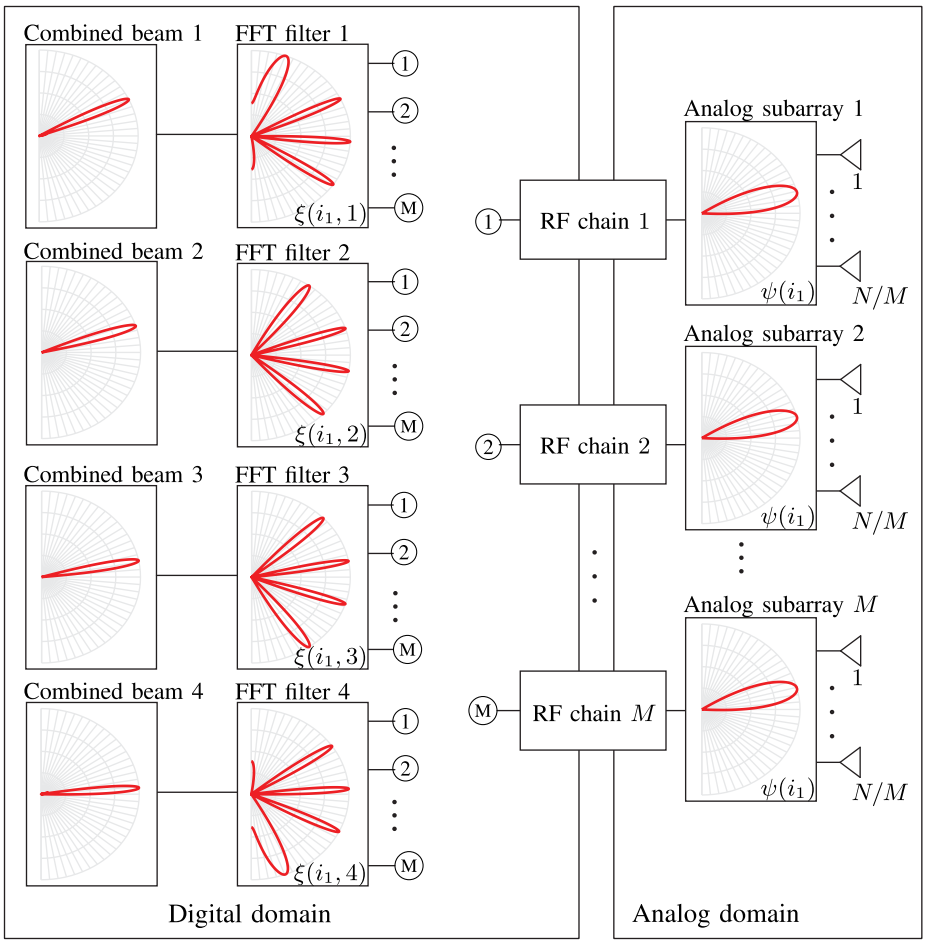


Fig. 4. Illustration of the proposed two-level hybrid beam search framework: $N = 16$, $M = 4$, $I_1 = I_2 = 4$, and $i_1 = 2$. (The output of the ⓜ-th RF chain is connected to the ⓜ-th input of each FFT filter.)

Algorithm 1 The Proposed Beam Search Method

Require: I_1, I_2, N_C where $N_C \leq N_{cp}$

1: **for** $i_1 = 1 : I_1$ **do**
2: Set the first-level subarray look-angle as $\psi(i_1)$.
3: **for** $i_2 = 1 : I_2$ **do**
4: Compute T^{i_1,i_2} in (25)
5: **end for**
6: $\hat{i}_2(i_1) = \ argmax_{1 \leq i_2 \leq I_2} T^{i_1,i_2}$
7: **end for**
8: Select the N_C pairs $(i_1, \hat{i}_2(i_1))$ yielding the N_C largest $T^{i_1,\hat{i}_2(i_1)}$ values.

where $\xi(i_1, i_2) \in B_{I_1,I_2}$. Note that the search (25) for the best beam direction $\xi(i_1, \hat{i}_2(i_1))$ for given $i_1 \in \{1, \cdots, I_1\}$ is done by post-processing and does not require actual phase shift at the analog subarrays. Since the combined beam pattern of the two-level phased array has a single narrow main lobe, the algorithm has the capability to identify multiple AoAs with resolvable angle separation. Hence, we select the N_C AoA estimates with the N_C largest sum received power values where $N_C \leq N_{cp}$. This can be used for multipath receiver combining.

Remark 1 (Interpretation of the Proposed Method): Note that we consider uplink beam training with an ULA of N elements with inter-element spacing $d = \lambda/2$. Suppose that we perform full digital beamforming with N RF chains, and consider N orthogonal beams as the training beams. Then, N orthogonal narrow beam training can be done all at one in parallel by digital processing. On the other hand, suppose that we perform full analog beamforming with a single RF chain. Then, all N orthogonal narrow beams should be swept for training. However, in the case of hybrid beamforming with M subarrays each with N/M elements, as we propose, we can construct the two-level phased array. Then, the beam pattern of the two-level phased array is the product of the first-level and second-level arrays and is similar to that of the phased array of N elements at a single level. Here, first-level coarse analog beams need to be swept but second-level fine beam search can be done by digital processing. Thus, we can achieve the performance of narrow beam training with the sweeping overhead of coarse first-level analog beam sweeping.

Remark 2 (Correction of Analog-Subarray Look-Angle): Once the beam training is performed with fine resolution on ξ, this fine-resolution AoA information can be fed back to the analog subarrays to correct the look-angle of the first-level

analog subarrays (coarsely determined in beam training) to the more accurate look-angle for data reception to increase the received power.

Remark 3 (Beam Refinement): When the channel direction is already known with the coarsely quantized spatial resolution B_{I_1} in (21), only the beam refinement procedure, i.e., Lines 3-6 of Algorithm 1 can be performed.

Remark 4 (Cross/Dual-Polarized Subarrays): Note that the proposed two-level beam search method is based on linear subarrays. In the case that linear subarrays are difficult to construct due to the form factor and subarrays are formed by cross/dual polarization with elements at the same locations, the phase-combined signal from each subarray does not have phase difference and the proposed method is not applicable.

IV. Analysis of the Proposed Method

In this section, we first analyze the unambiguity of the AoA determined by the proposed two-level phased array beam training method in the noiseless case. We then analyze the performance of the proposed method in terms of beam search latency and computational complexity.

A. Unambiguous Beam Detection

In the proposed two-level phased array method, the AF of the second-level virtual array $AF^{\xi}_{M,\frac{\lambda N}{2M}}(\theta_\ell)$ exhibits N/M major peaks composed of the wanted main and unwanted grating lobes at $\xi \in \left\{\theta_\ell + i'\frac{2M}{N} \in [-1,1) : i' \in \mathbb{Z}\right\}$ over the visible region $[-1,1)$ due to the undersampling virtual array spacing $\frac{\lambda}{2}\frac{N}{M}$. (Recall that the normalized angle interval between two adjacent grating lobes is $2M/N$.) This leads to ambiguity in AoA estimation at the second-level virtual array. (Other AoA estimation methods based on the outputs of the analog subarrays show similar ambiguity problems since the subarray spacing is larger than the critical spatial sampling rate $d = \lambda/2$ [23].) That is, even if we determine $\hat{\xi}$ for AoA from the second-level virtual array, not only the angle $\hat{\xi}$ but also other (grating lobe) angles $\hat{\xi} + i'\frac{2M}{N} \in [-1,1)$ for feasible non-zero integer i' are all AoA candidates with the same quality due to the beam pattern structure of the second-level virtual array, if we only consider the second-level array. (Please see Fig. 3(c).) However, the AF of the first-level subarray $AF^{\psi}_{N/M,\lambda/2}(\theta_\ell)$ shows a single dominant peak (i.e., the main lobe) in the visible region $[-1,1)$, and hence the second-level directional ambiguity is resolved without additional training or processing in the proposed two-level phased array method when the overall two-level phased array structure is considered. The following propositions elaborate this quantitatively in the noiseless case.

Proposition 3: Let $\hat{\xi}$ be the main-lobe look-angle of the fine training beam in B_{I_1,I_2} that maximizes the magnitude of the second-level AF. Then, for any grating-lobe look-angle $\xi' = \hat{\xi} + i'\frac{2M}{N} \in [-1,1)$ with non-zero i', we have

$$\left|AF^{\xi'}_{N/M,\lambda/2}(\theta_\ell)\right|^2 \le \left(\frac{\pi}{2}\right)^4 \left(\frac{1}{2I_2-1}\right)^2 \left|AF^{\hat{\xi}}_{N/M,\lambda/2}(\theta_\ell)\right|^2.$$

Hence, the grating-lobe ambiguity is suppressed by the proposed two-level phased array search method when (23) is satisfied with $I_2 \ge 2$.

Proof: With the condition (23) on I_1 and I_2, the fine beam search spacing satisfies $2/(I_1 I_2) \le 2/N$, and the fine beam search spacing is less than the HBW of each of the main and grating lobes of the second-level array. So, the AF of the second-level array $AF^{\xi}_{M,\frac{\lambda N}{2M}}(\theta_\ell)$ is maximized when choosing $\xi \in B_{I_1,I_2}$ that is closest to θ_ℓ. Let this ξ be denoted by $\hat{\xi}$. Define $\epsilon := \theta_\ell - \hat{\xi}$. Then, by the resolution of B_{I_1,I_2} in (22) and the condition (23) on I_1 and I_2, we have

$$|\epsilon| \le \frac{1}{I_1 I_2} \le \frac{M}{N}\frac{1}{I_2} \le \frac{1}{N} \le 1. \tag{26}$$

However, due to the presence of grating lobes, the same maximum value of the AF of the second-level virtual array is attained at $\xi' = \hat{\xi} + i'\frac{2M}{N} \notin \mathcal{R}_{i_1} := \left[\psi(i_1) - \frac{M}{N}, \psi(i_1) + \frac{M}{N}\right]$, where the feasible range for integer i' is given by

$$i' \in \left[\frac{N(-1-\hat{\xi})}{2M}, \frac{N(1-\hat{\xi})}{2M}\right) \setminus \{0\} \subset \left(-\tfrac{N}{M}, \tfrac{N}{M}\right) \setminus \{0\}. \tag{27}$$

Now, consider the AF values of the first-level phased array at ξ' and $\hat{\xi}$. At ξ', the AF of the first-level subarray is given by

$$\begin{aligned}\left|AF^{\xi'}_{N/M,\lambda/2}(\theta_\ell)\right|^2 &= \left(\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}(\epsilon - i'\frac{2M}{N})\right)}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\left(\epsilon - i'\frac{2M}{N}\right)\right)}\right)^2 \\ &= \left(\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}\epsilon\right)}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\left(\epsilon - i'\frac{2M}{N}\right)\right)}\right)^2,\end{aligned}$$

since $|\sin(x - n\pi)| = |\sin(x)|$ for any integer n. Then, we have

$$\begin{aligned}&\left|AF^{\xi'}_{N/M,\lambda/2}(\theta_\ell)\right|^2 \\ &\le \frac{M}{N}\left(\frac{\frac{N}{M}\frac{\pi}{2}\epsilon}{\sin\left(\frac{\pi}{2}\left(\epsilon - i'\frac{2M}{N}\right)\right)}\right)^2 &(28)\\ &\le \frac{M}{N}\left(\frac{N}{M}\right)^2\left(\frac{\pi}{2}\frac{M}{N}\frac{1}{I_2}\right)^2\left(\frac{1}{\sin\left(\frac{\pi}{2}\left(\frac{2M}{N} - |\epsilon|\right)\right)}\right)^2 &(29)\\ &\le \frac{N}{M}\left(\frac{\pi}{2}\frac{M}{N}\frac{1}{I_2}\right)^2\left(\frac{1}{\frac{2M}{N} - |\epsilon|}\right)^2 &(30)\\ &\le \frac{N}{M}\left(\frac{\pi}{2}\frac{M}{N}\frac{1}{I_2}\right)^2\left(\frac{1}{\frac{M}{N}\frac{2I_2-1}{I_2}}\right)^2 = \frac{N}{M}\left(\frac{\pi}{2}\right)^2\left(\frac{1}{2I_2-1}\right)^2. &(31)\end{aligned}$$

Here, (28) is valid since $|\sin x| \le |x|$; (29) follows from (26) and the fact that[4]

$$\min \ \sin^2\left(i'\frac{\pi M}{N} - \frac{\pi}{2}\epsilon\right) = \sin^2\left(\frac{\pi M}{N} - \frac{\pi}{2}|\epsilon|\right); \tag{32}$$

(30) holds because of $0 \le \frac{\pi M}{N}\left(1 - \frac{1}{2\ I_2}\right) \le \frac{\pi M}{N} - \frac{\pi}{2}|\epsilon| \le \frac{\pi M}{N} \le \pi/2$ and Jordan's inequality stating $(2/\pi)x \le \sin x$ for $0 \le x \le \pi/2$ [27]; and finally, (31) holds by $|\epsilon| \le \frac{1}{I_2}\frac{M}{N}$.

[4]Since $M/N \le 1/2$ and $|\epsilon| \le \frac{1}{N}$, the minimum of $\sin^2\left(\frac{\pi}{2}(\epsilon - i'\frac{2\ M}{N})\right)$ over the range of i' in (27) occurs when i' is 1 or -1 so that $|i'\frac{2\ M}{N} - \epsilon|$ is minimized. (The minimum absolute value is $\frac{2M}{N} - |\epsilon|$.) This is because $\sin^2(\frac{\pi}{2}x)$ is a monotone decreasing function of $|x|$ for $0 \le |x| \le 1$.

On the other hand, the magnitude of the AF of the first-level subarray at $\hat{\xi}$ is given by

$$\left|AF^{\hat{\xi}}_{N/M,\lambda/2}(\theta_\ell)\right|^2 = \left(\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}\epsilon\right)}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\epsilon\right)}\right)^2 \geq \left(\frac{\frac{2}{\pi}\frac{N}{M}\frac{\pi}{2}\epsilon}{\sqrt{N/M}\frac{\pi}{2}\epsilon}\right)^2 = \frac{N}{M}\left(\frac{2}{\pi}\right)^2, \quad (33)$$

where in (33), we used $(2/\pi)x \leq \sin x$ for $0 \leq x \leq \pi/2$ for the numerator and $\sin x \leq x$ for $0 \leq x \leq \pi/2$ for the denominator. Combining (31) and (33), we have (26). Finally, when $I_2 = 2$, we have $\left(\frac{\pi}{2}\right)^4 (1/(2\ I_2 - 1))^2 \approx 0.6765 < 1$. Hence, when $I_2 \geq 2$, the combined gain of any grating lobe angle is strictly less than that of the main lobe combined gain. Hence, the ambiguity associated with the second-level virtual array is resolved by the combined first-level beam pattern for the two-level approach. ■

In Proposition 3, it is implicitly assumed that $\psi(i_1) = \hat{\xi}$. However, this assumption is not required, and the following proposition states a relaxed version.

Proposition 4: If the AoA θ_ℓ under detection is located within the HBW of the first-level array look-angle $\psi(i_1)$, i.e., $\theta_\ell \in \mathcal{R}_{i_1} := \left[\psi(i_1) - \frac{M}{N}, \psi(i_1) + \frac{M}{N}\right]$, and $\hat{\xi}$ is in-between $\psi(i_1)$ and θ_ℓ with $|\epsilon| = |\theta_\ell - \hat{\xi}| \leq \frac{1}{I_1 I_2}$, then still the ambiguity caused by the grating lobes of the second-level array is resolved by the first-level array beam pattern.

Proof: Intuitively, this can be seen by noting the beam pattern of the critically sampled first-level array shown in Fig. 2. Recall the AF of the first-level array:

$$\left|AF^{\psi(i_1)}_{N/M,\lambda/2}(\theta_\ell)\right|^2 = \left(\frac{\sin\left(\frac{N}{M}\frac{\pi}{2}\tilde{\theta}\right)}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\tilde{\theta}\right)}\right)^2, \quad (34)$$

where $\tilde{\theta}$ is the angle distance from the first-level look-angle $\psi(i_1)$. First, note that $N/M \geq 2$. This is because each subarray has at least two elements. Otherwise, it is not called the subarray architecture. When $N/M = 2$, we have only the main lobe and no side lobes, and the main lobe gain monotonically decreases as the measurement angle deviates from the center. Hence, the first-level array gain for the grating lobe of the second-level array (outside of the HBW of the first-level array) is smaller than that for the main lobe of the second-level array (between $\psi(i_1)$ and θ_ℓ). Now suppose that $N/M \geq 3$. Then, the first side lobe peak occurs nearly at $\tilde{\theta} = 3M/N$ [23] since in the numerator of the RHS of (34), $\left|\sin\left(\frac{N}{M}\frac{\pi}{2}\tilde{\theta}\right)\right| = 1$ at $\tilde{\theta} = (2i+1)M/N$ for $i = 1, 2, \cdots$, and in the denominator of the RHS of (34), $\sin(\frac{\pi}{2}\tilde{\theta})$ is a monotone increasing function of $\tilde{\theta}$ for $0 \leq \tilde{\theta} \leq 1$. Furthermore, the peak at the first side lobe has the largest gain outside of the main lobe. Hence, the maximum beam gain of the first-level array at the grating lobe of the second-level array is upper bounded as

$$\left(\frac{\sin(\frac{N}{M}\frac{\pi}{2}\frac{3M}{N})}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\frac{3M}{N}\right)}\right)^2 \overset{(a)}{\leq} \frac{1}{\frac{N}{M}\left(\frac{2}{\pi}\frac{\pi}{2}\frac{3M}{N}\right)^2} = \frac{1}{3^2}\frac{N}{M}, \quad (35)$$

where in step (a) we used $(2/\pi)x \leq \sin x$ for $0 \leq x \leq \pi/2$ since $N/M \geq 3$ in the considered case. On the other hand, the minimum beam gain of the first-level array at the main-lobe look-angle $\hat{\xi}$ of the second-level array is lower bounded as

$$\left(\frac{\sin(\frac{N}{M}\frac{\pi}{2}\frac{M}{N})}{\sqrt{N/M}\sin\left(\frac{\pi}{2}\frac{M}{N}\right)}\right)^2 \overset{(b)}{\geq} \frac{1}{\frac{N}{M}\left(\frac{\pi}{2}\frac{M}{N}\right)^2} = \left(\frac{2}{\pi}\right)^2\frac{N}{M}, \quad (36)$$

if $\theta_\ell \in \mathcal{R}_{i_1} = \left[\psi(i_1) - \frac{M}{N}, \psi(i_1) + \frac{M}{N}\right]$, and $\hat{\xi}$ is in-between $\psi(i_1)$ and θ_l. This is because the main lobe beam gain decreases monotonically as $\tilde{\theta}$ increases from $0 \leq \tilde{\theta} \leq \frac{2M}{N}$, as seen in Fig. 2. In step (b), we used $|\sin x| \leq |x|$. From (35) and (36), it is seen that the minimum gain for the second-level main lobe is larger than the maximum gain for the second-level grating lobe. Hence, we have the claim. Note from (35) and (36) that at least the suppression $10\log_{10}((2/\pi)^2 3^2) \approx 5.62$ dB for the grating lobes of the second-level array by the first-level beam pattern is guaranteed under the condition of Proposition 4. ■

Since $I_1 \geq N/M$ in our design (23), the difference between $\psi(i_1)$ and $\psi(i_1+1)$ is less than or equal to $2M/N$, and hence $\theta_\ell \in \mathcal{R}_{i_1} = \left[\psi(i_1) - \frac{M}{N}, \psi(i_1) + \frac{M}{N}\right]$ for the best $\psi(i_1)$. Within the angle coverage of the best $\psi(i_1)$, $\hat{\xi} = \xi(i_1, \hat{i}_2(i_1))$ is found so that $|\theta_\ell - \hat{\xi}| \leq 1/(I_1 I_2)$. The grating lobe positions at $\hat{\xi} + i'\frac{2M}{N}$ for nonzero i' is suppressed by the first-level array beam pattern. Hence, in the noiseless case, there is no angle ambiguity for the proposed two-level phased array method.

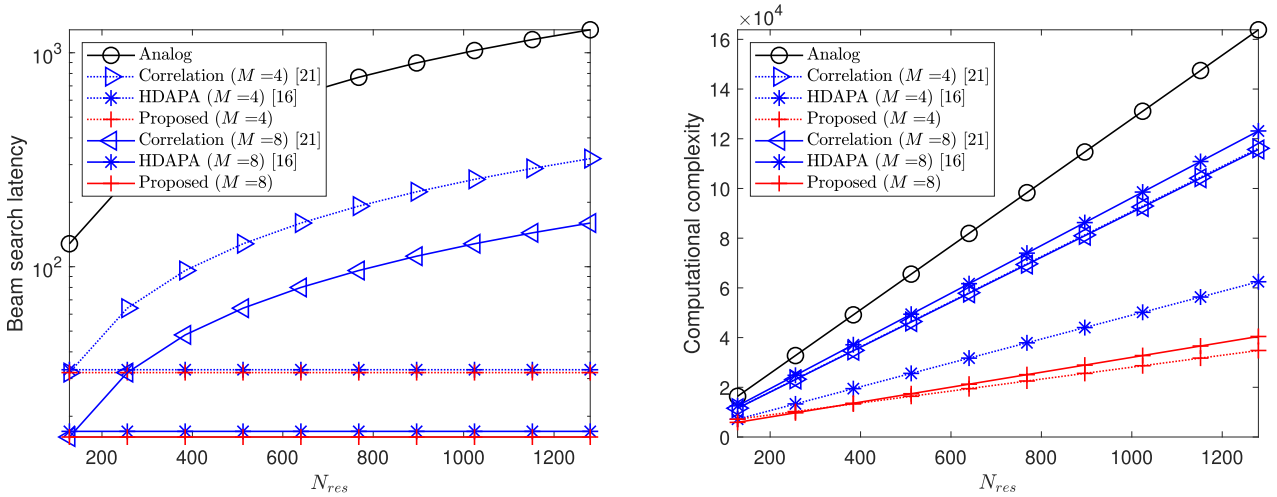
B. Beam Search Latency and Computational Complexity

We analyze the beam search latency and computational complexity of the proposed beam search method. In the proposed method, the implementation of digital filtering for AoA estimation at the second-level virtual array can be achieved with pipelined fast Fourier transform (FFT) filtering because the digital combiner in (16) is the DFT-based matched filtering. Table I shows the beam search latency and computational complexity of several beam search methods, where the beam search latency is measured by the number of required training symbols, the computational complexity is measured by the number of required complex multiplications occurring in both analog and digital domains for each algorithm, and $N_{res}(\geq N)$ is the target resolution factor whose inverse determines target resolution $2/N_{res} = \Delta_{angle}$. In the proposed method, $N_{res} = I_1 I_2$ and $\Delta_{angle} = |\xi(i_1, i_2) - \xi(i_1, i_2 + 1)|$. It is seen that for the analog beamforming and the correlation-based method, the number of training symbols increases linearly with respect to N_{res}. For HDAPA and the proposed method, on the other hand, the number of training symbols can be fixed regardless of the target angular resolution N_{res}. Note that the proposed method requires the least number of training beams among the compared algorithms. Fig. 5(a) shows the beam search latency measured by the number of required training beams versus the beam resolution N_{res} for two different values of M. Note that the latency gain by the proposed method can be significant for high resolution beam search. Fig. 5(b) shows the computational complexity as a function of N_{res}. It can be shown from Table I that the proposed method has the least complexity among the compared algorithms when $N_{\text{res}} \geq \frac{N^2}{KM}$ and $M > 2$. (The performance will be presented

TABLE I

COMPARISON OF BEAM SEARCH LATENCY AND COMPUTATIONAL COMPLEXITY. (THE NOTATIONS IN THE FIRST COLUMN FOLLOW THE REFERENCES)

Methods	Beam search latency (N_P)	Computational complexity
Analog beamforming	N_{res}	$N_{\text{res}}N$
Correlation [21] ($K=M, \Delta_i = I/2$)	$\frac{N_{\text{res}}}{M}$	$N_{\text{res}}\left(\frac{N}{M}+\frac{3}{2}K\frac{M-1}{M}\right)+N_{\text{res}}\left(K+K\log_2 M+K\frac{M-1}{M}+\frac{1}{M}\right)$
HDAPA [16] ($L=1$)	$\frac{N}{M}+1$	$N_{\text{res}}KM+\left(\frac{N}{M}\right)^2$
Proposed method ($I_1 = N/M$)	$\frac{N}{M}$	$N_{\text{res}}K\left(1+\frac{\log_2 M}{2}\right)+\frac{N^2}{M}$



(a) Beam search latency (b) Computational complexity

Fig. 5. Beam search latency and computational complexity: Comparison of several beam search methods for different M where $N = 128$ and $K = 12$.

in in SectionVI.) From Table I, we can see that the beam search latency and the computational complexity of the proposed method decreases as M increases, if $M \leq 2\ln 2\frac{N^2}{N_{res}K}$.

V. EXTENSIONS OF THE PROPOSED METHOD

In this section, we consider several extensions of the proposed method.

A. Beyond Partial Connection

First, we consider a structural extension beyond the partial connection architecture shown in Fig. 1. We can consider the M interleaved linear subarray architecture in which the elements forming an analog subarray is spread across the whole antena array. That is, in the M interleaved linear subarray architecture, each analog linear subarray has N/M antenna elements with inter-element spacing $M\lambda/2$ but the subarray spacing is now the same as the antenna element spacing $\lambda/2$. In this case, based on Proposition 1, we set the m-th column vector $\mathbf{f}_{RF,m}$ of $\mathbf{F}_{RF}$ and the digital combiner $\mathbf{f}_{BB,k}(\xi)$ as follows:

$$\begin{aligned}\mathbf{f}_{RF,m}(\psi) &= \sqrt{M/N}\mathbf{u}_{N/M,M\lambda/2}(\psi)\otimes\mathbf{e}_M(m)\\ \mathbf{f}_{BB,k}(\xi) &= (1/\sqrt{M})\mathbf{u}_{M,\lambda/2}(\xi).\end{aligned} \tag{37}$$

Then, the combined beamformer achieves the maximum beamforming gain described in Proposition 1. In the interleaved linear subarray case, we design the training beam sets $\tilde{B}_{I_1}$ for ψ and $\tilde{B}_{I_1,I_2}$ for ξ as follows:

$$\tilde{B}_{I_1} = \left\{ \psi(i_1) = -1 + \frac{2}{I_1 M}\left(i_1 - \frac{1}{2}\right) \middle| \, i_1 = 1,\cdots,I_1 \right\},$$

for some integer $I_1 \geq N/M$. The I_1 quantized angles in $\tilde{B}_{I_1}$ are evenly spaced over a normalized angular interval $\left[-1, -1+\frac{2}{M}\right)$. Note that the remaining angular intervals $[-1,1)\setminus\left[-1,-1+\frac{2}{M}\right)$ are covered by the grating lobes of the first-level subarrays. Then, we design $\tilde{B}_{I_1,I_2}$ as

$$\begin{aligned}\tilde{B}_{I_1,I_2} = \Bigg\{ \xi(i_1,i_2',i_2'') = \psi(i_1) + \frac{2}{I_1 I_2}\left(i_2' - \frac{1}{2}\left(\frac{I_2}{M}+1\right)\right)\\ + \frac{2}{M}(i_2''-1) \Bigg| \, \psi(i_1)\in\tilde{B}_{I_1}, i_1 = 1,\cdots,I_1,\\ i_2' = 1,\cdots,\frac{I_2}{M}, i_2'' = 1,\cdots,M \Bigg\},\end{aligned} \tag{38}$$

where $\xi\left(i_1-1,\frac{I_2}{M},i_2''\right)+\frac{2}{I_1I_2} = \xi\left(i_1,1,i_2''\right) \leq \xi\left(i_1,i_2',i_2''\right) \leq \xi\left(i_1,\frac{I_2}{M},i_2''\right) = \xi\left(i_1+1,1,i_2''\right) - \frac{2}{I_1I_2}$ and $I_2 \geq M$. Here, the index i_2' is to obtain a set of $\frac{I_2}{M}$ finely quantized angles around $\psi(i_1)$ and the index i_2'' is to set the main-lobe look angle of the second-level virtual array to the i_2''-th grating lobe

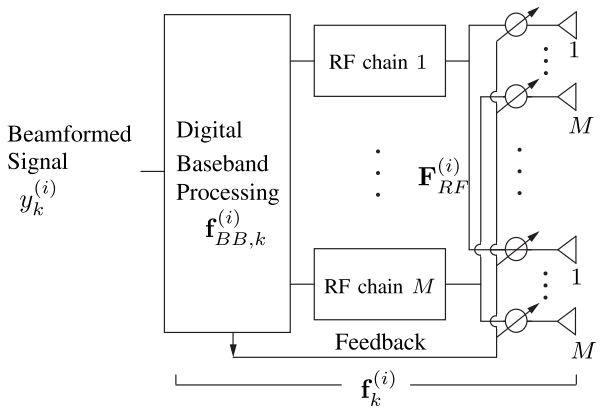
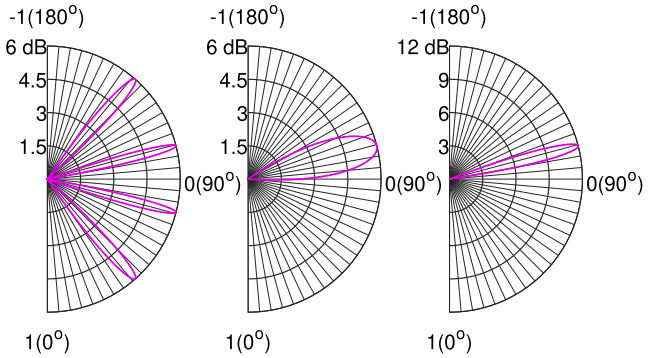


Fig. 6. Block diagram of M interleaved linear arrays.



(a) A beam pattern at the first-level sub-array using a single symbol ($i_1 = 2$)

(b) A beam pattern at the second-level array using a single symbol ($i_1 = 2$, $i_2' = 1, i_2'' = 2$)

(c) A beam pattern at the interleaved arrays using a single symbol ($i_1 = 4$, $i_2' = 1, i_2'' = 2$)

Fig. 7. Beam patterns for $\theta \in (-1, 1)$ where $N = 16$, $M = 2$, and $I_1 = I_2 = 4$: (a) the first-level subarray pattern $\left|AF_{N/M,M\lambda/2}^{\psi(i_1)}(\theta)\right|^2$, (b) the second-level virtual array pattern $\left|AF_{M,\lambda/2}^{\xi(i_1,i_2',i_2'')}(\theta)\right|^2$, and (c) the combined pattern $\left|AF_{N/M,M\lambda/2}^{\psi(i_1)}(\theta)AF_{M,\lambda/2}^{\xi(i_1,i_2',i_2'')}(\theta)\right|^2$. (Patterns below 0 dB are not drawn.)

of the first-level subarrays. The quantized angles of $\tilde{B}_{I_1,I_2}$ in (38) are evenly spaced over $[-1, 1)$ with spacing $2/(I_1 I_2)$.

Fig. 7 shows the beam patterns in the interleaved linear subarray case, where $N = 16$, $M = 4$, $I_1 = I_2 = 4$. It has a dual structure to the beam pattern in Fig. 3. The beam pattern of the first-level subarrays now has multiple M dominant peaks (i.e., grating lobes) due to space-undersampling inter-element spacing $M\lambda/2$ and the second-level virtual array has a single main lobe due to the critical subarray spacing $\lambda/2$.

Furthermore, the proposed method can be applied to the general interleaved linear subarray framework in which each of M analog subarrays has $\frac{N}{N_g M}$ groups of N_g consecutive antenna elements with inter-group spacing $MN_g\frac{\lambda}{2}$. The cases of $N_g = N/M$ and $N_g = 1$ correspond to the two-level phased array in Fig. 1 and the interleaved linear subarray architecture in Fig. 6, respectively. When $1 < N_g < N/M$, the proposed method requires a few additional symbols to resolve the directional ambiguity at the second-level virtual array.

B. Extension to the Multi-User Case

The system model (3) is for the single-user case. However, application of the proposed method to the multi-user case is possible by increasing the number of channel paths in the system model because the proposed algorithm can estimate multiple $N_C \geq 1$ dominant AoAs.

C. On the Joint Search of Transmit Beam from UE and Receive Beam at BS

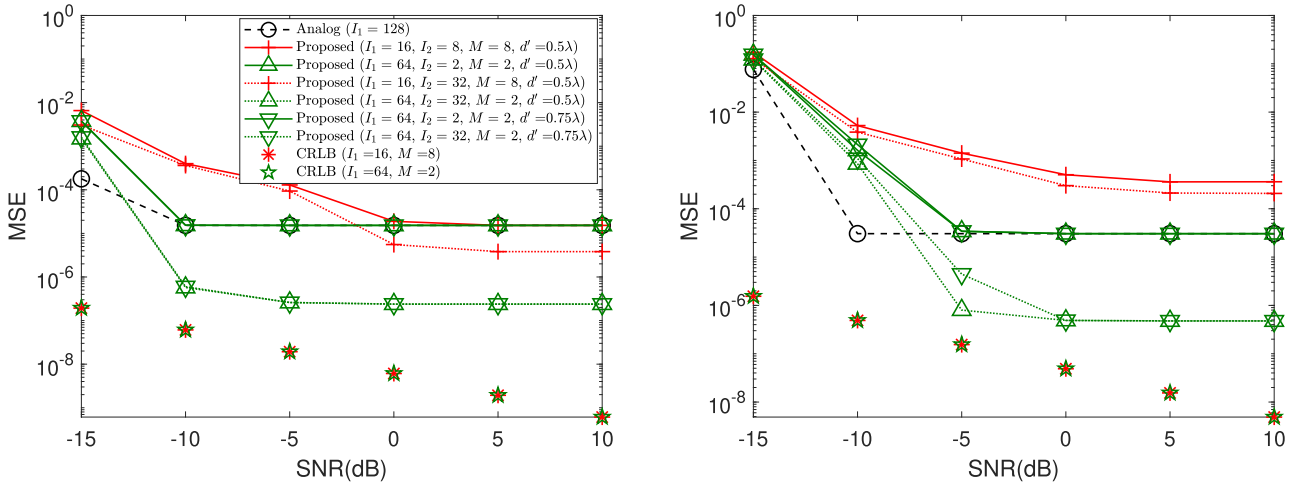
Note that the channel model (9) captures the single-input multiple-output (SIMO) uplink channel. In the case that a user equipment (UE) has multiple transmit antennas, transmit beam search is also required. Typically, the joint beam search is performed so that for each UE transmit beam, all BS receive beams are swept and this procedure is repeated by sweeping the UE transmit beam. Let the number of transmit beams to be probed at a UE be N_{UE} and let the numbers of antennas and RF chains at BS be N and M, respectively. When we apply full joint analog beam sweeping, the required beam search latency is $N_{UE}N$. On the other hand, if we apply the proposed receiver-side beam search for each of N_{UE} transmit beam, the overall joint beam search latency is reduced by factor M to yield the beam search latency of $N_{UE}\frac{N}{M}$. Furthermore, the proposed method can be extended to the case of AoA estimation at both BS and UE by dividing the beam search period into slots alternatively dedicated to uplink and downlink transmissions in a manner similar to that in [28], [29].

VI. Numerical Results

In this section, we provide numerical results to evaluate the proposed beam search method. Throughout simulation, the number of receive antennas was $N = 128$, and the SNR associated with the data model (3) was defined as ρ/σ_n^2. We assumed the channel model (8) with equal power profile $\alpha_\ell = 1$ for $1 \leq \ell \leq N_{cp}$ before normalization, where the number of paths was set to $N_{cp} \in \{1, 2, 3\}$, and the number of delay taps was set to $D = 3$. The pulse-shaping function $\mathbf{p}_\ell$ with $\|\mathbf{p}_\ell\|_2 = 1$ was set from the raised-cosine pulse with roll-off factor zero and the delay time of the ℓ-th channel path is assumed to be uniformly distributed between $(-1/(2T_s), 1/(2T_s))$, where $T_s = 1/\Delta f$ denotes the OFDM symbol duration with subcarrier spacing $\Delta f \in \{60, 120\}$ kHz.[5] For simplicity of simulation, we assumed that the true AoA θ_ℓ is quantized with a fine angular resolution, i.e., $\theta_\ell \in \left\{-1 + \frac{1}{N}, -1 + \frac{2}{N}, \cdots, 1 - \frac{1}{N}\right\}$ and that the angular distance between any two channel paths is larger than the HBW of the first-level array $2M/N$, i.e., $|\theta_\ell - \theta_{\ell'}| \geq 2M/N$ when $\ell \neq \ell'$.[6] The performance was measured by averaging the numerically estimated AoA and beamforming gain over $N_{MC} = 10^5$ Monte Carlo runs.

[5]In the 3GPP NR specification [30], the subcarrier spacing is specified as $\Delta f \in \{15, 30, 60\}$ kHz for Frequency Range 1 (FR1) including sub-6GHz frequency bands and as $\Delta f \in \{60, 120\}$ kHz for Frequency Range 2 (FR2) including frequency bands from 24.25 GHz to 52.6 GHz. A resource block (RB) is defined as 12 consecutive subcarriers [31].

[6]Note that at line 6 in Algorithm 1, we select only one i_2 for given i_1. Hence, Algorithm 1 considers only one path within the beamwidth of the first-level analog search beam.



(a) $N = 128$, $N_{cp} = 1$ (b) $N = 128$, $N_{cp} = 2$ (same legend as in (a))

Fig. 8. MSE versus SNR where $K = 12$.

Fig. 8 shows the AoA estimation performance in the case in which the inter-subarray spacing d_s satisfies the inequality $d_s \geq \frac{\lambda}{2}\frac{N}{M}$. Here, the inter-subarray spacing is defined as $d_s := \frac{\lambda}{2}\left(\frac{N}{M} - 1 + \frac{2}{\lambda}d'\right) \geq \frac{\lambda}{2}\frac{N}{M}$ and two cases $d' = \frac{\lambda}{2}, \frac{3\lambda}{4}$ are considered. The mean square error (MSE) is defined as $\text{MSE} = (1/N_{MC})\sum_{n=1}^{N_{MC}}\sum_{\ell=1}^{N_{cp}}|\theta_\ell - \xi_\ell(i_1, \hat{i}_2(i_1))|^2$, where $\xi_\ell(\cdot)$ is the estimate for the AoA of the ℓ-th path. The analog sweeping resolution I_1 in (21) is set as $I_1 = 16$, 64, and 128 for $\frac{N}{M} = 16$, 64, and 128, respectively. For comparison, we considered full analog beam sweeping with $I_1 = 128$. For the proposed method we considered $(I_1, I_2) = (64, 2)$ and $(I_1, I_2) = (16, 8)$ ($I_1 \times I_2 = 128$) both targeting the resolution of full analog beamforming with $I_1 = 128$. It is seen that the MSE performance of the proposed method with $(I_1, I_2) = (64, 2)$ and $(I_1, I_2) = (16, 8)$ achieves the performance of the full analog beam sweeping method with $I_1 = 128$. Thus, the proposed method with $(I_1, I_2) = (64, 2)$ and $(I_1, I_2) = (16, 8)$ achieves the beam training with almost the same resolution of full analog beam sweeping with $I_1 = 128$ only with $\frac{1}{2}$ and $\frac{1}{8}$ analog beam sweeping overhead as compared to the full analog beam sweeping, respectively. There exists a performance floor at high SNR. This is mainly because beam search is performed over quantized angular intervals,[7] and this effect can be mitigated by increasing the value of I_2. So, we increased I_2 to 32 for both $I_1 = 16$ and 64. Indeed, the floor level decreases as I_2 increases, as shown in Fig. 8. It is seen in Fig. 8(a) that there is a marginal difference in performance between the two cases of $d' = \frac{\lambda}{2}$ and $d' = \frac{3\lambda}{4}$. Fig. 8(b) shows that the increased d' degrades the MSE performance, especially at low SNR, because the sidelobe level increases as the peak-to-peak spacing of the array factor at the second-level virtual subarray decreases. This performance gap shrinks with increased SNRs above 0 dB. Furthermore, we included the Cramér–Rao lower bound (CRLB) [32] in Fig. 8 as reference. The CRLB was obtained under the assumption

[7]Note that the quantization level for true θ_ℓ and the quantization level (I_1, I_2) for estimation are different.

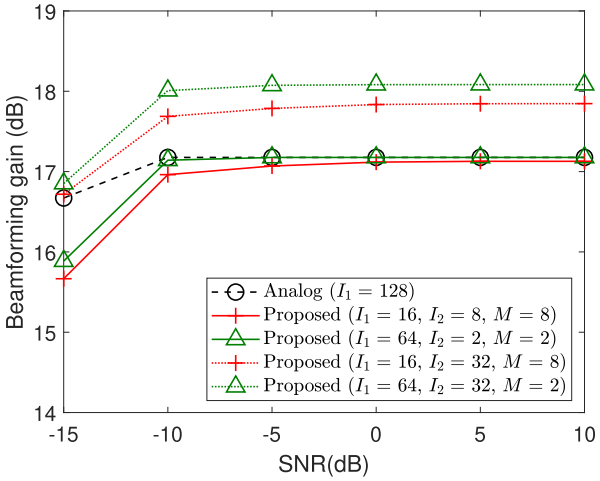
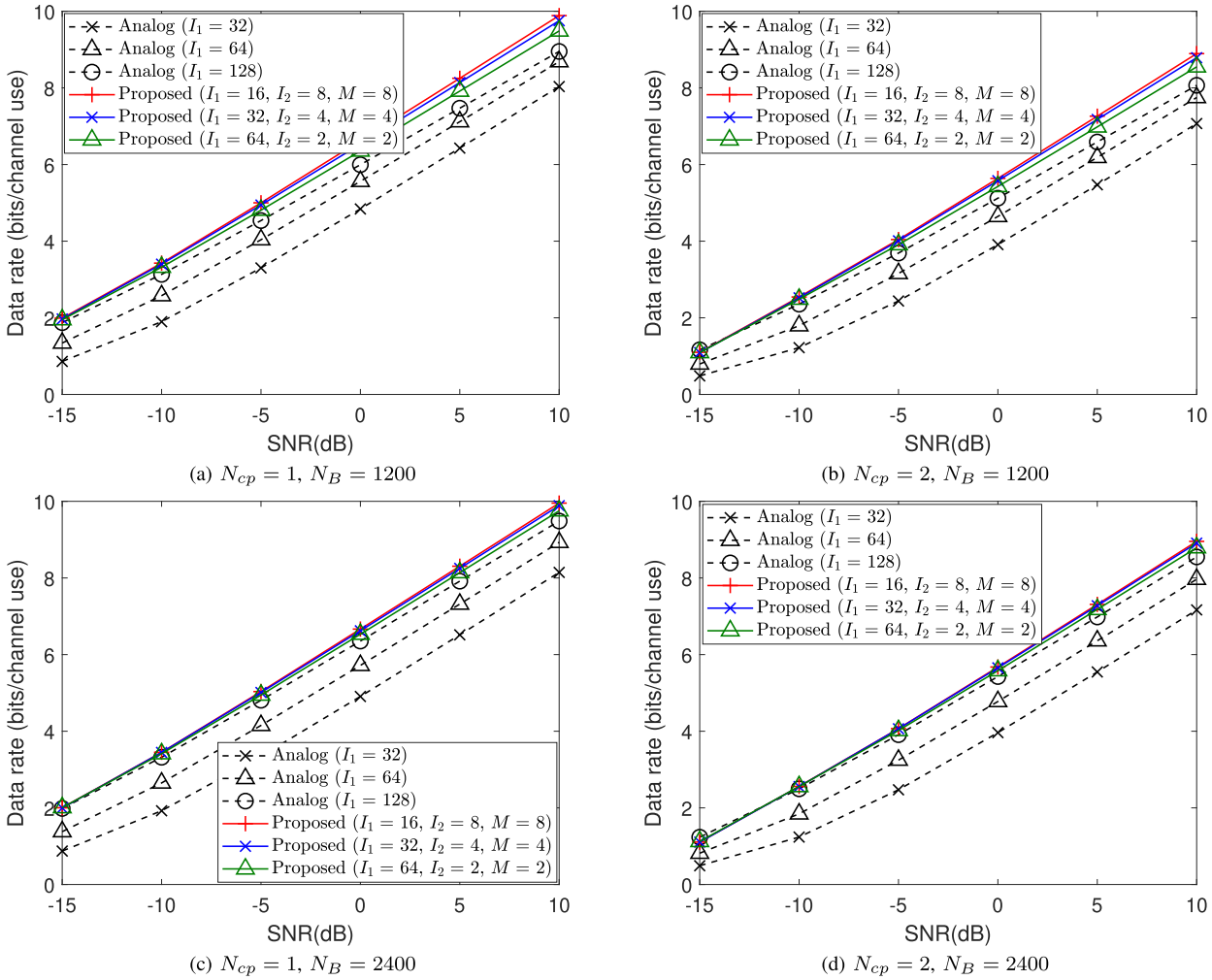


Fig. 9. Beamforming gain where $N_{cp} = 2$ and $K = 12$.

of using I_1 analog training beams evenly spaced over the normalized angle range $[-1, 1]$ and K subcarriers without any assumption on I_2. It is seen that the performance of the proposed method using $I_1 = 64$ training symbols and $I_2 = 32$ is closest to the CRLB. Note that given I_1, the proposed method estimates the AoA with different angular granularity by changing the value of I_2 but using the same I_1 training symbols. In addition to the gap at high SNR due to quantization, there exists a performance gap from the CRLB at low SNR. This is mainly due to non-coherent combining over multiple subcarriers and across training symbols shown in (24) of the proposed method. In the following simulation, we assume $d' = \lambda/2$.

Fig. 9 shows the beamforming gain performance of the proposed beam training method. The beamforming gain was measured as the absolute square of the inner product $|\hat{\mathbf{f}}_k^H \mathbf{h}_k|^2$ between the estimated beamformer $\hat{\mathbf{f}}_k = (1/\sqrt{N_{cp}})\sum_{\ell=1}^{N_{cp}} \mathbf{u}_N(\xi_\ell(i_1, \hat{i}_2(i_1)))$ and the channel vector $\mathbf{h}_k$. It is seen that the proposed method with $(I_1, I_2) \in \{(16, 8), (64, 2)\}$ ($I_1 \times I_2 = 128$) yields almost the same



(a) $N_{cp} = 1$, $N_B = 1200$

(b) $N_{cp} = 2$, $N_B = 1200$

(c) $N_{cp} = 1$, $N_B = 2400$

(d) $N_{cp} = 2$, $N_B = 2400$

Fig. 10. Data rate performance of the proposed method ($N = 128$ and $K = 12$).

performance as the full analog beam training with $I_1 = 128$ training symbols with much smaller analog sweeping overhead, i.e., $I_1 = 16$ and 64 training symbols. For $(I_1, I_2) \in \{(16, 32), (64, 32)\}$ ($I_1 \times I_2 \in \{512, 2048\}$, the proposed method outperforms the analog beam training at the expense of increased computational complexity due to the increased I_2, but I_1 can be kept to be small.

Fig. 10 shows the training-based rate performance of the proposed method and the full analog beam training method. The training-based achievable rate was measured as [9]

$$R_k = \frac{N_B - I_1}{N_B} \log\left(1 + \frac{\rho}{\sigma_n^2} |\hat{\mathbf{f}}_k^H \mathbf{h}_k|^2\right), \tag{39}$$

where the number of training symbols is given by I_1 and the pre-log factor incorporates the ratio of the number of pure data-carrying symbols to the whole transmission block length N_B. Based on 3GPP [33], we considered the cases of $N_B \in \{1200, 2400\}$.[8] First, we considered $N_B = 1200$ for $\Delta f = 60$ kHz and the result is shown in Figs. 10(a) and 10(b). It is seen in Fig. 10(a) that among the three cases of full analog beam training ($I_1 = 32, 64, 128$), the analog beam alignment method with $I_1 = 128$ shows a good performance. In Fig. 10(b), the analog beam alignment method with $I_1 = 64, 128$ shows balanced trade-off between the training overhead and the beamforming gain as compared to the case of $I_1 = 36$. Note that the training overhead directly affects the slope of the rate increased at high SNR as shown in (39). Hence, the reduction of training overhead is important at high SNR, and the full analog beam training method with $I_1 = 64$ starts to outperform the full analog beam training method with $I_1 = 128$ around 5 dB SNR in Fig. 10(b). It is seen that the proposed methods with $I_1 \times I_2 = 128$ outperforms the full analog beam training method with $I_1 = 128$ because the proposed methods achieves almost the same beamformig gain with much reduced training overhead. Next, N_B is increased from 1200 to 2400 for $\Delta f = 120$ kHz, and the corresponding results are shown in Figs. 10(c) and 10(d). Similar behavior to that in the case of $N_B = 1200$ is observed. In this case, even for the full analog beam training method with $I_1 = 128$, the training overhead is 8%, which is far smaller than com-

[8]In [33], the synchronization signal block (SSB) transmitted by using a single beam consists of four consecutive OFDM symbols with period of 20 ms by default. During the data transmission, we assume four-symbol mini-slot based scheduling. Therefore, given the subcarrier spacing Δf, we set the the block length as $N_B = 20\text{ms}/(4/\Delta f)$.

TABLE II
COMPARISON OF BEAM SEARCH LATENCY AND COMPUTATIONAL COMPLEXITY FOR THE SAME SETUP IN FIG. 11

Methods (Simulation setup)	Beam search latency (N_P)	Computational complexity
Analog beamforming ($I_1 = N_{res} = 128$)	128	16,384
Correlation [21] ($I_1 = 64, M = 8, K = 12, N_{res} = 512$)	64	46,272
HDAPA [16] ($I_1 = 17, M = 8, K = 12, N_{res} = 512$)	17	49,408
Proposed method		
($I_1 = 16, I_2 = 8, M = 8, K = 12, N_{res} = 128$)	16	5,888
($I_1 = 16, I_2 = 32, M = 8, K = 12, N_{res} = 512$)		17,408
($I_1 = 16, I_2 = 128, M = 8, K = 2, N_{res} = 2048$)		12,288

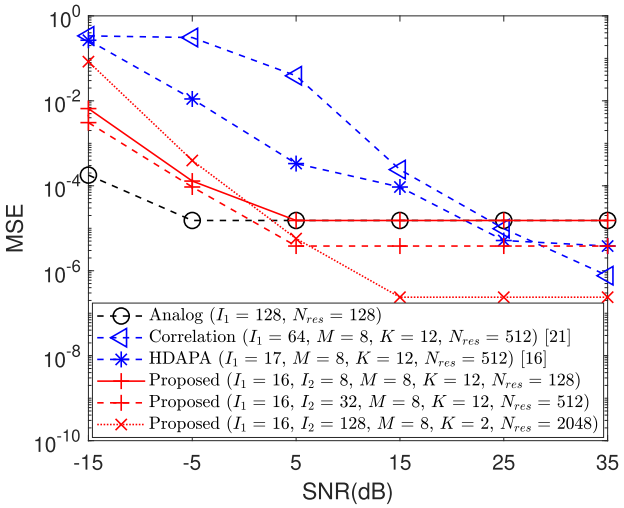


Fig. 11. MSE performance with $N = 128$, $M = 8$, $K \in \{2, 12\}$, and $N_{\text{res}} \in \{128, 512, 2048\}$. (The full analog beam sweeping method is added as a benchmark showing what can be achieved by using $I_1 = 128$ training symbols.)

monly considered training overhead ratio. In this case, for the considered range of SNR, the beamforming gain becomes the dominant factor for the data rate performance. Hence the full analog beam training method with $I_1 = 128$ outperforms the same method with $I_1 = 32$ and 64. It is seen that even in this case the proposed method outperforms the full analog beam training method. Compared to Figs. 10(a) and (c), the dara rate performance in Figs. 10(b), 10(d) is slightly degraded due to the increased MSE value as shown in Figs. 8(a) and 8(b). While the performance behavior of the beam search algorithm with $N_{cp} = 2$ is similar to that with $N_{cp} = 1$.

Finally, we compared the MSE performance and computational complexity of the proposed method with several existing beam search methods including analog, correlation-based, and HDAPA methods.[9] Fig. 11 and Table II show the results. It is seen that the proposed method with $I_1 = 16$ training symbols yields good performance compared to other methods, and outperforms the full analog beam search method using $I_1 = 128$ training symbols when the SNR value increases. It is seen that the performance of the proposed method is saturated at high SNR due to the finite target angular resolution N_{res}.

[9]For [21], we set the parameters as $\Delta_i = K/2$, subcarrier spacing 583 MHz, and carrier frequency 70 GHz.

Such a performance floor can be lowered by increasing I_2. Based on the analysis in Section IV-B, we computed the beam search latency and computational complexity required for the same setup used in Fig. 11. The result is shown in Table II. It is seen that the proposed method requires the smallest beam search latency and its computational complexity is comparable to that of analog beamforming and is less than that of the correlation-based and HDAPA methods.

VII. CONCLUSION

In this paper, we have proposed a fast beam search and refinement method for millimeter-wave hybrid beamforming systems, based on a two-level phased array approach. Exploiting the fact that the overall beam pattern of a two-level array is the product of the beam patterns of the two level subarrays, the proposed method searches AoAs by sweeping coarse-resolution analog training beams at the first-level phased array and matching the first-level subarray outputs with fine-resolution digital training beams by parallel FFT filtering. Thus, the overall beam search resolution of the proposed method is given by the fine resolution of the second-level virtual array only with the overhead of coarse beam sweeping at the first-level subarrays. We have analyzed the proposed method in terms of the directional ambiguity, beam search latency, and computational complexity. The proposed method can be extended to an interleaved linear array framework. We have provided numerical results showing the effectiveness of the proposed method.

REFERENCES

[1] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, 1st Quart., 2019.
[2] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5014–5029, Jun. 2017.
[3] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.
[4] I. Ahmed *et al.*, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.
[5] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
[6] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–2683, Jul. 2017.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "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.
[8] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.
[9] 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.
[10] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.
[11] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using OFDM waveforms," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 226–237, Feb. 2010.
[12] X. Huang, Y. Jay Guo, and J. D. Bunton, "A hybrid adaptive antenna array," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1770–1779, May 2010.
[13] X. Huang and Y. J. Guo, "Frequency-domain AoA estimation and beamforming with wideband hybrid arrays," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2543–2553, Aug. 2011.
[14] J. Zhang, X. Huang, V. Dyadyuk, and Y. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.
[15] S.-F. Chuang, W.-R. Wu, and Y.-T. Liu, "High-resolution AoA estimation for hybrid antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 63, no. 7, pp. 2955–2968, Jul. 2015.
[16] F. Shu *et al.*, "Low-complexity and high-resolution DOA estimation for hybrid analog and digital massive MIMO receive array," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2487–2501, Jun. 2018.
[17] H. Li, T. Q. Wang, X. Huang, and Y. Jay Guo, "Adaptive AoA and polarization estimation for receiving polarized mmWave signals," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 540–543, Apr. 2019.
[18] C. Qin, J. A. Zhang, X. Huang, and Y. J. Guo, "Virtual-subarray-based angle-of-arrival estimation in analog antenna arrays," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 194–197, Feb. 2020.
[19] J. A. Zhang, W. Ni, P. Cheng, and Y. Lu, "Angle-of-arrival estimation using different phase shifts across subarrays in localized hybrid arrays," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2205–2208, Nov. 2016.
[20] K. Wu, W. Ni, T. Su, R. P. Liu, and Y. J. Guo, "Robust unambiguous estimation of angle-of-arrival in hybrid array with localized analog subarrays," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2987–3002, May 2018.
[21] K. Wu, W. Ni, T. Su, R. P. Liu, and Y. J. Guo, "Fast and accurate estimation of angle-of-arrival for satellite-borne wideband communication system," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 314–326, Feb. 2018.
[22] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.
[23] S. J. Orfanidis. (2013). *Electromagnetic Waves and Antennas*. [Online]. Available: http://eceweb1.rutgers.edu/~orfanidi/ewa/
[24] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[25] S. Park, A. Alkhateeb, and R. W. Heath, Jr., "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.
[26] J. O. Smith, III, *Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications*. 2nd ed. Charleston, SC, USA: Booksurge Publishing, 2007.
[27] F. Yuefeng, "Proof without words: Jordan's inequality," *Math. Mag.*, vol. 69, no. 2, p. 126, Apr. 1996.
[28] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
[29] N. Akdim, C. N. Manchon, M. Benjillali, and E. de Carvalho, "Ping pong beam training for multi stream MIMO communications with hybrid antenna arrays," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–7.
[30] *Base Station (BS) Radio Transmission and Reception (Release 15)*, document TS 38.104 V15.4.0, 3GPP, Dec. 2018.
[31] *NR—Physical Channels and Modulation (Release 15)*, document TS 38.211 V15.7.0, 3GPP, Sep. 2019.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.
[33] *NR—Physical Layer Procedures for Control (Release 15)*, document TS 38.213 V15.4.0, 3GPP, Jan. 2019.

Song Noh (Member, IEEE) received the B.S. degree in electrical engineering from Soongsil University, Seoul, South Korea, in 2008, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015. From 2015 until 2018, he was a System Engineer with the Next Generation and Standards Group, Intel Corporation, Hillsboro, OR, USA, and participated in the design of PHY/MAC layer protocol for millimeter-wave mobile broadband systems. Since 2018, he has been on the faculty of the Department of Information Telecommunication Engineering, Incheon National University, Incheon, South Korea. His research interests include the design and analysis of millimeter-wave communication systems and adaptive signal processing for wireless communications. Along with coauthors, he was a co-recipient of the Communication Theory Symposium Best Paper Award at the 2014 IEEE Global Communications Conference (Globecom).

Jiho Song (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2009 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2017. He was a Senior Researcher with Motorola Mobility, Chicago, IL, USA. He is currently an Assistant Professor with the School of Electrical Engineering, University of Ulsan, Ulsan, South Korea. His research interests include the design and analysis of millimeter-wave communication, multiuser MIMO communication, and limited feedback strategies for massive MIMO systems. He was a recipient of the Bronze Prize in Samsung Electronics' 23rd Humantech Paper Contest in 2017.

Youngchul Sung (Senior Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2005. Before joining the Ph.D. program, he was with the LG Electronics Limited, Seoul, from 1995 to 2000. From 2005 to 2007, he was a Senior Engineer with the Corporate Research and Development Center of Qualcomm, Inc., San Diego, CA, USA, and participated in design of Qualcomm's 3GPP R6 WCDMA base station modem. Since 2007, he has been on the faculty with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea. His research interests include signal processing for communications, statistical signal processing, and statistical inference and learning with applications to wireless communications and related areas. He is currently the Vice Director of the IEEE Communications Society Asia–Pacific Board. He was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2012 to 2014 and a Guest Editor of the 2012 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue Theories and Methods for Advanced Wireless Relays. He is also an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.