

# Closed-Loop Sparse Channel Estimation for Wideband Millimeter-Wave Full-Dimensional MIMO Systems

Anwen Liao<sup>1</sup>, *Student Member, IEEE*, Zhen Gao<sup>1</sup>, *Member, IEEE*, Hua Wang<sup>1</sup>, Sheng Chen<sup>2</sup>, *Fellow, IEEE*, Mohamed-Slim Alouini<sup>3</sup>, *Fellow, IEEE*, and Hao Yin

**Abstract**—This paper proposes a closed-loop sparse channel estimation (CE) scheme for wideband millimeter-wave hybrid full-dimensional multiple-input multiple-output and time division duplexing based systems, which exploits the channel sparsity in both angle and delay domains. At the downlink CE stage, random transmit precoding matrix is designed at base station (BS) for channel sounding, and receive combining matrices at user devices (UDs) are designed whereby the hybrid array is visualized as a low-dimensional digital array for facilitating the multi-dimensional unitary ESPRIT (MDU-ESPRIT) algorithm to estimate respective angle-of-arrivals (AoAs). At the uplink CE stage, the estimated downlink AoAs, namely, uplink angle-of-departures (AoDs), are exploited to design multi-beam transmit precoding matrices at UD to enable BS to estimate the uplink AoAs, i.e., the downlink AoDs, and delays of different UD, whereby the MDU-ESPRIT algorithm is used based on the designed receive combining matrix at BS. Furthermore, a maximum likelihood approach is proposed to pair the channel parameters acquired at the two stages, and the path gains are then obtained using least squares estimator. According to spectrum estimation theory, our solution can acquire the super-resolution estimations of the AoAs/AoDs and delays of sparse multipath components with low training overhead. Simulation results verify the better CE performance and lower computational complexity of our solution over state-of-the-art approaches.

**Index Terms**—Wideband channel estimation, millimeter-wave, hybrid full-dimensional MIMO, super-resolution.

Manuscript received March 15, 2019; revised August 10, 2019; accepted September 12, 2019. Date of publication September 23, 2019; date of current version December 17, 2019. This work was supported by the National Natural Science Foundation of China under Grant no. 61701027, the Beijing Natural Science Foundation under Grant nos. 4182055 and L182024, Young Elite Scientists Sponsorship Program (2018-2021) by CAST, and Talent Innovation Project of BIT. The associate editor coordinating the review of this article and approving it for publication was I. H. Kim. (*Corresponding author: Zhen Gao.*)

A. Liao and H. Wang are with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing 100081, China.

Z. Gao is with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing 100081, China, and also with the Advanced Research Institute of Multidisciplinary Science, Beijing Institute of Technology (BIT), Beijing 100081, China (e-mail: gaozhen16@bit.edu.cn).

S. Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@soton.ac.uk).

M.-S. Alouini is with the Electrical Engineering Program, Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia.

H. Yin is with Institute of China Electronic System Engineering Corporation, Beijing 100141, China.

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/TCOMM.2019.2942911

## I. INTRODUCTION

MILLIMETER-WAVE (mmWave) communication with the aid of massive multiple-input multiple-output (MIMO) is an enabling technology for next-generation mobile communications, since the abundant spectrum resources at mmWave frequency band can boost the throughput by orders of magnitude [1], [2]. To mitigate the severe path loss for mmWave signal, massive MIMO is usually integrated into mmWave communications to form beams for directional signal transmission [3], [4]. However, the powerful fully-digital MIMO architecture, which requires a radio frequency (RF) chain for each antenna, is unaffordable for mmWave massive MIMO, due to the prohibitive hardware cost and power consumption of RF chains required [5]. The hybrid MIMO architecture with a much smaller number of RF chains than that of antennas offers a practical solution by using hybrid analog/digital beamforming [6]. Nonetheless, for such a hybrid MIMO system, it is challenging to estimate the high-dimensional mmWave channel from the low-dimensional effective measurements observed from the limited number of RF chains, since the training overhead for channel estimation (CE) can be excessively high [2]. Moreover, the low signal-to-noise ratio (SNR) before beamforming can further degrade the performance of channel state information (CSI) acquisition [7].

### A. Related Work

Several approaches were proposed in the literature to acquire CSI for narrowband mmWave communications, including codebook-based beam training [8]–[11] and compressed sensing (CS)-based CE [12], [13]. The beam training approaches were initially adopted in analog beamforming, e.g., IEEE standards 802.11ad [8] and 802.15.3c [2], where the transceiver exhaustively searches for the optimal beam pair from a predefined codebook to maximize the received SNR for improved transmission performance. To reduce the search dimension of codebooks for achieving lower training overhead, the multi-stage overlapped beam patterns were designed in [9], where the beam patterns can become narrow as the training stage increases. However, these schemes only consider the analog beamforming with single-stream transmission. For hybrid beamforming with multi-stream transmission, beam training

solutions with hierarchical multi-beam codebooks were proposed in [10], [11], where the optimal multi-beam pairs can be acquired after hierarchical beam search with gradually finer and narrower beams. However, the training overhead of a beam training scheme is usually proportional to the dimension of codebook, which is very large for full-dimensional (FD) MIMO with a large number of antennas. By exploiting the inherent angle-domain sparsity of mmWave MIMO channels, several CS-based CE schemes were proposed to reduce the CE overhead [12], [13]. In [12], the orthogonal matching pursuit (OMP) algorithm was considered to estimate sparse mmWave channels by formulating the CSI acquisition problem as a sparse signal recovery problem, where a redundant dictionary with non-uniformly quantized angle-domain grids was designed for improved performance. Furthermore, a Bayesian CS-based CE scheme was proposed in [13] by considering the impact of transceiver hardware impairments. Besides, by leveraging the low-rank property of mmWave channels, a CANDECOMP/PARAFAC decomposition-based CE scheme [14] was proposed with further improved performance.

The aforementioned solutions [9]–[14] only consider frequency-flat mmWave channels but practical mmWave channels can be frequency selective due to the very large system bandwidth in mmWave frequency band and the distinct delay spreads of multipath components (MPCs) [15]. A distributed grid matching pursuit (DGMP) algorithm was proposed in [16] to estimate time-dispersive channels, where orthogonal frequency division multiplexing (OFDM) is considered. An adaptive grid matching pursuit (AGMP) algorithm developed from the DGMP was proposed to reduce power leakage by using adaptive grid matching solution [17]. In [18], the sparse mmWave channels at different subcarriers were estimated separately by utilizing the OMP, but the computational complexity is high as the number of subcarriers is typically large. To reduce complexity, a simultaneous weighted (SW)-OMP based scheme was proposed in [19], which exploits the angle-domain common sparsity of channels at different subcarriers to improve performance. By leveraging the common sparsity of delay-domain channels among transceiver antenna pairs, a block CS-based CE solution was proposed for mmWave fully-digital MIMO system [20], where the training sequences are designed to improve CE performance. Based on the low-rank property of wideband mmWave channels, the training signal received can be formulated as a high-order tensor with the low-rank CANDECOMP/PARAFAC decomposition to estimate the dominated channel parameters, including angle-of-arrivals/angle-of-departures (AoAs/AoDs) and delays [21]. However, most CS-based CE schemes for wideband mmWave MIMO usually adopt discrete AoAs/AoDs grids in CS dictionary, but the practical AoAs/AoDs of MPCs are continuously distributed. This mismatch may degrade CE performance. Moreover, the state-of-the-art works [9]–[14], [16]–[22] usually focus on the ideal uniform linear array (ULA) while seldom investigate the practical uniform planar array (UPA). Compared to the ULA, the UPA offers more compact array with three-dimensional (3-D) beamforming in both horizontal and vertical directions [23], [24], leading to the FD-MIMO. Although mmWave FD-MIMO CE



Fig. 1. The air-ground mmWave channels between the UAV aerial-BS and UD exhibit sparsity in both angle-domain and delay-domain due to the limited significant scatterers [25], [26].

has been investigated in [23] and [24], they only considered either fully-digital MIMO or frequency-flat channels.

### B. Our Contributions

We propose a closed-loop sparse CE scheme for multi-user wideband mmWave FD-MIMO systems by exploiting the sparsity of MPCs in both angle and delay domains. To illustrate this sparsity, we consider the mmWave FD-MIMO based unmanned aerial vehicle (UAV) aerial-base station (BS), as shown in Fig. 1, which has flexible deployment capacity to serve user devices (UDs) in hotspot areas [27]. Different from the terrestrial BS in 3GPP or QuaDRiGa [28], [29], UAV aerial-BS usually works at the height of hundreds of meters, where fewer MPCs corresponding to dominant scatterers could establish the communications links between the aerial-BS and UD. Therefore, the air-ground mmWave channels in such aerial-BS based systems exhibit inherent sparsity in both angle and delay domains due to the limited significant scatterers. By carefully designing the transmit beamforming or precoding and receive combining in the training stage, our solution is capable of acquiring the super-resolution estimates of AoAs, AoDs, and delays<sup>1</sup> based on spectrum estimation theory [30], [31] with low training overhead and computational complexity. In terms of sparse mmWave UAV air-ground channels, the proposed CE scheme can obtain a better performance of parametric CE. To clearly show the novelty and new contribution of our proposed solution as well as to contrast it with the existing solutions, below we conceptually explain our proposed closed-loop sparse CE scheme.

Our closed-loop solution includes the downlink CE stage followed by the uplink CE stage as illustrated in Fig. 2, where the channel reciprocity in time division duplex (TDD) based systems is exploited [32]–[34]. In TDD based systems, downlink AoAs (AoDs) are uplink AoDs (AoAs). The frame structure of our solution is further depicted in Fig. 3. As shown in Figs. 2 and 3, at the downlink CE stage, the horizontal/vertical AoAs of sparse MPCs are first estimated at each UD and they are fed back to the BS with limited

<sup>1</sup>By contrast, the state-of-the-art CS-based solutions [16]–[19] only focus on the angle-domain sparsity and design the limited resolution CS dictionary with quantized angle-domain grids. This quantization error limits the achievable CE performance.

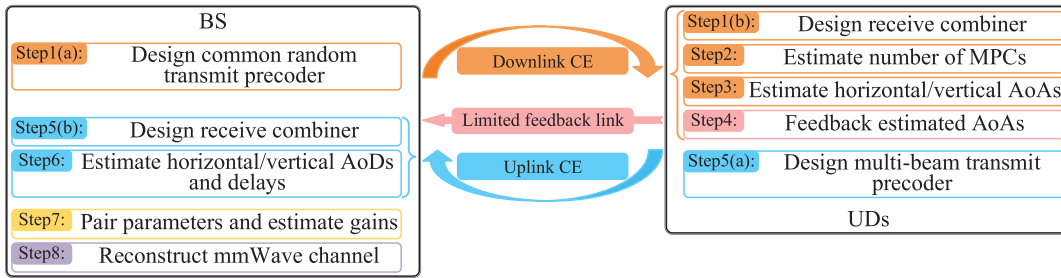


Fig. 2. Procedure of the proposed closed-loop sparse CE solution, where the limited feedback is realized via the low-frequency control link.

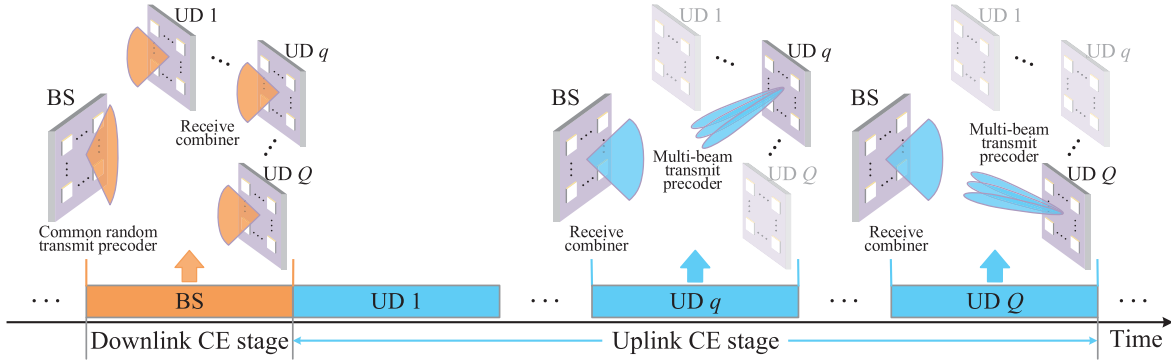


Fig. 3. Frame structure of the proposed closed-loop sparse CE solution.

quantization accuracy through the feedback link. At this stage, we design a common random transmit precoding matrix at the BS to transmit the training signals for omnidirectional channel sounding and we design the receive combining matrix at each UD to visualize the high-dimensional hybrid array as a low-dimensional digital array, which facilitates the use of the multi-dimensional unitary ESPRIT (MDU-ESPRIT) algorithm to estimate channel parameters. Similarly, at the uplink CE stage, the horizontal/vertical AoDs and delays associated with different UDs are successively estimated by using the MDU-ESPRIT algorithm at the BS. Owing to the channel reciprocity, the AoAs estimated at UD side can be utilized as *a priori* to design the multi-beam transmit precoding matrix to improve the received SNR for uplink CE. A maximum likelihood (ML) approach is adopted at the BS to pair the channel parameters acquired at these two stages and, consequently, the associated path gains can readily be obtained using the least squares (LS) estimator. Finally, the mmWave channel associated with each UD can be separately reconstructed based on the dominant channel parameters estimated above.

In contrast to the existing solutions, our main contributions are summarized as follows:

- **We propose a closed-loop sparse CE solution with the common downlink CE stage for all UDs and the dedicated uplink CE stage for each UD.** At the downlink CE stage, the BS fully exploits its large transmit power to allow multiple UDs simultaneously perform CE for reduced training overhead, and only horizontal/vertical AoAs estimation with low computational complexity is required at UDs. At the uplink CE stage, the multi-beam transmit precoding matrices are

designed at UDs to enhance CE accuracy, and the BS with high computational capacity can jointly estimate horizontal/vertical AoDs and delays. By contrast, the state-of-the-art CS-based solutions [16]–[19] are all based on open-loop approach, which imposes high computational complexity and storage requirements on the receiver.<sup>2</sup> It is worth emphasizing that by exploiting the horizontal/vertical AoAs estimated at the downlink CE stage, the multi-beam transmit precoding matrices designed at UDs significantly improve received SNR, which works even for the case that the number of MPCs is larger than that of RF chains, i.e., the number of beams can be larger than that of RF chains.

- **We design the receive combining matrices at UDs and BS for visualizing the hybrid array as a digital array, to enable the application of spectrum estimation techniques.** For hybrid MIMO, it is challenging to directly apply spectrum estimation techniques to estimate channel parameters, since the shift-invariance structure of array response matrix observed from the digital baseband domain does not hold [31], [35]. Our solution sheds light on how to apply spectrum estimation techniques, e.g., ESPRIT-type algorithms, to hybrid MIMO, so that the super-resolution estimation of channel parameters can be acquired with low training overhead

<sup>2</sup>More specifically, to improve the CE accuracy, CS-based solutions [16]–[19] adopt the redundant dictionary, whose dimension is very high for mmWave massive MIMO. Therefore, downlink open-loop CE imposes prohibitive storage of redundant dictionary and the associated computational complexity on UDs, while uplink open-loop CE suffers from the low received SNR and thus poor CE performance due to the limited transmit power of UDs.



and computational complexity. By contrast, to achieve the high-resolution estimations of channel parameters, the existing CS-based solutions [16]–[19] usually rely on the high-dimensional angle-domain or delay-domain redundant dictionary, which poses the excessively high computational complexity and storage requirements for FD-MIMO systems with large-scale antenna arrays.

- **The double sparsity of MPCs in both angle and delay domains is harnessed in our proposed CE scheme.** By leveraging the double sparsity, our scheme formulates the CE problem as a multi-dimensional spectrum estimation problem, where the super-resolution estimations of horizontal/vertical AoAs/AoDs and delays can be obtained simultaneously. By comparison, the existing CE solutions [16], [17], [19] only consider the angle-domain sparsity of mmWave channels. Moreover, the delay-domain CE approaches of [18], [20] have to estimate the effective delay-domain channel impulse response (CIR), which includes the time-domain pulse shaping filter (PSF) that can weaken the delay-domain channel sparsity. By contrast, the super-resolution estimation of delays in our solution is immune to PSF.

Throughout this paper, boldface lower and upper-case symbols denote column vectors and matrices, respectively.  $(\cdot)^*$ ,  $(\cdot)^T$ ,  $(\cdot)^H$ ,  $(\cdot)^{-1}$ ,  $[\cdot]$  and  $\lfloor \cdot \rfloor$  denote the conjugate, transpose, Hermitian transpose, matrix inversion, integer ceiling and integer floor operators, respectively.  $\|\mathbf{a}\|_1$  and  $\|\mathbf{a}\|_2$  are  $\ell_1$ -norm and  $\ell_2$ -norm of  $\mathbf{a}$ , respectively, while  $\|\mathbf{A}\|_F$  is Frobenius norm of  $\mathbf{A}$ , and  $|\mathcal{Q}|_c$  is the cardinality of the set  $\mathcal{Q}$ . The Kronecker and Khatri-Rao product operations are denoted by  $\otimes$  and  $\odot$ , respectively.  $\mathbf{I}_n$  denotes the  $n \times n$  identity matrix and  $\mathbf{O}_{m \times n}$  is the null matrix of size  $m \times n$ , while  $\mathbf{1}_n$  ( $\mathbf{0}_n$ ) denotes the vector of size  $n$  with all the elements being 1 (0).  $\text{diag}(\mathbf{a})$  is the diagonal matrix with the elements of  $\mathbf{a}$  at its diagonal entries,  $\text{vdiag}(\mathbf{A})$  denotes the vector consisting of the main diagonal elements of  $\mathbf{A}$ , and  $\text{Bdiag}([\mathbf{A}_1 \cdots \mathbf{A}_n])$  denotes the block diagonal matrix with  $\mathbf{A}_1, \cdots, \mathbf{A}_n$  as its block diagonal entries. The expectation and determinant operators are denoted by  $\mathbb{E}(\cdot)$  and  $\det(\cdot)$ , respectively. The modulo operation  $\text{mod}(m, n)$  returns the remainder of dividing  $m$  by  $n$ , and  $\text{mod}(\mathcal{Q}, n)$  returns the set containing  $\text{mod}(m, n) \forall m \in \mathcal{Q}$  of the ordered set  $\mathcal{Q}$ . The operator  $\text{find}(\mathbf{a} \neq 0)$  returns the set containing the indices of nonzero elements of  $\mathbf{a}$ , and  $\text{mat}(\mathbf{a}; m, n)$  converts the vector  $\mathbf{a}$  of size  $mn$  into the matrix of size  $m \times n$  by successively selecting every  $m$  elements of  $\mathbf{a}$  as its columns. The operator  $\text{vec}(\mathbf{A})$  stacks the columns of  $\mathbf{A}$  on top of each another,  $[\mathbf{A}]_{m, n}$  denotes the  $m$ th-row and  $n$ th-column element of  $\mathbf{A}$ , and  $\mathbf{a}_{[m:n]}$  is the vector consisting the  $m$ th to  $n$ th elements of  $\mathbf{a}$ , while  $\mathbf{A}_{[:, m:n]}$  is the sub-matrix containing the  $m$ th to  $n$ th columns of  $\mathbf{A}$ .  $\mathbf{A}_{[\mathcal{Q}, :]}$  denotes the sub-matrix containing the rows of  $\mathbf{A}$  indexed in the ordered set  $\mathcal{Q}$ , and  $\mathbf{A}_{[\mathcal{Q}, i]}$  is the  $i$ th column of  $\mathbf{A}_{[\mathcal{Q}, :]}$ . Finally,  $\Re\{\cdot\}$  and  $\Im\{\cdot\}$  denote the real part and imaginary part of the argument, respectively.

## II. DOWNLINK CHANNEL ESTIMATION STAGE

Consider the mmWave FD-MIMO system with hybrid beamforming, where the BS and  $Q$  UD's are all equipped with

UPA, and OFDM with  $K$  subcarriers is adopted, while  $N_s^d$  independent signal streams are transmitted on each subcarrier [16]. The BS (UD) employs  $N_{BS} = N_{BS}^h N_{BS}^v$  ( $N_{UD} = N_{UD}^h N_{UD}^v$ ) antennas and  $N_{BS}^{\text{RF}} \ll N_{BS}$  ( $N_{UD}^{\text{RF}} \ll N_{UD}$ ) RF chains, where  $N_{BS}^h$  ( $N_{UD}^h$ ) and  $N_{BS}^v$  ( $N_{UD}^v$ ) are the numbers of antennas in horizontal and vertical directions at the BS (UD), respectively.

### A. Downlink Channel Estimation Signal Model

The downlink CE stage lasts  $N_d$  time slots and each time slot contains  $N_o^d$  OFDM symbols. The signal  $\mathbf{y}_q[k, i, m] \in \mathbb{C}^{N_s^d}$  received by the  $q$ th UD at the  $k$ th subcarrier of the  $i$ th OFDM symbol in the  $m$ th time slot can be expressed as

$$\mathbf{y}_q[k, i, m] = \mathbf{W}_{d,q}^H[k, m] \mathbf{H}_q[k] \mathbf{F}_d[k, m] \mathbf{s}[k, i, m] + \mathbf{W}_{d,q}^H[k, m] \mathbf{n}_q[k, i, m], \quad (1)$$

for  $1 \leq q \leq Q$ ,  $0 \leq k \leq K - 1$ ,  $1 \leq i \leq N_o^d$  and  $1 \leq m \leq N_d$ . In (1), the UD's receive combining matrix  $\mathbf{W}_{d,q}[k, m] = \mathbf{W}_{\text{RF},q}[m] \mathbf{W}_{\text{BB},q}[k, m] \in \mathbb{C}^{N_{UD} \times N_s^d}$  in which  $\mathbf{W}_{\text{RF},q}[m] \in \mathbb{C}^{N_{UD} \times N_{UD}^{\text{RF}}}$  and  $\mathbf{W}_{\text{BB},q}[k, m] \in \mathbb{C}^{N_{UD}^{\text{RF}} \times N_s^d}$  are the analog and digital receive combining matrices, and the BS's transmit precoding matrix  $\mathbf{F}_d[k, m] = \mathbf{F}_{\text{RF},d}[m] \mathbf{F}_{\text{BB},d}[k, m] \in \mathbb{C}^{N_{BS} \times N_s^d}$  in which  $\mathbf{F}_{\text{RF},d}[m] \in \mathbb{C}^{N_{BS} \times N_{BS}^{\text{RF}}}$  and  $\mathbf{F}_{\text{BB},d}[k, m] \in \mathbb{C}^{N_{BS}^{\text{RF}} \times N_s^d}$  are the analog and digital transmit precoding matrices, respectively, while  $\mathbf{H}_q[k] \in \mathbb{C}^{N_{UD} \times N_{BS}}$  is the corresponding downlink channel matrix,  $\mathbf{s}[k, i, m] \in \mathbb{C}^{N_s^d}$  is the training signal with  $\mathbb{E}(\mathbf{s}[k, i, m] \mathbf{s}^H[k, i, m]) = \frac{1}{N_s^d} \mathbf{I}_{N_s^d}$ , and  $\mathbf{n}_q[k, i, m] \in \mathbb{C}^{N_{UD}}$  is the complex additive white Gaussian noise (AWGN) vector with the covariance matrix  $\sigma_n^2 \mathbf{I}_{N_{UD}}$ , i.e.,  $\mathbf{n}_q[k, i, m] \sim \mathcal{CN}(\mathbf{0}_{N_{UD}}, \sigma_n^2 \mathbf{I}_{N_{UD}})$ . Due to the constant modulus of the phase shift network (PSN),  $[\mathbf{F}_{\text{RF},d}[m]]_{j_1, j_2} = \frac{1}{\sqrt{N_{BS}}} e^{j\vartheta_{1, j_1, j_2}}$  and  $[\mathbf{W}_{\text{RF},q}[m]]_{j_1, j_2} = \frac{1}{\sqrt{N_{UD}}} e^{j\vartheta_{2, j_1, j_2}}$  with  $\vartheta_{1, j_1, j_2}, \vartheta_{2, j_1, j_2} \in \mathcal{A}$ , and  $\mathcal{A}$  is the quantized phase set of the PSN with the resolution  $N_q^{\text{ps}}$ , given by

$$\mathcal{A} = \left\{ -\pi, -\pi + \frac{2\pi}{2^{N_q^{\text{ps}}}}, -\pi + 2 \cdot \frac{2\pi}{2^{N_q^{\text{ps}}}}, \dots, \pi - \frac{2\pi}{2^{N_q^{\text{ps}}}} \right\}. \quad (2)$$

Also  $\|\mathbf{F}_d[k, m]\|_F^2 \leq N_{BS}^{\text{RF}}$  to guarantee the constraint on the total transmit power [5]. Here at the CE stage, some elegant solutions [39], [42] can be used to acquire the robust synchronization of burst training signals without the knowledge of noise/interference power even at low SNR.

Due to the obviously resolvable delay spread for each MPC caused by the large bandwidth, according to the typical mmWave channel model [16]–[21], [25], the downlink delay-domain continuous channel matrix  $\mathbf{H}_q(\tau) \in \mathbb{C}^{N_{UD} \times N_{BS}}$  with  $L_q$  MPCs can be expressed as

$$\mathbf{H}_q(\tau) = \beta_q \sum_{l=1}^{L_q} \mathbf{H}_{q,l} p(\tau - \tau_{q,l}), \quad (3)$$

where  $\beta_q = \sqrt{N_{UD} N_{BS} / L_q}$  is the normalization factor,  $\tau_{q,l}$  is the delay of the  $l$ th MPC, and  $p(\tau)$  denotes the equivalent PSF, while the complex gain matrix  $\mathbf{H}_{q,l} \in \mathbb{C}^{N_{UD} \times N_{BS}}$  is given by

$$\mathbf{H}_{q,l} = \alpha_{q,l} \mathbf{a}_{UD}(\mu_{q,l}^{\text{UD}}, \nu_{q,l}^{\text{UD}}) \mathbf{a}_{BS}^H(\mu_{q,l}^{\text{BS}}, \nu_{q,l}^{\text{BS}}), \quad (4)$$

where  $\alpha_{q,l} \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the associated complex path gain,  $\mu_{q,l}^{\text{UD}} = \pi \sin(\theta_{q,l}^{\text{UD}}) \cos(\varphi_{q,l}^{\text{UD}})$  ( $\mu_{q,l}^{\text{BS}} = \pi \sin(\theta_{q,l}^{\text{BS}}) \cos(\varphi_{q,l}^{\text{BS}})$ ) and  $\nu_{q,l}^{\text{UD}} = \pi \sin(\varphi_{q,l}^{\text{UD}})$  ( $\nu_{q,l}^{\text{BS}} = \pi \sin(\varphi_{q,l}^{\text{BS}})$ ) denote the horizontally and vertically spatial frequencies with half-wavelength antenna spacing at the  $q$ th UD (the BS), respectively. Here,  $\theta_{q,l}^{\text{UD}}$  ( $\theta_{q,l}^{\text{BS}}$ ) and  $\varphi_{q,l}^{\text{UD}}$  ( $\varphi_{q,l}^{\text{BS}}$ ) are the downlink horizontal and vertical AoAs (AoDs) of the  $l$ th MPC associated with the UPA, respectively. The array response vector at UD is given by  $\mathbf{a}_{\text{UD}}(\mu_{q,l}^{\text{UD}}, \nu_{q,l}^{\text{UD}}) = \mathbf{a}_v(\nu_{q,l}^{\text{UD}}) \otimes \mathbf{a}_h(\mu_{q,l}^{\text{UD}}) \in \mathbb{C}^{N_{\text{UD}}}$  [24], [30], [36], in which

$$\mathbf{a}_h(\mu_{q,l}^{\text{UD}}) = \frac{1}{\sqrt{N_{\text{UD}}^h}} [1 e^{j\mu_{q,l}^{\text{UD}}} \dots e^{j(N_{\text{UD}}^h-1)\mu_{q,l}^{\text{UD}}}]^T, \quad (5)$$

$$\mathbf{a}_v(\nu_{q,l}^{\text{UD}}) = \frac{1}{\sqrt{N_{\text{UD}}^v}} [1 e^{j\nu_{q,l}^{\text{UD}}} \dots e^{j(N_{\text{UD}}^v-1)\nu_{q,l}^{\text{UD}}}]^T. \quad (6)$$

are the steering vectors associated with the horizontal and vertical directions, respectively. Similarly, the array response vector at BS is given by  $\mathbf{a}_{\text{BS}}(\mu_{q,l}^{\text{BS}}, \nu_{q,l}^{\text{BS}}) = \mathbf{a}_v(\nu_{q,l}^{\text{BS}}) \otimes \mathbf{a}_h(\mu_{q,l}^{\text{BS}}) \in \mathbb{C}^{N_{\text{BS}}}$ , where the horizontal and vertical direction steering vectors  $\mathbf{a}_h(\mu_{q,l}^{\text{BS}}) \in \mathbb{C}^{N_{\text{BS}}^h}$  and  $\mathbf{a}_v(\nu_{q,l}^{\text{BS}}) \in \mathbb{C}^{N_{\text{BS}}^v}$  are given respectively by substituting  $\mu_{q,l}^{\text{UD}}$  and  $N_{\text{UD}}^h$  with  $\mu_{q,l}^{\text{BS}}$  and  $N_{\text{BS}}^h$  in (5) as well as by substituting  $\nu_{q,l}^{\text{UD}}$  and  $N_{\text{UD}}^v$  with  $\nu_{q,l}^{\text{BS}}$  and  $N_{\text{BS}}^v$  in (6).

The frequency-domain channel matrix  $\mathbf{H}_q[k]$  at the  $k$ th subcarrier can then be expressed as

$$\begin{aligned} \mathbf{H}_q[k] &= \beta_q \sum_{l=1}^{L_q} \mathbf{H}_{q,l} e^{-j\frac{2\pi k f_s \tau_{q,l}}{K}} \\ &= \beta_q \sum_{l=1}^{L_q} \alpha_{q,l} \mathbf{a}_{\text{UD}}(\mu_{q,l}^{\text{UD}}, \nu_{q,l}^{\text{UD}}) \mathbf{a}_{\text{BS}}^H(\mu_{q,l}^{\text{BS}}, \nu_{q,l}^{\text{BS}}) e^{-j\frac{2\pi k f_s \tau_{q,l}}{K}}, \end{aligned} \quad (7)$$

where  $f_s = 1/T_s$  denotes the system bandwidth, and  $T_s$  is the sampling period. The derivation of the first equation in (7) is shown in Appendix. Observe that  $\mathbf{H}_q[k]$  does not depend on the PSF, and it exhibits the sparsity in delay domain due to small  $L_q$  but large normalized delay spread. Recall that the existing CS-based solutions of [18], [20] have to estimate the effective delay-domain CIRs that include the PSF, and this PSF will destroy the delay-domain sparsity of mmWave channels when the order of PSF is large.  $\mathbf{H}_q[k]$  in (7) can be rewritten as

$$\mathbf{H}_q[k] = \mathbf{A}_{\text{UD},q} \mathbf{D}_q[k] \mathbf{A}_{\text{BS},q}^H, \quad (8)$$

where  $\mathbf{D}_q[k] = \text{diag}(\mathbf{d}_q[k]) \in \mathbb{C}^{L_q \times L_q}$  is the diagonal matrix in which  $\mathbf{d}_q[k] = \text{diag}(\alpha_q) \boldsymbol{\tau}_q[k]$  with  $\alpha_q = \beta_q [\alpha_{q,1} \dots \alpha_{q,L_q}]^T$  and  $\boldsymbol{\tau}_q[k] = [e^{-j2\pi k f_s \tau_{q,1}/K} \dots e^{-j2\pi k f_s \tau_{q,L_q}/K}]^T$ , and the array response matrix associated with the AoAs of the  $q$ th UD  $\mathbf{A}_{\text{UD},q} \in \mathbb{C}^{N_{\text{UD}} \times L_q}$  can be expressed as  $\mathbf{A}_{\text{UD},q} = \mathbf{A}_{\text{UD},q}^v \odot \mathbf{A}_{\text{UD},q}^h$  in which  $\mathbf{A}_{\text{UD},q}^h = [\mathbf{a}_h(\mu_{q,1}^{\text{UD}}) \dots \mathbf{a}_h(\mu_{q,L_q}^{\text{UD}})] \in \mathbb{C}^{N_{\text{UD}}^h \times L_q}$  and  $\mathbf{A}_{\text{UD},q}^v = [\mathbf{a}_v(\nu_{q,1}^{\text{UD}}) \dots \mathbf{a}_v(\nu_{q,L_q}^{\text{UD}})] \in \mathbb{C}^{N_{\text{UD}}^v \times L_q}$  are the steering matrices corresponding to the horizontally and vertically spatial frequencies, respectively, while  $\mathbf{A}_{\text{BS},q} = \mathbf{A}_{\text{BS},q}^v \odot \mathbf{A}_{\text{BS},q}^h \in \mathbb{C}^{N_{\text{BS}} \times L_q}$  is the array response

matrix associated with the AoDs in which the steering matrices  $\mathbf{A}_{\text{BS},q}^h \in \mathbb{C}^{N_{\text{BS}}^h \times L_q}$  and  $\mathbf{A}_{\text{BS},q}^v \in \mathbb{C}^{N_{\text{BS}}^v \times L_q}$  have the similar form as  $\mathbf{A}_{\text{UD},q}^h$  and  $\mathbf{A}_{\text{UD},q}^v$ , respectively.

### B. Obtain Horizontal/Vertical AoAs at UD

The downlink CE corresponds to Step 1 to Step 4 of Fig. 2, where the horizontal and vertical AoAs are estimated. We first assume that the training signal  $\mathbf{s}[i, m]$  is independent of subcarriers, and its  $j_1$ th element can be designed as  $[s[i, m]]_{j_1} = \frac{1}{N_s^d} e^{j2\pi\phi_{j_1}}$  with  $\phi_{j_1}$  randomly and uniformly selected from the interval  $[0, 1]$ , i.e.,  $\phi_{j_1} \sim \mathcal{U}[0, 1]$ . Second, a predefined frequency-domain scrambling code  $\mathbf{x}_d \in \mathbb{C}^K$  with its  $k$ th element being  $x_d[k]$  for  $0 \leq k \leq K-1$  can be introduced to effectively avoid the high peak-to-average power ratio (PAPR) resulted from the same training signal used at all subcarriers.<sup>3</sup> Then, we can obtain the scrambled training signal  $\mathbf{s}[k, i, m]$  at the  $k$ th subcarrier, i.e.,  $\mathbf{s}[k, i, m] = x_d[k] \mathbf{s}[i, m]$ . The signals received at the UD will be first descrambled by multiplying the conjugate of scrambling code  $\mathbf{x}_d^*$ , which indicates that the scrambling code  $\mathbf{x}_d$  does not affect the subsequent signal processing. Moreover, the same digital transmit precoding/receive combining matrices are adopted at every subcarrier, i.e.,  $\mathbf{F}_{\text{BB},d}[k, m] = \mathbf{F}_{\text{BB},d}[m]$  and  $\mathbf{W}_{\text{BB},q}[k, m] = \mathbf{W}_{\text{BB},q}[m]$ , for  $0 \leq k \leq K-1$ . The number of independent signal streams associated with each subcarrier in each OFDM symbol is  $N_s^d \leq N_{\text{UD}}^{\text{RF}}$ . We can visualize a low-dimensional digital  $M_{\text{UD}}^h \times M_{\text{UD}}^v$  sub-UPA, in which  $M_{\text{UD}}^h$  and  $M_{\text{UD}}^v$  are the numbers of antennas in horizontal and vertical directions, from the high-dimensional hybrid analog/digital  $N_{\text{UD}}^h \times N_{\text{UD}}^v$  UPA. Given  $N_{\text{UD}}^{\text{sub}} = M_{\text{UD}}^h M_{\text{UD}}^v$ , the BS only requires  $N_d = \lceil N_{\text{UD}}^{\text{sub}} / N_s^d \rceil$  time slots to broadcast training signals, with each time slot containing  $N_o^d$  OFDM symbols. The choice of  $M_{\text{UD}}^h$ ,  $M_{\text{UD}}^v$  and  $N_o^d$  trades off estimation accuracy with training overhead,<sup>4</sup> because larger  $M_{\text{UD}}^h$ ,  $M_{\text{UD}}^v$  and  $N_o^d$  lead to better estimation accuracy but higher training overhead, and vice versa. Since the signals received by all UDs have the same form, we can focus on the  $q$ th UD and the user index  $q$  can be omitted from  $\mathbf{y}_q[k, i, m]$ ,  $\mathbf{W}_{d,q}[m]$ ,  $\mathbf{H}_q[k]$ ,  $\mathbf{n}_q[k, i, m]$ ,  $\mathbf{A}_{\text{UD},q}$ ,  $\mathbf{D}_q[k]$ ,  $\mathbf{A}_{\text{BS},q}$  and other relevant variables for clarity.

By collecting the received signals of (1) associated with the  $k$ th subcarrier over all the  $N_o^d$  OFDM symbols of the  $m$ th time slot into the signal matrix  $\mathbf{Y}_m[k] = [\mathbf{y}[k, 1, m] \dots \mathbf{y}[k, N_o^d, m]] \in \mathbb{C}^{N_s^d \times N_o^d}$ , we have

$$\mathbf{Y}_m[k] = \mathbf{x}_d^*[k] \mathbf{W}_d^H[m] \mathbf{H}[k] \mathbf{F}_d[m] \mathbf{S}_d[k, m] + \mathbf{W}_d^H[m] \mathbf{N}_m[k], \quad (9)$$

where  $\mathbf{S}_d[k, m] = [s[k, 1, m] \dots s[k, N_o^d, m]] = x_d[k] \mathbf{S}_d[m]$  with  $\mathbf{S}_d[m] = [s[1, m] \dots s[N_o^d, m]] \in \mathbb{C}^{N_s^d \times N_o^d}$ , and  $\mathbf{N}_m[k] = [\mathbf{n}[k, 1, m] \dots \mathbf{n}[k, N_o^d, m]] \in \mathbb{C}^{N_{\text{UD}} \times N_o^d}$ . Since the BS transmits the common random signal  $\mathbf{F}_d[m] \mathbf{S}_d[m]$ ,

<sup>3</sup>Each element in the predefined scrambling code  $\mathbf{x}_d$  should satisfy  $x_d^*[k] x_d[k] = 1, 0 \leq k \leq K-1$ . To achieve the low PAPR of training signals, we can adopt the constant-module Zadoff-Chu sequence as the scrambling code  $\mathbf{x}_d$ .

<sup>4</sup>In this paper, the training overhead is defined as the number of OFDM symbols required at the CE stage. In terms of downlink CE stage, the training duration is  $N_d N_o^d$  OFDM symbols.

the transmit precoding matrix  $\mathbf{F}_d[m] = \mathbf{F}_{\text{RF},d}[m]\mathbf{F}_{\text{BB},d}[m]$  should be a random matrix. This is achieved by designing  $\mathbf{F}_{\text{RF},d}[m]$  as  $[\mathbf{F}_{\text{RF},d}[m]]_{j_1,j_2} = \frac{1}{\sqrt{N_{\text{BS}}}} e^{j\vartheta_{3,j_1,j_2}}$  with  $\vartheta_{3,j_1,j_2}$  randomly and uniformly selected from  $\mathcal{A}$ , and designing  $\mathbf{F}_{\text{BB},d}[m]$  as  $[\mathbf{F}_{\text{BB},d}[m]]_{j_1,j_2} = e^{j2\pi a_{j_1,j_2}}$  with  $a_{j_1,j_2} \sim \mathcal{U}[0,1]$ . The BS can use the same transmit precoding matrix  $\mathbf{F}_d = \mathbf{F}_d[m]$  to send the same sounding signal  $\mathbf{S}_d = \mathbf{S}_d[m]$  for every time slot. By stacking the received signal matrices  $\mathbf{Y}_m[k]$  of (9) over the  $N_d$  time slots into  $\tilde{\mathbf{Y}}_d[k] = [\mathbf{Y}_1^T[k] \cdots \mathbf{Y}_{N_d}^T[k]]^T \in \mathbb{C}^{N_d N_s^d \times N_o^d}$ , we have

$$\tilde{\mathbf{Y}}_d[k] = \tilde{\mathbf{W}}_d^H \mathbf{A}_{\text{UD}} \mathbf{D}[k] \mathbf{A}_{\text{BS}}^H \mathbf{F}_d \mathbf{S}_d + \text{Bdiag}(\tilde{\mathbf{W}}_d) \tilde{\mathbf{N}}_d[k], \quad (10)$$

where  $\tilde{\mathbf{W}}_d = [\mathbf{W}_d[1] \cdots \mathbf{W}_d[N_d]] \in \mathbb{C}^{N_{\text{UD}} \times N_d N_s^d}$  aggregates the downlink receive combining matrices used in the  $N_d$  time slots, and  $\text{Bdiag}(\tilde{\mathbf{W}}_d) = \text{Bdiag}([\mathbf{W}_d^H[1] \cdots \mathbf{W}_d^H[N_d]]) \in \mathbb{C}^{N_d N_s^d \times N_d N_{\text{UD}}}$ , while  $\tilde{\mathbf{N}}_d[k] = [\mathbf{N}_1^T[k] \cdots \mathbf{N}_{N_d}^T[k]]^T \in \mathbb{C}^{N_d N_{\text{UD}} \times N_o^d}$  is the corresponding noise matrix.

Multiplying  $\tilde{\mathbf{Y}}_d[k]$  with  $\mathbf{J}_d = [\mathbf{I}_{N_{\text{UD}}^{\text{sub}}} \mathbf{O}_{N_{\text{UD}}^{\text{sub}} \times (N_d N_s^d - N_{\text{UD}}^{\text{sub}})}]$  and aggregating the resulting signals over all the  $K$  subcarriers lead to the signal matrix  $\tilde{\mathbf{Y}}_d = [\mathbf{J}_d \tilde{\mathbf{Y}}_d[0] \mathbf{J}_d \tilde{\mathbf{Y}}_d[1] \cdots \mathbf{J}_d \tilde{\mathbf{Y}}_d[K-1]] \in \mathbb{C}^{N_{\text{UD}}^{\text{sub}} \times K N_o^d}$  as

$$\tilde{\mathbf{Y}}_d = \tilde{\mathbf{A}}_{\text{UD}} \tilde{\mathbf{S}}_d + \tilde{\mathbf{N}}_d, \quad (11)$$

where  $\tilde{\mathbf{N}}_d = \mathbf{J}_d \text{Bdiag}(\tilde{\mathbf{W}}_d) [\tilde{\mathbf{N}}_d[0] \tilde{\mathbf{N}}_d[1] \cdots \tilde{\mathbf{N}}_d[K-1]]$ ,  $\tilde{\mathbf{A}}_{\text{UD}} = \mathbf{J}_d \tilde{\mathbf{W}}_d^H \mathbf{A}_{\text{UD}}$ , and  $\tilde{\mathbf{S}}_d = [\tilde{\mathbf{S}}_d[0] \tilde{\mathbf{S}}_d[1] \cdots \tilde{\mathbf{S}}_d[K-1]]$  with  $\tilde{\mathbf{S}}_d[k] = \mathbf{D}[k] \mathbf{A}_{\text{BS}}^H \mathbf{F}_d \mathbf{S}_d$ . Observe from (10) and (11) that we cannot directly apply powerful spectrum estimation techniques [30], [37] to estimate the horizontal/vertical AoAs from  $\tilde{\mathbf{Y}}_d$ , since the shift-invariance structure of the array response matrix  $\mathbf{A}_{\text{UD}}$  does not hold in hybrid receive array [31], [35]. We propose to visualize the high-dimensional hybrid array as a low-dimensional digital array by designing appropriate aggregated receive combining matrix  $\tilde{\mathbf{W}}_d$  so that the shift-invariance structure of array response can be reconstructed and therefore super-resolution CE based on spectrum estimation techniques can be harnessed.

### C. Design Receive Combining Matrix at UD

Without loss of generality, we consider  $N_s^d = N_{\text{UD}}^{\text{RF}} - 1$  independent signal streams. First, we utilize a unitary matrix  $\mathbf{U}_{N_{\text{UD}}^{\text{RF}}} = [\mathbf{u}_1 \cdots \mathbf{u}_{N_{\text{UD}}^{\text{RF}}}] \in \mathbb{C}^{N_{\text{UD}}^{\text{RF}} \times N_{\text{UD}}^{\text{RF}}}$  to design the digital receive combining matrix  $\mathbf{W}_{\text{BB}}[m] \in \mathbb{C}^{N_{\text{UD}}^{\text{RF}} \times N_s^d}$  of the  $m$ th time slot's receive combining matrix  $\mathbf{W}_d[m] = \mathbf{W}_{\text{RF}}[m] \mathbf{W}_{\text{BB}}[m]$  for  $1 \leq m \leq N_d$ . Specifically,  $\mathbf{W}_{\text{BB}}[m] = \mathbf{U}_{N_{\text{UD}}^{\text{RF}}[:,1:N_s^d]}$ . To design the analog receive combining matrix  $\mathbf{W}_{\text{RF}}[m] \in \mathbb{C}^{N_{\text{UD}} \times N_{\text{UD}}^{\text{RF}}}$ , we construct the matrix  $\tilde{\mathbf{E}}_d \in \mathbb{R}^{N_{\text{UD}} \times N_d N_s^d}$  as

$$\tilde{\mathbf{E}}_d = \begin{bmatrix} \mathbf{I}_{M_{\text{UD}}^{\text{v}}+1} \otimes \mathbf{B} \\ \mathbf{O}_{(N_{\text{UD}} - N_{\text{UD}}^{\text{h}})(M_{\text{UD}}^{\text{v}}+1)} \times M_{\text{UD}}^{\text{h}}(M_{\text{UD}}^{\text{v}}+1) \end{bmatrix} \times \begin{bmatrix} \mathbf{I}_{N_s^d N_d} \\ \mathbf{O}_{(M_{\text{UD}}^{\text{h}}(M_{\text{UD}}^{\text{v}}+1) - N_s^d N_d) \times N_s^d N_d} \end{bmatrix}, \quad (12)$$

where  $\mathbf{B} = [\mathbf{I}_{M_{\text{UD}}^{\text{h}}} \mathbf{O}_{M_{\text{UD}}^{\text{h}} \times (N_{\text{UD}}^{\text{h}} - M_{\text{UD}}^{\text{h}})}]^T \in \mathbb{R}^{N_{\text{UD}}^{\text{h}} \times M_{\text{UD}}^{\text{h}}}$ . Then we take the sub-matrix  $\tilde{\mathbf{E}}_{d,m}^{\text{sub}} = \tilde{\mathbf{E}}_d[:,(m-1)N_s^d+1:mN_s^d] \in$

### Algorithm 1 Proposed Receive Combining Matrix Design

**Input:**  $N_d, N_s^d, N_{\text{UD}}^{\text{RF}}, N_{\text{UD}}^{\text{h}}, N_{\text{UD}}^{\text{v}}, M_{\text{UD}}^{\text{h}}, M_{\text{UD}}^{\text{v}}$

**Output:**  $\tilde{\mathbf{W}}_d$

- 1: Generate unitary matrix  $\mathbf{U}_{N_{\text{UD}}^{\text{RF}}} = [\mathbf{u}_1 \cdots \mathbf{u}_{N_{\text{UD}}^{\text{RF}}}]$
- 2: Construct index matrix  $\tilde{\mathbf{E}}_d$  of (12)
- 3: **for**  $m = 1, 2, \dots, N_d$  **do**
- 4:  $\mathbf{W}_{\text{BB}}[m] = \mathbf{U}_{N_{\text{UD}}^{\text{RF}}[:,1:N_s^d]}$ , and initialize  $\mathbf{W}_{\text{RF}}[m] = \mathbf{1}_{N_{\text{UD}}} \otimes \mathbf{u}_{N_{\text{UD}}^{\text{RF}}}^{\text{H}}$
- 5: Extract  $\tilde{\mathbf{E}}_{d,m}^{\text{sub}} = \tilde{\mathbf{E}}_d[:,(m-1)N_s^d+1:mN_s^d]$ , and obtain  $\tilde{\xi}_{d,m} = \text{vec}(\tilde{\mathbf{E}}_{d,m}^{\text{sub}})$
- 6: Obtain ordered index set  $\mathcal{I}_m = \text{mod}(\text{find}(\tilde{\xi}_{d,m} \neq 0), N_{\text{UD}})$
- 7: Replace  $\mathbf{W}_{\text{RF}}[m]_{[\mathcal{I}_m,:]} \leftarrow \mathbf{W}_{\text{BB}}^{\text{H}}[m]$
- 8: Quantize phase values of  $\mathbf{W}_{\text{RF}}[m]$  based on  $\mathcal{A}$
- 9:  $\mathbf{W}_d[m] = \mathbf{W}_{\text{RF}}[m] \mathbf{W}_{\text{BB}}[m]$
- 10: **end for**
- 11: **return**  $\tilde{\mathbf{W}}_d = [\mathbf{W}_d[1] \cdots \mathbf{W}_d[N_d]]$

$\mathbb{R}^{N_{\text{UD}} \times N_s^d}$  and define  $\tilde{\xi}_{d,m} = \text{vec}(\tilde{\mathbf{E}}_{d,m}^{\text{sub}})$  to construct the ordered index set  $\mathcal{D}_m = \text{find}(\tilde{\xi}_{d,m} \neq 0)$  with  $|\mathcal{D}_m|_c = N_s^d$ . Next we perform the modulo operation on  $\mathcal{D}_m$  with  $N_{\text{UD}}$  to get the ordered index set  $\mathcal{I}_m = \text{mod}(\mathcal{D}_m, N_{\text{UD}})$  with  $|\mathcal{I}_m|_c = N_s^d$ . The rows of  $\mathbf{W}_{\text{RF}}[m]$  whose indices correspond to  $\mathcal{I}_m$  are determined by  $\mathbf{W}_{\text{BB}}[m]$  as  $\mathbf{W}_{\text{RF}}[m]_{[\mathcal{I}_m,:]} = \mathbf{W}_{\text{BB}}^{\text{H}}[m]$ , while the rest rows of  $\mathbf{W}_{\text{RF}}[m]$  consist of the  $(N_{\text{UD}} - N_s^d)$  identical  $\mathbf{u}_{N_{\text{UD}}^{\text{RF}}}^{\text{H}}$ . The phase value of arbitrary element in the designed  $\mathbf{W}_{\text{RF}}[m]$ , denoted by  $\vartheta_d$ , is then quantized to  $\vartheta \in \mathcal{A}$  by minimizing the Euclidean distance according to  $\arg \min_{\vartheta \in \mathcal{A}} \|\vartheta_d - \vartheta\|_2$ . Thus, the  $m$ th receive combining matrix can be obtained as  $\mathbf{W}_d[m] = \mathbf{W}_{\text{RF}}[m] \mathbf{W}_{\text{BB}}[m]$  for  $1 \leq m \leq N_d$ .

The proposed design for  $\tilde{\mathbf{W}}_d$  is summarized in Algorithm 1. Since the number of RF chains is usually the power of 2, we can adopt Hadamard matrix for  $N_q^{\text{ps}} \geq 1$  or discrete Fourier transform (DFT) matrix for  $N_q^{\text{ps}} \geq 2$  to construct  $\mathbf{U}_{N_{\text{UD}}^{\text{RF}}}$ .<sup>5</sup> Clearly, our design can be used for the PSN with arbitrary  $N_q^{\text{ps}}$ , even the extremely low resolution PSN with  $N_q^{\text{ps}} = 1$ . With the designed  $\tilde{\mathbf{W}}_d$ , we have

$$\tilde{\mathbf{A}}_{\text{UD}} = \mathbf{J}_d \tilde{\mathbf{W}}_d^H (\mathbf{A}_{\text{UD}}^{\text{v}} \odot \mathbf{A}_{\text{UD}}^{\text{h}}) = \tilde{\mathbf{A}}_{\text{UD}}^{\text{v}} \odot \tilde{\mathbf{A}}_{\text{UD}}^{\text{h}}. \quad (13)$$

Clearly,  $\tilde{\mathbf{A}}_{\text{UD}}^{\text{h}} \in \mathbb{C}^{M_{\text{UD}}^{\text{h}} \times L}$  ( $\tilde{\mathbf{A}}_{\text{UD}}^{\text{v}} \in \mathbb{C}^{M_{\text{UD}}^{\text{v}} \times L}$ ) is the matrix containing the first  $M_{\text{UD}}^{\text{h}}$  ( $M_{\text{UD}}^{\text{v}}$ ) rows of  $\mathbf{A}_{\text{UD}}^{\text{h}}$  ( $\mathbf{A}_{\text{UD}}^{\text{v}}$ ). Thus,  $\tilde{\mathbf{A}}_{\text{UD}}$  maintains the double shift-invariance structure of the original array response matrix  $\mathbf{A}_{\text{UD}}$  for both horizontal and vertical AoAs [35], and the designed  $\tilde{\mathbf{W}}_d$  can be used to visualize the high-dimensional hybrid analog/digital array as a low-dimensional digital array. Therefore, we can utilize the MDU-ESPRIT algorithm detailed in Section IV to obtain the super-resolution estimates of horizontal/vertical AoAs at UD. Since the ESPRIT-type algorithms [30], [31], [35], [37] require the knowledge of the number of MPCs, we next turn to the task

<sup>5</sup>Note that the phase value of every entry of the quantized Hadamard or DFT matrices still belongs to the set  $\mathcal{A}$ , and therefore we ensure the columns of the selected  $\mathbf{U}_{N_{\text{UD}}^{\text{RF}}}$  to be mutually orthogonal.



of acquiring the number of MPCs at the receiver, i.e., Step 2 of Fig. 2.

#### D. EVD-Based Estimate for Number of MPCs

In OFDM systems, the channels of multiple adjacent subcarriers within coherence bandwidth are highly correlated. If the maximum delay spread is  $\tau_{\max} = N_c T_s$  with  $N_c$  delay taps, the channel coherence bandwidth is  $B_c \approx \frac{1}{\tau_{\max}} = \frac{f_s}{N_c}$ . Then we can jointly use the measurements of  $P \leq \frac{B_c}{\Delta f} = \frac{K}{N_c}$  adjacent subcarriers to estimate the number of MPCs, where  $\Delta f = \frac{f_s}{K}$  is the subcarrier's bandwidth. Specifically, by dividing  $K$  signal matrices  $\{J_d \tilde{Y}_d[k]\}_{k=0}^{K-1}$  into  $N_P = \lfloor K/P \rfloor$  groups, we can obtain the  $n_p$ th measurement matrix  $\tilde{Y}_d[n_p] \in \mathbb{C}^{N_{\text{UD}}^{\text{sub}} \times N_o^d}$ , as the average of the measurements in the  $n_p$ th group

$$\tilde{Y}_d[n_p] = \frac{1}{P} \sum_{k=(n_p-1)P}^{n_p P-1} J_d \tilde{Y}_d[k], \quad 1 \leq n_p \leq N_P. \quad (14)$$

The  $N_P$  average measurements are collected as  $\tilde{Y}_d = [\tilde{Y}_d[1] \cdots \tilde{Y}_d[N_P]] \in \mathbb{C}^{N_{\text{UD}}^{\text{sub}} \times N_o^d N_P}$ , and the covariance matrix of  $\tilde{Y}_d$  is  $\mathbf{R}_d = \frac{1}{N_o^d N_P} \tilde{Y}_d \tilde{Y}_d^H$ . According to the eigenvalue decomposition (EVD), we obtain  $\mathbf{R}_d = [\mathbf{U}_s \mathbf{U}_n] \text{diag}(\boldsymbol{\lambda}_d) [\mathbf{U}_s \mathbf{U}_n]^H$ , where  $\boldsymbol{\lambda}_d = [\lambda_1 \cdots \lambda_L \lambda_{L+1} \cdots \lambda_{N_{\text{UD}}^{\text{sub}}}]^T = [\boldsymbol{\lambda}_s^T \boldsymbol{\lambda}_n^T]^T$  is the eigenvalue vector with the eigenvalues arranged in descending order,  $\mathbf{U}_s$  and  $\mathbf{U}_n$  are the eigenvector matrices corresponding to the signal and noise subspaces, respectively, while  $\boldsymbol{\lambda}_s = [\lambda_1 \cdots \lambda_L]^T$  and  $\boldsymbol{\lambda}_n = [\lambda_{L+1} \cdots \lambda_{N_{\text{UD}}^{\text{sub}}}]^T$  are the eigenvalue vectors related to  $\mathbf{U}_s$  and  $\mathbf{U}_n$ , respectively. The number of MPCs  $L$  is the dimension of  $\boldsymbol{\lambda}_s$ .

To obtain an accurate estimate of  $L$ , we first construct  $\tilde{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_s^T \mathbf{0}_{N_{\text{UD}}^{\text{sub}}-L}^T]^T \in \mathbb{C}^{N_{\text{UD}}^{\text{sub}}}$ . The optimal estimate of  $\tilde{\boldsymbol{\lambda}}$  can be acquired by solving the following optimization problem

$$\tilde{\boldsymbol{\lambda}}^* = \arg \min_{\tilde{\boldsymbol{\lambda}} \geq \mathbf{0}_{N_{\text{UD}}^{\text{sub}}}} \frac{1}{2} \|\tilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_d\|_2^2 + \varepsilon \|\tilde{\boldsymbol{\lambda}}\|_1, \quad (15)$$

where  $\varepsilon$  is the threshold parameter related to the AWGN power, which is determined experimentally. Clearly, the solution to the optimization problem (15) is [38]

$$\tilde{\lambda}_i^* = \begin{cases} \lambda_i - \varepsilon, & \lambda_i \geq \varepsilon \\ 0, & \lambda_i < \varepsilon, \end{cases} \quad (16)$$

where  $\tilde{\lambda}_i^*$  is the  $i$ th element of  $\tilde{\boldsymbol{\lambda}}^*$ . From the estimate  $\tilde{\boldsymbol{\lambda}}^*$ , we obtain the estimate of the number of MPCs, denoted by  $\hat{L}$ , which is the input to the MDU-ESPRIT algorithm for estimating  $\hat{L}$  pairs of horizontal and vertical AoAs. The resulting estimates  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  are quantized as  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  with  $N_q^{\text{ang}}$  angle quantized bits in  $[-\pi/2, \pi/2]$ .

Finally, only the few bits of the quantized angle estimates are fed back to BS through the low-frequency control link with limited resource [2]. Thus, since very little data needs to be transmitted via the feedback link, the feedback overhead

at the AoAs feedback stage can be ignored in our proposed closed-loop sparse CE scheme.<sup>6</sup>

### III. UPLINK CHANNEL ESTIMATION STAGE

#### A. Obtain Horizontal/Vertical AoDs and Delays at BS

At the uplink CE stage, the BS jointly estimates the horizontal/vertical AoDs and delays for each UD. Due to the channel reciprocity in TDD systems [32]–[34], the uplink channel matrix for the  $q$ th UD is given by  $\mathbf{H}^T[k] = \mathbf{A}_{\text{BS}}^* \mathbf{D}[k] \mathbf{A}_{\text{UD}}^T \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{UD}}}$ , where again the user index  $q$  is omitted. We employ  $N_s^u = N_{\text{BS}}^{\text{RF}} - 1$  independent signal streams, and a low-dimensional digital  $M_{\text{BS}}^h \times M_{\text{BS}}^v$  sub-UPA with  $M_{\text{BS}}^h$  and  $M_{\text{BS}}^v$  antennas in horizontal and vertical directions is visualized from the high-dimensional hybrid analog/digital  $N_{\text{BS}}^h \times N_{\text{BS}}^v$  UPA at the BS. Each UD requires  $N_u = \lceil N_{\text{BS}}^{\text{sub}} / N_s^u \rceil$  time slots with  $N_{\text{BS}}^{\text{sub}} = M_{\text{BS}}^h M_{\text{BS}}^v$  to transmit the training signals, and each time slot consists of  $N_o^u$  OFDM symbols. Hence, the uplink CE for  $Q$  UDs has a training overhead of  $Q N_u N_o^u$ , and the total training overhead of the proposed closed-loop sparse CE scheme is  $T_{\text{CE}} = N_d N_o^d + Q N_u N_o^u$ . Similar to (10), after the frequency-domain scrambling/descrambling operation, the signal matrix  $\tilde{Y}_u[k] \in \mathbb{C}^{N_u N_s^u \times N_o^u}$  received by the BS at the  $k$ th subcarrier and over the  $N_u$  time slots can be expressed as

$$\tilde{Y}_u[k] = \tilde{\mathbf{W}}_u^H \mathbf{A}_{\text{BS}}^* \mathbf{D}[k] \mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u + \text{Bdiag}(\tilde{\mathbf{W}}_u) \tilde{\mathbf{N}}_u[k], \quad (17)$$

where  $\tilde{\mathbf{W}}_u = [\mathbf{W}_u[1] \cdots \mathbf{W}_u[N_u]] \in \mathbb{C}^{N_{\text{BS}} \times N_u N_s^u}$  with  $\mathbf{W}_u[m] \in \mathbb{C}^{N_{\text{BS}} \times N_s^u}$  being the uplink receive combining matrix used in the  $m$ th time slot for  $1 \leq m \leq N_u$ ,  $\tilde{\mathbf{W}}_u = [\mathbf{W}_u^H[1] \cdots \mathbf{W}_u^H[N_u]]$ , and  $\mathbf{F}_u \in \mathbb{C}^{N_{\text{UD}} \times N_s^u}$  is the multi-beam transmit precoding matrix at UD, while  $\mathbf{S}_u \in \mathbb{C}^{N_s^u \times N_o^u}$  is the uplink training signal matrix,

and  $\tilde{\mathbf{N}}_u[k]$  is the uplink noise matrix.  $\tilde{Y}_u[k]$  is multiplied by  $\mathbf{J}_u = [\mathbf{I}_{N_{\text{BS}}^{\text{sub}}} \mathbf{0}_{N_{\text{BS}}^{\text{sub}} \times (N_s^u N_u - N_{\text{BS}}^{\text{sub}})}] \in \mathbb{R}^{N_{\text{BS}}^{\text{sub}} \times N_s^u N_u}$  and the result is converted into the vector  $\tilde{\mathbf{y}}_u[k] = \text{vec}((\mathbf{J}_u \tilde{Y}_u[k])^T)$ , i.e.,

$$\tilde{\mathbf{y}}_u[k] = (\bar{\mathbf{A}}_{\text{BS}} \odot (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u)^T) \text{diag}(\boldsymbol{\alpha}) \boldsymbol{\tau}[k] + \tilde{\mathbf{n}}_u[k], \quad (18)$$

where we have used the identity  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \odot \mathbf{A}) \mathbf{b}$  with  $\mathbf{B} = \text{diag}(\mathbf{b})$  [40],  $\bar{\mathbf{A}}_{\text{BS}} = \mathbf{J}_u \tilde{\mathbf{W}}_u^H \mathbf{A}_{\text{BS}}^*$ , and  $\tilde{\mathbf{n}}_u[k]$  is the corresponding noise vector. Furthermore, by collecting  $\tilde{\mathbf{y}}_u[k] \in \mathbb{C}^{N_{\text{BS}}^{\text{sub}} N_o^u}$  for  $0 \leq k \leq K-1$ , we obtain the aggregated signal matrix  $\tilde{\mathbf{Y}}_u = [\tilde{\mathbf{y}}_u[0] \tilde{\mathbf{y}}_u[1] \cdots \tilde{\mathbf{y}}_u[K-1]] \in \mathbb{C}^{N_{\text{BS}}^{\text{sub}} N_o^u \times K}$  given by

$$\tilde{\mathbf{Y}}_u = (\bar{\mathbf{A}}_{\text{BS}} \odot (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u)^T) \text{diag}(\boldsymbol{\alpha}) \mathbf{A}_{\boldsymbol{\tau}}^T + \tilde{\mathbf{N}}_u, \quad (19)$$

where  $\mathbf{A}_{\boldsymbol{\tau}} = [\boldsymbol{\tau}[0] \boldsymbol{\tau}[1] \cdots \boldsymbol{\tau}[K-1]]^T \in \mathbb{C}^{K \times L}$ , and  $\tilde{\mathbf{N}}_u$  is the aggregated noise matrix. Recalling

<sup>6</sup>In the open-loop CE schemes [18], [19], the support sets and channel gains for every subcarrier estimated at the receiver also need to be fed back to transmitter to perform the following signal processing such as beamforming design or channel equalization [2], [15]. Compared with these schemes, our proposed closed-loop CE scheme only feeds back the dominant channel parameters estimated by the BS and UD to each other, and thus, its feedback overhead is almost negligible.

$\tau[k] = [e^{-j2\pi kf_s \tau_1/K} \dots e^{-j2\pi kf_s \tau_L/K}]^T$ , we have  $\mathbf{A}_\tau = [\mathbf{a}_\tau(\mu_1^T) \dots \mathbf{a}_\tau(\mu_L^T)]$ , in which  $\mathbf{a}_\tau(\mu_l^T) = [1 e^{j\mu_l^T} \dots e^{j(K-1)\mu_l^T}]^T \in \mathbb{C}^K$  with  $\mu_l^T = -2\pi f_s \tau_l/K$ . Observe that  $\mathbf{A}_\tau$  can be considered as the steering matrix associated with the delays  $\{\tau_l\}_{l=1}^L$ . Taking the vectorization of  $\tilde{\mathbf{Y}}_u$ , i.e.,  $\tilde{\mathbf{y}}_u = \text{vec}(\tilde{\mathbf{Y}}_u)$ , leads to

$$\tilde{\mathbf{y}}_u = ((\mathbf{A}_\tau \odot \bar{\mathbf{A}}_{\text{BS}}) \odot (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u)^T) \boldsymbol{\alpha} + \tilde{\mathbf{n}}_u, \quad (20)$$

where we have used the identity  $\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$  [40], and  $\tilde{\mathbf{n}}_u = \text{vec}(\tilde{\mathbf{N}}_u)$ . We further reshape  $\tilde{\mathbf{y}}_u \in \mathbb{C}^{KN_{\text{BS}}^{\text{sub}} N_o^u}$  as the matrix  $\tilde{\mathbf{Y}}_u = \text{mat}(\tilde{\mathbf{y}}_u; N_o^u, KN_{\text{BS}}^{\text{sub}}) \in \mathbb{C}^{N_o^u \times KN_{\text{BS}}^{\text{sub}}}$ :

$$\tilde{\mathbf{Y}}_u = (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u)^T \text{diag}(\boldsymbol{\alpha}) (\mathbf{A}_\tau \odot \bar{\mathbf{A}}_{\text{BS}})^T + \tilde{\mathbf{N}}_u, \quad (21)$$

where  $\tilde{\mathbf{N}}_u = \text{mat}(\tilde{\mathbf{n}}_u; N_o^u, KN_{\text{BS}}^{\text{sub}})$ . Hence,  $\bar{\mathbf{Y}}_u = \tilde{\mathbf{Y}}_u^T \in \mathbb{C}^{KN_{\text{BS}}^{\text{sub}} \times N_o^u}$  can be written as

$$\bar{\mathbf{Y}}_u = (\mathbf{A}_\tau \odot \bar{\mathbf{A}}_{\text{BS}}) \text{diag}(\boldsymbol{\alpha}) (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u) + \tilde{\mathbf{N}}_u^T. \quad (22)$$

From  $\bar{\mathbf{A}}_{\text{BS}} = \mathbf{J}_u \bar{\mathbf{W}}_u^H \mathbf{A}_{\text{BS}}^*$ , we observe that  $\bar{\mathbf{W}}_u$  may destroy the shift-invariance structure of  $\mathbf{A}_{\text{BS}}$ . Similar to downlink CE, we can design  $\bar{\mathbf{W}}_u$  using Algorithm 1 by replacing the input parameters  $N_d, N_s^d, N_{\text{UD}}^{\text{RF}}, N_{\text{UD}}^h, N_{\text{UD}}^v, M_{\text{UD}}^h$  and  $M_{\text{UD}}^v$  for UD with  $N_u, N_s^u, N_{\text{BS}}^{\text{RF}}, N_{\text{BS}}^h, N_{\text{BS}}^v, M_{\text{BS}}^h$  and  $M_{\text{BS}}^v$  for BS. By substituting the designed  $\bar{\mathbf{W}}_u$  into (22), we obtain

$$\bar{\mathbf{Y}}_u = \mathbf{A}_{\tau \text{BS}} \bar{\mathbf{S}}_u + \tilde{\mathbf{N}}_u^T, \quad (23)$$

where  $\bar{\mathbf{S}}_u = \text{diag}(\boldsymbol{\alpha}) (\mathbf{A}_{\text{UD}}^T \mathbf{F}_u \mathbf{S}_u)$ , and  $\mathbf{A}_{\tau \text{BS}} \in \mathbb{C}^{KM_{\text{BS}}^v M_{\text{BS}}^h \times L}$  is given by

$$\begin{aligned} \mathbf{A}_{\tau \text{BS}} &= \mathbf{A}_\tau \odot (\mathbf{J}_u \bar{\mathbf{W}}_u^H ((\mathbf{A}_{\text{BS}}^v)^* \odot (\mathbf{A}_{\text{BS}}^h)^*)) \\ &= \mathbf{A}_\tau \odot \bar{\mathbf{A}}_{\text{BS}}^v \odot \bar{\mathbf{A}}_{\text{BS}}^h. \end{aligned} \quad (24)$$

In (24),  $\bar{\mathbf{A}}_{\text{BS}}^\mu \in \mathbb{C}^{M_{\text{BS}}^h \times L}$  ( $\bar{\mathbf{A}}_{\text{BS}}^v \in \mathbb{C}^{M_{\text{BS}}^v \times L}$ ) is the sub-matrix consisting of the first  $M_{\text{BS}}^h$  ( $M_{\text{BS}}^v$ ) rows of  $(\mathbf{A}_{\text{BS}}^\mu)^*$  ( $(\mathbf{A}_{\text{BS}}^v)^*$ ). In this way,  $\bar{\mathbf{W}}_u$  visualizes the high-dimensional hybrid analog/digital array as a low-dimensional digital array, and  $\mathbf{A}_{\tau \text{BS}}$  holds the triple shift-invariance structure for horizontal/vertical AoDs and delays [35]. Therefore, we can apply the MDU-ESPRIT algorithm to obtain the super-resolution estimates of horizontal/vertical AoDs and delays,  $\{\hat{\theta}_l^{\text{BS}}, \hat{\varphi}_l^{\text{BS}}, \hat{\tau}_l\}_{l=1}^{\hat{L}}$ .

### B. Design Multi-Beam Transmit Precoding Matrix at UD

We design the transmit precoding matrix  $\mathbf{F}_u = \mathbf{F}_{\text{RF},u} \mathbf{F}_{\text{BB},u}$  at UD by exploiting the estimate  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  obtained at the downlink CE stage so that the UD with the limited transmit power can transmit directional multi-beam signals for improving the received SNR at the BS.

We first consider the analog transmit precoding matrix  $\mathbf{F}_{\text{RF},u} \in \mathbb{C}^{N_{\text{UD}} \times N_{\text{UD}}^{\text{RF}}}$ . The estimate  $\hat{\mathbf{A}}_{\text{UD}}$  of the array response matrix  $\mathbf{A}_{\text{UD}}$  can be calculated given the estimated AoAs  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$ . To fully exploit the acquired  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$ , the multi-beam transmit precoding matrix should align its  $\hat{L}$  beams with the  $\hat{L}$  estimated AoAs. Specifically, the phase shifters of the PSN at UD can be divided

into the  $\hat{L}$  groups as equally as possible, depending on  $\hat{L}$  and  $N_{\text{UD}}^{\text{RF}}$ .

*Case I:*  $\hat{L} > N_{\text{UD}}^{\text{RF}}$ . This is the case that the number of beams transmitted by UD is larger than that of RF chains. For the UD equipped with the hybrid array with the fully-connected PSN, there are  $N_{\text{PS}} = N_{\text{UD}}^{\text{RF}} N_{\text{UD}}$  phase shifters. Let the number of phase shifters assigned to the  $l$ th group be  $n_{\text{ps},l}$  with  $1 \leq l \leq \hat{L}$ . We introduce the  $\hat{L}$ -dimensional vector  $\mathbf{v}_{\text{ps}}$  as

$$\mathbf{v}_{\text{ps}} = [n_{\text{ps},1} \dots n_{\text{ps},\hat{L}}]^T = n_{\text{ps}} \mathbf{1}_{\hat{L}} + [\mathbf{1}_{n_{\text{re}}}^T \mathbf{0}_{\hat{L}-n_{\text{re}}}^T]^T, \quad (25)$$

where  $n_{\text{ps}} = \lfloor N_{\text{PS}}/\hat{L} \rfloor$  and  $n_{\text{re}} = \text{mod}(N_{\text{PS}}, \hat{L})$ . By defining the index vector  $\mathbf{p} = [1 \ 2 \dots N_{\text{PS}}]^T$ , the ordered index set  $\mathcal{P}_l$  of the  $l$ th group can be obtained as  $\mathcal{P}_l = \mathbf{p}[\sum_{i=1}^{l-1} n_{\text{ps},i+1} : \sum_{i=1}^l n_{\text{ps},i}]$  with  $|\mathcal{P}_l|_c = n_{\text{ps},l}$ . Then, we can define the vector  $\mathbf{f} = [\mathbf{f}_1^T \dots \mathbf{f}_{\hat{L}}^T]^T \in \mathbb{C}^{N_{\text{PS}}}$ , where  $\mathbf{f}_l = \hat{\mathbf{A}}_{\text{UD}}^*[\text{mod}(\mathcal{P}_l, N_{\text{UD}}), l] \in \mathbb{C}^{n_{\text{ps},l}}$  for  $1 \leq l \leq \hat{L}$ , to obtain  $\mathbf{F}_{\text{RF},u} = \text{mat}(\mathbf{f}; N_{\text{UD}}, N_{\text{UD}}^{\text{RF}})$ .

*Case II:*  $\hat{L} \leq N_{\text{UD}}^{\text{RF}}$  and  $N_{\text{UD}}^{\text{RF}}$  can be divided exactly by  $\hat{L}$ . The  $N_{\text{UD}}^{\text{RF}}$  RF chains can be equally allocated to the  $\hat{L}$  groups, and we can choose  $\mathbf{F}_{\text{RF},u} = \mathbf{1}_{N_{\text{rep}}}^T \otimes \hat{\mathbf{A}}_{\text{UD}}^*$ , with  $N_{\text{rep}} = N_{\text{UD}}^{\text{RF}}/\hat{L}$ .

*Case III:*  $\hat{L} < N_{\text{UD}}^{\text{RF}}$  and  $N_{\text{UD}}^{\text{RF}}$  cannot be divided exactly by  $\hat{L}$ . In this case, we have  $\mathbf{F}_{\text{RF},u} = [\mathbf{F}_{\text{RF},u}^1 \ \mathbf{F}_{\text{RF},u}^2]$ , where  $\mathbf{F}_{\text{RF},u}^1 \in \mathbb{C}^{N_{\text{UD}} \times \hat{L} N_{\text{rep}}}$  with  $N_{\text{rep}} = \lfloor N_{\text{UD}}^{\text{RF}}/\hat{L} \rfloor$  and  $\mathbf{F}_{\text{RF},u}^2 \in \mathbb{C}^{N_{\text{UD}} \times N_{\text{UD}}^{\text{RF},\text{re}}}$  with  $N_{\text{UD}}^{\text{RF},\text{re}} = \text{mod}(N_{\text{UD}}^{\text{RF}}, \hat{L})$ . We can design  $\mathbf{F}_{\text{RF},u}^1 = \mathbf{1}_{N_{\text{rep}}}^T \otimes \hat{\mathbf{A}}_{\text{UD}}^*$  similar to Case II, and we can choose  $\mathbf{F}_{\text{RF},u}^2 = \text{mat}([\tilde{\mathbf{f}}_1^T \dots \tilde{\mathbf{f}}_{\hat{L}}^T]^T; N_{\text{UD}}, N_{\text{UD}}^{\text{RF},\text{re}})$  similar to Case I, where  $\tilde{\mathbf{f}}_l$  can be acquired by using  $N_{\text{UD}}^{\text{RF},\text{re}}$  instead of  $N_{\text{UD}}^{\text{RF}}$  in Case I.

Due to the limited resolution of the PSN, the phase value of every element in the designed  $\mathbf{F}_{\text{RF},u}$  is quantized to the nearest value in the phase set  $\mathcal{A}$ . As for the digital transmit precoding matrix  $\mathbf{F}_{\text{BB},u} \in \mathbb{C}^{N_{\text{UD}}^{\text{RF}} \times N_s^u}$ , we can design its element as  $[\mathbf{F}_{\text{BB},u}]_{j_1, j_2} = e^{j2\pi b_{j_1, j_2}}$  with  $b_{j_1, j_2} \sim \mathcal{U}[0, 1]$ . Finally, we obtain the multi-beam transmit precoding matrix  $\mathbf{F}_u = \mathbf{F}_{\text{RF},u} \mathbf{F}_{\text{BB},u}$ .

To intuitively compare  $\mathbf{F}_d$  of the BS designed at the downlink CE stage and  $\mathbf{F}_u$  of the UD designed at the uplink CE stage, we provide the comparison of beam patterns in Fig. 4, where we have  $N_{\text{BS}}^{\text{RF}} = N_{\text{UD}}^{\text{RF}} = 4$  RF chains at the BS and each UD, and the AoAs of the 5 MPCs are known to the UD. Specifically, Fig. 4(a) depicts the beam pattern of the random transmit precoding matrix  $\mathbf{F}_d$  for the  $8 \times 8$  UPA, while Fig. 4(b) and Fig. 4(c) plot the beam patterns of the multi-beam transmit precoding matrices  $\mathbf{F}_u$  for the  $8 \times 8$  and  $16 \times 16$  UPAs, respectively. Compared with the beam pattern of Fig. 4(a), the beam pattern in Fig. 4(b) has 5 mainlobes aligned with the directions of the AoAs of the 5 MPCs, which can significantly improve the SNR at receiver. Moreover, by comparing Fig. 4(b) with Fig. 4(c), it can be observed that the sidelobes of the multi-beam signals are further suppressed when the array size increases. In a nutshell, the proposed multi-beam transmit precoding matrix design enables the UD with limited transmit power to form the directional signals



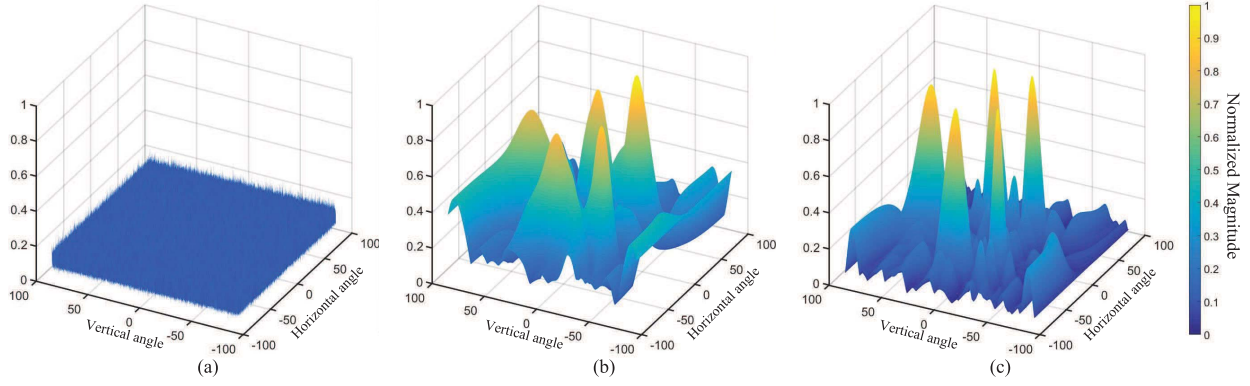


Fig. 4. Comparison of beam patterns, where  $N_{\text{BS}}^{\text{RF}} = N_{\text{UD}}^{\text{RF}} = 4$  and the AoAs of the 5 MPCs are known to the UD: (a) Random transmit precoding matrix with  $8 \times 8$  UPA, (b) Multi-beam transmit precoding matrix with  $8 \times 8$  UPA, and (c) Multi-beam transmit precoding matrix with  $16 \times 16$  UPA.

with multiple beams aligned with the estimated AoAs of the MPCs for improving the uplink CE performance.

#### IV. MDU-ESPRIT ALGORITHM

Because the aggregated receive combining matrix  $\widetilde{\mathbf{W}}_d$  ( $\widetilde{\mathbf{W}}_u$ ) designed at the UD (BS) reconstructs the double (triple) shift-invariance structure of array response, spectrum estimation techniques can be utilized to estimate the channel parameters. We consider  $R$ -dimensional ( $R$ -D) unitary ESPRIT algorithm with  $R \geq 2$ . Without loss of generality, we define a general signal transmission model for the channel consisting of  $L$  MPCs and  $R$  sets of spatial frequencies as

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{N}, \quad (26)$$

where  $\mathbf{Y} \in \mathbb{C}^{M \times N}$  is the received data matrix aggregated over  $N$  snapshots,  $M = \prod_{r=1}^R M_r$ , with  $M_r$  being the dimension of the parameter vector associated with the  $r$ th spatial frequency for  $1 \leq r \leq R$ , and  $\mathbf{S} \in \mathbb{C}^{L \times N}$  and  $\mathbf{N} \in \mathbb{C}^{M \times N}$  are the transmit signal and noise matrices, respectively, while the array response matrix  $\mathbf{A} \in \mathbb{C}^{M \times L}$  is given by

$$\begin{aligned} \mathbf{A} &= \mathbf{A}_{\mu_R} \odot \cdots \odot \mathbf{A}_{\mu_2} \odot \mathbf{A}_{\mu_1} \\ &= [\mathbf{a}(\mu_1^1, \mu_1^2, \dots, \mu_1^R) \cdots \mathbf{a}(\mu_L^1, \mu_L^2, \dots, \mu_L^R)]. \end{aligned} \quad (27)$$

In (27),  $\mathbf{A}_{\mu_r} = [\mathbf{a}(\mu_1^r) \cdots \mathbf{a}(\mu_L^r)] \in \mathbb{C}^{M_r \times L}$  is the steering matrix related to the  $r$ th set of spatial frequencies  $\{\mu_l^r\}_{l=1}^L$ , with  $\mathbf{a}(\mu_l^r) = [1 e^{j\mu_l^r} \cdots e^{j(M_r-1)\mu_l^r}]^T \in \mathbb{C}^{M_r}$  being the  $l$ th steering vector, while the array response vector  $\mathbf{a}(\mu_l^1, \mu_l^2, \dots, \mu_l^R) \in \mathbb{C}^M$  related to the  $l$ th MPC is given by

$$\mathbf{a}(\mu_l^1, \mu_l^2, \dots, \mu_l^R) = \mathbf{a}(\mu_l^R) \otimes \cdots \otimes \mathbf{a}(\mu_l^2) \otimes \mathbf{a}(\mu_l^1). \quad (28)$$

The MDU-ESPRIT algorithm, which acquires the super-resolution estimates of the  $R$  sets of spatial frequencies from (26), denoted by  $\{\hat{\mu}_l^r\}_{l=1}^L$  for  $1 \leq r \leq R$ , consists of the five steps.

1) *R-D Spatial Smoothing Preprocessing (SSP)*: In order to take into account the insufficient measurement dimension  $N$  caused by the limited training overhead, we will exploit the spatial smoothing technique [31] to preprocess the original data matrix  $\mathbf{Y}$  of (26). This preprocessing can mitigate the

influence of other coherent signals and avoid the rank deficiency of the covariance matrix of  $\mathbf{S}$  to enhance robustness. Specifically, we first define the  $R$  spatial smoothing parameters  $\{G_r\}_{r=1}^R$  with  $1 \leq G_r \leq M_r$ , and obtain the sub-dimensions  $\{M_r^{\text{sub}}\}_{r=1}^R$  corresponding to  $\{M_r\}_{r=1}^R$  as  $M_r^{\text{sub}} = M_r - G_r + 1$  for  $1 \leq r \leq R$ . Thus the size of total sub-dimension is  $M_{\text{sub}} = \prod_{r=1}^R M_r^{\text{sub}}$ . To obtain the  $R$ -D selection matrix, we next define the  $g_r$ th ‘1-D’ selection matrix as  $\mathbf{J}^{(g_r)} = [\mathbf{O}_{M_r^{\text{sub}} \times (g_r-1)} \quad \mathbf{I}_{M_r^{\text{sub}}} \quad \mathbf{O}_{M_r^{\text{sub}} \times (G_r-g_r)}] \in \mathbb{R}^{M_r^{\text{sub}} \times M_r}$  for  $1 \leq g_r \leq G_r$ . Then, we can obtain  $G = \prod_{r=1}^R G_r$  ‘ $R$ -D’ selection matrices, with the  $(g_1, g_2, \dots, g_R)$ th ‘ $R$ -D’ selection matrix given by  $\mathbf{J}_{g_1, g_2, \dots, g_R} = \mathbf{J}^{(g_R)} \otimes \cdots \otimes \mathbf{J}^{(g_2)} \otimes \mathbf{J}^{(g_1)} \in \mathbb{R}^{M_{\text{sub}} \times M}$ . By applying these  $R$ -D selection matrices to  $\mathbf{Y}$ , the smoothed complex-valued data matrix  $\bar{\mathbf{Y}} \in \mathbb{C}^{M_{\text{sub}} \times NG}$  is obtained as

$$\begin{aligned} \bar{\mathbf{Y}} &= [(\mathbf{J}_{1,1,\dots,1,1} \mathbf{Y}) \cdots (\mathbf{J}_{1,1,\dots,1,G_R} \mathbf{Y}) (\mathbf{J}_{1,1,\dots,2,1} \mathbf{Y}) \\ &\quad (\mathbf{J}_{1,1,\dots,2,2} \mathbf{Y}) \cdots (\mathbf{J}_{G_1,G_2,\dots,G_{R-1},G_R} \mathbf{Y})]. \end{aligned} \quad (29)$$

2) *Real-Valued Processing (RVP)*: To reduce the computational complexity, the forward backward averaging technique [30] is utilized to transform  $\bar{\mathbf{Y}}$  into the real-valued matrix  $\bar{\mathbf{Y}}_{\text{re}} \in \mathbb{R}^{M_{\text{sub}} \times 2NG}$

$$\bar{\mathbf{Y}}_{\text{re}} = \mathbf{Q}_{M_{\text{sub}}}^H [\bar{\mathbf{Y}} (\mathbf{\Pi}_{M_{\text{sub}}} \bar{\mathbf{Y}}^* \mathbf{\Pi}_{NG})] \mathbf{Q}_{2NG}, \quad (30)$$

where  $\mathbf{\Pi}_n$  is the exchange matrix of size  $n \times n$  that permutes the row order of  $\mathbf{I}_n$ , and  $\mathbf{Q}_n \in \mathbb{C}^{n \times n}$  is a sparse unitary matrix satisfying  $\mathbf{\Pi}_n \mathbf{Q}_n^H = \mathbf{Q}_n$ .

3) *Signal Subspace Approximation (SSA)*: To extract the information of spatial frequencies from the real-valued matrix  $\bar{\mathbf{Y}}_{\text{re}}$ , we introduce the transform steering matrix  $\mathbf{K}$  which satisfies [30]

$$\Re \{ \mathbf{Q}_{m_r}^H \mathbf{J}_r \mathbf{Q}_{M_{\text{sub}}} \} \mathbf{K} \mathbf{A}_r = \Im \{ \mathbf{Q}_{m_r}^H \mathbf{J}_r \mathbf{Q}_{M_{\text{sub}}} \} \mathbf{K}, \quad (31)$$

where  $1 \leq r \leq R$ ,  $m_r = M_{\text{sub}} (M_r^{\text{sub}} - 1) / M_r^{\text{sub}}$ , and  $\mathbf{A}_r = \text{diag}([\tan(\mu_1^r/2) \cdots \tan(\mu_L^r/2)]^T)$  is the real-valued diagonal matrix involving the desired spatial frequencies  $\{\mu_l^r\}_{l=1}^L$ , while  $\mathbf{J}_r = \mathbf{I}_{\prod_{i=r+1}^R M_i^{\text{sub}}} \otimes \tilde{\mathbf{J}}^{(r)} \otimes \mathbf{I}_{\prod_{i=1}^{r-1} M_i^{\text{sub}}} \in \mathbb{R}^{m_r \times M_{\text{sub}}}$ , with  $\tilde{\mathbf{J}}^{(r)} = [\mathbf{0}_{M_r^{\text{sub}}-1} \quad \mathbf{I}_{M_r^{\text{sub}}-1}]$ . Note that  $\mathbf{K}$  is related to the approximate signal subspace matrix  $\mathbf{E}_s \in \mathbb{R}^{M_{\text{sub}} \times L}$  corresponding to the underlying signal subspace. Specifically,

since the columns of  $\mathbf{K}$  and  $\mathbf{E}_s$  span the same  $L$ -dimensional signal subspace [31], [35],  $\mathbf{K} = \mathbf{E}_s \mathbf{T}$ , where  $\mathbf{T} \in \mathbb{R}^{L \times L}$  is a non-singular matrix. To determine  $\mathbf{E}_s$ , we take the left singular vectors corresponding to the largest  $L$  singular values of  $\bar{\mathbf{Y}}_{\text{re}}$  as  $\mathbf{E}_s$ . Specifically, from the real-valued partial singular values decomposition (SVD) of  $\bar{\mathbf{Y}}_{\text{re}} = \mathbf{U}_{\text{re}} \boldsymbol{\Sigma}_{\text{re}} \mathbf{V}_{\text{re}}^H$ , we have  $\mathbf{E}_s = \mathbf{U}_{\text{re}[:,1:L]}$ .

4) *Shift-Invariance Equation Solving (SIES)*: Based on the acquired approximate signal subspace  $\mathbf{E}_s$ , we use  $\mathbf{K} = \mathbf{E}_s \mathbf{T}$  in (31) to obtain the  $R$  shift-invariance equations

$$\Re \{ \mathbf{Q}_{m_r}^H \mathbf{J}_r \mathbf{Q}_{M_{\text{sub}}} \} \mathbf{E}_s \boldsymbol{\Phi}_r = \Im \{ \mathbf{Q}_{m_r}^H \mathbf{J}_r \mathbf{Q}_{M_{\text{sub}}} \} \mathbf{E}_s, \quad (32)$$

where  $\boldsymbol{\Phi}_r = \mathbf{T} \boldsymbol{\Lambda}_r \mathbf{T}^{-1} \in \mathbb{R}^{L \times L}$  for  $1 \leq r \leq R$ . To estimate the diagonal matrices  $\{\boldsymbol{\Lambda}_r\}_{r=1}^R$ , we first obtain the estimates of the  $R$  real-valued matrices  $\{\boldsymbol{\Phi}_r\}_{r=1}^R$ , denoted as  $\{\hat{\boldsymbol{\Phi}}_r\}_{r=1}^R$ , by applying the LS or total least squares (TLS) estimator to solve the  $R$  shift-invariance equations of (32).

5) *R-D Joint Diagonalization (JD)*: From the estimated  $\hat{\boldsymbol{\Phi}}_r = \mathbf{T} \hat{\boldsymbol{\Lambda}}_r \mathbf{T}^{-1}$  with  $\hat{\boldsymbol{\Lambda}}_r$  denoting the estimate of  $\boldsymbol{\Lambda}_r$  for  $1 \leq r \leq R$ , we exploit the following  $R$ -D joint diagonalization to obtain the paired estimates of the spatial frequencies  $\{\hat{\mu}_l^r\}_{l=1}^L$  from  $\{\hat{\boldsymbol{\Lambda}}_r\}_{r=1}^R$ . Specifically, we consider the two cases of  $R = 2$  and  $R \geq 3$ . For  $R = 2$ , namely, the 2-D case, since  $\hat{\boldsymbol{\Phi}}_1$  and  $\hat{\boldsymbol{\Phi}}_2$  share the same eigenvector matrix  $\mathbf{T}$ , we can calculate the EVD of the complex-valued matrix  $\boldsymbol{\Psi} = \hat{\boldsymbol{\Phi}}_1 + j\hat{\boldsymbol{\Phi}}_2$  to obtain  $\hat{\boldsymbol{\Lambda}}_1$  and  $\hat{\boldsymbol{\Lambda}}_2$ , specifically,  $\boldsymbol{\Psi} = \mathbf{T} \boldsymbol{\Delta} \mathbf{T}^{-1}$  with  $\boldsymbol{\Delta} = \hat{\boldsymbol{\Lambda}}_1 + j\hat{\boldsymbol{\Lambda}}_2$ , and  $\hat{\boldsymbol{\Lambda}}_1 = \Re\{\boldsymbol{\Delta}\}$  and  $\hat{\boldsymbol{\Lambda}}_2 = \Im\{\boldsymbol{\Delta}\}$ . For  $R \geq 3$ , the noise-corrupted matrices  $\{\hat{\boldsymbol{\Phi}}_r\}_{r=1}^R$  do not always exactly share the same  $\mathbf{T}$ . Hence, we exploit the simultaneous Schur decomposition (SSD) algorithm [37], which is developed from the real Schur decomposition [40] for multi-parameter estimation and pairing. By utilizing the SSD algorithm, we obtain the  $R$  approximate upper-triangular matrices  $\{\mathbf{T}_r\}_{r=1}^R$  so that  $\{\hat{\boldsymbol{\Lambda}}_r\}_{r=1}^R$  are acquired as the main diagonal elements of  $\{\mathbf{T}_r\}_{r=1}^R$ , i.e.,  $\hat{\boldsymbol{\Lambda}}_r = \text{diag}(\text{vdiag}(\mathbf{T}_r))$ , for  $1 \leq r \leq R$ . Finally, the  $R$  paired super-resolution estimates of the spatial frequencies  $\{\hat{\mu}_l^r\}_{l=1}^L$  can be calculated from  $\hat{\boldsymbol{\Lambda}}_r$  as  $\hat{\mu}_l^r = 2 \arctan([\hat{\boldsymbol{\Lambda}}_r]_{l,l})$  for  $1 \leq l \leq L$  and  $1 \leq r \leq R$ .

This MDU-ESPRIT algorithm is summarized in Algorithm 2. At the downlink CE stage, we estimate  $\{\hat{\mu}_l^{\text{UD}}, \hat{\nu}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  based on  $\bar{\mathbf{Y}}_d$  of (11) by applying the 2-D ( $R = 2$ ) unitary ESPRIT algorithm. Furthermore, based on  $\bar{\mathbf{Y}}_u$  of (23),  $\{\hat{\mu}_l^{\text{BS}}, \hat{\nu}_l^{\text{BS}}, \hat{\tau}_l\}_{l=1}^{\hat{L}}$  are estimated using the 3-D ( $R = 3$ ) unitary ESPRIT algorithm at the uplink CE stage. Hence, the spatial smoothing parameters for Algorithm 2 in the 2-D and 3-D cases are  $\{G_1^d, G_2^d\}$  and  $\{G_1^u, G_2^u, G_3^u\}$ , respectively. The corresponding total sub-dimension sizes at the UD and BS are  $M_{\text{sub}}^{\text{UD}} = (M_{\text{UD}}^h - G_1^d + 1)(M_{\text{UD}}^v - G_2^d + 1)$  and  $M_{\text{sub}}^{\text{BS}} = (M_{\text{BS}}^h - G_1^u + 1)(M_{\text{BS}}^v - G_2^u + 1)(K - G_3^u + 1)$ , respectively.

## V. ML PAIRING AND PATH GAINS ESTIMATION

At the downlink CE stage, the BS obtains the estimated horizontal/vertical AoAs  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  fed back by the UD. It then estimates the horizontal/vertical AoDs and delays  $\{\hat{\theta}_l^{\text{BS}}, \hat{\varphi}_l^{\text{BS}}, \hat{\tau}_l\}_{l=1}^{\hat{L}}$  at the uplink CE stage.

### Algorithm 2 MDU-ESPRIT Algorithm

**Input:** Data matrix  $\mathbf{Y}$ , number of MPCs  $L$ , sub-dimensions  $\{M_r\}_{r=1}^R$ , spatial smoothing parameters  $\{G_r\}_{r=1}^R$

Super-resolution

- Output:** estimates of spatial frequencies  $\{\hat{\mu}_l^r\}_{l=1}^L, 1 \leq r \leq R$
- 1: Obtain smoothed data matrix  $\bar{\mathbf{Y}}$  (29) using  $R$ -D spatial smoothing preprocessing
  - 2: Obtain real-valued data matrix  $\bar{\mathbf{Y}}_{\text{re}}$  (30) using forward backward averaging
  - 3: Determine approximate signal subspace matrix  $\mathbf{E}_s$  through SVD
  - 4: Solve shift-invariance equations (32) to obtain  $R$  real-valued matrices  $\{\hat{\boldsymbol{\Phi}}_r\}_{r=1}^R$
  - 5: Perform  $R$ -D joint diagonalization to estimate diagonal matrices  $\{\hat{\boldsymbol{\Lambda}}_r\}_{r=1}^R$ : i)  $R = 2$ , calculate EVD of  $\boldsymbol{\Psi} = \hat{\boldsymbol{\Phi}}_1 + j\hat{\boldsymbol{\Phi}}_2 = \mathbf{T} \boldsymbol{\Delta} \mathbf{T}^{-1}$  to obtain  $\hat{\boldsymbol{\Lambda}}_1 = \Re\{\boldsymbol{\Delta}\}$  and  $\hat{\boldsymbol{\Lambda}}_2 = \Im\{\boldsymbol{\Delta}\}$ ; ii)  $R \geq 3$ , obtain  $R$  diagonal matrices  $\{\hat{\boldsymbol{\Lambda}}_r\}_{r=1}^R$  via SSD algorithm
  - 6: Extract  $R$  paired  $\{\hat{\mu}_l^r\}_{l=1}^L$  from  $\{\hat{\boldsymbol{\Lambda}}_r\}_{r=1}^R$

Since  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$  and  $\{\hat{\theta}_l^{\text{BS}}, \hat{\varphi}_l^{\text{BS}}, \hat{\tau}_l\}_{l=1}^{\hat{L}}$  are acquired in the two different ends of the channel at two different stages, it is necessary to pair them. Furthermore, the path gain vector  $\boldsymbol{\alpha}$  needs to be estimated. We propose to apply an ML approach to pair the channel parameters and to estimate the path gains, which corresponds to Step 7 of Fig. 2 at the BS.

Specifically, according to (24), we construct the equivalent steering matrix associated with  $\{\hat{\theta}_l^{\text{BS}}, \hat{\varphi}_l^{\text{BS}}, \hat{\tau}_l\}_{l=1}^{\hat{L}}$  as  $\hat{\mathbf{A}}_{\tau\text{BS}}$ , where  $\{\hat{\tau}_l\}_{l=1}^{\hat{L}}$  are arranged in ascending order. Based on the estimated horizontal/vertical AoAs fed back to the BS,  $\{\hat{\theta}_l^{\text{UD}}, \hat{\varphi}_l^{\text{UD}}\}_{l=1}^{\hat{L}}$ , we can reconstruct the estimated multi-beam transmit precoding matrix, denoted by  $\hat{\mathbf{F}}_u$ , similar to the construction of  $\mathbf{F}_u$  given in Section III-B. Clearly, there are a total of  $J_c = \hat{L}!$  possible ordered combinations or pairs  $\{\hat{\theta}_{l_j}^{\text{UD}}, \hat{\varphi}_{l_j}^{\text{UD}}\}_{l_j=1}^{\hat{L}}$  that can pair with  $\hat{\mathbf{A}}_{\tau\text{BS}}$  or  $\{\hat{\theta}_i^{\text{BS}}, \hat{\varphi}_i^{\text{BS}}, \hat{\tau}_i\}_{i=1}^{\hat{L}}$ , where  $j \in \mathcal{J}$  and the size of the ordered set  $\mathcal{J}$  is  $|\mathcal{J}|_c = J_c$ . For each  $\{\hat{\theta}_{l_j}^{\text{UD}}, \hat{\varphi}_{l_j}^{\text{UD}}\}_{l_j=1}^{\hat{L}}$ , we can establish the corresponding array response matrix  $\mathbf{A}_{\text{UD},j}$ , which is denoted as  $\hat{\mathbf{A}}_{\text{UD},j}$ . Thus, for each pair of the AoDs and AoAs, we have  $\hat{\mathbf{A}}_{\tau\text{BS}}$ ,  $\hat{\mathbf{A}}_{\text{UD},j}$  and  $\hat{\mathbf{F}}_u$ . Substituting them into (20) yields  $\check{\mathbf{y}}_u = \hat{\mathbf{A}}_j \boldsymbol{\alpha}_j + \check{\mathbf{n}}_u$ , where  $\hat{\mathbf{A}}_j = \hat{\mathbf{A}}_{\tau\text{BS}} \odot (\hat{\mathbf{A}}_{\text{UD},j}^T \hat{\mathbf{F}}_u \mathbf{S}_u)^T$  while  $\boldsymbol{\alpha}_j$  is the path gain vector corresponding to the  $j$ th pair of the AoDs and AoAs with  $j \in \mathcal{J}$ . The LS estimate of  $\boldsymbol{\alpha}_j$  is readily given as

$$\hat{\boldsymbol{\alpha}}_j = (\hat{\mathbf{A}}_j^H \hat{\mathbf{A}}_j)^{-1} \hat{\mathbf{A}}_j^H \check{\mathbf{y}}_u. \quad (33)$$

From the estimate  $\hat{\boldsymbol{\alpha}}_j$ , we can estimate  $\check{\mathbf{y}}_u$  according to  $\hat{\mathbf{y}}_{u,j} = \hat{\mathbf{A}}_j \hat{\boldsymbol{\alpha}}_j$  with the residual  $\|\check{\mathbf{y}}_u - \hat{\mathbf{y}}_{u,j}\|_2^2$ . We can then find the optimal pair index  $j^*$  by solving the following optimization problem

$$j^* = \arg \min_{j \in \mathcal{J}} \|\check{\mathbf{y}}_u - \hat{\mathbf{y}}_{u,j}\|_2^2. \quad (34)$$

Hence, the optimal estimate of the path gains is given by  $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}_{j^*} = \beta [\widehat{\alpha}_1 \cdots \widehat{\alpha}_{\widehat{L}}]^T$  and we have the optimal ordered mmWave channel parameter estimate  $\{\widehat{\theta}_l^{\text{UD}}, \widehat{\varphi}_l^{\text{UD}}, \widehat{\theta}_l^{\text{BS}}, \widehat{\varphi}_l^{\text{BS}}, \widehat{\tau}_l, \widehat{\alpha}_l\}_{l=1}^{\widehat{L}}$ .

By substituting  $\{\widehat{\theta}_l^{\text{UD}}, \widehat{\varphi}_l^{\text{UD}}, \widehat{\theta}_l^{\text{BS}}, \widehat{\varphi}_l^{\text{BS}}, \widehat{\tau}_l, \widehat{\alpha}_l\}_{l=1}^{\widehat{L}}$  into (7), we obtain the optimally estimated frequency-domain channel matrix  $\widehat{\mathbf{H}}[k]$  at the  $k$ th subcarrier as

$$\widehat{\mathbf{H}}[k] = \beta \sum_{l=1}^{\widehat{L}} \widehat{\alpha}_l \mathbf{a}_{\text{UD}}(\widehat{\mu}_l^{\text{UD}}, \widehat{\nu}_l^{\text{UD}}) \mathbf{a}_{\text{BS}}^H(\widehat{\mu}_l^{\text{BS}}, \widehat{\nu}_l^{\text{BS}}) e^{-j2\pi \frac{k f_s}{K} \widehat{\tau}_l}, \quad (35)$$

where  $\widehat{\mu}_l^{\text{UD}} = \pi \sin(\widehat{\theta}_l^{\text{UD}}) \cos(\widehat{\varphi}_l^{\text{UD}})$  and  $\widehat{\nu}_l^{\text{UD}} = \pi \sin(\widehat{\varphi}_l^{\text{UD}})$ , while  $\widehat{\mu}_l^{\text{BS}} = \pi \sin(\widehat{\theta}_l^{\text{BS}}) \cos(\widehat{\varphi}_l^{\text{BS}})$  and  $\widehat{\nu}_l^{\text{BS}} = \pi \sin(\widehat{\varphi}_l^{\text{BS}})$ .

## VI. PERFORMANCE EVALUATION

An extensive simulation investigation is carried out to evaluate the CE performance and computational complexity of the proposed closed-loop CE scheme. In simulations, the carrier frequency is  $f_c = 30$  GHz with the bandwidth  $f_s = 200$  MHz, the numbers of RF chains at BS and UD are both 4, i.e.,  $N_{\text{BS}}^{\text{RF}} = N_{\text{UD}}^{\text{RF}} = 4$ , and the numbers of horizontal and vertical antennas at BS and UD are both 12, i.e.,  $N_{\text{BS}}^{\text{h}} = N_{\text{BS}}^{\text{v}} = N_{\text{UD}}^{\text{h}} = N_{\text{UD}}^{\text{v}} = 12$ , and the quantization accuracy of the PSN is defined by  $N_q^{\text{ps}} = 3$  bits, while the feedback quantization accuracy for AoAs is specified by  $N_q^{\text{ang}} = 10$  bits. Without loss of generality, the case of single UD  $Q = 1$  is considered. From Fig. 3, it is clear that for the generic case of  $Q > 1$ , the uplink training overhead becomes  $N_u N_o^u$  instead of  $Q N_u N_o^u$  for the case of  $Q = 1$ . The channel model is simulated as follows. Each of the path gains  $\{\alpha_l\}_{l=1}^L$  is generated according to  $\mathcal{CN}(0, 1)$ , while the other channel parameters  $\{\tau_l, \theta_l^{\text{UD}}, \theta_l^{\text{BS}}, \varphi_l^{\text{UD}}, \varphi_l^{\text{BS}}\}_{l=1}^L$  all follow uniform distribution, specifically,  $\tau_l \sim \mathcal{U}[0, \tau_{\text{max}}]$  and  $\theta_l^{\text{UD}}, \theta_l^{\text{BS}}, \varphi_l^{\text{UD}}, \varphi_l^{\text{BS}} \sim \mathcal{U}[-\pi/3, \pi/3]$  for the  $l$ th MPC. The maximum multipath delay is set to  $\tau_{\text{max}} = 16 T_s$ , i.e.,  $N_c = 16$ . The number of subcarriers is set to  $K = 128$  with the length of cyclic prefix (CP) being 32, and perfect frame synchronization is assumed. In our proposed solution, the sizes of low-dimensional digital sub-arrays visualized from the high-dimensional hybrid arrays are set to  $M_{\text{BS}}^{\text{h}} = M_{\text{BS}}^{\text{v}} = M_{\text{UD}}^{\text{h}} = M_{\text{UD}}^{\text{v}} = 8$ . Hence, the numbers of downlink and uplink training time slots are  $N_d = 22$  and  $N_u = 22$ , respectively, given the number of downlink independent signal streams  $N_s^d = N_{\text{UD}}^{\text{RF}} - 1 = 3$  and the number of uplink independent signal streams  $N_s^u = N_{\text{BS}}^{\text{RF}} - 1 = 3$ . Additionally,  $P = 8$  adjacent subcarriers are jointly employed to estimate the number of MPCs, with the threshold parameter  $\varepsilon$  empirically set to 1.54, 0.50, 0.16, 0.05, 0.016, and 0.005, respectively, at the SNR of  $-15$  dB,  $-10$  dB,  $-5$  dB,  $0$  dB,  $5$  dB, and  $10$  dB. The spatial smoothing parameters used for Algorithm 2 are  $G_1^d = G_2^d = G_1^u = G_2^u = 2$  and  $G_3^u = K/2$ . The downlink and uplink SNRs are both defined as  $\rho \sigma_\alpha^2 / \sigma_n^2$ , where  $\rho$  and  $\sigma_n^2$  are the transmit power and receiver noise variance, respectively.

The state-of-the-art OMP-based frequency-domain scheme [18]<sup>7</sup> and the SW-OMP-based scheme [19] are adopted as two benchmarks. In order to be consistent with [18] and [19], their digital transmit precoding/receive combining matrices are taken as the identity matrix, while the design of analog counterparts is similar to the construction of  $\mathbf{F}_{\text{RF},d}[m]$  given in Section II-B. The sizes of the quantized angle-domain grids associated with horizontal/vertical AoAs/AoDs, denoted by  $G_{\text{BS}}^{\text{h}}, G_{\text{BS}}^{\text{v}}, G_{\text{UD}}^{\text{h}}$  and  $G_{\text{UD}}^{\text{v}}$ , are set to twice the numbers of antennas in the horizontal and vertical directions of UPA, respectively, according to [18], [19], i.e.,  $G_{\text{BS}} = G_{\text{BS}}^{\text{h}} \times G_{\text{BS}}^{\text{v}} = 2 N_{\text{BS}}^{\text{h}} \times 2 N_{\text{BS}}^{\text{v}} = 24 \times 24$  and  $G_{\text{UD}} = G_{\text{UD}}^{\text{h}} \times G_{\text{UD}}^{\text{v}} = 2 N_{\text{UD}}^{\text{h}} \times 2 N_{\text{UD}}^{\text{v}} = 24 \times 24$ . Furthermore, all CE schemes adopt the same training overhead, which is equal to the required number of training frames [18], [19], to ensure the fairness of comparison.

### A. CE Performance Evaluation

First the CE performance is evaluated using the normalized mean square error (NMSE) metric given by

$$\text{NMSE} = \mathbb{E} \left( \frac{\sum_{k=0}^{K-1} \|\mathbf{H}[k] - \widehat{\mathbf{H}}[k]\|_F^2}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_F^2} \right). \quad (36)$$

Fig. 5 compares the NMSE performance of the proposed closed-loop scheme with those of the OMP and SW-OMP based schemes for different SNRs, given the numbers of MPCs  $L = 3$  and  $L = 5$ . For our closed-loop scheme, the numbers of OFDM symbols in each downlink time slot and uplink time slot are  $N_o^d = N_o^u = 3$ . Therefore, the training overhead of our closed-loop scheme is  $T_{\text{CE}} = N_d N_o^d + N_u N_o^u = 132$ . In Fig. 5, the NMSE curve labeled as ‘Proposed Close-Loop’ is our proposed closed-loop CE scheme, which also estimates the number of MPCs  $L$ , while the curve labeled with ‘Benchmark of Closed-Loop’ is the closed-loop scheme given the perfect knowledge of  $L$ , which provides a lower bound NMSE. It can be seen that the CE accuracy achieved by our closed-loop scheme with no knowledge of  $L$  is very close to this lower bound, which demonstrates the super-resolution accuracy of our solution. Additionally, our closed-loop CE scheme adopting the random transmit precoding matrix  $\mathbf{F}_u$  is also illustrated in Fig. 5, where it is observed to suffer from around 5 dB and 3 dB performance losses in the cases of  $L = 3$  and  $5$ , respectively. This clearly demonstrates the effectiveness of the proposed multi-beam transmit precoding matrix design which fully exploits the estimated horizontal/vertical AoAs to optimize the received SNR for improving CE performance. Furthermore, the results of Fig. 5 show that our proposed closed-loop CE scheme dramatically outperforms the two CS-based schemes, in terms of CE accuracy. In particular, the OMP and SW-OMP based schemes seem to suffer from the NMSE floor at high SNR. By adopting larger discrete

<sup>7</sup>The redundant dictionary of OMP-based time-domain method in [18] is generated by the quantized grids at the delay and angle domains, which imposes the unaffordable computational complexity and storage requirements. Hence, we just consider the frequency-domain scheme in simulations.



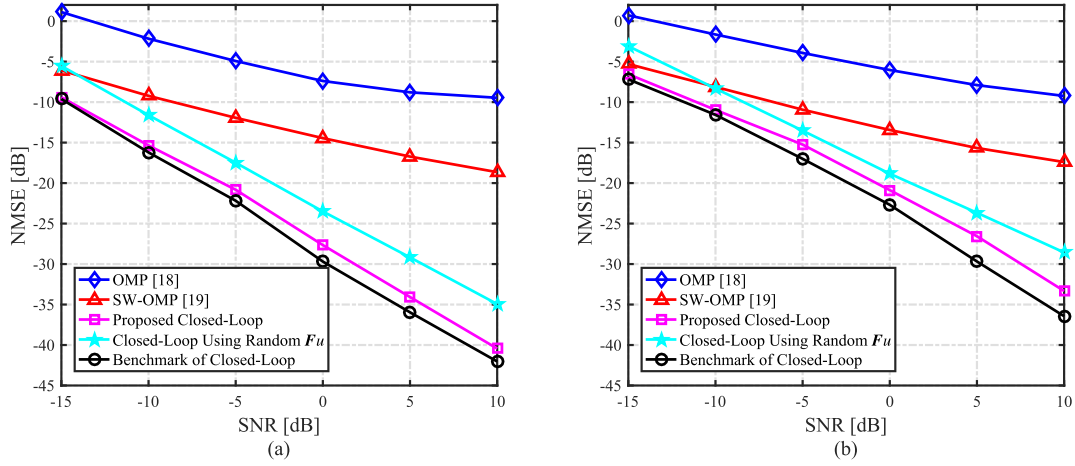


Fig. 5. NMSE performance comparison of different CE schemes versus SNRs with the same training overhead  $T_{CE} = 132$ : (a) the number of MPCs  $L = 3$ ; and (b)  $L = 5$ .

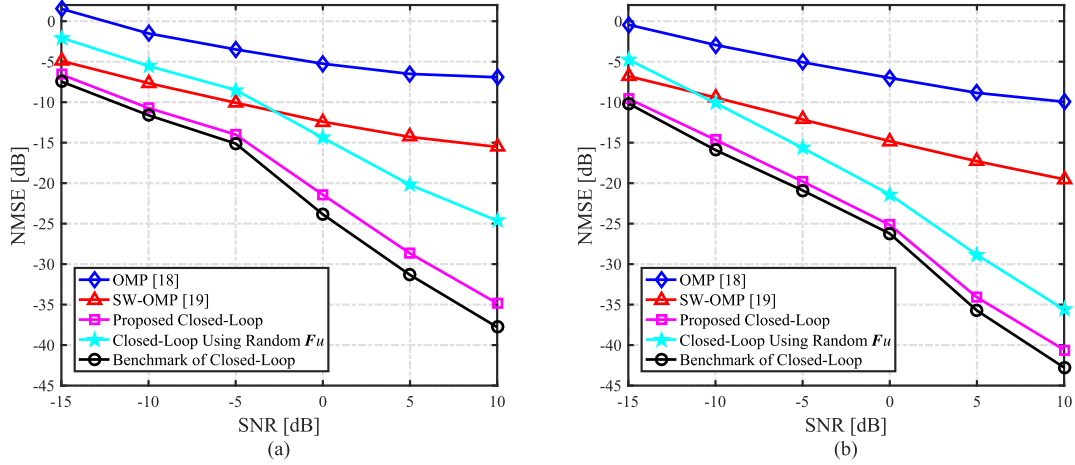


Fig. 6. NMSE performance comparison of different CE schemes versus SNRs with the same number of MPCs  $L = 4$ : (a) training overhead  $T_{CE} = 88$ ; and (b)  $T_{CE} = 176$ .

angle-domain grids to achieve larger quantized CS dictionary, the performance of these CS-based schemes can be improved [18], [19] at the expense of significantly increased computational complexity, which becomes unaffordable for FD-MIMO systems with massive antenna array.

Fig. 6 compares the NMSE performance of different CE schemes against different SNRs, given two training overheads with the same number of MPCs  $L = 4$ .  $T_{CE} = 88$  in Fig. 6a and  $T_{CE} = 176$  in Fig. 6b correspond to choosing  $N_o^d = N_o^u = 2$  and  $N_o^d = N_o^u = 4$  in our scheme, respectively. From Fig. 6, similar conclusions to those observed for Fig. 5 can be obtained. In particular, it can be seen that our closed-loop CE scheme considerably outperforms the two CS-based schemes.

Fig. 7 compares the NMSE performance of different CE schemes versus the number of MPCs  $L$  given two SNR values of 0 dB and 10 dB as well as two training overheads of  $T_{CE} = 88$  and  $T_{CE} = 176$ . In simulations, we adopted the same parameter settings as in Fig. 6 except the number of MPCs. From Fig. 7, the good performance of the proposed multi-beam transmit precoding matrix design and EVD-based approach for MPCs' number estimation is evident. Again, the proposed closed-loop scheme significantly outperforms two

other CS-based schemes. Observe that the performance gain of our scheme over the two other schemes increases for sparser mmWave channels, i.e., having smaller number of MPCs. Moreover, although the performance gap between the proposed solution and the CS-based methods is gradually reduced at low SNRs as the number of MPCs  $L$  increases, the proposed scheme can achieve the considerable performance gain over the CS schemes at high SNRs. Hence, the proposed scheme is suitable for sparse mmWave channels, and more performance gain can be obtained for sparser channels.

Next we consider the average spectral efficiency (ASE) performance metric [41] defined as

$$\text{ASE} = \frac{1}{K} \sum_{k=0}^{K-1} \log_2 \det (\mathbf{I}_{N_s} + \frac{1}{N_s} \mathbf{R}_n^{-1}[k] \mathbf{W}_c^H[k] \mathbf{H}[k] \mathbf{F}_p[k] \mathbf{F}_p^H[k] \mathbf{H}^H[k] \mathbf{W}_c[k]), \quad (37)$$

where  $\mathbf{R}_n[k] = \sigma_n^2 \mathbf{W}_c^H[k] \mathbf{W}_c[k]$ ,  $\mathbf{F}_p[k] = \mathbf{F}_{\text{RF},p} \mathbf{F}_{\text{BB},p}[k]$  and  $\mathbf{W}_c[k] = \mathbf{W}_{\text{RF},c} \mathbf{W}_{\text{BB},c}[k]$  are the transmit precoding and receive combining matrices used during data transmission, respectively, while  $N_s$  is the number of transmit data stream.

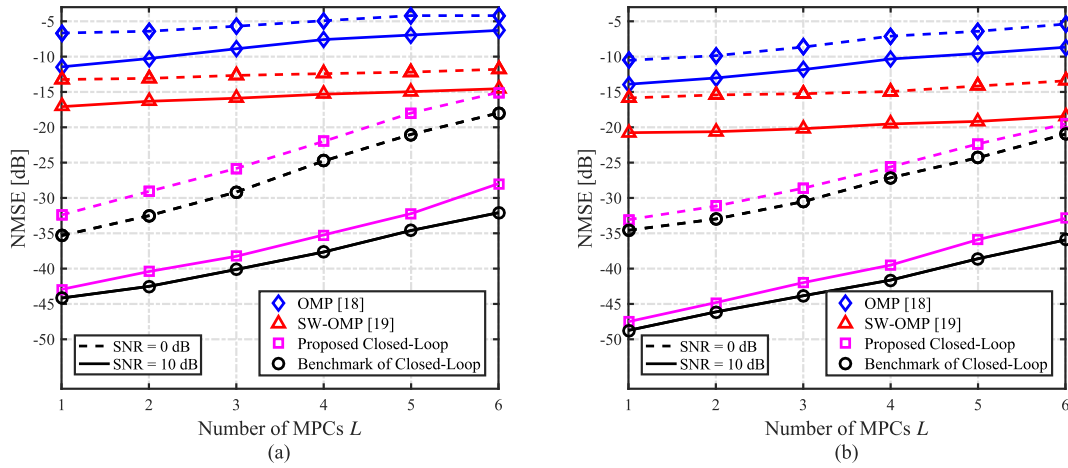


Fig. 7. NMSE performance comparison of different CE schemes versus the number of MPCs  $L$  with SNR = 0 dB and 10 dB: (a) training overhead  $T_{CE} = 88$ ; and (b)  $T_{CE} = 176$ .

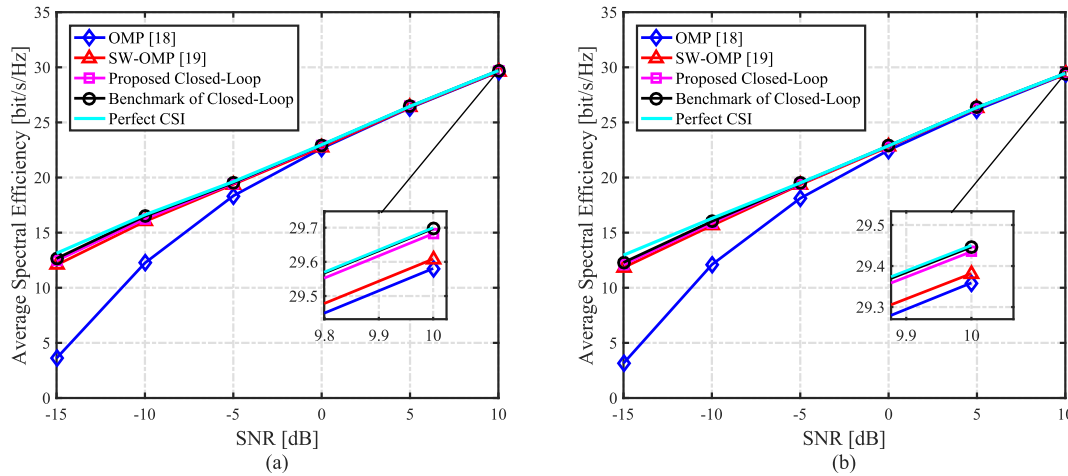


Fig. 8. ASE performance comparison of different CE schemes versus SNRs: (a) the number of MPCs  $L = 3$ ; and (b)  $L = 5$ .

The principle component analysis (PCA)-based hybrid beamforming scheme proposed in [41] is used to evaluate the ASE performance, where the CSI is based on the estimated channels. Besides, the spectral efficiency of the PCA-based hybrid beamforming scheme with the perfect CSI known both to the BS and UD is adopted as the performance upper bound. Here, the same simulation parameters used in Fig. 5 are considered, and the number of transmit data streams used is  $N_s = 2$  in (37). Fig. 8 compares the ASE performance of different CE schemes against different SNRs. It can be observed from Fig. 8 that the ASE performance using the CSI estimated by the proposed scheme closely matches to the performance upper bound obtained using the perfect CSI at both the BS and UD. It can also be seen that the ASE performance gain achieved by the proposed scheme over the two CS-based schemes is 0.1 [bit/s/Hz] at high SNR conditions. At low SNRs, this gain is clearly larger. Note that the ASE performance of the OMP based scheme is particularly poor when  $\text{SNR} \leq 0$  dB.

### B. Computational Complexity Evaluation

The computational complexity analysis of our closed-loop CE scheme is detailed in Table I, where the notation  $\mathcal{O}(N)$

stands for ‘on order of  $N$ ’. The computational requirements of Step 1(a) and Step 5(a) are omitted, since they are much smaller, compared with the requirements of other steps. Clearly, Step 4 does not involve computation. It can be seen that the computational requirements are dominated by Step 6 (corresponding to Algorithm 2 with  $R = 3$ ) and Step 7. Also observe that the complexity of the CE scheme increases fast as  $\hat{L}$  increases, since the computational complexity of Step 6 and Step 7 are proportional to  $\hat{L}^4$  and  $\hat{L}!$ , respectively.

The computational complexity of the two CS-based CE schemes is given in Table II for comparison, where the numbers of iterations for the OMP algorithm at the  $k$ th subcarrier and the SW-OMP algorithm are denoted by  $I_k$  and  $I$ , respectively. Note that the values of  $I_k$  are different for different subcarriers. It can be seen that the computational complexity of these two CS-based schemes increase fast as the quantized grids  $G_{BS}$  and  $G_{UD}$  increase. Also the complexity of the OMP scheme is around  $K$  times of the SW-OMP scheme, because the  $K$  subchannels at  $K$  subcarriers are independently estimated in the OMP scheme but they are jointly estimated in the SW-OMP scheme. Due to the power leakage caused by the mismatch between the discrete

TABLE I  
COMPUTATIONAL COMPLEXITY OF PROPOSED CLOSED-LOOP SCHEME

Operation	Complexity
Step 1(b)&5(b)	$O(N_d N_{UD} N_{UD}^{RF} (1 + N_s^d) + N_u N_{BS} N_{BS}^{RF} (1 + N_s^u))$
Step 2	$O((N_{UD}^{sub})^3 + (N_{UD}^{sub})^2 N_o^d N_P)$
Step 3 ( $R = 2$ )	$O\left(\underbrace{M_{sub}^{UD} K N_o^d G_1^d G_2^d}_{2-D\ SSP} + \underbrace{8 M_{sub}^{UD} K N_o^d G_1^d G_2^d + \frac{1}{4} M_{sub}^{UD} \widehat{L}^2}_{RVP} + \frac{1}{2} \left(\widehat{L}^3 + 2\widehat{L}^2 M_{sub}^{UD} + 2(\widehat{L}+1)(M_{sub}^{UD})^2\right)}_{SSA} + \underbrace{\frac{1}{4} \widehat{L}^3}_{SIES} + \underbrace{\frac{1}{4} \widehat{L}^3}_{2-D\ JD}\right)$
Step 6 ( $R = 3$ )	$O\left(\underbrace{M_{sub}^{BS} N_o^u G_1^u G_2^u G_3^u}_{3-D\ SSP} + \underbrace{8 M_{sub}^{BS} N_o^u G_1^u G_2^u G_3^u + \frac{1}{4} M_{sub}^{BS} \widehat{L}^2}_{RVP} + \frac{3}{4} \left(\widehat{L}^3 + 2\widehat{L}^2 M_{sub}^{BS} + 2(\widehat{L}+1)(M_{sub}^{BS})^2\right)}_{SSA} + \underbrace{\frac{3}{4} \widehat{L}^4}_{SIES} + \underbrace{\frac{3}{4} \widehat{L}^4}_{3-D\ JD}\right)$
Step 7	$O(\widehat{L}!(\widehat{L}^3 + 2\widehat{L}^2 K N_{BS}^{sub} N_o^u))$
Step 8	$O(K \widehat{L} N_{BS} N_{UD})$

TABLE II  
COMPUTATIONAL COMPLEXITY OF TWO CS-BASED CE SCHEMES

Operation	OMP-based Scheme [18]	SW-OMP-based Scheme [19]
Measurement matrix	$O(K T_{CE} N_{BS}^{RF} N_{BS} N_{UD} G_{BS} G_{UD})$	$O(T_{CE} N_{BS}^{RF} N_{BS} N_{UD} G_{BS} G_{UD})$
Whitening	NA	$O(T_{CE} N_{BS}^{RF} ((T_{CE} N_{BS}^{RF})^2 + K + G_{BS} G_{UD}))$
Correlation	$O(T_{CE} N_{BS}^{RF} G_{BS} G_{UD} (\sum_{k=1}^K I_k))$	$O(T_{CE} N_{BS}^{RF} G_{BS} G_{UD} K I)$
Project subspace	$O(\sum_{k=1}^K (\frac{1}{4} I_k^2 (I_k + 1)^2 + \frac{1}{3} I_k (I_k + 1) (2I_k + 1) T_{CE} N_{BS}^{RF}))$	$O(\frac{1}{4} I^2 (I + 1)^2 + \frac{1}{3} K I (I + 1) (2I + 1) T_{CE} N_{BS}^{RF})$
Update residual	$O(T_{CE} N_{BS}^{RF} (\sum_{k=1}^K \frac{I_k}{2} (I_k + 1)))$	$O(T_{CE} N_{BS}^{RF} K \frac{I}{2} (I + 1))$
Compute MSE	$O(T_{CE} N_{BS}^{RF} (\sum_{k=1}^K I_k))$	$O(T_{CE} N_{BS}^{RF} K^2 I)$
Reestablishment	$O(N_{BS} N_{UD} (\sum_{k=1}^K I_k))$	$O(N_{BS} N_{UD} K I)$

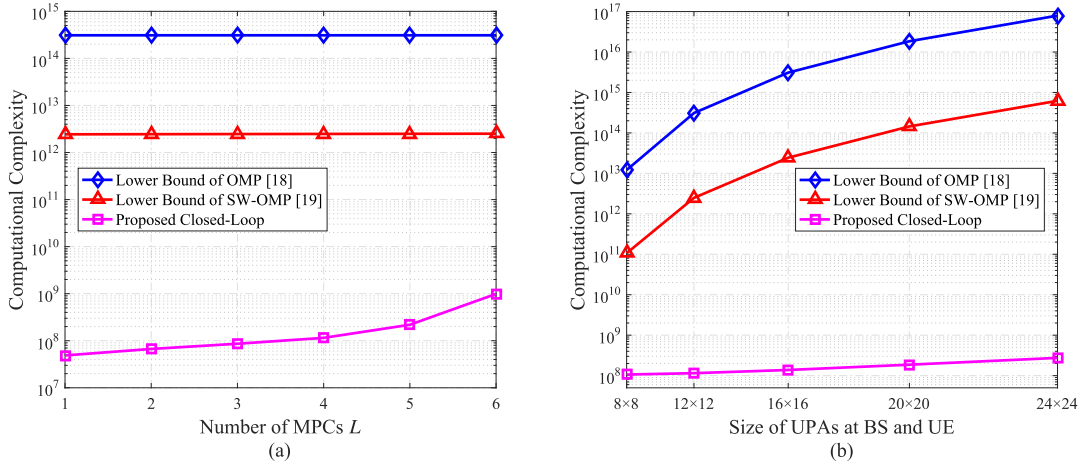


Fig. 9. Computational complexity comparison of different CE schemes given the training overhead  $T_{CE} = 88$ : (a) sizes of UPAs at BS and UD are both  $12 \times 12$ ; and (b) number of MPCs  $L = 4$ .

CS angle-domain dictionary and continuously distributed AoAs/AoDs of channels, the number of effective MPCs represented in the redundant CS dictionary are usually greater than  $L$ . Hence, the value of  $I_k$  in the OMP scheme and the value of  $I$  in the SW-OMP scheme are not fixed and they are usually greater than  $L$ . Therefore, we can use  $I = I_k = L$  to provide the lower bounds of the computational complexity for the two CS-based schemes.

Fig. 9 compares the computational complexity of our closed-loop CE scheme with those of the two CS-based

schemes given the training overhead  $T_{CE} = 88$  corresponding to  $N_o^d = N_o^u = 2$  in our scheme. From Fig. 9a, we observe that the computational complexity of the proposed CE solution increases slightly as the number of MPCs increase. Most strikingly, however, given the size of UPA as  $12 \times 12$ , the complexity of our solution is at least 3 orders of magnitude lower than the SW-OMP scheme and at least 5 orders of magnitude lower than the OMP scheme. The results of Fig. 9b indicate that given  $L = 4$ , the complexity of our solution is almost immune to the size of UPA at the BS and UD.



TABLE III  
COMPARISON OF ADVANTAGES AND DISADVANTAGES OF DIFFERENT CE SCHEMES

		Proposed closed-loop scheme	OMP [18] and SW-OMP [19] schemes	
			CE in Downlink	CE in Uplink
Overhead	Training overhead	<b>Medium</b> ; $Q$ users share downlink CE but perform respective uplink CE	<b>Small</b> ; $Q$ users share downlink CE	<b>Large</b> ; $Q$ users have to perform respective uplink CE
	Feedback overhead	<b>Small</b> ; Quantized dominant channel parameters are fed back/forward	<b>Large</b> ; Support sets and channel gains for every subcarrier are fed back	<b>Large</b> ; Support sets and channel gains for every subcarrier are fed forward
Storage requirements		<b>Small</b> storage at UDs in downlink; <b>Medium</b> storage required at BS in uplink	<b>Excessive high</b> storage of large redundant dictionary required at UDs	<b>High</b> storage of large redundant dictionary required at BS
Computational complexity (CC)		<b>Lower</b> CC imposed on UDs in downlink; <b>Medium</b> CC imposed on BS in uplink	<b>Excessive high</b> CC caused by large matrix operations at UDs	<b>High</b> CC due to large matrix operations at BS
Received SNR at CE		<b>High</b> ; Large transmit power in downlink; Directional multi-beam design in uplink	<b>High</b> ; Large transmit power at BS	<b>Low</b> ; Limited transmit power of UDs

By contrast, the complexity of the two CS-based schemes increase considerably as the number of antennas increases. Again, the complexity of our solution is several orders of magnitude lower than the other two schemes. It should also be reiterated that to mitigate the power leakage, the CS-based schemes adopt the high-dimensional redundant dictionary, which results in unaffordable storage space requirements when the number of antennas is large. Clearly, for FD-MIMO systems with massive number of antennas, the proposed closed-loop scheme offers considerable advantage over the CS-based schemes, in terms of both computational complexity and storage requirements.

The advantages and disadvantages of our proposed solution and two other CS-based CE schemes are given in Table III, where the training/feedback overhead, storage requirements, computational complexity and received SNR at CE are compared.

## VII. CONCLUSION

We have proposed a closed-loop sparse CE scheme for multi-user wideband mmWave FD-MIMO systems with hybrid beamforming. By exploiting the sparsity of mmWave channels in both angle and delay domains and by visualizing high-dimensional hybrid arrays as low-dimensional digital arrays, the proposed scheme is capable of obtaining the super-resolution estimates of horizontal/vertical AoDs/AoAs and delays based on the MDU-ESPRIT algorithm. Specifically, at the downlink CE stage, we design the common random transmit precoding matrix at the BS and the receive combining matrix at each UD to estimate the horizontal/vertical AoAs of sparse MPCs. At the uplink CE stage, based on the designed receive combining matrix at the BS, we estimate horizontal/vertical AoDs and delays. Furthermore, the AoAs estimated at each UD are utilized to design the multi-beam transmit precoding matrix for further enhancing CE performance. We also propose an ML approach at the BS to pair the channel parameters acquired at the two stages and to optimally estimate the path gains. Simulation results have demonstrated that the proposed closed-loop CE scheme offers considerable advantages over state-of-the-art CS-based CE schemes, in terms of providing significantly more accurate CSI estimate while imposing dramatically lower computational complexity and storage requirements.

## APPENDIX

Sampling the delay-domain continuous  $\mathbf{H}_q(\tau)$  in (3) with the sampling period  $T_s$  yields

$$\begin{aligned} \mathbf{H}_q(nT_s) &= \beta_q \sum_{l=1}^{L_q} \mathbf{H}_{q,l} p(\tau - \tau_{q,l}) \otimes \sum_{n=-\infty}^{\infty} \delta(\tau - nT_s) \\ &= \beta_q \sum_{n=-\infty}^{\infty} \sum_{l=1}^{L_q} \mathbf{H}_{q,l} p(nT_s - \tau_{q,l}), \end{aligned} \quad (38)$$

where  $\otimes$  and  $\delta(\cdot)$  represent the linear convolution operation and Dirac delta function, respectively. The Fourier transform of  $\mathbf{H}_q(nT_s)$  is then given by

$$\mathbf{H}_q(f) = \frac{\beta_q}{T_s} \sum_{l=1}^{L_q} \sum_{n=-\infty}^{\infty} \mathbf{H}_{q,l} P(f) e^{-j2\pi f \tau_{q,l}} \delta(f - n f_s), \quad (39)$$

where  $P(f)$  is the Fourier transform of  $p(\tau)$ . Obviously,  $\mathbf{H}_q(f)$  exhibits periodicity with period  $f_s$ . Thus,  $\mathbf{H}_q(f)$  within a period of  $f \in [-f_s/2, f_s/2]$  can be expressed as

$$\begin{aligned} \mathbf{H}_q(f) &= \frac{\beta_q}{T_s} \sum_{l=1}^{L_q} P(f) \mathbf{H}_{q,l} e^{-j2\pi f \tau_{q,l}} \\ &\approx \frac{\beta_q}{T_s} \sum_{l=1}^{L_q} C \mathbf{H}_{q,l} e^{-j2\pi f \tau_{q,l}}. \end{aligned} \quad (40)$$

The approximation in (40) is valid because the PSF  $p(\tau)$  is designed to realize the ideal passband filter characteristics of  $P(f) = C$  for  $f \in [-f_s/2, f_s/2]$  and  $P(f) \approx 0$  for  $f \notin [-f_s/2, f_s/2]$ . For convenience, we consider  $C = T_s$ . Therefore, the frequency-domain channel matrix  $\mathbf{H}_q[k]$  at the  $k$ th subcarrier, where  $0 \leq k \leq K - 1$ , can be written as

$$\mathbf{H}_q[k] = \mathbf{H}_q\left(\frac{k f_s}{K}\right) = \beta_q \sum_{l=1}^{L_q} \mathbf{H}_{q,l} e^{-j \frac{2\pi k f_s \tau_{q,l}}{K}}. \quad (41)$$

## REFERENCES

- [1] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

- [2] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shaker, "mmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
- [3] W. Ma and C. Qi, "Beamspace channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4839–4853, Sep. 2018.
- [4] Y. Huang, J. Zhang, and M. Xiao, "Constant envelope hybrid precoding for directional millimeter-wave communications," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 845–859, Apr. 2018.
- [5] A. Liu and V. K. N. Lau, "Impact of CSI knowledge on the codebook-based hybrid beamforming in massive MIMO," *IEEE Trans. Signal Process.*, vol. 64, no. 24, pp. 6545–6556, Dec. 2016.
- [6] A. Liu, V. K. N. Lau, and M.-J. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 432–444, Jun. 2018.
- [7] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, Nov. 2017.
- [8] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard 802.11ad, 2012.
- [9] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, "Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 601–616, Feb. 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] Z. Xiao, P. Xia, and X.-G. Xia, "Channel estimation and hybrid precoding for millimeter-wave MIMO systems: A low-complexity overall solution," *IEEE Access*, vol. 5, pp. 16100–16110, Aug. 2017.
- [12] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [13] Y. Wu, Y. Gu, and Z. Wang, "Channel estimation for mmWave MIMO with transmitter hardware impairments," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 320–323, Feb. 2018.
- [14] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.
- [15] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [16] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [17] Y. Dong, C. Chen, N. Yi, G. Lu, and Y. Jin, "Channel estimation using low-resolution PSs for wideband mmWave systems," in *Proc. VTC-Spring*, Sydney, NSW, Australia, Jun. 2017, pp. 1–5.
- [18] K. Venugopal, A. Alkhatieb, N. González-Prelcic, and R. W. Heath, Jr., "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. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, Jr., "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] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1486–1497, Jul. 2017.
- [21] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.
- [22] 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.
- [23] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6150–6163, Dec. 2018.
- [24] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.
- [25] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeter-wave communication: Potentials and approaches," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 66–73, May 2016.
- [26] C. Zhang, W. Zhang, W. Wang, L. Yang, and W. Zhang, "Research challenges and opportunities of UAV millimeter-wave communications," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 58–62, Feb. 2019.
- [27] P. Yu *et al.*, "Capacity enhancement for 5G networks using mmWave aerial base stations: Self-organizing architecture and approach," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 58–64, Aug. 2018.
- [28] S. Jaekel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "Quasi deterministic radio channel generator: User manual and documentation," Fraunhofer Inst. Telecommun., Berlin, Germany, Tech. Rep. v2.2.0, 2019.
- [29] S. Jaekel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
- [30] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb. 1996.
- [31] A.-J. van der Veen, C. M. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
- [32] 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.
- [33] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.
- [34] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.
- [35] M. C. Vanderveen, A.-J. van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.
- [36] J. Mao, Z. Gao, Y. Wu, and M.-S. Alouini, "Over-sampling codebook-based hybrid minimum sum-mean-square-error precoding for millimeter-wave 3D-MIMO," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 938–941, Dec. 2018.
- [37] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 161–169, Jan. 1998.
- [38] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [39] Z. Xiao, X. G. Xia, and L. Bai, "Achieving antenna and multipath diversities in GLRT-based burst packet detection," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1832–1845, Apr. 2015.
- [40] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.
- [41] Y. Sun, Z. Gao, H. Wang, and D. Wu, "Wideband hybrid precoding for next-generation backhaul/fronthaul based on mmWave FD-MIMO," in *Proc. IEEE Globecom Workshops*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.
- [42] Z. Xiao, C. Zhang, D. Jin, and N. Ge, "GLRT approach for robust burst packet acquisition in wireless communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1127–1137, Mar. 2013.



**Anwen Liao** (S'19) received the B.S. and M.S. degrees from the Beijing Institute of Technology, Beijing, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include massive MIMO systems, millimeter-wave communications, hybrid beamforming, and unmanned aerial vehicle communications.



**Zhen Gao** received the B.S. degree in information engineering from the Beijing Institute of Technology, Beijing, China, in 2011, and the Ph.D. degree in communication and signal processing with the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, China, in 2016.

He is currently an Assistant Professor with the Beijing Institute of Technology. His research interests are in wireless communications, with a focus on multi-carrier modulations, multiple antenna systems, and sparse signal processing. He was a recipient of the IEEE Broadcast Technology Society 2016 Scott Helt Memorial Award (Best Paper), the Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS in 2016, *IET Electronics Letters* Premium Award (Best Paper) 2016, and the Young Elite Scientists Sponsorship Program (2018–2021) from China Association for Science and Technology.



**Hua Wang** received the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 1999.

From February 2009 to January 2010, he was a Visiting Professor with the Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include the fields of communication theory and signal processing, wireless networking, modem design, and implementation for satellite communications.



**Sheng Chen** (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from the City, University of London, in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., where he is currently a Professor of Intelligent Systems and Signal Processing. He has published more than 650 research articles. He has 13400 plus Web of Science Citations and 28000 plus Google Scholar Citations. His research interests include neural network and machine learning, adaptive signal processing, and wireless communications.

Dr. Chen is a fellow of the Royal Academy of Engineering, U.K., a fellow of the IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an original ISI highly cited Researcher in engineering in March 2004.



**Mohamed-Slim Alouini** (S'94–M'98–SM'03–F'09) was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998.

He was a Faculty Member with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, and then the Electrical and Computer Engineering Program at the Texas A&M University at Qatar, Doha, Qatar. Since July 2009, he has been with the Electrical Engineering Program, Division of Physical and Chemical Sciences and Engineering, King Abdullah University of Sciences and Technology (KAUST), Thuwal, Saudi Arabia. His current research interests include the modeling, design, and performance analysis of wireless communication systems.



**Hao Yin** received the B.S. degree in microwave communication and the M.S. degree in communication and information systems from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1987, respectively, and the Ph.D. degree in communication and information systems from the Beijing Institute of Technology, Beijing, China, in 1999.

He is currently an Adjunct Professor with PLA Army Engineering University, Nanjing, and a Researcher with the Institute of China Electronic System Engineering. His research interests include wireless communication networks and information systems.