MMV-Based Sequential AoA and AoD Estimation for Millimeter Wave MIMO Channels

Wei Zhang, *Member, IEEE*, Miaomiao Dong, and Taejoon Kim, *Senior Member, IEEE*

***Abstract*— The fact that the millimeter-wave (mmWave) multiple-input multiple-output (MIMO) channel has sparse support in the spatial domain has motivated recent compressed sensing (CS)-based mmWave channel estimation methods, where the angles of arrivals (AoAs) and angles of departures (AoDs) are quantized using angle dictionary matrices. However, the existing CS-based methods usually obtain the estimation result through one-stage channel sounding that have two limitations: (i) the requirement of large-dimensional dictionary and (ii) unresolvable quantization error. These two drawbacks are irreconcilable; improvement of the one implies deterioration of the other. To address these challenges, we propose, in this paper, a two-stage method to estimate the AoAs and AoDs of mmWave channels. In the proposed method, the channel estimation task is divided into two stages, Stage I and Stage II. Specifically, in Stage I, the AoAs are estimated by solving a multiple measurement vectors (MMV) problem. In Stage II, based on the estimated AoAs, the receive sounders are designed to estimate AoDs. The dimension of the angle dictionary in each stage can be reduced, which in turn reduces the computational complexity substantially. We then analyze the successful recovery probability (SRP) of the proposed method, revealing the superiority of the proposed framework over the existing one-stage CS-based methods. We further enhance the reconstruction performance by performing resource allocation between the two stages. We also overcome the unresolvable quantization error issue present in the prior techniques by applying the atomic norm minimization method to each stage of the proposed two-stage approach. The simulation results illustrate the substantially improved performance with low complexity of the proposed two-stage method.**

***Index Terms*— Millimeter wave communications, compressed sensing, channel estimation, multiple-input multiple-output system, support recovery, and sequential estimation.**

Manuscript received September 2, 2021; revised February 22, 2022; accepted April 10, 2022. Date of publication April 20, 2022; date of current version June 16, 2022. The work of Taejoon Kim was supported in part by the National Science Foundation (NSF) under Grant CNS1955561 and in part by the Office of Naval Research (ONR) under Grant N00014-21-1-2472. The associate editor coordinating the review of this article and approving it for publication was I. Guvenc. *(Corresponding author: Wei Zhang.)*

Wei Zhang was with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong. He is now with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: weizhang@ntu.edu.sg).

Miaomiao Dong is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong (e-mail: miao4600@163.com).

Taejoon Kim is with the Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045 USA (e-mail: taejoonkim@ku.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3168886.

Digital Object Identifier 10.1109/TCOMM.2022.3168886

I. Introduction

THE spectrum-rich millimeter-wave (mmWave) frequencies between $30-300$ GHz have the potential to alleviate the current spectrum crunch in sub-6GHz bands that service providers are already experiencing. This major potential of the mmWave band has made it one of the most important components of future mobile cellular and emerging WiFi networks. However, due to significant differences between systems operating in mmWave and legacy sub-6 GHz bands, providing reliable and low-delay communication in the mmWave bands is extremely challenging. Specifically, to achieve the high spectral efficiency of mmWave communications, accurate channel state information (CSI) is the key [1]–[5], which is, however, challenging due to the high dimensionality of the channel as well as the mmWave hardware constraints.

Nevertheless, the mmWave multiple-input multiple-output (MIMO) channel exhibits sparse property [6], [7], facilitating the sparse channel representation by using small numbers of the angles of arrivals (AoAs), angles of departures (AoDs), and path gains. Typically, by approximating the AoAs and AoDs to be on quantized angle grids, the compressed sensing (CS)-based approaches transform the AoA and AoD estimation problem to a sparse signal recovery problem [8], [9], where the transmitter sends the channel sounding beams to the receiver and the receiver jointly estimates AoAs and AoDs. We refer to this method as the one-stage channel sounding scheme. In particular, due to easy implementation and amenability for analysis, the orthogonal matching pursuit (OMP) has been widely studied [9]–[13]. The OMP iteratively searches a pair of AoA and AoD over an over-complete dictionary. However, the computational complexity of OMP increases quadratically with the sizes of the dictionaries, i.e., $O(LKG_rG_t)$, where K is the number of channel uses for the channel sounding, L is the number of channel paths, and G_r and G_t are the dimensions of angle dictionaries for AoA and AoD, respectively. It is worth pointing out that when the dimensions of the over-complete dictionaries, i.e., G_r and G_t, increase, the complexity of the one-stage CS-based methods such as OMP becomes exceedingly impractical.

The over-complete dictionary and high computational complexity issues have been addressed in an adaptive-CS point-of-view with the primary focus on the sensing vector adaptation to the previous observations [3], [8], [14]. Theoretically, it has been shown that the adaptive CS can be beneficial in low SNR [15]. The multi-level (hierarchical) AoA and AoD search techniques [3], [8] leveraged the feedback, where the receiver

conveys a feedback to the transmitter to guide the next level angle dictionary design. It is worth noting that these adaptation methods [3], [8] need multiple feedbacks and its performance critically relies on the reliability of the feedback. To reduce the feedback overhead, a two-stage CS was proposed in [14], where the first stage is to obtain a coarse estimation of the support set and the second stage refines the result of the first stage. This method [14] only requires one-time feedback, but achieves compatible estimation performance in low SNR.

A. Our Contributions

We newly study a sequential, two-stage AoA and AoD estimation framework for reduced computational complexity and improved estimation performance. Specifically, in Stage I, the support set of AoAs is recovered at the receiver by solving a multiple measurement vectors (MMV) problem. Leveraging the shared sparse set, it has been found that the MMV approach can provide improved estimation performance compared to the single measurement vector (SMV) approach [16]–[18]. In Stage II, the receiver estimates the AoDs of the channel by exploiting the estimated AoAs from Stage I. Importantly, the estimated AoAs guide the design of receive sounding signals, which saves the channel use overhead and improves the accuracy of AoD estimation. In each stage, since we only estimate AoAs or AoDs, the dimensions of the signal and angle dictionary are much smaller than those of the one-stage joint AoA and AoD estimation [9], [11], [12], readily reducing the computational complexity substantially. This can be viewed as of converting the multiplicative channel sounding overhead (e.g., $\mathcal{O}(G_r G_t)$ of OMP) to an additive overhead.

By analyzing the MMV statistics, we present a lower bound for the successful probability of recovering the support sets. Furthermore, based on the successful recovery probability (SRP) analysis of the proposed two-stage method, a resource allocation (between Stage I and Stage II) strategy is newly proposed to improve SRPs for both AoA and AoD estimation. The numerical results validate the efficacy of the proposed resource allocation method.

Finally, in order to address the issue of unresolvable quantization error, we extend the proposed two-stage method to the one with super resolution. Specifically, in each stage of AoA or AoD estimation, we reformulate the MMV problem as an atomic norm minimization problem [19]–[21], which is solved by using alternating direction method of multipliers (ADMM). Compared to the dictionary-based methods, the atomic norm minimization can be thought of as the case when the infinite dictionary matrix is employed. We demonstrate through simulations that the quantization error of the two-stage method with super resolution can be effectively reduced.

B. Paper Organization and Notations

The paper is organized as follows. In Section II, we introduce the signal model and the CS-based channel estimation problem. In Section III, based on the angular-domain channel representation, the proposed sequential AoA and AoD estimation method is presented. In Section IV, we analyze the proposed method in terms of SRP and introduce the resource allocation strategy. In Section V, the atomic norm-based design is described, which resolves the quantization error in the estimated AoAs and AoDs. The simulation results and conclusion are presented in Section VI and Section VII, respectively.

Notations: A bold lower case letter $\mathbf{a}$ is a vector and a bold capital letter $\mathbf{A}$ is a matrix. $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\mathrm{tr}(\mathbf{A})$, $|\mathbf{A}|$, $\|\mathbf{A}\|_F$ and $\|\mathbf{a}\|_2$ are, respectively, the transpose, conjugate, Hermitian, inverse, trace, determinant, Frobenius norm of $\mathbf{A}$, and ℓ_2-norm of $\mathbf{a}$. $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ denotes the pseudo inverse of a tall matrix $\mathbf{A}$. $[\mathbf{A}]_{:,i}$, $[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{i,j}$, and $[\mathbf{a}]_i$ are, respectively, the ith column, ith row, ith row and jth column entry of $\mathbf{A}$, and ith entry of vector $\mathbf{a}$. $\mathrm{vec}(\mathbf{A})$ stacks the columns of $\mathbf{A}$ and forms a long column vector. $\mathrm{diag}(\mathbf{a})$ returns a square diagonal matrix with the vector $\mathbf{a}$ on the main diagonal. $\mathbf{I}_M \in \mathbb{R}^{M\times M}$ is the M-dimensional identity matrix. The $\mathbf{1}_{M,N} \in \mathbb{R}^{M\times N}$ and $\mathbf{0}_{M,N} \in \mathbb{R}^{M\times N}$ are the all one matrix, and zero matrix, respectively. $\mathcal{R}(\mathbf{F})$ denotes the subspace spanned by the columns of matrix $\mathbf{F}$. $\mathbf{A}\otimes\mathbf{B}$ and $\mathbf{A}\circ\mathbf{B}$ denote the Kronecker product and Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$, respectively. The $\lceil x \rceil$ returns the smallest integer greater than or equal to x.

II. System Model and General Statement of Techniques

A. Channel Model

The mmWave transmitter and receiver are equipped with N_t and N_r antennas, respectively. Suppose that the number of separable paths between the transmitter and receiver is L, where $L \ll \min\{N_r, N_t\}$. The physical mmWave channel representation based on the uniform linear array [9], [22]–[24] is given by,[1]

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_r(f_{r,l})\mathbf{a}_t^H(f_{t,l}), \tag{1}$$

where $\mathbf{a}_t(\cdot) \in \mathbb{C}^{N_t\times 1}$ and $\mathbf{a}_r(\cdot) \in \mathbb{C}^{N_r\times 1}$ are the array response vectors of the transmit and receive antenna arrays. Specifically, $\mathbf{a}_t(f)$ and $\mathbf{a}_r(f)$ are given by $\mathbf{a}_t(f) = \frac{1}{\sqrt{N_t}}\left[1, e^{j2\pi f}, \ldots, e^{j2\pi(N_t-1)f}\right]^T$ and $\mathbf{a}_r(f) = \frac{1}{\sqrt{N_r}}\left[1, e^{j2\pi f}, \ldots, e^{j2\pi(N_r-1)f}\right]^T$, where $f \in [0, 1)$ is the normalized spatial angle. Here we assume $f_{r,l}$ and $f_{t,l}$ in (1) are independent and uniformly distributed in $[0, 1)$, and the gain of the lth path α_l follows the complex Gaussian distribution, i.e., $\alpha_l \sim \mathcal{CN}(0, \sigma_l^2)$. Angular domain representation of the channel in (1) can be rewritten as

$$\mathbf{H} = \mathbf{A}_r \,\mathrm{diag}(\mathbf{h})\mathbf{A}_t^H, \tag{2}$$

where $\mathbf{A}_r = [\mathbf{a}_r(f_{r,1}), \ldots, \mathbf{a}_r(f_{r,L})] \in \mathbf{C}^{N_r\times L}$, $\mathbf{A}_t = [\mathbf{a}_t(f_{t,1}), \ldots, \mathbf{a}_t(f_{t,L})] \in \mathbf{C}^{N_t\times L}$, and $\mathbf{h} = [h_1, \ldots, h_L] \in \mathbb{C}^{L\times 1}$ with $h_l = \sqrt{\frac{N_r N_t}{L}}\alpha_l$, $l = 1, \ldots, L$.

[1] In wideband communication systems, one can model the channel as constant AoA/AoD and varying path gains [25], [26]. Here we could also assume a narrow band block fading channel where the channel is static during the channel coherence time. The CSI acquisition and data transfer are framed to happen within the channel coherence time [9], [22]–[24].

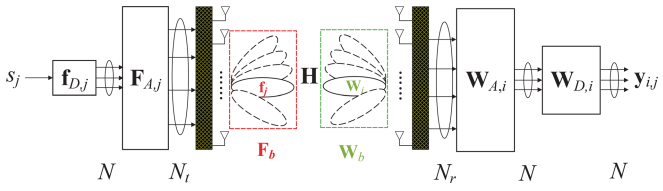

Fig. 1. Conventional one-stage mmWave channel sounding.

B. Channel Sounding

Fig. 1 illustrates the conventional one-stage mmWave channel sounding operation, where the transmitter and receiver are equipped with the large-dimensional hybrid analog-digital MIMO arrays that are driven by a limited number of RF chains, i.e., $N \ll \min\{N_t, N_r\}$. In each channel use of downlink channel sounding, the transmitter generates a beam conveying the pilot signal and the receiver simultaneously generates N separate beams, using the N RF chains, to obtain a N-dimensional observation. We let the numbers of the transmit sounding beams (TSBs) and receive sounding beams (RSBs) for channel estimation be B_t and B_r, respectively. For convenience, we assume that B_r is an integer multiple of N. The total number of channel uses for the conventional one-stage sounding process is then $K = B_r B_t/N$. Specifically, the RSB matrix in Fig. 1 is given by

$$\mathbf{W}_b = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{B_r/N}] \in \mathbf{C}^{N_r \times B_r}, \tag{3}$$

where $\mathbf{W}_i \in \mathbf{C}^{N_r \times N}$ for $i = 1, 2, \ldots, B_r/N$, and $\mathbf{W}_i = \mathbf{W}_{A,i}\mathbf{W}_{D,i}$ with $\mathbf{W}_{A,i} \in \mathbf{C}^{N_r \times N}$ and $\mathbf{W}_{D,i} \in \mathbf{C}^{N \times N}$ being the receive analog and digital sounders, respectively. Similarly, the TSB matrix is given by

$$\mathbf{F}_b = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_{B_t}] \in \mathbf{C}^{N_t \times B_t}, \tag{4}$$

where $\mathbf{f}_j \in \mathbf{C}^{N_t \times 1}$ for $j = 1, 2, \cdots, B_t$ is the jth transmit sounder, and $\mathbf{f}_j = \mathbf{F}_{A,j}\mathbf{f}_{D,j}s_j$ with $\mathbf{F}_{A,j} \in \mathbf{C}^{N_t \times N}$ and $\mathbf{f}_{D,j} \in \mathbf{C}^{N \times 1}$ being the transmit analog and digital sounders, respectively. Each observation $\mathbf{y}_{i,j} \in \mathbb{C}^{N \times 1}$ in Fig. 1, associated with the ith RSB and jth TSB, $i \in \{1, \ldots, B_r/N\}$ and $j \in \{1, 2, \ldots, B_t\}$, can be expressed as

$$\mathbf{y}_{i,j} = \mathbf{W}_i^H \mathbf{H} \mathbf{f}_j s_j + \mathbf{W}_i^H \mathbf{n}_j. \tag{5}$$

The s_j denotes the training signal and without loss of generality, we let $s_j = 1$. It is worth noting that only phase shifters are employed to constitute the analog arrays for power saving, where $|[\mathbf{W}_{A,i}]_{m,n}| = 1/\sqrt{N_r}$, and $|[\mathbf{F}_{A,j}]_{m,n}| = 1/\sqrt{N_t}, \forall m, n$. Moreover, the power constraint $\|\mathbf{f}_j\|_2^2 = p$ is imposed to the transmit sounding beam at each channel use with p being the power budget, and the noise vector follows $\mathbf{n}_j \sim \mathcal{CN}(\mathbf{0}_{N_r}, \sigma^2 \mathbf{I}_{N_r})$. Thus, the signal to noise ratio is p/σ^2.

We collect all observations in (5) by using $\mathbf{W}_b$ in (3) and $\mathbf{F}_b$ in (4) as

$$\mathbf{Y} = \mathbf{W}_b^H \mathbf{H} \mathbf{F}_b + \mathbf{W}_b^H \mathbf{N}, \tag{6}$$

where $\mathbf{Y} \in \mathbb{C}^{B_r \times B_t}$ and $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_{B_t}] \in \mathbb{C}^{N_r \times B_t}$. For example, $\mathbf{W}_b$ and $\mathbf{F}_b$ in (6) can be generated randomly [4] or designed as a partial discrete Fourier transform (DFT) matrix [9]. We assume that the number of observations is strictly lower than the dimension of the channel matrix, i.e., $B_r B_t \ll N_r N_t$. The channel estimation task is to utilize the observations in (5) (equivalently, (6)) to obtain the estimate of the channel matrix $\mathbf{H}$ in (2). Encountering (2), the channel estimation task boils down to reconstructing $\{f_{r,1}, \ldots, f_{r,L}\}$, $\{f_{t,1}, \ldots, f_{t,L}\}$ and $\{h_1, \ldots, h_L\}$ from the observations.

1) Oracle Estimator: The oracle estimator that we will utilize for benchmark[2] is obtained by assuming perfect knowledge of AoAs and AoDs in (2). The oracle channel estimate only needs to estimate the path gain $\mathbf{h}$, thus the channel estimate is expressed as $\widehat{\mathbf{H}} = \mathbf{A}_r \operatorname{diag}(\widehat{\mathbf{h}})\mathbf{A}_t^H$, where $\operatorname{diag}(\widehat{\mathbf{h}}) \in \mathbb{C}^{L \times 1}$ is the solution to the following problem:

$$\widehat{\mathbf{h}} = \arg\min_{\mathbf{h}} \|\mathbf{Y} - \mathbf{W}_b^H \mathbf{A}_r \operatorname{diag}(\mathbf{h}) \mathbf{A}_t^H \mathbf{F}_b\|_F^2. \tag{7}$$

Because (7) is convex, the optimal solution is $\widehat{\mathbf{h}} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H \operatorname{vec}(\mathbf{Y})$, where $\mathbf{X} \in \mathbb{C}^{B_r B_t \times L}$ is given by $\mathbf{X} = \left[\operatorname{vec}\left([\mathbf{W}_b^H\mathbf{A}_r]_{:,1}[\mathbf{A}_t^H\mathbf{F}_b]_{1,:}\right), \ldots, \operatorname{vec}\left([\mathbf{W}_b^H\mathbf{A}_r]_{:,L}[\mathbf{A}_t^H\mathbf{F}_b]_{L,:}\right)\right]$. Because we have $B_r B_t \gg L$, $\mathbf{X}^H\mathbf{X}$ is invertible.

C. Compressed Sensing-Based Channel Estimation

Recalling the channel model in (2), a typical CS framework restricts the normalized spatial angles $f_{r,l}, f_{t,l}$, $l = 1, 2, \ldots, L$, to be chosen from the discrete angle dictionaries, $f_{r,l} \in [0, 1/G_r, \ldots, (G_r - 1)/G_r]$, and $f_{t,l} \in [0, 1/G_t, \ldots, (G_t - 1)/G_t]$, where $G_r = \lceil sN_r \rceil$ and $G_t = \lceil sN_t \rceil$ with $s \geq 1$ are, respectively, the cardinalities of the receive and transmit spatial angle dictionaries. The transmit and receive array response dictionaries are then given by

$$\bar{\mathbf{A}}_r = \left[\mathbf{a}_r(0), \mathbf{a}_r\left(\frac{1}{G_r}\right), \ldots, \mathbf{a}_r\left(\frac{G_r - 1}{G_r}\right)\right] \in \mathbf{C}^{N_r \times G_r}$$

and

$$\bar{\mathbf{A}}_t = \left[\mathbf{a}_t(0), \mathbf{a}_t\left(\frac{1}{G_t}\right), \ldots, \mathbf{a}_t\left(\frac{G_t - 1}{G_t}\right)\right] \in \mathbf{C}^{N_t \times G_t}.$$

For the latter array response dictionaries, the channel model in (2) can be rewritten as

$$\mathbf{H} = \bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H + \mathbf{E}, \tag{8}$$

where $\bar{\mathbf{H}}_a \in \mathbf{C}^{G_r \times G_t}$ is an L-sparse matrix with L non-zero entries corresponding to the positions of AoAs and AoDs on their respective angle grids, and $\mathbf{E} \in \mathbb{C}^{N_r \times N_t}$ denotes the quantization error.

Because the dictionary matrices $\bar{\mathbf{A}}_r$ and $\bar{\mathbf{A}}_t$ are known, the channel estimation task is equivalent to estimating the non-zero entries in $\bar{\mathbf{H}}_a$. Plugging the model in (8) into (6) gives

$$\mathbf{Y} = \mathbf{W}_b^H \bar{\mathbf{A}}_r (\bar{\mathbf{H}}_a + \mathbf{E}) \bar{\mathbf{A}}_t^H \mathbf{F}_b + \mathbf{W}_b^H \mathbf{N}. \tag{9}$$

Vectorizing $\mathbf{Y}$ in (9) yields

$$\operatorname{vec}(\mathbf{Y}) = (\mathbf{F}_b^T \bar{\mathbf{A}}_t^* \otimes \mathbf{W}_b^H \bar{\mathbf{A}}_r)(\operatorname{vec}(\bar{\mathbf{H}}_a + \mathbf{E})) + \operatorname{vec}(\mathbf{W}_b^H \mathbf{N}). \tag{10}$$

[2]Both Cramer-Rao lower bound (CRLB) [27] and the oracle estimator [9] can be utilized to evaluate the accuracy of estimation algorithms. Since the CRLB can only be calculated for one-stage method, in this work we use the oracle estimator as the benchmark instead.

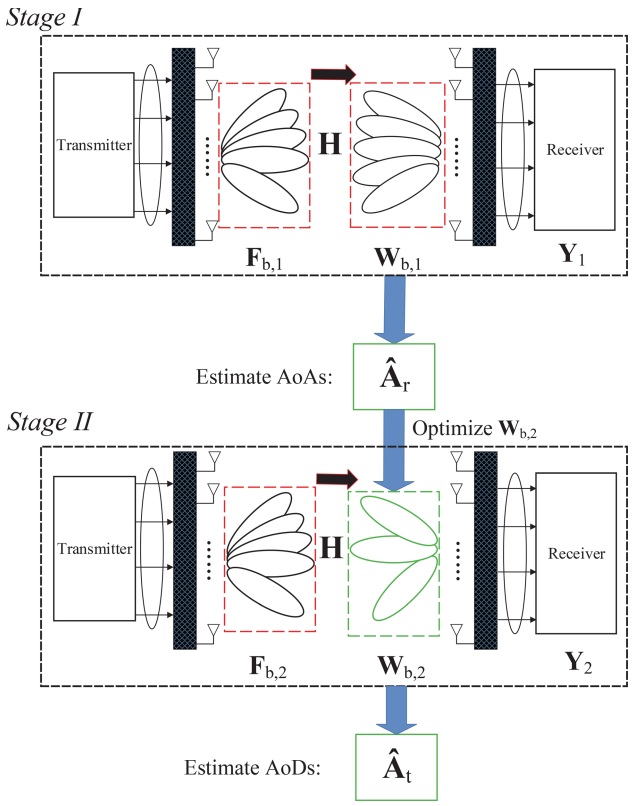

Fig. 2. Illustration of the proposed two-stage AoA and AoD estimation.

Denoting $\mathbf{D} = \mathbf{F}_b^T \bar{\mathbf{A}}_t^* \otimes \mathbf{W}_b^H \bar{\mathbf{A}}_r \in \mathbf{C}^{B_r B_t \times G_r G_b}$ and $\bar{\mathbf{n}} = \mathbf{D}\,\mathrm{vec}(\mathbf{E}) + \mathrm{vec}(\mathbf{W}_b^H \mathbf{N}) \in \mathbf{C}^{B_r B_t \times 1}$ gives $\mathrm{vec}(\mathbf{Y}) = \mathbf{D}\,\mathrm{vec}(\bar{\mathbf{H}}_a) + \bar{\mathbf{n}}$. Hence, the estimation of $\mathrm{vec}(\bar{\mathbf{H}}_a)$ from (10) can be stated as a sparse signal reconstruction problem:

$$\min_{\bar{\mathbf{H}}_a} \|\,\mathrm{vec}(\mathbf{Y}) - \mathbf{D}\,\mathrm{vec}(\bar{\mathbf{H}}_a)\|_2, \ \text{s.t.}\ \|\,\mathrm{vec}(\bar{\mathbf{H}}_a)\|_0 = L, \quad (11)$$

where $\|\cdot\|_0$ is the ℓ_0-norm that returns the number of non-zero coordinates of a vector. The problem in (11) can be solved by using standard CS methods [28], [29].

The number of required observations to reconstruct L-sparse vector $\mathrm{vec}(\bar{\mathbf{H}}_a) \in \mathbb{C}^{G_r G_t \times 1}$ in (11) has previously characterized as $O\left(L \cdot \log(G_r G_t)\right)$ [28], which is much smaller than $O(N_r N_t)$. However, the computational complexity for estimating $\mathrm{vec}(\bar{\mathbf{H}}_a)$ in (11) by using OMP, for example, is $O(L B_r B_t G_r G_t)$. Though the quantization error associated with using dictionaries can be made small by increasing the sizes of the dictionaries, the growing computational complexity remains a critical challenge. Instead of developing another one-stage channel sounding method (as in Fig. 1), we propose a new two-stage channel sounding and estimation framework to overcome the large overhead and complexity drawbacks.

III. Two-Stage AoA and AoD Estimation

A conceptual diagram of the proposed two-stage AoA and AoD estimation framework is presented in Fig. 2. The proposed sequential technique has constituent two stages of channel sounding, where each stage exclusively exploits much low-dimensional dictionary compared to the one-stage channel sounding in Fig. 1.

Under the similar definitions of one-stage method in (6), in Stage I of the two-stage framework of Fig. 2, the transmit and receive sounding beams are represented by $\mathbf{F}_{b,1} \in \mathbf{C}^{N_t \times B_{t,1}}$ and $\mathbf{W}_{b,1} \in \mathbf{C}^{N_r \times B_{r,1}}$, respectively. The AoA estimates of Stage I produce the estimation of array response matrix $\mathbf{A}_r$ in (2), i.e., $\widehat{\mathbf{A}}_r \in \mathbb{C}^{N_t \times L}$. In Stage II, the transmit and receive sounding beams are denoted by $\mathbf{F}_{b,2} \in \mathbf{C}^{N_t \times B_{t,2}}$ and $\mathbf{W}_{b,2} \in \mathbf{C}^{N_r \times B_{r,2}}$, respectively. In particular, the receive sounding beams $\mathbf{W}_{b,2}$ is optimized based on the estimated AoA array response matrix $\widehat{\mathbf{A}}_r$ at Stage I, which leads to improved estimation accuracy as our analysis and simulation show. The total number of observations is given by $N_p = B_{t,1}B_{r,1} + B_{t,2}B_{r,2}$. Accordingly, the total number of channel uses is $K = (B_{t,1}B_{r,1} + B_{t,2}B_{r,2})/N$.

A. Stage I: AoA Estimation

We rewrite the channel model in (8) as $\mathbf{H} = \bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H + \mathbf{E} = \bar{\mathbf{A}}_r \mathbf{Q}_r + \mathbf{E}$, where $\mathbf{Q}_r \in \mathbf{C}^{G_r \times N_t}$ has L non-zero rows, whose indices are collected into the support set $\Omega_r \subset \{1, 2, \ldots, G_r\}$ and $|\Omega_r| = L$. Using Ω_r, the $\mathbf{A}_r$ in (2) can be written using the columns of $\bar{\mathbf{A}}_r$ indexed by Ω_r as $[\bar{\mathbf{A}}_r]_{:,\Omega_r} = \mathbf{A}_r$.

To estimate the AoAs, we need to recover the support set Ω_r. Similar to the one-stage sounding in (6), at Stage I in Fig. 2, the observations $\mathbf{Y}_1 \in \mathbb{C}^{B_{r,1} \times B_{t,1}}$ is expressed as

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{W}_{b,1}^H \mathbf{H} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1 \\ &= \mathbf{W}_{b,1}^H \bar{\mathbf{A}}_r \mathbf{Q}_r \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{E} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1 \\ &= \mathbf{\Phi}_1 \mathbf{C}_1 + \mathbf{W}_{b,1}^H \mathbf{E} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1, \end{aligned} \quad (12)$$

where $\mathbf{\Phi}_1 = \mathbf{W}_{b,1}^H \bar{\mathbf{A}}_r \in \mathbb{C}^{B_{r,1} \times G_r}$, $\mathbf{C}_1 = \mathbf{Q}_r \mathbf{F}_{b,1} \in \mathbb{C}^{G_r \times B_{t,1}}$, and $\mathbf{N}_1 \in \mathbf{C}^{N_r \times B_{t,1}}$ is the noise matrix with independent and identically distributed (i.i.d.) entries according to $[\mathbf{N}_1]_{i,j} \sim \mathcal{CN}(0, \sigma^2)$, $\forall i, j$. Due to the row sparsity of $\mathbf{Q}_r$, it is clear that $\mathbf{C}_1$ also has L non-zero rows indexed by Ω_r. If $B_{t,1} = 1$, the recovery of $\mathbf{C}_1$ in (12) can be formulated as a common SMV CS problem. When $B_{t,1} > 1$, it becomes an MMV CS problem [30], where the multiple columns of $\mathbf{C}_1$ in (12) shares a common support. The optimization problem estimating the row support of $\mathbf{C}_1$ for MMV is now given by

$$\widehat{\mathbf{C}}_1 = \arg\min_{\mathbf{C}_1} \|\mathbf{Y}_1 - \mathbf{\Phi}_1 \mathbf{C}_1\|_F^2, \ \text{s.t.}\ \|\mathbf{C}_1\|_{r,0} \le L, \quad (13)$$

where $\|\mathbf{C}_1\|_{r,0}$ is defined as the number of non-zero rows of $\mathbf{C}_1$. Using a similar method as the OMP, the problem in (13) can be solved by simultaneous OMP (SOMP) [31] that is described in Algorithm 1. The output is the estimated support set $\widehat{\Omega}_r$.[3] For notational simplicity, we omit the subscripts in $\mathbf{Y}_1$ and $\mathbf{\Phi}_1$ in Algorithm 1.

It should be emphasized that the choice of the measurement matrix $\mathbf{\Phi}_1$ and $\mathbf{C}_1$ has a profound impact on the recovery performance of SOMP [31]. Observing (12), the TSB $\mathbf{F}_{b,1}$ is incorporated in $\mathbf{C}_1$, and the RSB $\mathbf{W}_{b,1}$ is included in the

[3]Here, we assume the number of paths is known as a priori for convenience of performance analysis in Section IV. When the number of paths is unavailable as a priori, a threshold can be introduced to compare with the power of the residual matrix $\mathbf{R}^{(l)}$ in Step 8 at each iteration [32], [33]. When the power of $\mathbf{R}^{(l)}$ is less then the threshold, Algorithm 1 terminates, which generates the estimate of number of paths.

Algorithm 1 Simultaneous OMP: SOMP($\mathbf{Y}, \mathbf{\Phi}, L$)

1: Input: Observations $\mathbf{Y}$, measurement matrix $\mathbf{\Phi}$, sparsity level L.
2: Initialization: Support set $\widehat{\Omega}^{(0)} = \emptyset$, residual matrix $\mathbf{R}^{(0)} = \mathbf{Y}$.
3: **for** $l = 1$ to L **do**
4: Calculate the coefficient matrix: $\mathbf{S} = \mathbf{\Phi}^H \mathbf{R}^{(l-1)}$.
5: Select the largest index $\eta = \underset{i=1,\cdots,G_r}{\arg\max} \|[\mathbf{S}]_{i,:}\|_2$.
6: Update the support set: $\widehat{\Omega}^{(l)} = \widehat{\Omega}^{(l-1)} \bigcup \eta$.
7: Update the recovery of matrix: $\widehat{\mathbf{C}} = ([\mathbf{\Phi}]_{:,\widehat{\Omega}^{(l)}})^{\dagger} \mathbf{Y}$.
8: Update the residual matrix: $\mathbf{R}^{(l)} = \mathbf{Y} - [\mathbf{\Phi}]_{:,\widehat{\Omega}^{(l)}} \widehat{\mathbf{C}}$.
9: **end for**
10: Output: $\widehat{\Omega}^{(L)}, \widehat{\mathbf{C}}$.

measurement matrix $\mathbf{\Phi}_1$. Thus, in what follows, the design of RSB $\mathbf{W}_{b,1}$ and TSB $\mathbf{F}_{b,1}$, is of interest.

1) RSB and TSB Design: Firstly, we focus on the design of TSB $\mathbf{F}_{b,1}$. Considering $\mathbf{C}_1 = \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H \mathbf{F}_{b,1}$, in order to guarantee that $\mathbf{F}_{b,1}$ is unbiased for each item (column) in $\bar{\mathbf{A}}_t$, we design $\mathbf{F}_{b,1}$ by maximizing the minimum correlation between $\mathbf{F}_{b,1}$ and each column in $\bar{\mathbf{A}}_t$, which yields

$$\max_{\mathbf{F}_{b,1}} \min_i \|\mathbf{F}_{b,1}^H \left[\bar{\mathbf{A}}_t\right]_{:,i}\|_2, \text{ s.t. } \mathbf{F}_{b,1}^H \mathbf{F}_{b,1} = p_1 \mathbf{I}_{B_{t,1}}, \quad (14)$$

where p_1 is the power allocation of Stage I. After taking the constraint into account, the optimal solution to the problem in (14) should ideally satisfy the following $\|\mathbf{F}_{b,1}^H \left[\bar{\mathbf{A}}_t\right]_{:,i}\|_2 = \sqrt{p_1 B_{t,1}/N_t}, \quad i = 1, \ldots, G_t$. It means that $\mathbf{F}_{b,1}$ is isometric to all columns of $\bar{\mathbf{A}}_t$, which is obtained by

$$\mathbf{F}_{b,1} = \sqrt{p_1} \left[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{B_{t,1}}\right], \quad (15)$$

where $\mathbf{e}_i$ the ith column of $\mathbf{I}_{N_t}$. The construction of $\mathbf{e}_j, \ j = 1, \ldots, B_{t,1}$ in (15) using the hybrid analog-digital array is possible due to the fact that any vector can be constructed by linearly combining $N(\geq 2)$ RF chains [34]. To be more specific, there exists $\mathbf{F}_{A,j} \in \mathbb{C}^{N_t \times N}$, $\mathbf{f}_{D,j} \in \mathbb{C}^{N \times 1}$, and $s_j = 1$ such that $\mathbf{e}_j = \mathbf{F}_{A,j} \mathbf{f}_{D,j} s_j$, i.e.,

$$\mathbf{e}_j = \underbrace{\frac{1}{\sqrt{N_t}} [\mathbf{1}_{N_t} \ \tilde{\mathbf{1}}_{N_t}^{(j)} \ \mathbf{1}_{N_t} \cdots \mathbf{1}_{N_t}]}_{\triangleq \mathbf{F}_{A,j}} \underbrace{\frac{\sqrt{N_t}}{2} [1, -1, 0, \cdots, 0]^T}_{\triangleq \mathbf{f}_{D,j}} \times 1, \quad (16)$$

where $\tilde{\mathbf{1}}_{N_t}^{(j)} \in \mathbb{R}^{N_t \times 1}$ is defined as the all one vector $\mathbf{1}_{N_t} \in \mathbb{R}^{N_t \times 1}$ other than the jth entry being -1.

For the measurement matrix $\mathbf{\Phi}_1 = \mathbf{W}_{b,1} \bar{\mathbf{A}}_r$, we optimize $\mathbf{W}_{b,1}$ by incorporating the isometric CS measurement matrix design criterion [35]–[37]:

$$\min_{\mathbf{\Phi}_1} \left\| \mathbf{\Phi}_1^H \mathbf{\Phi}_1 - \mathbf{I}_{G_r} \right\|_F^2. \quad (17)$$

After performing standard algebraic manipulations and exploiting the fact $\bar{\mathbf{A}}_r \bar{\mathbf{A}}_r^H = \frac{G_r}{N_r} \mathbf{I}_{N_r}$, the optimality condition for (17) is that the columns of $\mathbf{W}_{b,1}$ are orthogonal. Accounting for the analog-digital array constraint into $\mathbf{W}_{b,1}$ and setting $B_{r,1} = N_r$, we use the DFT matrix $\mathbf{S}_{N_r} \in \mathbb{C}^{N_r \times N_r}$ such that

$$\mathbf{W}_{b,1} = \mathbf{S}_{N_r}, \quad (18)$$

where $[\mathbf{S}_{N_r}]_{m,n} = \frac{1}{\sqrt{N_r}} e^{-j \frac{2\pi(m-1)(n-1)}{N_r}}, \forall m, n$.

Based on the RSB in (18), in the following, the distribution of the noise term in (12) is discussed.

Proposition 1: For any semi-orthogonal matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ with $\mathbf{A}\mathbf{A}^H = \mathbf{I}$ and random vector $\mathbf{n} \in \mathbb{C}^{n \times 1}$ with i.i.d. entries according to $\mathcal{CN}(0, \sigma^2)$, then if we denote $\mathbf{b} = \mathbf{A}\mathbf{n}$, and the entries in $\mathbf{b}$ are also i.i.d. $\mathcal{CN}(0, \sigma^2)$.

Proof: The covariance matrix of $\mathbf{b}$ is given by $\mathbb{E}[\mathbf{A}\mathbf{n}\mathbf{n}^H\mathbf{A}^H] = \sigma^2\mathbf{I}$. Because the entries in $\mathbf{b}$ are obviously complex Gaussian, thus, from the property of Gaussian distribution, the entries in $\mathbf{b}$ are also i.i.d. $\mathcal{CN}(0, \sigma^2)$. □

Remark 1: Due to the semi-orthogonality of $\mathbf{W}_{b,1}$ in (18), according to Proposition 1, the effective noise matrix $\mathbf{W}_{b,1}^H \mathbf{N}_1 \in \mathbb{C}^{N_r \times B_{t,1}}$ in (12) has i.i.d. Gaussian entries, i.e., $[\mathbf{W}_{b,1}^H \mathbf{N}_1]_{i,j} \sim \mathcal{CN}(0, \sigma^2), \forall i, j$. Moreover, since $\mathbf{\Phi}_1 = \mathbf{W}_{b,1}^H \bar{\mathbf{A}}_r$, we have $\|[\mathbf{\Phi}_1]_{:,i}\|_2 = 1, \forall i$.

The algorithmic procedure estimating AoAs are described in Algorithm 2. Given the estimated support set $\widehat{\Omega}_r$ from Algorithm 1, the output of Algorithm 2 is the estimated AoA array response matrix $\widehat{\mathbf{A}}_r = [\bar{\mathbf{A}}_r]_{:,\widehat{\Omega}_r} \in \mathbb{C}^{N_r \times L}$. Overall, the number of channel uses for the AoA estimation is $K_1 = B_{t,1} \frac{N_r}{N}$.

Algorithm 2 AoA Estimation Algorithm

1: Input: Channel dimension N_r, N_t, number of RF chains N, channel paths L, power allocation p_1, receive array response dictionary $\bar{\mathbf{A}}_r \in \mathbb{C}^{N_r \times G_r}$.
2: Initialization: Generate the TSB $\mathbf{F}_{b,1} = \sqrt{p_1}[\mathbf{e}_1, \ldots, \mathbf{e}_{B_{t,1}}]$ in (15) according to (16) and the RSB $\mathbf{W}_{b,1} = \mathbf{S}_{N_r}$ in (18).
3: Collect the observations $\mathbf{Y}_1 = \mathbf{W}_{b,1}^H \mathbf{H} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1$.
4: Solve the problem in (13) by using Algorithm 1 with the sparsity level L and $\mathbf{\Phi}_1 = \mathbf{W}_{b,1}^H \bar{\mathbf{A}}_r$,

$$(\widehat{\Omega}_r, \widehat{\mathbf{C}}_1) = \text{SOMP}(\mathbf{Y}_1, \mathbf{\Phi}_1, L).$$

5: Output: Estimation of AoA array response matrix $\widehat{\mathbf{A}}_r = [\bar{\mathbf{A}}_r]_{:,\widehat{\Omega}_r}$.

B. Stage II: AoD Estimation

To attain the estimation of AoDs, we can utilize the similar method as Stage I. Similar to the one-stage sounding in (6), the observations of Stage II in Fig. 2 is expressed as $\mathbf{Y}_2 \in \mathbb{C}^{B_{r,2} \times B_{t,2}}$,

$$\mathbf{Y}_2 = \mathbf{W}_{b,2}^H \mathbf{H} \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{N}_2, \quad (19)$$

where $\mathbf{W}_{b,2} \in \mathbf{C}^{N_r \times B_{r,2}}$ and $\mathbf{F}_{b,2} \in \mathbf{C}^{N_t \times B_{t,2}}$ are the RSB and TSB of the Stage II, respectively. The $\mathbf{N}_2 \in \mathbf{C}^{N_r \times B_{t,2}}$ is the noise matrix with i.i.d. entries according to $\mathcal{CN}(0, \sigma^2)$.

Recall from (2) and (8), the channel matrix is rewritten as

$$\mathbf{H} = \bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H + \mathbf{E}. \quad (20)$$

One can find that $\bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \in \mathbf{C}^{N_r \times G_t}$ has L non-zero columns, indexed by Ω_t with $|\Omega_t| = L$. Then, plugging (20) into (19) and taking conjugate transpose give

$$\begin{aligned} \mathbf{Y}_2^H &= \underbrace{\mathbf{F}_{b,2}^H \bar{\mathbf{A}}_t}_{\triangleq \mathbf{\Phi}_2} \underbrace{\bar{\mathbf{H}}_a^H \bar{\mathbf{A}}_r^H \mathbf{W}_{b,2}}_{\triangleq \mathbf{C}_2} + \mathbf{F}_{b,2}^H \mathbf{E}^H \mathbf{W}_{b,2} + \mathbf{N}_2^H \mathbf{W}_{b,2} \\ &= \mathbf{\Phi}_2 \mathbf{C}_2 + \mathbf{F}_{b,2}^H \mathbf{E}^H \mathbf{W}_{b,2} + \mathbf{N}_2^H \mathbf{W}_{b,2}, \end{aligned} \quad (21)$$

where $\mathbf{\Phi}_2 = \mathbf{F}_{b,2}^H \bar{\mathbf{A}}_t \in \mathbb{C}^{B_{t,2} \times G_t}$, and $\mathbf{C}_2 = \bar{\mathbf{H}}_a^H \bar{\mathbf{A}}_r^H \mathbf{W}_{b,2} \in \mathbb{C}^{G_t \times B_{r,2}}$. It is straightforward that the $\mathbf{C}_2$ has only L non-zero rows indexed by Ω_t. Similar to (13) in Stage I, the support set Ω_t estimation problem can be formulated as

$$\widehat{\mathbf{C}}_2 = \arg\min_{\mathbf{C}_2} \left\| \mathbf{Y}_2^H - \mathbf{\Phi}_2 \mathbf{C}_2 \right\|_F^2, \text{ s.t. } \|\mathbf{C}_2\|_{r,0} \leq L, \quad (22)$$

which is solved by Algorithm 1. In what follows, the design of RSB $\mathbf{W}_{b,2}$ and TSB $\mathbf{F}_{b,2}$ for Stage II is of interest.

1) RSB and TSB Design: For the design of RSB $\mathbf{W}_{b,2}$, we leverage the estimated AoAs from Stage I to formulate

$$\max_{\mathbf{W}_{b,2}} \min_i \|\mathbf{W}_{b,2}^H [\widehat{\mathbf{A}}_r]_{:,i}\|_2. \quad (23)$$

If $\mathbf{W}_{b,2}$ is semi-unitary, i.e., $\mathbf{W}_{b,2}^H \mathbf{W}_{b,2} = \mathbf{I}_{B_{r,2}}$, the objective value in (23) satisfies $\|\mathbf{W}_{b,2}^H [\widehat{\mathbf{A}}_r]_{:,i}\|_2 \leq 1, \forall i$ with the equality holding if

$$\mathcal{R}(\mathbf{W}_{b,2}) = \mathcal{R}(\widehat{\mathbf{A}}_r). \quad (24)$$

One can check (24) holds only if $B_{r,2} \geq L$. Without loss of optimality and to save the number of sounders, we set $B_{r,2} = L$. One solution to (24) is attained when the columns of $\mathbf{W}_{b,2}$ are the orthonormal basis of $\widehat{\mathbf{A}}_r$. For example, we let $\mathbf{W}_{b,2}$ be the $\mathbf{Q}$-matrix of the QR decomposition[4] of $\widehat{\mathbf{A}}_r$ such that

$$\mathbf{W}_{b,2} = \mathrm{QR}(\widehat{\mathbf{A}}_r), \quad (25)$$

where $\mathrm{QR}(\cdot)$ returns the $\mathbf{Q}$-matrix of a given matrix.

Remark 2: Due to the semi-orthogonality of $\mathbf{W}_{b,2}$ and the conclusions in Proposition 1, the effective noise matrix $\mathbf{W}_{b,2}^H \mathbf{N}_2 \in \mathbb{C}^{B_{r,2} \times B_{t,2}}$ in (19) has i.i.d. Gaussian entries, i.e., $[\mathbf{W}_{b,2}^H \mathbf{N}_2]_{i,j} \sim \mathcal{CN}(0, \sigma^2), \forall i, j$.

As for the design of $\mathbf{F}_{b,2}$, we exploit the isometric CS measurement matrix design criterion,

$$\min_{\mathbf{\Phi}_2} \|\mathbf{\Phi}_2^H \mathbf{\Phi}_2 - \mathbf{I}_{G_t}\|_F^2. \quad (26)$$

After similar manipulations as (17), the optimality condition for $\mathbf{F}_{b,2}$ of (26) is that the columns of $\mathbf{F}_{b,2}$ are orthogonal.

Then, following the same procedure as (15) and (16), we obtain the design of TSP $\mathbf{F}_{b,2}$ below,

$$\mathbf{F}_{b,2} = \sqrt{p_2}[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{B_{t,2}}], \quad (27)$$

where p_2 is the power coefficient of Stage II.

The algorithmic procedure of estimating AoDs are described in Algorithm 3. Provided the estimated support set $\widehat{\Omega}_t$, the output of Algorithm 3 is the estimated AoD array response matrix $\widehat{\mathbf{A}}_t = [\bar{\mathbf{A}}_t]_{:,\widehat{\Omega}_t} \in \mathbb{C}^{N_t \times L}$. The number of channel uses for the AoD estimation in Stage II is $K_2 = B_{t,2}$, and the overall number of channel uses for two stages is

$$K = K_1 + K_2 = B_{t,1} \frac{N_r}{N} + B_{t,2}. \quad (28)$$

Remark 3: Recall that the number of observations for the conventional one-stage channel sounding in Fig. 1 is $\mathcal{O}(L \cdot \log(G_r G_t / L))$ [28]. As a comparison, since the proposed two-stage channel sounding in Fig. 2 only estimates AoA in Stage I, and estimates AoD in Stage II, the number of required observations is $\mathcal{O}(L \cdot \log(G_r/L))$ in Stage I, and $\mathcal{O}(L \cdot \log(G_t/L))$ in Stage II. The total number of required observations for the proposed two-stage channel sounding is $\mathcal{O}(L \cdot \log(G_r/L)) + \mathcal{O}(L \cdot \log(G_t/L)) = \mathcal{O}(L \cdot \log(G_t G_r / L^2))$, which is less than the conventional one-stage sounding.

[4]The QR decomposition is a decomposition of a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ into the product $\mathbf{A} = \mathbf{QR}$ of an orthonormal matrix $\mathbf{Q} \in \mathbb{C}^{m \times n}$ and an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{n \times n}$.

Algorithm 3 AoD Estimation Algorithm

1: Input: Channel dimension N_r, N_t, number of RF chains N, channel paths L, power allocation p_2, output of AoA estimation $\widehat{\mathbf{A}}_r$, transmit array response dictionary $\bar{\mathbf{A}}_t \in \mathbb{C}^{N_t \times G_t}$.
2: Initialization: Generate the TSB $\mathbf{F}_{b,2} = \sqrt{p_2}[\mathbf{e}_1, \ldots, \mathbf{e}_{B_{t,2}}]$ in (27) and RSB $\mathbf{W}_{b,2} = \mathrm{QR}(\widehat{\mathbf{A}}_r)$ in (25).
3: Collect the observations $\mathbf{Y}_2 = \mathbf{W}_{b,2}^H \mathbf{H} \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{N}_2$.
4: Solve the problem in (22) by using Algorithm 1 with the sparsity level L and $\mathbf{\Phi}_2 = \mathbf{F}_{b,2}^H \bar{\mathbf{A}}_t$,

$$(\widehat{\Omega}_t, \widehat{\mathbf{C}}_2) = \mathrm{SOMP}(\mathbf{Y}_2^H, \mathbf{\Phi}_2, L).$$

5: Output: Estimation of AoD array response matrix $\widehat{\mathbf{A}}_t = [\bar{\mathbf{A}}_t]_{:,\widehat{\Omega}_t}$.

Remark 4: About happening of the design RSB and TSB, in Stage I, one can find that the design of RSB in (18) and TSB in (15) are completed before the channel estimation, which are then utilized by the transmitter and receiver. Like the fact that the training pilots are known for the transmitter and receiver in advance before the task of channel estimation, here we also assume that the TSB and RSB are known as a priori. In Stage II, the TSB $\mathbf{F}_{b,2}$ in (27) is also designed in advance, while the RSB $\mathbf{W}_{b,2}$ in (25) is designed and employed at the receiver side, which requires no feedback to the transmitter. Overall, the proposed method requires no feedback during the whole procedures of the channel estimation.

C. Channel Estimation

Recalling the channel representation in (2) and after estimating $\widehat{\mathbf{A}}_r \in \mathbb{C}^{N_r \times L}$ in Algorithm 2 and $\widehat{\mathbf{A}}_t \in \mathbb{C}^{N_t \times L}$ in Algorithm 3, we can express the channel estimate as

$$\widehat{\mathbf{H}} = \widehat{\mathbf{A}}_r \widehat{\mathbf{R}} \widehat{\mathbf{A}}_t^H, \quad (29)$$

where $\widehat{\mathbf{R}} \in \mathbb{C}^{L \times L}$ denotes the estimation of $\mathrm{diag}(\mathbf{h})$ in (2). In the following, we will discuss how to obtain the estimate $\widehat{\mathbf{R}}$. It is worth noting that unlike (2) we do not restrict $\widehat{\mathbf{R}}$ to be a diagonal matrix because of the possible permutations in the columns of $\widehat{\mathbf{A}}_r$ and $\widehat{\mathbf{A}}_t$.

Recall the observations of each stage, i.e., $\mathbf{Y}_1 = \mathbf{W}_{b,1}^H \bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{E} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1$, and $\mathbf{Y}_2 = \mathbf{W}_{b,2}^H \bar{\mathbf{A}}_r \bar{\mathbf{H}}_a \bar{\mathbf{A}}_t^H \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{E} \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{N}_1$. Since $\mathbf{W}_{b,1}^H \mathbf{N}_1$ and $\mathbf{W}_{b,2}^H \mathbf{N}_2$ are i.i.d. Gaussian, incorporating the expressions of channel estimate in (29), the estimation of $\widehat{\mathbf{R}}$ is given by

$$\widehat{\mathbf{R}} = \arg\min_{\mathbf{R}} \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Y}_1) \\ \mathrm{vec}(\mathbf{Y}_2) \end{bmatrix} - \begin{bmatrix} \mathrm{vec}(\mathbf{W}_{b,1}^H \widehat{\mathbf{A}}_r \mathbf{R} \widehat{\mathbf{A}}_t^H \mathbf{F}_{b,1}) \\ \mathrm{vec}(\mathbf{W}_{b,2}^H \widehat{\mathbf{A}}_r \mathbf{R} \widehat{\mathbf{A}}_t^H \mathbf{F}_{b,2}) \end{bmatrix} \right\|_F^2,$$

where the optimal solution is given by $\mathrm{vec}(\widehat{\mathbf{R}}) = \left(\mathbf{A}_1^H \mathbf{A}_1 + \mathbf{A}_2^H \mathbf{A}_2\right)^{-1} \left(\mathbf{A}_1^H \mathrm{vec}(\mathbf{Y}_1) + \mathbf{A}_2^H \mathrm{vec}(\mathbf{Y}_2)\right)$, where $\mathbf{A}_1 = (\widehat{\mathbf{A}}_t^H \mathbf{F}_{b,1})^T \otimes \mathbf{W}_{b,1}^H \widehat{\mathbf{A}}_r \in \mathbb{C}^{N_r B_{t,1} \times L^2}$ and $\mathbf{A}_2 = (\widehat{\mathbf{A}}_t^H \mathbf{F}_{b,2})^T \otimes \mathbf{W}_{b,2}^H \widehat{\mathbf{A}}_r \in \mathbb{C}^{L B_{t,2} \times L^2}$. Because $N_r B_{t,1} \gg L^2$ and $B_{t,2} \gg L$, the matrix $\mathbf{A}_1^H \mathbf{A}_1 + \mathbf{A}_2^H \mathbf{A}_2 \in \mathbb{C}^{L^2 \times L^2}$ is always invertible.

Remark 5: After $\widehat{\mathbf{R}}$ is estimated, the pairing of AoAs and AoDs can be obtained by selecting positions of the largest L entries in $\widehat{\mathbf{R}}$. Then, the path gain $h_l, l = 1, 2, \cdots, L$, can be calculated by solving a problem like the oracle estimator in (7), where the two-stage RSBs and TSBs are utilized.

IV. Performance Analysis and Resource Allocation

In this section, we discuss the reconstruction probability of AoAs and AoDs of the proposed two-stage method in Section III. Moreover, we further enhance the reconstruction performance by performing power and channel use allocation to each stage.

A. Successful Recovery Probability Analysis

1) SRP of AoA Estimation: As a starting point, we focus on the SRP of Algorithm 1. An SRP bound of SOMP was previously studied in [38], where the analysis was based on the restricted isometry property constant of the measurement matrix $\mathbf{\Phi}$. In this work, we instead analyze the recovery performance of Algorithm 1, based on the mutual incoherence property (MIP) constant[5] [39] of $\mathbf{\Phi}$.

Lemma 1: Suppose $\mathbf{C} \in \mathbb{C}^{N\times d}$ is a row sparse matrix, where L ($\ll N$) rows of $\mathbf{C}$, indexed by Ω, are non-zero. We consider the observation $\mathbf{Y} = \mathbf{\Phi C} + \mathbf{N}$, where $\mathbf{Y} \in \mathbb{C}^{M\times d}$, $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ is the measurement matrix with $L \le M \ll N$ and $\|[\mathbf{\Phi}]_{:,i}\|_2 = 1, \forall i$, and $\mathbf{N} \in \mathbb{C}^{M\times d}$ is the noise matrix with each entry i.i.d. according to complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$. Given that the MIP constant μ of the measurement matrix $\mathbf{\Phi}$ is $\mu < 1/(2L-1)$, the SRP of Algorithm 1 satisfies

$$\Pr(\mathcal{V}_S) \ge F_2\left(\frac{(1-(2L-1)\mu)^2 C_{\min}^2 - 4\sigma^2\mu_{M,d}}{4\sigma^2\sigma_{M,d}}\right), \quad (30)$$

where $\mathcal{V}_S$ is the event of successful reconstruction of Algorithm 1, $C_{\min} = \min_{i\in\Omega} \|[\mathbf{C}]_{i,:}\|_2$, $\mu_{M,d} = (M^{1/2} + d^{1/2})^2$, $\sigma_{M,d} = (M^{1/2} + d^{1/2})(M^{-1/2} + d^{-1/2})^{1/3}$, and the function $F_2(\cdot)$[6] is the cumulative distribution function (CDF) of Tracy-Widom law [40], [41].

[5]The MIP constant of matrix $\mathbf{\Phi}$ is quantified by a variable $\mu = \max_{i\neq j} |\langle [\mathbf{\Phi}]_{:,i}, [\mathbf{\Phi}]_{:,j}\rangle|$, where $\langle\cdot,\cdot\rangle$ denotes the inner product.

[6]The CDF of Tracy-Widom law [40], [41] $F_2(\cdot)$ is expressed as

$$F_2(s) = \exp\left(\int_s^\infty (x-s)q(x)dx\right),$$

where $q(x)$ is the solution of Painlevé equation of type II:

$$q''(x) = xq(x) + 2q(x)^3,\ q(x) \sim \mathrm{Ai}(x), x \to \infty,$$

where $\mathrm{Ai}(x)$ is the Airy function [40], [41]. To save computational complexity, we admit the table lookup method [42] to obtain the value of $F_2(\cdot)$.

Proof: See Appendix A. □

Proposition 2: Suppose the signal model provided in Lemma 1 and, given the quantization error, the observation model $\mathbf{Y} = \mathbf{\Phi C} + \tilde{\mathbf{N}}$, where effective noise $\tilde{\mathbf{N}} = \mathbf{E} + \mathbf{N}$ with quantization error $\mathbf{E}$ and Gaussian noise $\mathbf{N}$ of i.i.d. $\mathcal{CN}(0, \sigma^2)$ entries. If μ is the MIP constant of the measurement matrix $\mathbf{\Phi}$ with $\mu < 1/(2L-1)$, the SRP of Algorithm 1 is given by

$$\Pr(\mathcal{V}_S) \ge F_2\left(\frac{((1-(2L-1)\mu)C_{\min} - 2\|\mathbf{E}\|_2)^2 - 4\sigma^2\mu_{M,d}}{4\sigma^2\sigma_{M,d}}\right), \quad (31)$$

where $C_{\min} = \min_{i\in\Omega} \|[\mathbf{C}]_{i,:}\|_2$, $\mu_{M,d} = (M^{1/2} + d^{1/2})^2$, and $\sigma_{M,d} = (M^{1/2} + d^{1/2})(M^{-1/2} + d^{-1/2})^{1/3}$.

Proof: See Appendix B. □

As a direct consequence of Proposition 2, Theorem 1 below quantifies the SRP of AoA estimation in Algorithm 2.

Theorem 1: Assume the MIP constant of the measurement matrix $\mathbf{\Phi}_1$ in Algorithm 2 satisfies $\mu_1 < 1/(2L-1)$. Then, the SRP of Algorithm 2 is lower bounded by, (32) and (33), as shown at the bottom of the page, where $\widetilde{\mu}_1 = 1 - (2L-1)\mu_1$, $\mathcal{A}_S$ is the event of successful reconstruction of AoA, $h_{\min} = \min_{l\le L} |h_l|$ with h_l being the lth entry of $\mathbf{h}$ in (2), $\mu_{N_r,B_{t,1}} = (N_r^{1/2} + B_{t,1}^{1/2})^2$, $\sigma_{N_r,B_{t,1}} = (N_r^{1/2} + B_{t,1}^{1/2})(N_r^{-1/2} + B_{t,1}^{-1/2})^{1/3}$, and $\mathbf{E}_1 = \mathbf{W}_{b,1}^H \mathbf{E} \mathbf{F}_{b,1}$. In (33), the SRP bound is denoted as a function of $(p_1, B_{t,1})$.

Proof: Recalling the observation model in (12) with the TSB and RSB in (15) and (18), respectively, the effective TSB matrix $\mathbf{C}_1$ in (12) satisfies $\|[\mathbf{C}_1]_{r_l,:}\|_2 = \sqrt{\frac{p_1 B_{t,1}}{N_t}}\,|h_l|$, where $r_l \in \Omega_r$ is the index of the lth path of $\mathbf{A}_r$ in $\bar{\mathbf{A}}_r$ such that $[\bar{\mathbf{A}}_r]_{:,r_l} = [\mathbf{A}_r]_{:,l}$, $l = 1, \ldots, L$. According to Proposition 2, substituting $C_{\min} = \min_{r_l\in\Omega_r} \|[\mathbf{C}_1]_{r_l,:}\|_2 = \sqrt{\frac{p_1 B_{t,1}}{N_t}}\,|h_{\min}|$ in (31) results in the inequality for $\Pr(\mathcal{A}_S)$. The approximation in (32) is obtained by neglecting the quantization term $\mathbf{E}_1$, which completes the proof. □

Remark 6: According to Theorem 1, when the power p_1 of Stage I is fixed and the number of transmit sounding beams $B_{t,1}$($\ll N_r$) increases, the SRP of AoA increases accordingly. Interestingly, it is more efficient to increase the power allocation p_1 than the number of transmit sounding beams $B_{t,1}$ to achieve a higher SRP of AoA. This can be understood through the two cases where p_1 or $B_{t,1}$ grow at the same rate. Compared to the case of p_1, both $\mu_{N_r,B_{t,1}}$ and $\sigma_{N_r,B_{t,1}}$ increase slowly as $B_{t,1}$ grows, resulting in lower SRP in (32). This aspect will be clearer in the next subsection when we optimize the allocation of p_1 and $B_{t,1}$.

$$\Pr(\mathcal{A}_S) \ge F_2\left(\frac{\left(\widetilde{\mu}_1 h_{\min}\sqrt{\frac{p_1 B_{t,1}}{N_t}} - 2\|\mathbf{E}_1\|_2\right)^2 - 4\sigma^2\mu_{N_r,B_{t,1}}}{4\sigma^2\sigma_{N_r,B_{t,1}}}\right)$$

$$\approx F_2\left(\frac{\widetilde{\mu}_1^2 h_{\min}^2 \frac{p_1 B_{t,1}}{N_t} - 4\sigma^2\mu_{N_r,B_{t,1}}}{4\sigma^2\sigma_{N_r,B_{t,1}}}\right) \quad (32)$$

$$\triangleq P_1(p_1, B_{t,1}) \quad (33)$$

2) SRP of AoD Estimation: Regarding the SRP of Algorithm 3, we assume for tractability that the AoA estimation in Stage I was perfect. The following theorem quantifies the SRP of AoD estimation in Algorithm 3.

Theorem 2: Provided the perfect AoA knowledge known a priori and MIP constant μ_2 of matrix $\sqrt{N_t/(p_2 B_{t,2})}\mathbf{\Phi}_2$ satisfying $\mu_2 < 1/(2L-1)$, the SRP of Algorithm 3 is lower bounded by, (34) and (35), as shown at the bottm of the page, where $\widetilde{\mu}_2 = 1-(2L-1)\mu_2$, $\mathcal{D}_S$ denotes the event of successful AoD reconstruction, $h_{\min} = \min_{l\le L}|h_l|$ with h_l being the lth entry of $\mathbf{h}$ in (2), $\mu_{B_{t,2},L} = (L^{1/2}+B_{t,2}^{1/2})^2$, $\sigma_{B_{t,2},L} = (L^{1/2}+B_{t,2}^{1/2})(L^{-1/2}+B_{t,2}^{-1/2})^{1/3}$, and $\mathbf{E}_2 = \frac{N_t}{p_2 B_{t,2}}\mathbf{F}_{b,2}^H \mathbf{E}\mathbf{W}_{b,2}$. In (35), the SRP lower bound is substituted as a function of $(p_2, B_{t,2})$.

Proof: See Appendix C. □

B. Power and Channel Use Allocation

We recall that in the proposed two-stage method, the transmit sounding beams at Stage I and II are, respectively, $\mathbf{F}_{b,1} = \sqrt{p_1}[\mathbf{e}_1,\ldots,\mathbf{e}_{B_{t,1}}]$ in (15) and $\mathbf{F}_{b,2} = \sqrt{p_2}[\mathbf{e}_1,\ldots,\mathbf{e}_{B_{t,2}}]$ in (27). The total power budget E is therefore defined by

$$E = \underbrace{p_1 B_{t,1} N_r/N}_{\triangleq E_1} + \underbrace{p_2 B_{t,2}}_{\triangleq E_2}, \tag{36}$$

where E_1 and E_2 are the power budgets at the Stage I and Stage II, respectively.

We let $\eta_1 > 0$ and $\eta_2 > 0$ be the target SRP values at Stage I and Stage II, respectively. The SRP-guaranteed power budget minimization problem[7] is then formulated as

$$\min_{p_1,p_2,B_{t,1},B_{t,2}} E_1 + E_2 \tag{37a}$$

$$\text{s.t. } P_{\text{I}}(p_1, B_{t,1}) \ge \eta_1, \quad P_{\text{II}}(p_2, B_{t,2}) \ge \eta_2, \tag{37b}$$

$$E_1 = p_1 B_{t,1} N_r/N, \quad E_2 = p_2 B_{t,2}, \tag{37c}$$

$$B_{t,1} \ge \widetilde{B}_{t,1}, \quad B_{t,2} \ge \widetilde{B}_{t,2}, \tag{37d}$$

where $\widetilde{B}_{t,1}$ and $\widetilde{B}_{t,2}$ are the minimum numbers of allowed transmit beams at Stage I and Stage II, respectively. The problem in (37) optimizes the power allocation p_1 and p_2, and the number of transmit beams $B_{t,1}$ and $B_{t,2}$ to minimizes the total power budget subject to the SRP requirements at Stage I and Stage II. It is worth noting that that because the problem in (37) is separable, thus (37) is equivalent to the following two sub-problems,

$$\min_{p_1,B_{t,1}} E_1 \tag{38a}$$

$$\text{s.t. } P_{\text{I}}(p_1, B_{t,1}) \ge \eta_1, \quad E_1 = p_1 B_{t,1}\frac{N_r}{N}, B_{t,1} \ge \widetilde{B}_{t,1}, \tag{38b}$$

and

$$\min_{p_2,B_{t,2}} E_2 \tag{39a}$$

$$\text{s.t. } P_{\text{II}}(p_2, B_{t,2}) \ge \eta_2, \quad E_2 = p_2 B_{t,2}, B_{t,2} \ge \widetilde{B}_{t,2}. \tag{39b}$$

First of all, we focus on the sub-problem of Stage I in (38). It is worth noting that directly solving (38) is difficult due to the coupled constraints. Thus, we first maximize the SRP, i.e., $P_{\text{I}}(p_1, B_{t,1})$, with arbitrary power budget E_1,

$$\max_{p_1,B_{t,1}} P_{\text{I}}(p_1, B_{t,1}) \tag{40a}$$

$$\text{s.t. } p_1 B_{t,1} N_r/N = E_1, \quad B_{t,1} \ge \widetilde{B}_{t,1}. \tag{40b}$$

Prior to showing how to solve the problem in (40), we first elaborate the relation between the problem in (38) and (40). It is easy to observe that as E_1 increases the achievable SRP of the objective function in (40) also increases. Thus, the minimum E_1 in (38) is achieved when the SRP constraint in (38b), i.e., $P_{\text{I}}(p_1, B_{t,1}) \ge \eta_1$, holds as the equality. Moreover, given any arbitrary power budget E_1 in problem (40), the interrelation between the power allocation p_1 and the number of transmit sounding beams $B_{t,1}$ points to a fundamental tradeoff between them, which is demonstrated in the following theorem.

Theorem 3: Consider the following non-linear programming

$$(\hat{p}_1, \widehat{B}_{t,1}) = \arg\max_{p_1,B_{t,1}} P_{\text{I}}(p_1, B_{t,1}) \tag{41a}$$

$$\text{s.t. } p_1 B_{t,1} N_r/N = E_1, \quad B_{t,1} \ge \widetilde{B}_{t,1}, \tag{41b}$$

where E_1 is an arbitrary power budget. The solution to (41) is given by $\widehat{B}_{t,1} = \widetilde{B}_{t,1}$ and $p_1 = \frac{E_1 N}{\widetilde{B}_{t,1} N_r}$.

Proof: Substituting constraint $p_1 = \frac{E_1 N}{B_{t,1} N_r}$ in (41b) into the objective function in (41a), we first show that $P_{\text{I}}(\frac{E_1 N}{B_{t,1} N_r}, B_{t,1})$ in (41a) is a monotonically decreasing function of the number of transmit sounding beams $B_{t,1}$ for a fixed E_1. Specifically, substituting $\mu_{N_r,B_{t,1}} = (N_r^{1/2}+B_{t,1}^{1/2})^2$ and $\sigma_{N_r,B_{t,1}} = (N_r^{1/2}+B_{t,1}^{1/2})(N_r^{-1/2}+B_{t,1}^{-1/2})^{1/3}$ of (32) into

[7] In (37), we present the SRP-constrained power minimization problem for optimizing power and channel use allocations. For instance, this criterion can be thought of as a prudent alternative of the performance maximization subject to power constraints in the MIMO literature because it provides a guarantee on the achievable performance [43]. Multiple variants of the performance-guaranteed power minimization problem can be found in the context of MIMO resource allocation [44], [45].

$$\Pr(\mathcal{D}_S) \ge F_2\left(\frac{(\widetilde{\mu}_2 h_{\min} - 2\|\mathbf{E}_2\|_2)^2 - 4\sigma^2\frac{N_t}{p_2 B_{t,2}}N_t\mu_{B_{t,2},L}}{4N_t\sigma^2\frac{N_t}{p_2 B_{t,2}}\sigma_{B_{t,2},L}}\right)$$

$$\approx F_2\left(\frac{\widetilde{\mu}_2^2 h_{\min}^2 - 4\sigma^2\frac{N_t}{p_2 B_{t,2}}N_t\mu_{B_{t,2},L}}{4N_t\sigma^2\frac{N_t}{p_2 B_{t,2}}\sigma_{B_{t,2},L}}\right) \tag{34}$$

$$\triangleq P_{\text{II}}(p_2, B_{t,2}) \tag{35}$$

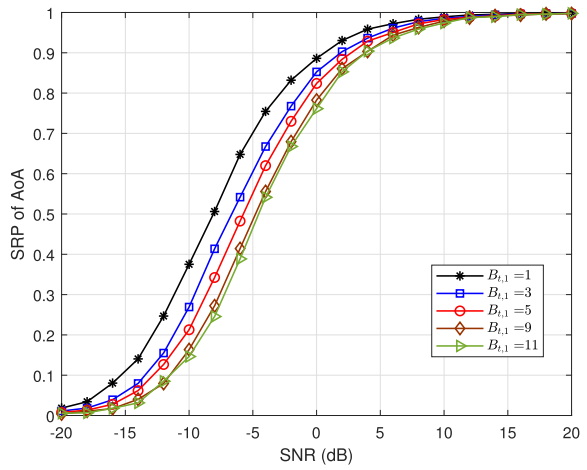

Fig. 3. SRP of AoA vs. SNR (dB) ($N_r = 20, N_t = 64, L = 4, N = 4, s = 1, E_1 = 10, \widetilde{B}_{t,1} = 1$).

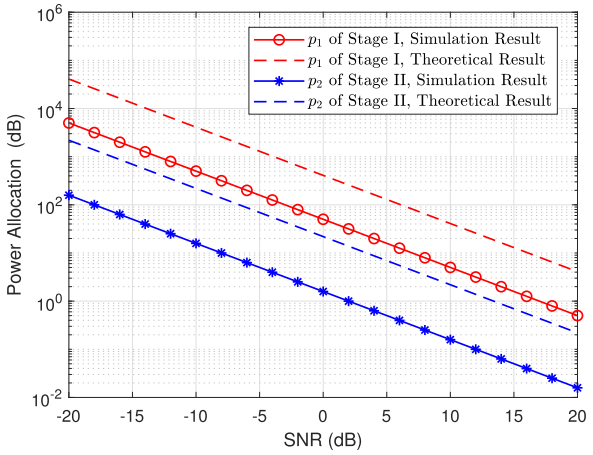

Fig. 4. Power allocation to achieve the required SRP vs. SNR (dB) ($N_r = 20, N_t = 64, L = 4, N = 4, s = 1, \widetilde{B}_{t,1} = 1, \eta_1 = \eta_2 = 0.95$).

$P_1(\frac{E_1 N}{B_{t,1} N_r}, B_{t,1})$ gives

$$P_I\left(\frac{E_1 N}{B_{t,1} N_r}, B_{t,1}\right) = F_2\left(\frac{h_{\min}^2 \widetilde{\mu}_1^2 E_1 N - 4N_t N_r \sigma^2 (N_r^{\frac{1}{2}} + B_{t,1}^{\frac{1}{2}})^2}{4N_t N_r \sigma^2 (N_r^{\frac{1}{2}} + B_{t,1}^{\frac{1}{2}})(N_r^{-\frac{1}{2}} + B_{t,1}^{-\frac{1}{2}})^{\frac{1}{3}}}\right). \quad (42)$$

Taking the first derivative of the argument inside $F_2(\cdot)$ in (42) with respect to $B_{t,1}$ reveals that the argument is a decreasing function of $B_{t,1}$. This implies that the $P_I(\frac{E_1 N}{B_{t,1} N_r}, B_{t,1})$ in (42) is a monotonically decreasing function of $B_{t,1}$. Hence, (41) is maximized when $B_{t,1} = \widetilde{B}_{t,1}$, which completes the proof. □

Therefore, based on Theorem 3, the maximum SRP of AoA estimation for a given E_1 is given by

$$P_I\left(\frac{E_1 N}{\widetilde{B}_{t,1} N_r}, \widetilde{B}_{t,1}\right) = F_2\left(\frac{h_{\min}^2 \widetilde{\mu}_1^2 E_1 N / N_r - 4\sigma^2 N_t \mu_{N_r, \widetilde{B}_{t,1}}}{4N_t \sigma^2 \sigma_{N_r, \widetilde{B}_{t,1}}}\right). \quad (43)$$

We demonstrate Theorem 3 via numerical simulations in Fig. 3, in which the SRP of AoA is evaluated for different numbers of channel uses $B_{t,1} \in \{1, 3, 5, 9, 11\}$. The simulation parameters $N_r = 20$, $N_t = 64$, $L = 4$, $N = 4$, $s = 1$, $E_1 = 10$, and $\widetilde{B}_{t,1} = 1$ are assumed. The curves clearly show that the highest SRP is achieved when $B_{t,1} = 1$.

Now, based on Theorem 3, the solution to (38) is readily obtained as follows. In order to make SRP of AoA higher than η_1 in (38), we solve the inverse function in (43) with respect to E_1 and conclude that the resource allocation of Stage I should meet the following conditions:

$$\begin{cases} E_1 = \dfrac{4\sigma^2 N_t N_r (F_2^{-1}(\eta_1)\sigma_{N_r, \widetilde{B}_{t,1}} + \mu_{N_r, \widetilde{B}_{t,1}})}{h_{\min}^2 \widetilde{\mu}_1^2 N}, & (44a) \\ B_{t,1} = \widetilde{B}_{t,1}, & (44b) \\ p_1 = \dfrac{E_1 N}{\widetilde{B}_{t,1} N_r}, & (44c) \end{cases}$$

where $F_2^{-1}(\cdot)$ is the inverse function of $F_2(\cdot)$. By using similar procedures of the proof of Theorem 3, we observe the following more general result about the number of vectors d in the signal model stated Lemma 1.

Corollary 1: The bound in (30) is a monotonically decreasing function of the number of measurement vectors d.

Remark 7: Corollary 1 states the effect of d on the recovery performance of SOMP. It can be interpreted in the following way. The increase of the number of measurement vectors d has an effect of increasing the number of columns of $\mathbf{C}$ in Lemma 1 while keeping the $C_{\min}$ unchanged. This leads to the increase of the noise power due to the increase in the dimension of $\mathbf{N}$, which in turn reduces SRP.

When it comes to the number of channel uses $B_{t,2}$ at Stage II, we cannot reach the same conclusion as Theorem 3 because the constant μ_2 in (34) changes with $B_{t,2}$. Therefore, Given $B_{t,1} = \widetilde{B}_{t,1}$ and the total number of channel uses K for channel sounding, $B_{t,2}$ is determined by (28), i.e., $K = \widetilde{B}_{t,1} N_r / N + B_{t,2}$. Then, the solution to (39) is given by

$$\begin{cases} E_2 = \dfrac{4\sigma^2 N_t (F_2^{-1}(\eta_2)\sigma_{B_{t,2},L} + \mu_{B_{t,2},L})}{h_{\min}^2 \widetilde{\mu}_2^2}, & (45a) \\ B_{t,2} = K - \widetilde{B}_{t,1} N_r / N, & (45b) \\ p_2 = \dfrac{E_2}{K - \widetilde{B}_{t,1} N_r / N}. & (45c) \end{cases}$$

In summary, after solving the two-subproblems in (38) and (39), we successfully solve the problem in (37). The specific resource allocations for two stages are shown in (44) and (45), respectively. In particular, when the total power budget $E \geq E_1 + E_2$, the joint SRP of AoA and AoD is at least $\eta_1 \eta_2$.

In Fig. 4, we illustrate the designed resource allocations in (44) and (45) with the simulation results. The parameters are set as $\eta_1 = \eta_2 = 0.95$. The curves of theoretical results calculate the power allocations p_1 and p_2 through (44c) and (45c). The curves of simulation results are the required power allocations to achieved SRPs of η_1 and η_2. The simulation parameters $N_r = 20$, $N_t = 64$, $L = 4$, $N = 4$, $s = 1$ are assumed. In Fig. 4, to achieve the same

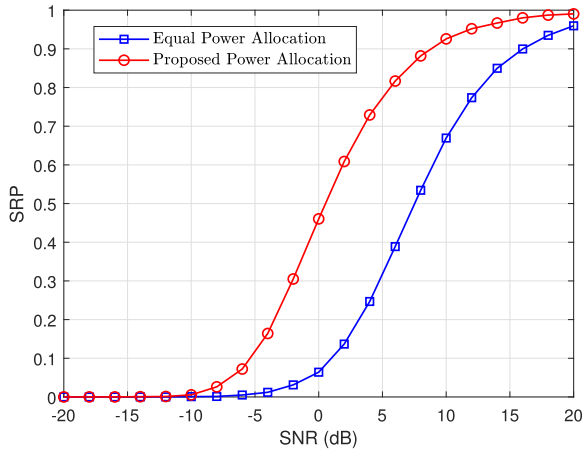

Fig. 5. Evaluation of the power allocation strategy with equal power allocation ($N_r = 20, N_t = 64, L = 4, N = 4, s = 1, \widetilde{B}_{t,1} = 1, \eta_1 = \eta_2 = 0.95$).

required SRP, i.e., $\eta_1 = \eta_2 = 0.95$, Stage II requires less power allocation than Stage I. This is because the design of the sounding beams for Stage II saves the power consumption. Overall, the trend of the theoretical results is consistent with that of the simulation results, which validates the proposed resource allocation strategies in (44) and (45).

In Fig. 5, we demonstrate the SRP of AoA and AoD achieved by the power allocations in (44) and (45) compared to the equal power allocation. The power allocations p_1 and p_2 are calculated by setting $\eta_1 = \eta_2 = 0.95$ and $\sigma = 0.1$ in (44) and (45). The simulation parameters are $N_r = 20$, $N_t = 64$, $L = 4$, $N = 4$, $s = 1$. As we can see from Fig. 5, the proposed power allocation achieves much higher SRP than that of the equal power allocation, which verifies the effectiveness of the proposed power allocation strategy.

V. Extension to Two-Stage Method With Super Resolution

In this section, we extend the proposed two-stage method to the one with super resolution, through which we aim to address the issue of unresolvable quantization errors. Among the existing works, there are two directions to solve the quantization error for off-grid effect. Firstly, the works in [46]–[48] model the response vector as the summation of on-grid part and the approximation error, in which the sparse Bayesian inference is utilized to estimate the approximation error. Secondly, the atomic norm minimization has been proposed in [19]–[21], which can be viewed as the case when the infinite dictionary matrix is employed. Based on atomic norm minimization, the sparse signal recovery is reformulated as a semidefinite programming. Compared to the sparse Bayesian inference, one advantage of atomic norm minimization is that the recovery guarantee is analyzable [19]–[21]. Following the methodology of the atomic norm minimization, in this section, we aim to estimate the AoAs and AoDs, i.e., $\{f_{r,1}, \ldots, f_{r,L}\}$ and $\{f_{t,1}, \ldots, f_{t,L}\}$, under the proposed two-stage framework.

A. Super Resolution AoA Estimation

The sounding beams of Stage I, i.e., $\mathbf{F}_{b,1}$ and $\mathbf{W}_{b,1}$, are designed according to (15) and (18). By using the exact expression of $\mathbf{H}$ in (2) rather than the quantized version in (8), the observations for Stage I is given by

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{W}_{b,1}^H \mathbf{H} \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1 \\ &= \mathbf{W}_{b,1}^H \mathbf{A}_r \operatorname{diag}(\mathbf{h}) \mathbf{A}_t^H \mathbf{F}_{b,1} + \mathbf{W}_{b,1}^H \mathbf{N}_1 \\ &= \mathbf{W}_{b,1}^H \mathbf{A}_r \mathbf{C}_r + \mathbf{W}_{b,1}^H \mathbf{N}_1, \end{aligned} \tag{46}$$

where $\mathbf{Y}_1 \in \mathbb{C}^{N_r \times B_{t,1}}$ and $\mathbf{C}_r = \operatorname{diag}(\mathbf{h}) \mathbf{A}_t^H \mathbf{F}_{b,1} \in \mathbb{C}^{L \times B_{t,1}}$. Since $\mathbf{W}_{b,1} = \mathbf{S}_{N_r}$ in (18), projecting $\mathbf{Y}_1$ onto $\mathbf{W}_{b,1}$ yields

$$\widetilde{\mathbf{Y}}_1 = \mathbf{W}_{b,1} \mathbf{Y}_1 = \mathbf{A}_r \mathbf{C}_r + \mathbf{N}_1. \tag{47}$$

The observation in (47) is rewritten by explicitly involving the array response vectors,

$$\widetilde{\mathbf{Y}}_1 = [\mathbf{a}_r(f_{r,1}), \ldots, \mathbf{a}_r(f_{r,L})] \mathbf{C}_r + \mathbf{N}_1 = \mathbf{R}_1 + \mathbf{N}_1, \tag{48}$$

where $\mathbf{R}_1 = [\mathbf{a}_r(f_{r,1}), \ldots, \mathbf{a}_r(f_{r,L})] \mathbf{C}_r \in \mathbb{C}^{N_r \times B_{t,1}}$. The atom $\mathbf{A}_r(f, \mathbf{b}) \in \mathbf{C}^{N_r \times B_{t,1}}$ is defined in [19], [20] as $\mathbf{A}_r(f, \mathbf{b}) = \mathbf{a}_r(f) \mathbf{b}^H$, where $f \in [0, 1)$ and $\mathbf{b} \in \mathbb{C}^{B_{t,1} \times 1}$ with $\|\mathbf{b}\|_2 = 1$. We let the collection of all such atoms be the set $\mathcal{A}_r = \{\mathbf{A}_r(f, \mathbf{b}) : f \in [0, 1), \|\mathbf{b}\|_2 = 1\}$. Obviously, the cardinality of $\mathcal{A}_r$ is infinite. The matrix $\mathbf{R}_1$ in (48) can be written as the linear combination among the atoms from the atomic set $\mathcal{A}_r$,

$$\mathbf{R}_1 = \sum_{l=1}^{L} [\mathbf{c}_r]_l \mathbf{A}_r(f_{r,l}, \mathbf{b}_l) = \sum_{l=1}^{L} [\mathbf{c}_r]_l \mathbf{a}_r(f_{r,l}) \mathbf{b}_l^H, \tag{49}$$

where $\mathbf{c}_r \in \mathbb{R}^{L \times 1}$ is the coefficient vector with $[\mathbf{c}_r]_l \geq 0$, and it has the relationship $[\mathbf{C}_r]_{l,:} = [\mathbf{c}_r]_l \mathbf{b}_l^H$, $\forall l = 1, \ldots, L$. Observing (49), the dimension of vector $\mathbf{c}_r$, i.e., L, can be interpreted as the sparest representation of $\mathbf{R}_1$ in the context of the atomic set $\mathcal{A}_r$. Therefore, in order to seek the sparsest representation, after taking the noise in (48) into account, the reconstruction problem is formulated by

$$\min_{\mathbf{R}_1} \|\mathbf{R}_1\|_{\mathcal{A}_r,0} + \frac{\lambda_1}{2} \|\widetilde{\mathbf{Y}}_1 - \mathbf{R}_1\|_F^2, \tag{50}$$

where $\lambda_1 > 0$ is the penalty parameter, and $\|\mathbf{R}_1\|_{\mathcal{A}_r,0}$ is defined as

$$\|\mathbf{R}_1\|_{\mathcal{A}_r,0} = \inf_{\mathbf{c}_r} \|\mathbf{c}_r\|_0 \tag{51a}$$

$$\text{s.t. } \mathbf{R}_1 = \sum_{l=1}^{L} [\mathbf{c}_r]_l \mathbf{A}_r(f_{r,l}, \mathbf{b}_l), \tag{51b}$$

$$\mathbf{A}_r(f_{r,l}, \mathbf{b}_l) \in \mathcal{A}_r, \quad [\mathbf{c}_r]_l \geq 0, \tag{51c}$$

with $\|\mathbf{R}_1\|_{\mathcal{A}_r,0}$ revealing the minimal number of atoms in $\mathbf{R}_1$. When the sparest representation of $\mathbf{R}_1$, i.e., $\{[\mathbf{c}_r]_l \mathbf{a}_r(f_{r,l}) \mathbf{b}_l^H\}_{l=1}^L$, is found by solving (50), the AoAs $\{f_{r,l}\}_{l=1}^L$ can be obtained from the atomic decomposition in (49). However, since the minimization problem in (51) is combinatorial, it is not tractable to calculate the value of $\|\mathbf{R}_1\|_{\mathcal{A}_r,0}$. To overcome the challenge, the problem in (50) is relaxed as,

$$\min_{\mathbf{R}_1} \|\mathbf{R}_1\|_{\mathcal{A}_r,1} + \frac{\lambda_1}{2} \|\widetilde{\mathbf{Y}}_1 - \mathbf{R}_1\|_F^2, \tag{52}$$

where $\|\mathbf{R}_1\|_{\mathcal{A}_r,1}$ is the atomic norm of $\mathbf{R}_1$ defined by

$$\|\mathbf{R}_1\|_{\mathcal{A}_r,1} = \inf_{\mathbf{c}_r} \|\mathbf{c}_r\|_1 \tag{53a}$$

$$\text{s.t.} \quad \mathbf{R}_1 = \sum_{l=1}^{L} [\mathbf{c}_r]_l \mathbf{A}_r(f_{r,l}, \mathbf{b}_l), \tag{53b}$$

$$\mathbf{A}_r(f_{r,l}, \mathbf{b}_l) \in \mathcal{A}_r, \quad [\mathbf{c}_r]_l \geq 0. \tag{53c}$$

It is noted that in (53), the atomic norm $\|\mathbf{R}_1\|_{\mathcal{A}_r,1}$ is to minimize the summation of entries in $\mathbf{c}_r$ instead of the number of non-zero elements in (51).

Different from the intractability of (51), the problem in (53) can be efficiently solved by semi-definite programming [19]:

$$\|\mathbf{R}_1\|_{\mathcal{A}_r,1} = \inf_{\mathbf{u},\mathbf{Z}} \frac{1}{2}\text{tr}\left(\text{Toeplitz}(\mathbf{u})\right) + \frac{1}{2}\text{tr}(\mathbf{Z}) \tag{54a}$$

$$\text{s.t.} \begin{bmatrix} \text{Toeplitz}(\mathbf{u}) & \mathbf{R}_1 \\ \mathbf{R}_1^H & \mathbf{Z} \end{bmatrix} \succeq \mathbf{0}, \tag{54b}$$

where $\mathbf{u} \in \mathbf{C}^{N_r \times 1}, \mathbf{Z} \in \mathbf{C}^{B_{t,1} \times B_{t,1}}$, and $\text{Toeplitz}(\mathbf{u}) \in \mathbf{C}^{N_r \times N_r}$ denotes the Hermitian Toeplitz matrix generated by the vector $\mathbf{u}$. Plugging (54) into (52) gives

$$\inf_{\mathbf{u},\mathbf{Z},\mathbf{R}_1} \text{tr}\left(\text{Toeplitz}(\mathbf{u})\right) + \text{tr}(\mathbf{Z}) + \lambda_1 \|\widetilde{\mathbf{Y}}_1 - \mathbf{R}_1\|_F^2 \tag{55a}$$

$$\text{s.t.} \mathbf{X} = \begin{bmatrix} \text{Toeplitz}(\mathbf{u}) & \mathbf{R}_1 \\ \mathbf{R}_1^H & \mathbf{Z} \end{bmatrix}, \quad \mathbf{X} \succeq \mathbf{0}. \tag{55b}$$

It is straightforward to find that (55) is convex, where ADMM can be employed to accelerate the computation. The augmented Lagrangian of (55) is expressed as

$$\begin{aligned} &\mathcal{L}(\mathbf{u}, \mathbf{Z}, \mathbf{R}_1, \mathbf{X}, \mathbf{\Lambda}) \\ &= \text{tr}\left(\text{Toeplitz}(\mathbf{u})\right) + \text{tr}(\mathbf{Z}) + \lambda_1 \|\widetilde{\mathbf{Y}}_1 - \mathbf{R}_1\|_F^2 \\ &\quad + \left\langle \mathbf{\Lambda}, \mathbf{X} - \begin{bmatrix} \text{Toeplitz}(\mathbf{u}) & \mathbf{R}_1 \\ \mathbf{R}_1^H & \mathbf{Z} \end{bmatrix} \right\rangle \\ &\quad + \frac{\rho}{2} \left\| \mathbf{X} - \begin{bmatrix} \text{Toeplitz}(\mathbf{u}) & \mathbf{R}_1 \\ \mathbf{R}_1^H & \mathbf{Z} \end{bmatrix} \right\|_F^2, \end{aligned} \tag{56}$$

where $\mathbf{X} \in \mathbb{C}^{(N_r+B_{t,1}) \times (N_r+B_{t,1})}$ and $\mathbf{\Lambda} \in \mathbb{C}^{(N_r+B_{t,1}) \times (N_r+B_{t,1})}$ are Hermitian matrices, and ρ is the Lagrangian multiplier. Then, with t being the iteration index, we iteratively update the variables in (56) as follows:

$$\begin{aligned} &(\mathbf{u}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{R}_1^{t+1}) \\ &\quad = \arg\min_{\mathbf{u},\mathbf{Z},\mathbf{R}_1} \mathcal{L}(\mathbf{u}, \mathbf{Z}, \mathbf{R}_1, \mathbf{X}^t, \mathbf{\Lambda}^t), \end{aligned} \tag{57}$$

$$\mathbf{X}^{t+1} = \arg\min_{\mathbf{X} \succeq \mathbf{0}} \mathcal{L}(\mathbf{u}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{R}_1^{t+1}, \mathbf{X}, \mathbf{\Lambda}^t), \tag{58}$$

$$\mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t + \rho \left(\mathbf{X}^{t+1} - \begin{bmatrix} \text{Toeplitz}(\mathbf{u}^{t+1}) & \mathbf{R}_1^{t+1} \\ (\mathbf{R}_1^{t+1})^H & \mathbf{Z}^{t+1} \end{bmatrix} \right). \tag{59}$$

The solutions of the (57) and (58) are respectively

$$[\mathbf{u}^{t+1}]_i = \begin{cases} \dfrac{V_i + \rho S_i}{(N_r - t)\rho + N_r}, & i = 1 \\ \dfrac{V_i + \rho S_i}{(N_r - t)\rho}, & i = 2, \ldots, N_r, \end{cases}$$

$$\text{with } V_i = \sum_{k=1}^{N_r+1-i} [\mathbf{\Lambda}^t]_{k,k-1+i}, \quad S_i = \sum_{k=1}^{N_r+1-i} [\mathbf{X}^t]_{k,k-1+i},$$

$$\begin{aligned} \mathbf{R}_1^{t+1} &= \frac{1}{\lambda_1 + \rho}(\lambda_1 \widetilde{\mathbf{Y}}_1 + \rho[\mathbf{X}^t]_{1:N_r, N_r+1:\text{end}} \\ &\quad + [\mathbf{\Lambda}^t]_{1:N_r, N_r+1:\text{end}}), \\ \mathbf{Z}^{t+1} &= \frac{1}{\rho}([\mathbf{\Lambda}^t]_{N_r+1:\text{end}, N_r+1:\text{end}} \\ &\quad + \rho[\mathbf{X}^t]_{N_r+1:\text{end}, N_r+1:\text{end}} - \mathbf{I}_{B_{t,1}}), \\ \mathbf{X}^{t+1} &= \begin{bmatrix} \text{Toeplitz}(\mathbf{u}^{t+1}) & \mathbf{R}_1^{t+1} \\ (\mathbf{R}_1^{t+1})^H & \mathbf{Z}^{t+1} \end{bmatrix} - \frac{1}{\rho}\mathbf{\Lambda}^t. \end{aligned}$$

It is worth noting that in order to guarantee $\mathbf{X} \succeq 0$ as shown in (58), we can set the negative eigenvalues of $\mathbf{X}^{t+1}$ to 0. When the iterative process converges, the result $\text{Toeplitz}(\mathbf{u})$ can be utilized to obtain the estimation of AoAs. Specifically, we can take Vandermonde decomposition [19] for $\text{Toeplitz}(\mathbf{u})$, $\text{Toeplitz}(\mathbf{u}) = \mathbf{V}\mathbf{D}\mathbf{V}^H$, where $\mathbf{V} = [\mathbf{a}_r(\hat{f}_{r,1}), \ldots, \mathbf{a}_r(\hat{f}_{r,L})] \in \mathbb{C}^{N_r \times L}$ with $\{\hat{f}_{r,l}\}_{l=1}^L$ being the estimated AoAs and $\mathbf{D} = \text{diag}([d_1, \ldots, d_L]) \in \mathbb{C}^{L \times L}$. In practice, it is not necessary to calculate the Vandermonde decomposition of $\text{Toeplitz}(\mathbf{u})$ explicitly. Since the column subspace of $\text{Toeplitz}(\mathbf{u})$ is equal to $\mathcal{R}(\mathbf{V})$, the set of AoAs can be estimated from $\text{Toeplitz}(\mathbf{u})$ efficiently by spectrum estimation algorithms such as MUSIC or ESPRIT [19], [20].

B. Super Resolution AoD Estimation

Similarly, the observations for the second stage is given by

$$\begin{aligned} \mathbf{Y}_2 &= \mathbf{W}_{b,2}^H \mathbf{H} \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{N}_2 \\ &= \mathbf{W}_{b,2}^H \mathbf{A}_r \,\text{diag}(\mathbf{h}) \mathbf{A}_t^H \mathbf{F}_{b,2} + \mathbf{W}_2^H \mathbf{N}_2 \\ &= \mathbf{C}_t \mathbf{A}_t^H \mathbf{F}_{b,2} + \mathbf{W}_{b,2}^H \mathbf{N}_2, \end{aligned} \tag{60}$$

where we let $\mathbf{C}_t = \mathbf{W}_{b,2}^H \mathbf{A}_r \,\text{diag}(\mathbf{h}) \in \mathbb{C}^{L \times L}$. At Stage II, the observation $\mathbf{Y}_2$ in (60) is rewritten as

$$\mathbf{Y}_2^H = \mathbf{F}_{b,2}^H \mathbf{A}_t \mathbf{C}_t^H + \mathbf{N}_2^H \mathbf{W}_{b,2} = \mathbf{R}_2^H + \mathbf{N}_2^H \mathbf{W}_{b,2}, \tag{61}$$

where we let $\mathbf{R}_2 = \mathbf{F}_{b,2}^H \mathbf{A}_t \mathbf{C}_t^H \in \mathbb{C}^{B_{t,2} \times L}$. Due to the design of $\mathbf{F}_{b,2}$ in (27), we have

$$\mathbf{F}_{b,2}^H \mathbf{A}_t = \sqrt{p_2}[\mathbf{A}_t]_{1:B_{t,2},:} = \sqrt{p_2}[\mathbf{a}_t(f_{t,1}), \ldots, \mathbf{a}_t(f_{t,L})]_{1:B_{t,2},:}.$$

For convenience, we define $\widetilde{\mathbf{a}}_t(f) = [\mathbf{a}_t(f)]_{1:B_{t,2}} \in \mathbb{C}^{B_{t,2} \times 1}$ and $\widetilde{\mathbf{A}}_t = [\widetilde{\mathbf{a}}_t(f_{t,1}), \ldots, \widetilde{\mathbf{a}}_t(f_{t,L})] \in \mathbb{C}^{B_{t,2} \times L}$. The AoD estimation boils down to extracting L parameters $\{f_{t,l}\}_{l=1}^L$ in $\widetilde{\mathbf{A}}_t$. We let $\mathbf{A}_t(f, \mathbf{b}) \in \mathbf{C}^{B_{t,2} \times L}$ be $\mathbf{A}_t(f, \mathbf{b}) = \widetilde{\mathbf{a}}_t(f)\mathbf{b}^H$, where $f \in [0, 1)$, $\mathbf{b} \in \mathbb{C}^{L \times 1}$ with $\|\mathbf{b}\|_2 = 1$, and the atomic set $\mathcal{A}_t$ is defined by $\mathcal{A}_t = \{\mathbf{A}_t(f, \mathbf{b}) : f \in [0, 1], \|\mathbf{b}\|_2 = 1\}$, Similarly, $\mathbf{R}_2^H$ in (61) can be written as the linear combination of the atoms from the set $\mathcal{A}_t$,

$$\mathbf{R}_2^H = \sum_{l=1}^{L} [\mathbf{c}_t]_l \mathbf{A}_t(f_{t,l}, \mathbf{b}_l) = \sum_{l=1}^{L} [\mathbf{c}_t]_l \mathbf{a}_t(f_{t,l}) \mathbf{b}_l^H,$$

where $\mathbf{c}_t \in \mathbb{R}^{L\times 1}$ is the coefficient vector with $[\mathbf{c}_t]_l \geq 0$. Therefore, using the similar approach as AoA estimation in (52), the AoD estimation problem is given by

$$\min_{\mathbf{R}_2} \left\|\mathbf{R}_2^H\right\|_{\mathcal{A}_t,1} + \frac{\lambda_2}{2}\left\|\mathbf{Y}_2 - \mathbf{R}_2\right\|_F^2, \tag{62}$$

where λ_2 is a penalty parameter. The problem in (62) can also be solved in a similar manner as (52), and the estimation for AoDs, i.e., $\{\hat{f}_{t,l}\}_{l=1}^L$, can be obtained.

Furthermore, after the AoAs $\{\hat{f}_{r,l}\}_{l=1}^L$ and AoDs $\{\hat{f}_{t,l}\}_{l=1}^L$ are estimated, we can easily calculate the AoA and AoD array response matrix $\widehat{\mathbf{A}}_r$ and $\widehat{\mathbf{A}}_t$. Then, by using the channel estimation technique provided in Section III-C, the final channel estimation result is obtained.

VI. Simulation Results

In this section, we evaluate the performance of the proposed two-stage AoA and AoD estimation method and two-stage method with super resolution. For comparison, we take the OMP-based mmWave channel estimation method [9] as our benchmark. Also, we include the oracle estimator as we discussed in (7). The parameter settings for evaluation are as follows. We assume throughout the simulation $N_r = 20$, $N_t = 64$, and the channel model is given by (1). We let the dimensions of the angle grids for the proposed two-stage method and OMP [9] be $G_r = sN_r$ and $G_t = sN_t$. The number of paths is $L = 4$. The variance of the path gain is $\sigma_l^2 = 1, \forall l$. The number of RF chains is $N = 4$. The number of channel uses for the estimation task is $K = 50$. The minimum allowed transmit beams at Stage I are $\widetilde{B}_{t,1} = 1$. Without loss of generality, for the proposed two-stage framework, the power budget $E = E_1 + E_2$, where E_1 and E_2 are, respectively, given by the resource allocations in (44) and (45) with $\eta_1 = \eta_2 = 0.95$ and SNR $= 20$dB.

To evaluate the estimation performance, we use three performance metrics:

- The first metric is the SRP. The error of the estimated angles are defined as

$$\epsilon = \frac{1}{2L}\sum_{l=1}^{L}\left(|f_{r,l} - \hat{f}_{r,l}|^2 + |f_{t,l} - \hat{f}_{t,l}|^2\right).$$

  We declare the reconstruction is successful if $\epsilon \leq 10^{-3}$. Precisely, SRP is defined as

$$\text{SRP} = \frac{\text{number of trials with } \epsilon \leq 10^{-3}}{\text{number of total trials}}.$$

- The second metric is MSE of angle estimation defined as

$$\text{MSE} = \mathbb{E}\left[\sum_{l=1}^{L}\left(|f_{r,l} - \hat{f}_{r,l}|^2 + |f_{t,l} - \hat{f}_{t,l}|^2\right)\right].$$

- The third metric is NMSE of channel estimation defined as

$$\text{NMSE} = \mathbb{E}[\|\mathbf{H} - \widehat{\mathbf{H}}\|_F^2/\|\mathbf{H}\|_F^2],$$

  where $\widehat{\mathbf{H}}$ is the channel estimate.

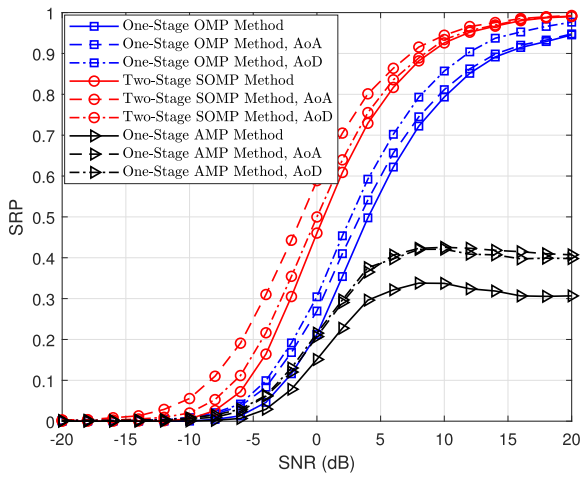

Fig. 6. SRP vs. SNR (dB) with discrete angles ($N_r = 20, N_t = 64$, $L = 4, N = 4, K = 50, \widetilde{B}_{t,1} = 1, s = 1$).

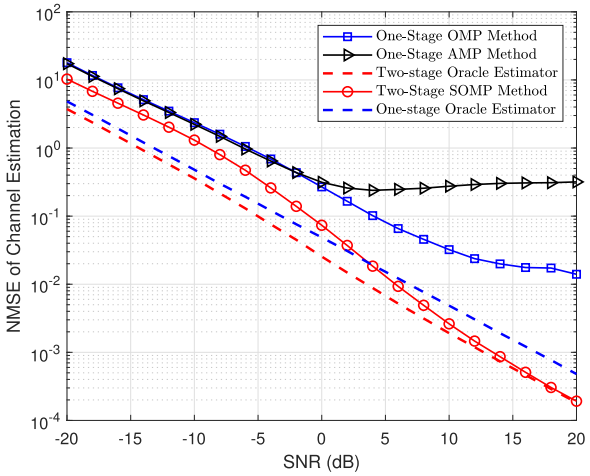

Fig. 7. NMSE vs. SNR (dB) with discrete angles ($N_r = 20, N_t = 64$, $L = 4, N = 4, K = 50, \widetilde{B}_{t,1} = 1, s = 1$).

A. Channel Estimation Performance of Two-Stage Method With Discrete Angles

For the simulations with discrete angles in Figs. 6-7, the $f_{t,l}$ and $f_{r,l}$ in (1) are uniformly distributed on the grids of size $G_t = N_t$ and $G_r = N_r$, respectively. Three methods are compared, which are proposed two-stage SOMP method, one-stage OMP method [9], AMP method [49], and oracle method in (7). We show the SRP in Fig. 6 and NMSE in Fig. 7.

In Fig. 6, considering that oracle method assumes that AoAs and AoDs are known as a priori, we do not illustrate the performance of the oracle method when comparing the SRP. As can be seen in Fig. 6, the proposed two-stage SOMP method achieves a higher SRP compared to the benchmarks. It is worth noting that the AMP-based method require the minimal measurements to guarantee the convergence [49]. When the number of channel uses is limited, the AMP-based method can not achieve the near one SRP even if the SNR is high. Also, the SRPs of AoA and AoD of the proposed two-stage SOMP method are both higher than those of one-stage OMP method. The improvement of SRP of AoD is because we optimize the sounding beams of the second stage based on the estimated AoA result. For the improvement of SRP of

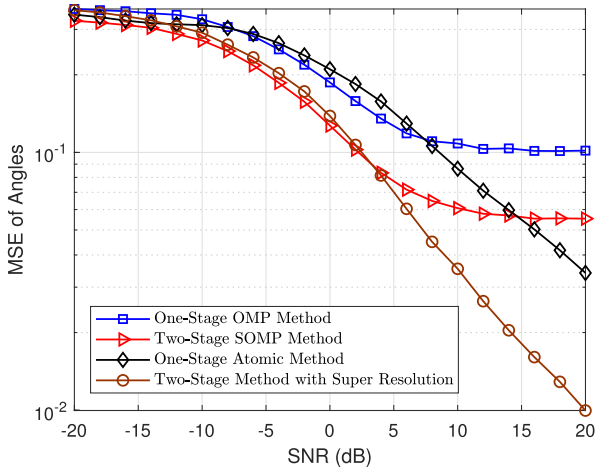

Fig. 8. MSE vs. SNR (dB) with continuous angles ($N_r = 20, N_t = 64, L = 4, N = 4, K = 50, \widetilde{B}_{t,1} = 1, s = 2$).

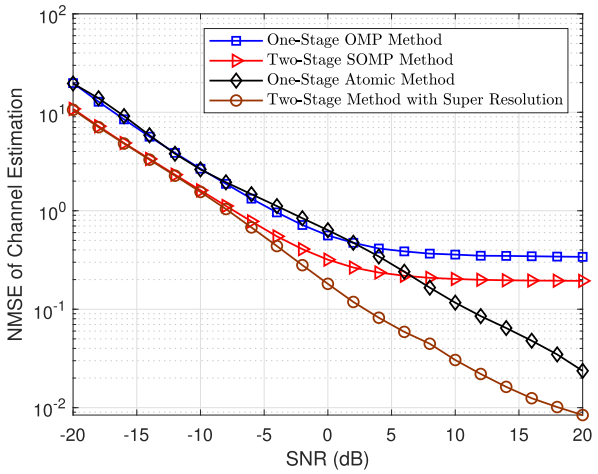

Fig. 9. NMSE vs. SNR (dB) with continuous angles ($N_r = 20, N_t = 64, L = 4, N = 4, K = 50, \widetilde{B}_{t,1} = 1, s = 2$).

AoA, it is because we allocate more power budget to Stage I according to the proposed resource allocation strategy.

Similarly, in Fig. 7, the proposed two-stage SOMP method has lower NMSE than the one-stage OMP and AMP methods. In addition, we can find from Fig. 7 that the proposed two-stage SOMP method converges to the performance of the oracle method as SNR grows. Overall, Figs. 6-7 verify that the proposed two-stage method outperforms the one-stage OMP in the scenario of discrete angles.

B. Channel Estimation Performance of Two-Stage Method With Continuous Angles

For this set of simulations in Fig. 8-9, we assume the $f_{t,l}$ and $f_{r,l}$ in (1) are uniformly distributed in $[0, 1)$. Four methods are compared, which are the proposed two-stage SOMP method, two-stage method with super resolution, one-stage OMP method [9], and one-stage atomic method [21]. When implementing the two-stage SOMP method and one-stage OMP method with the defined angle grids, the estimated angles are located on the defined grids. Fig. 8 illustrates the MSE and Fig. 9 illustrates the NMSE of channel estimation.

In Fig. 8, the proposed two-stage SOMP method and two-stage method with super resolution outperform the one-stage OMP and one-stage atomic method, respectively. Interestingly, the two-stage SOMP method achieves the minimal MSE when SNR is low. This is because when SNR is low, i.e., SNR $\leq$ 5dB, the noise power is higher than that of the quantization error. Therefore, using the quantized model could reduce the complexity of problem and achieve near-optimal performance. When SNR is high, i.e., SNR $\geq$ 5dB, the two-stage method with super resolution achieves the minimal MSE. This is because when SNR is high, the quantization error will become dominant, which can not be handled by the grid-based methods. Nevertheless, the Fig. 8 verifies that by dividing the estimation into two stages, the estimation of AoAs and AoDs is improved compared to the one-stage estimation.

Likewise, in Fig. 9, the proposed two-stage SOMP method and two-stage method with super resolution also achieve lower NMSE than the one-stage OMP and one-stage atomic methods. Similarly, when SNR is high, the two-stage method with super resolution shows the minimum NMSE.

C. Analysis of Computational Complexity

For two-stage method, the computational complexity for the first stage is $\mathcal{O}(LN_rG_r) = \mathcal{O}(sLN_r^2)$, and the complexity for the second stage is $\mathcal{O}(LB_{t,2}G_t) = \mathcal{O}(sL(K - N_r/N)N_t) = \mathcal{O}(sLKN_t)$ with K being the number of channel uses. Therefore, the total computational complexity for two-stage method is $\mathcal{O}(sLN_r^2) + \mathcal{O}(sLKN_t) = \mathcal{O}(sLKN_t)$. However, for the one-stage OMP method, the computational complexity is $\mathcal{O}(LKNG_tG_r) = \mathcal{O}(s^2LKNN_tN_r)$. It is obvious that the two-stage method has much lower computational complexity compared to the one-stage OMP by $\mathcal{O}(sNN_r)$ times.

For the two-stage method with super resolution, in Stage I, the computational complexity of ADMM per iteration is dominated by the eigenvalue decomposition of $\mathbf{X}^{t+1}$, i.e., $\mathcal{O}(N_r^3)$. Similarly, for Stage II, each iteration has the computational complexity of $\mathcal{O}(B_{t,2}^3) = \mathcal{O}((K - N_r/N)^3) = \mathcal{O}(K^3)$. Given the number of iteration T and $K \geq N_r$, the total computational complexity of the super resolution method is $\mathcal{O}(TN_r^3) + \mathcal{O}(TK^3) = \mathcal{O}(TK^3)$. In order to compare the complexities of the two-stage method with super resolution and one-stage OMP, we consider a simple example as follows. In particular, if $N_r = N_t$ and $K = \mathcal{O}(N_r)$, the complexity of the proposed two-stage method with super resolution is $\mathcal{O}(s^2LN/T)$ times lower than that of the one-stage OMP.

VII. Conclusion

In this paper, the two-stage method for the mmWave channel estimation was proposed. By sequentially estimating AoAs and AoDs of large-dimensional antenna arrays, the proposed two-stage method saved the computational complexity as well as channel use overhead compared to the existing methods. Theoretically, we analyzed the SRPs of AoA and AoD of the proposed two-stage method. Based on the analyzed SRP, we designed the resource allocation strategy among two stages to guarantee the accurate AoA and AoD estimation. In addition, to resolve the issue of quantization error, we extended the

proposed two-stage method to a version with super resolution. The numerical simulations showed that the proposed two-stage method achieves more accurate channel estimation result than the one-stage method.

APPENDIX A
PROOF OF LEMMA 1

For an arbitrary random noise matrix $\mathbf{N}$, the SRP of SOMP has been characterized in [33]. This result is general to be extended to the case in Lemma 1, where the entries in $\mathbf{N}$ are i.i.d. complex Gaussian.

Theorem 4 (SRP of SOMP With Arbitrary Random Noise [33]): Suppose the signal model provided in Lemma 1. Given the measurement matrix $\mathbf{\Phi}$ with its MIP constant satisfying $\mu < 1/(2L+1)$ and the cumulative distribution function (CDF) of $\|\mathbf{N}\|_2$ satisfying

$$\Pr(\|\mathbf{N}\|_2 \leq x) = F_N(x), \tag{63}$$

the SRP of SOMP in Algorithm 1 satisfies

$$\Pr(\mathcal{V}_S) \geq F_N\left(\frac{C_{\min}(1-(2L-1)\mu)}{2}\right), \tag{64}$$

where $\mathcal{V}_S$ is the event of successful reconstruction of Algorithm 1, $C_{\min} = \min\limits_{i\in\Omega} \|[\mathbf{C}]_{i,:}\|_2$.

According to the results in Theorem 4, the SRP of SOMP is characterized by the CDF of $\|\mathbf{N}\|_2$. Thus, in order to extend the result provided in Theorem 4 to the case in Lemma 1, the CDF of $\|\mathbf{N}\|_2$ is of interest when the entries of $\mathbf{N} \in \mathbb{C}^{M\times d}$ are i.i.d. $\mathcal{CN}(0,\sigma^2)$. Fortunately, according to [40], [41], the CDF of the largest singular value of $\mathbf{N}$ converges in distribution to the Tracy-Widom law as M, d tend to ∞,

$$\Pr(\|\mathbf{N}\|_2 \leq x) \approx F_2\left(\frac{x^2/\sigma^2 - \mu_{M,d}}{\sigma_{M,d}}\right), \tag{65}$$

where the function $F_2(\cdot)$ is the CDF of Tracy-Widom law [40], [41], $\mu_{M,d} = (M^{1/2}+d^{1/2})^2$, and $\sigma_{M,d} = (M^{1/2}+d^{1/2})(M^{-1/2}+d^{-1/2})^{1/3}$. Finally, after plugging the expression in (65) into (64) of Theorem 4, we obtain Lemma 1, which completes the proof. □

APPENDIX B
PROOF OF PROPOSITION 2

One can write the effective noise as $\tilde{\mathbf{N}} = \mathbf{E} + \mathbf{N}$ where the entries in $\mathbf{N}$ are i.i.d. with $\mathcal{CN}(0,\sigma^2)$. Therefore, we have the following probability bound,

$$\begin{aligned}\Pr\left(\|\tilde{\mathbf{N}}\|_2 \leq x\right) &\overset{(a)}{\leq} \Pr\left(\|\mathbf{E}\|_2 + \|\mathbf{N}\|_2 \leq x\right)\\ &\overset{(b)}{\approx} F_2\left(\frac{(x-\|\mathbf{E}\|_2)^2/\sigma^2 - \mu_{M,d}}{\sigma_{M,d}}\right),\end{aligned} \tag{66}$$

where the inequality (a) is due to the triangular inequality, and the approximation (b) holds from (65). Then, according to Theorem 4, plugging the expression (66) into (64) leads to

$$\Pr(\mathcal{V}_S) \geq F_2\left(\frac{\left((1-(2L-1)\mu)C_{\min} - 2\|\mathbf{E}\|_2\right)^2 - 4\sigma^2\mu_{M,d}}{4\sigma^2\sigma_{M,d}}\right),$$

where $C_{\min} = \min\limits_{i\in\Omega} \|[\mathbf{C}]_{i,:}\|_2$. This concludes the proof. □

APPENDIX C
PROOF OF THEOREM 2

Plugging RSB in (25) and TSB in (27) into (21) gives $\|[\mathbf{\Phi}_2]_{:,j}\|_2 = \sqrt{p_2 B_{t,2}/N_t}$, $j = 1,\ldots,G_t$, and $C_{\min} = \min_{t_l\in\Omega_t} \|[\mathbf{C}_2]_{t_l,:}\|_2 = |h_{\min}|$ with t_l being the index of the lth path of $\mathbf{A}_t$ in $\bar{\mathbf{A}}_t$ such that $[\bar{\mathbf{A}}_t]_{:,t_l} = [\mathbf{A}_t]_{:,l}$, $l = 1,\ldots,L$. Hence, incorporating the latter $C_{\min}$ and $\|[\mathbf{\Phi}_2]_{:,j}\|_2$ into Proposition 2, and neglecting the quantization term can conclude the proof. □

REFERENCES

[1] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[2] 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.

[3] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[4] W. Zhang, T. Kim, D. J. Love, and E. Perrins, "Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5814–5827, Nov. 2018.

[5] W. Zhang, T. Kim, and S.-H. Leung, "A sequential subspace method for millimeter wave MIMO channel estimation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5355–5368, May 2020.

[6] N. Docomo *et al.*, "White paper on 5G channel model for bands up to 100 GHz," Tech. Rep., 2016. [Online]. Available: http://www.5gworkshops.com/5GCM.html

[7] S. Hur *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 454–469, Apr. 2016.

[8] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[9] 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.

[10] A. Manoj and A. P. Kannu, "Channel estimation strategies for multi-user mm wave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5678–5690, Nov. 2018.

[11] X. Rao, V. K. N. Lau, and X. Kong, "CSIT estimation and feedback for FDD multi-user massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3157–3161.

[12] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[13] Q. Duan, T. Kim, L. Dai, and E. Perrins, "Coherence statistics of structured random ensembles and support detection bounds for OMP," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1638–1642, Nov. 2019.

[14] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 860–864.

[15] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4001–4012, Jul. 2014.

[16] J. Chen and X. Ho, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[17] E. van den Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516–2527, May 2010.

[18] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.

[19] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[20] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.
[21] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Super-resolution channel estimation for arbitrary arrays in hybrid millimeter-wave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 947–960, Sep. 2019.
[22] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2017.
[23] Z. Wan, Z. Gao, B. Shim, K. Yang, G. Mao, and M.-S. Alouini, "Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2337–2342, Feb. 2019.
[24] R. Zhang, B. Shim, and H. Zhao, "Downlink compressive channel estimation with phase noise in massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5534–5548, Sep. 2020.
[25] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9435–9448, Oct. 2018.
[26] S. Park and R. W. Heath, Jr., "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8047–8062, Dec. 2018.
[27] I. Ahmed and R. Annavajjala, "Bayesian CRLB for joint AoA, AoD, and channel estimation using UPA in millimeter-wave communications," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.
[28] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Jan. 2006.
[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Jan. 2007.
[30] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.
[31] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, p. v-721.
[32] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[33] W. Zhang and T. Kim, "Successful recovery performance guarantees of SOMP under the ℓ_2-norm of noise," 2021, *arXiv:2108.13855*.
[34] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.
[35] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. 18th Eur. Signal Process. Conf.*, 2010, pp. 427–431.
[36] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
[37] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.
[38] J.-F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "On the noise robustness of simultaneous orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 864–875, Feb. 2017.
[39] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
[40] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 2001.
[41] K. Johansson, "Shape fluctuations and random matrices," *Commun. Math. Phys.*, vol. 209, no. 2, pp. 437–476, Feb. 2000.
[42] B. Nadler, "On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix," *J. Multivariate Anal.*, vol. 102, no. 2, pp. 363–371, Feb. 2011.
[43] M. Bengtsson, "A pragmatic approach to multi-user spatial multiplexing," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, 2002, pp. 130–134.
[44] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Dec. 2005.
[45] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.
[46] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.
[47] H. Tang, J. Wang, and L. He, "Off-grid sparse Bayesian learning-based channel estimation for mmWave massive MIMO uplink," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, Feb. 2018.
[48] B. Qi, W. Wang, and B. Wang, "Off-grid compressive channel estimation for mm-Wave massive MIMO with hybrid precoding," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 108–111, Jan. 2018.
[49] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Oct. 2009.

Wei Zhang (Member, IEEE) received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2019. From 2019 to 2020, he was a Research Associate with the City University of Hong Kong. Since 2020, he has been a Post-Doctoral Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include channel estimation, signal processing in millimeter wave MIMO systems, reconfigurable intelligent surfaces, and matrix completion.

Miaomiao Dong received the B.Eng. degree from Northwestern Polytechnical University in 2010, the M.Eng. degree from Xidian University in 2013, and the Ph.D. degree from the City University of Hong Kong in 2020. Since 2020, he has been a Researcher with the Theory Laboratory, Central Research Institute, 2012 Laboratories, Huawei Technologies Company Ltd. His current research interests include intelligent reconfigurable surface communications, massive MIMO precoder design, and wireless localization.

Taejoon Kim (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA. Prior to joining the University of Kansas (KU) as an Assistant Professor, he was a Senior Researcher with the Nokia Bell Laboratories, Berkeley, CA, USA, a Postdoctoral Researcher at KTH, Stockholm, Sweden, and an Assistant Professor with the City University of Hong Kong. His research interests include 5G-and-beyond wireless systems, multiple-input multiple-output (MIMO) communications, statistical signal processing, security, and machine learning. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and previously served as a Guest Editor for IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He holds 29 issued U.S. patents.

He was a recipient of the Miller Faculty Award from the KU School of Engineering and The President's Award from the City University of Hong Kong. Along with the coauthors, we won The IEEE Communications Society Stephen O. Rice Prize in 2016 and IEEE PIMRC 2012 Best Paper Award.