Random SBT Precoding for Angle Estimation of mmWave Massive MIMO Systems Using Sparse Arrays Spacing

YUAN-PEI LIN, (Senior Member, IEEE),
AND TING-MING YANG, (Graduate Student Member, IEEE)
National Chiao Tung University, Hsinchu 30010, Taiwan

Corresponding author: Yuan-Pei Lin (ypl@mail.nctu.edu.tw)

ABSTRACT In this article, we consider the estimation of angle of arrival (AoA) and angle of departure (AoD) for mmWave MIMO channels. For mmWave systems, there is typically a large number of antennas but limited radio frequency (RF) chains. The RF chain limitation indirectly restricts the effective number of antennas or the effective array size. In radar applications, it is known that a larger array size leads to more accurate AoA estimation or more resolvable paths. It is also known that with sparse arrays, which places antennas in a nonuniform sparse manner, an enlarged virtual array can be constructed when the sources are uncorrelated. We show that the idea of sparse arrays can also be used in mmWave systems to have augmented array sizes. The main difficulty in applying sparse arrays to mmWave MIMO channels lies in the fact the signals of different paths can be correlated in mmWave system, but uncorrelatedness is necessary for applying sparse array results. We show that the paths can be decorrelated by employing random precoders and combiners that are submatrices of banded Toeplitz matrices (SBT). When the precoder and combiner are SBT that are designed according to the antenna spacing of a sparse array, we can construct enlarged virtual transmit and receive arrays. With enlarged virtual arrays, accurate estimates can be obtained, as will be demonstrated by simulation examples.

INDEX TERMS Angle estimation, hybrid precoder, massive MIMO, mmWave, submatrix of a banded Toeplitz, sparse arrays.

I. INTRODUCTION

Recent advances show that it is feasible to pack a large number of antennas in a small area, particularly in millimeter wave (mmWave) communication systems that use small wavelengths. However cost and power constraints often prohibit having one dedicated radio frequency (RF) chain for each antenna [1]–[3]. Hybrid precoding that employs RF analog processing and baseband digital processing has been proposed to overcome the RF limitation [4]. The RF limitation also places a restriction on the effective number of antennas, i.e., effective array size, for the estimation of angle of arrival (AoA) and angle of departure (AoD). In radar applications, it is known that a larger array size can improve the estimation accuracy [5] or increase the number of resolvable paths [6], [7]. Array augmentation

can be achieved through the use of sparse arrays when the sources are uncorrelated [8], [9]. By placing N antennas in a nonuniform sparse manner, a virtual ULA array with a size in the order of N^2 can be constructed [6]. The ULA structure of the virtual array lends itself to subspace-based methods for angle estimation such as MUSIC [10] or ESPRIT [11]. The use of sparse arrays usually needs a large number of training vectors as it requires the sources be uncorrelated and sample covariance matrix of the sources be close to a diagonal matrix. Having insufficient training vectors leads to partial correlation in the sample covariance matrix, which in turns causes interference in subsequent angle estimation. The problem addressed in sparse linear arrays bears some similarities to the problem of direction finding in mmWave systems, but there are important differences. In the former case the number of antennas is limited, whereas in the latter case there is usually a large number of antennas but limited RF chains. For mmWave channels, signals of different paths

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Khawaja.

can be correlated, so the condition of uncorrelated signals that is assumed in sparse linear array literature is not satisfied. Furthermore, in addition to the estimation of AoA, the AoD needs to be estimated as well for mmWave systems.

Many methods have been proposed to estimate AoA/AoD subject to hybrid structure [12]–[24]. In these works, the channels are assumed to be either time-invariant or time-varying. In the case of a time-invariant channel, the path gains and angles remain the same throughout training, i.e., training completed within the channel coherence time [12]–[19]. The precoder and combiner are designed in [12] so that an augmented matrix consisting of the received vectors becomes a submatrix of the channel and 2D unitary ESPRIT is applied for subsequent channel estimation. The channel reciprocity in time division duplexing systems is exploited to estimate the signal subspace in [13]. A two-step AoA estimation is given in [14] by first applying 2D-DFT to obtain a coarse estimation, which is then improved through angle rotation on a finer grid. Hierarchical multi-resolution estimation using successive beams of different beamwidth is considered in [15]. It is shown in [16] that auxiliary beam pairs can be used to provide high resolution channel estimates. With the feedback of initial AoD estimates, beamforming can be applied to improve the SNR for high resolution estimation [17]. Compressed sensing based methods have been shown to be a useful tool for angle estimation [18]–[20]. Accurate channel estimation is achieved using orthogonal matching pursuit (OMP) in [19]. Based on OMP, generalized block OMP is developed in [20] to take advantage of the block sparsity property of mmWave channels for better channel estimation. In the case of a time-varying channel [21]–[24], the angles remain the same while the path gains are time-varying due to fading that is caused by user movement. Such a channel model is valid if the training period is longer than the channel coherence time. An ESPRIT based estimation of AoA, AoD and path delays is given in [21] for MIMO-OFDM systems. A channel estimation method using subspace fitting of signal subspace is proposed in [22]. Simultaneous estimation of AoA and AoD using 2D beamspace MUSIC method is proposed in [23] by exploiting uncorrelated path gains. When the path gains are uncorrelated, the received covariance matrix has a Toeplitz structure, which has been shown to be useful for denoising and improving the estimation accuracy [24]. For estimation methods developed for a time-invariant channel, the performance can be degraded when there is fading. On the other hand, for methods designed for a time-varying channel, the performance can be considerably affected when there is little user movement and the channel is almost time-invariant.

In this article, we show that sparse arrays can be incorporated in the estimation of AoA and AoD for mmWave channels. The precoder is a submatrix of a banded Toeplitz (SBT). In this case, the coefficients in each column vector of the precoder are obtained by shifting the coefficients of a prototype beamformer. We show that when the shifts are designed according to the antenna spacing of a sparse array, array augmentation can be achieved. Suppose linear arrays are used, and there are $\bar{N}_r$ RF chains at the receiver and $\bar{N}_t$ RF chains at the transmitter. We can construct a virtual uniform linear array (ULA) of size in the order of $\bar{N}_t^2$ at the transmitter, and a virtual ULA of size in the order of $\bar{N}_r^2$ at the receiver. By designing the prototype beamformer to have random phase, we can decorrelate signals of different paths and apply sparse array results whether the channel is time-invariant or time-varying. This means angle estimation can be decoupled from the channel coherence time. Like earlier sparse array results, the decorrelation of the paths needs a large number of training vectors as correlation among the paths leads to interference in angle estimation. When a large number of training vectors is not available, we develop an interference cancellation scheme based on initial estimates of the angles. With the aid of interference cancellation, more accurate angle estimates can be obtained, as will be demonstrated in simulations.

Notation: The variance of a random variable x is denoted as σ_x^2 and the expectation of x by $E[x]$. The 2-norm of a vector $\mathbf{f}$ is denoted as $||\mathbf{f}||$ and the k-th entry of $\mathbf{f}$ as $[\mathbf{f}]_k$. The notation $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^H$ denote, respectively, the transpose, the conjugate, and conjugate transpose of a matrix $\mathbf{A}$. $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. Given an $m \times n$ matrix $\mathbf{X}$, the notation $vec(\mathbf{X})$ denotes the $mn \times 1$ vectorized version of $\mathbf{X}$. The notation $\mathbf{A} \odot \mathbf{B}$ denotes the Khatri-Rao product and $\mathbf{A} \otimes \mathbf{B}$ the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$. A useful property concerning the two products is as follows. Let $\mathbf{X} = \mathbf{ABC}$, then $vec(\mathbf{X}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$. When $\mathbf{B}$ is a diagonal matrix, we have $vec(\mathbf{X}) = (\mathbf{C}^T \odot \mathbf{A})\mathbf{b}$, where $\mathbf{b}$ is a column vector that consists of the diagonal elements of $\mathbf{B}$.

Outline of the Paper: A system model is given in Sec. II. Sec. III introduces SBT precoding. In Sec. IV, a review of sparse arrays is given and the proposed array augmentation using sparse array spacing for mmWave MIMO systems is presented. Simulation results are given in Sec. V and a conclusion in Sec. VI.
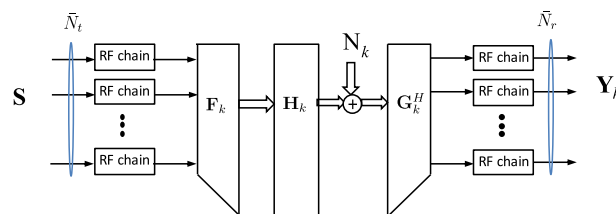

FIGURE 1. A MIMO communication system with hybrid precoding.

II. SYSTEM MODEL

Consider the wireless system with N_t transmit antennas and N_r receive antennas in Fig. 1. The transmitter is equipped with $\bar{N}_t$ RF chains and the receiver $\bar{N}_r$ RF chains. For the kth training block, the channel is modeled by an $N_r \times N_t$ matrix $\mathbf{H}_k$. We adopt the geometric channel representation that is useful for modeling mmWave propagation. Suppose

the channel has L paths, then [25]

$$\mathbf{H}_k = \mathbf{A}_r \mathbf{D}_{\alpha,k} \mathbf{A}_t^H, \tag{1}$$

where $\mathbf{D}_{\alpha,k}$ is an $L \times L$ diagonal matrix with the i-th diagonal element $[\mathbf{D}_{\alpha,k}]_{i,i} = \alpha_{k,i}$, which is the complex gain of the i-th path. The two matrices $\mathbf{A}_t$ and $\mathbf{A}_r$ consist of antenna response vectors at, respectively, the transmitter and receiver sides. In particular, $\mathbf{A}_r = \begin{bmatrix} \mathbf{a}_{r,1} & \mathbf{a}_{r,2} & \cdots & \mathbf{a}_{r,L} \end{bmatrix}$ and $\mathbf{A}_t = \begin{bmatrix} \mathbf{a}_{t,1} & \mathbf{a}_{t,2} & \cdots & \mathbf{a}_{t,L} \end{bmatrix}$, are of dimensions, respectively, $N_r \times L$ and $N_t \times L$, where $\mathbf{a}_{r,i}$ and $\mathbf{a}_{t,i}$ are antenna response vectors that depends on the array. We assume ULA is used. For an ULA, $\mathbf{a}_{r,i}$ can be given in terms of the angles of arrival (AoA) θ_i^r and $\mathbf{a}_{t,i}$ in terms of the angles of departure (AoD) θ_i^t. Define the $N \times 1$ vector

$$\mathbf{a}_N(u) = \begin{bmatrix} 1 & e^{ju} & e^{j2u} & \cdots & e^{j(N-1)u} \end{bmatrix}^T, \tag{2}$$

then $\mathbf{a}_{r,i} = \mathbf{a}_{N_r}(u_i)$ and $\mathbf{a}_{t,i} = \mathbf{a}_{N_t}(p_i)$, where $u_i = \xi \cos \theta_i^r$, $p_i = \xi \cos \theta_i^t$, $\xi = 2\pi d/\lambda$, d is the antenna spacing and λ the wavelength.

Assumptions: We assume the channel stays the same during one training block of $\bar{N}_t$ training vectors. The AoA and AoD remain unchanged during the training period. For the path gains, we consider two cases. (1) The path gains are fixed during training, i.e., $\alpha_{k,i} = \alpha_i$. In this case we say the channel is time-invariant. (2) The path gains $\alpha_{k,i}$ vary from block to block, in which case the channel is said to be time-varying. The channel is time-invariant if the entire training period is less than the channel coherence time.

For the kth training block, the transmit precoding matrix $\mathbf{F}_k = \begin{bmatrix} \mathbf{f}_{k,0} & \mathbf{f}_{k,1} & \cdots \mathbf{f}_{k,\bar{N}_t-1} \end{bmatrix}$ is of dimensions $N_t \times \bar{N}_t$ and $||\mathbf{f}_{k,\ell}|| = 1$. The receive combining matrix is $\mathbf{G}_k = \begin{bmatrix} \mathbf{g}_{k,0} & \mathbf{g}_{k,1} & \cdots \mathbf{g}_{k,\bar{N}_r-1} \end{bmatrix}$, of size $N_r \times \bar{N}_r$, and $||\mathbf{g}_{k,\ell}|| = 1$. Consider a block of $\bar{N}_t$ training vectors $\mathbf{S} = \begin{bmatrix} \mathbf{s}_0 & \mathbf{s}_1 & \cdots & \mathbf{s}_{\bar{N}_t-1} \end{bmatrix}$, which is used for all training blocks. The kth received block $\mathbf{Y}_k = \begin{bmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{\bar{N}_t-1} \end{bmatrix}$ is given by[1]

$$\mathbf{Y}_k = \mathbf{G}_k^H \mathbf{A}_r \mathbf{D}_{\alpha,k} \mathbf{A}_t^H \mathbf{F}_k \mathbf{S} + \mathbf{G}_k^H \mathbf{N}_k, \tag{3}$$

where $\mathbf{N}_k$ is the $N_r \times \bar{N}_t$ channel noise matrix. The channel noise is assumed to be additive white Gaussian with zero mean and variance σ_n^2.

III. SBT PRECODING

Consider the following class of matrices. Let $\mathbf{B}$ be an $N \times M$ matrix with $N \geq M$ and $n_0, n_1, \cdots, n_{M-1}$ be integers such that $0 = n_0 < n_1 < n_2 < \cdots < n_{M-1} \leq N - 1$. Suppose $\mathbf{B}$ satifies $[\mathbf{B}]_{n,0} = 0$ for $n > N - 1 - n_{M-1}$ and

$$[\mathbf{B}]_{n,i} = \begin{cases} [\mathbf{B}]_{n-n_i,0}, & n \geq n_i, \\ 0, & 0 \leq n < n_i, \end{cases} \quad \text{for } i = 1, 2, \cdots, M-1. \tag{4}$$

Such a matrix is lower triangular and it is an $N \times M$ submatrix of an $N \times N$ banded Toeplitz (SBT) by keeping only the column vectors corresponding to the integers $n_0, n_1, \cdots, n_{M-1}$. Fig. 2 gives an example of an SBT precoder with $\bar{N}_t = 3$,
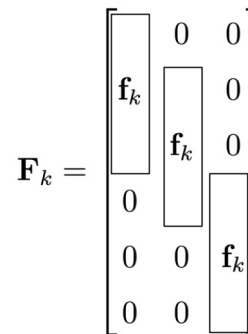
---

[1]In the presence of mutual coupling among the antennas, the output of one antenna can be affected by other antennas. We assume there is no mutual coupling.

FIGURE 2. An SBT precoder, where the vector $\mathbf{f}_k$ denotes the prototype beamformer for the kth precoder.

$n_1 = 1$, and $n_2 = 3$, i.e., 3 column vectors. The nonzero coefficients of each column vector are identical. The nonzero part, denoted as $\mathbf{f}_k$ in Fig. 2, will be referred to as the *prototype beamformer*. For example, the first column vector of $\mathbf{F}_k$ is equal to $\mathbf{f}_{k,0} = \begin{bmatrix} \mathbf{f}_k^T & 0 \end{bmatrix}^T$. When the precoder $\mathbf{F}_k$ is SBT with $\{n_0, n_1, \cdots, n_{\bar{N}_t-1}\}$, the i-th column vector $\mathbf{f}_{k,i}$ is the same as the prototype beamformer $\mathbf{f}_k$ except for a shift of n_i. Similarly, when the combiner $\mathbf{G}_k$ is SBT with $\{m_0, m_1, \cdots, m_{\bar{N}_r-1}\}$, the k-th column vector $\mathbf{g}_{k,i}$ is the same as the prototype $\mathbf{g}_k$ except for a shift of m_i. Let $F_k(\omega) = \sum_{n=0}^{N_t-1} [\mathbf{f}_k]_n e^{-jn\omega}$ and $G_k(\omega) = \sum_{n=0}^{N_r-1} [\mathbf{g}_k]_n e^{-jn\omega}$ be, respectively, the discrete-time Fourier transform of the beamforming prototype $\mathbf{f}_k$ and the combining prototype $\mathbf{g}_k$. We have the following lemma.

Lemma 1: When the precoder $\mathbf{F}_k$ is SBT with $\{n_0, n_1, \cdots, n_{\bar{N}_t-1}\}$, then

$$\mathbf{A}_t^H \mathbf{F}_k = \mathbf{D}_{f,k} \bar{\mathbf{A}}_t^H, \tag{5}$$

where $\mathbf{D}_{f,k}$ is a diagonal matrix with the i-th diagonal element given by $[\mathbf{D}_{f,k}]_{i,i} = F_k(p_i)$. The $\bar{N}_t \times L$ matrix $\bar{\mathbf{A}}_t$ has the ith column given by $\bar{\mathbf{a}}_{t,i} = \begin{bmatrix} 1 & e^{jp_i n_1} & \cdots & e^{jp_i n_{\bar{N}_t-1}} \end{bmatrix}^T$. Likewise, when the combiner $\mathbf{G}_k$ is SBT with $\{m_0, m_1, \cdots, m_{\bar{N}_r-1}\}$, then

$$\mathbf{A}_r^H \mathbf{G}_k = \mathbf{D}_{g,k} \bar{\mathbf{A}}_r^H, \tag{6}$$

where $\mathbf{D}_{g,k}$ is a diagonal matrix with $[\mathbf{D}_{g,k}]_{i,i} = G_k(u_i)$. The matrix $\bar{\mathbf{A}}_r$ is of dimensions $\bar{N}_r \times L$ and the i-th column is given by $\bar{\mathbf{a}}_{r,i} = \begin{bmatrix} 1 & e^{ju_i m_1} & \cdots & e^{ju_i m_{\bar{N}_r-1}} \end{bmatrix}^T$.

Proof. The (i, ℓ)th entry of $\mathbf{A}_t^H \mathbf{F}_k$ is equal to $\mathbf{a}_{t,i}^H \mathbf{f}_{k,\ell}$ and $\mathbf{a}_{t,i}^H \mathbf{f}_{k,\ell} = [\mathbf{f}_{k,\ell}]_0 + e^{-jp_i}[\mathbf{f}_{k,\ell}]_1 + \cdots + e^{-j(N_t-1)p_i}[\mathbf{f}_{k,\ell}]_{N_t-1}$. It can be rewritten as $e^{-jn_\ell p_i} F_k(p_i)$ as $\mathbf{f}_{k,\ell}$ is a shift of the prototype by n_ℓ. It follows that the i-th row of $\mathbf{A}_t^H \mathbf{F}_k$ can be written as

$$\mathbf{a}_{t,i}^H \begin{bmatrix} \mathbf{f}_{k,0} & \mathbf{f}_{k,1} & \cdots \mathbf{f}_{k,\bar{N}_t-1} \end{bmatrix}$$
$$= F_k(p_i) \begin{bmatrix} 1 & e^{-jp_i n_1} & \cdots & e^{-jp_i n_{\bar{N}_t-1}} \end{bmatrix}^T.$$

The right hand side of the above equation is $F_k(p_i)\bar{\mathbf{a}}_{t,i}^H$, where $\bar{\mathbf{a}}_{t,i}$ is as defined in Lemma 1, and (5) follows. We can derive (6) in a similar manner. △△△

Using the result in Lemma 1, the received block in (3) can be simplified as

$$\mathbf{Y}_k = \bar{\mathbf{A}}_r \mathbf{D}_k \bar{\mathbf{A}}_t^H + \mathbf{G}_k^H \mathbf{N}_k, \tag{7}$$

where we have used $\mathbf{S} = \sigma_s \mathbf{I}_{\bar{N}_t}$ and $\mathbf{D}_k = \sigma_s \mathbf{D}_{g,k}^H \mathbf{D}_{\alpha,k} \mathbf{D}_{f,k}$ with $\mathbf{D}_{f,k}$ and $\mathbf{D}_{g,k}$ as defined in (5) and (6). Notice that the signal part $\bar{\mathbf{A}}_r \mathbf{D}_k \bar{\mathbf{A}}_t^H$ is the same as that of a system with $\bar{N}_r$ receive antennas, $\bar{N}_t$ transmit antennas and equivalent path gains $[\mathbf{D}_k]_{i,i} = d_{k,i}$, given by

$$d_{k,i} = \sigma_s \alpha_{k,i} F_k(p_i) G_k^*(u_i). \tag{8}$$

With the use of SBT precoders and combiners, the system behaves as if there is no precoding and combining. The effect of precoding and combining is absorbed into the equivalent path gains. In other words, we can use SBT precoders and combiners to affect the equivalent path gains. Let us vectorize the received matrix $\mathbf{Y}_k$ and call it $\mathbf{y}_k$, we have

$$\mathbf{y}_k = \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r \right) \mathbf{d}_k + \mathbf{n}_k, \tag{9}$$

where $\mathbf{n}_k = vec(\mathbf{G}^H \mathbf{N}_k)$ and $\mathbf{d}_k = \begin{bmatrix} d_{k,1} & d_{k,2} & \cdots & d_{k,L} \end{bmatrix}^T$ is an $L \times 1$ vector. Consider the autocorrelation matrix $\mathbf{R}_y = E[\mathbf{y}_k \mathbf{y}_k^H]$. It is given by

$$\mathbf{R}_y = \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r \right) \mathbf{R}_d \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r \right)^H + \mathbf{R}_n, \tag{10}$$

where $\mathbf{R}_d = E[\mathbf{d}_k \mathbf{d}_k^H]$ and $\mathbf{R}_n = E[\mathbf{n}_k \mathbf{n}_k^H]$. When $\mathbf{G}_k$ is statistically semi-unitary, i.e, $E[\mathbf{G}_k^H \mathbf{G}_k] = \mathbf{I}_{\bar{N}_r}$, the noise $\mathbf{n}_k$ in (9) is uncorrelated and $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{\bar{N}_t \bar{N}_r}$.

Uniform SBT: From Lemma 1, we see that $\bar{\mathbf{A}}_t$ is an antenna response matrix corresponding to a hypothetical array, whose elements are placed according to $\{n_0, n_1, \cdots, n_{\bar{N}_t-1}\}$. Similarly $\bar{\mathbf{A}}_r$ is an antenna response matrix corresponding to a hypothetical array, whose elements are placed according to $\{m_0, m_1, \cdots, m_{\bar{N}_r-1}\}$. In the special case that the hypothetical arrays are uniform, i.e., $n_i = i$ and $m_i = i$, both $\bar{\mathbf{A}}_t$ and $\bar{\mathbf{A}}_r$ are antenna response matrices corresponding to ULA of sizes $\bar{N}_t$ and $\bar{N}_r$, respectively. In this case, the estimation of angles from $\mathbf{R}_y$ is a two-dimensional direction finding problem [27] with an effective $\bar{N}_t$-element ULA at the transmitter and $\bar{N}_r$-element ULA at the receiver. In particular, we can compute the eigen decomposition of $\mathbf{R}_y$,

$$\mathbf{R}_y = \begin{bmatrix} \mathbf{U}_s & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_s & 0 \\ 0 & \boldsymbol{\Sigma}_n \end{bmatrix} \begin{bmatrix} \mathbf{U}_s & \mathbf{U}_n \end{bmatrix}^H. \tag{11}$$

The signal subspace can be expressed as $\mathbf{U}_s = \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r \right) \mathbf{W}$, for some invertible matrix $\mathbf{W}$. Based on the signal subspace $\mathbf{U}_s$, the angles can be estimated using 2D MUSIC or 2D ESPRIT. The number of identifiable paths is up to $\bar{N}_r \bar{N}_t - 1$ with 2D MUSIC [32] and up to $\bar{N}_r \bar{N}_t - \max(\bar{N}_r, \bar{N}_t)$ with 2D ESPRIT [33]. In the following section, we will see that the effective ULA size can be increased to $O(\bar{N}_t^2)$ at the transmitter and $O(\bar{N}_r^2)$ at the receiver if we use sparse array spacing to design the SBT precoder and combiner.

IV. DIRECTION FINDING USING SBT

It is known that for a uniform linear array (ULA) with N antennas, subspace-based methods such as MUSIC can identify up to $N - 1$ uncorrelated sources [10]. The number of identifiable sources can be significantly increased through the use of sparse linear arrays. By judiciously placing N antennas, a virtual ULA of size in the order of N^2 can be constructed and the number of resolvable sources increased. For mmWave massive MIMO systems, the antenna array can be large but the number of RF chains is limited. In the following, we will see that for mmWave channels we can construct virtual arrays of sizes in the order of $\bar{N}_t^2$ for the transmitter and $\bar{N}_r^2$ for the receiver. This can be done by designing SBT precoders and combiners according to antenna spacing in sparse linear arrays. To make the paper more self-contained, we review in Sec. 4.1 related material on sparse linear arrays [6], [7] that is essential for our subsequent discussion. In Sec. 4.2, we show how to take advantage of sparse array results to increase the effective array sizes for hybrid mmWave MIMO systems.

A. REVIEW OF SPARSE LINEAR ARRAYS

Consider N antennas that are placed at the following locations $\{m_0 d, m_1 d, \cdots, m_{N-1} d\}$, where d is the smallest distance between any two antennas and m_i, for $i = 0, 1, \cdots, N - 1$, are integers. Let $\mathcal{S} = \{m_0, m_1, \cdots, m_{N-1}\}$.

Definition([6], [7]): The *difference coarray* of $\mathcal{S}$ is defined as the set $\mathcal{D} = \{m_i - m_j, 0 \le i, j < N\}$. Suppose $\mathcal{D}$ contains consecutive integers $-M_r, -M_r + 1, \cdots, M_r$. Then the sparse array is said to have *degree of freedom (DoF)* equal to $2M_r + 1$.

For example, let $N = 4$ and $\mathcal{S} = \{0, 1, 4, 6\}$. Then $\mathcal{D} = \{n : -6 \le n \le 6\}$; it contains 13 consecutive integers, from -6 to 6 and DoF $= 13$. Due to the consecutive integers in the coarray, a virtual ULA of size $2M_r + 1$ can be constructed. Suppose there are L sources and the antennas are placed according to $\mathcal{S}$. The received vector is

$$\mathbf{y}_k = \mathbf{A}_r \boldsymbol{\alpha}_k + \mathbf{n}_k,$$

where $\boldsymbol{\alpha}_k$ is the $L \times 1$ source vector, $\mathbf{A}_r$ is the antenna response matrix with the i-th column given by $\mathbf{a}_{r,i} = \begin{bmatrix} e^{jm_0 u_i} & e^{jm_1 u_i} & \cdots & e^{jm_{N-1} u_i} \end{bmatrix}^T$, u_i is the AoA parameter of the i-th source and $\mathbf{n}_k$ is the channel noise, assumed to be white and uncorrelated with the sources. The autocorrelation matrix $\mathbf{R}_y = E[\mathbf{y}_k \mathbf{y}_k^H]$ is given by $\mathbf{R}_y = \mathbf{A}_r \mathbf{R}_\alpha \mathbf{A}_r^H + \sigma_n^2 \mathbf{I}_N$, where $\mathbf{R}_\alpha = E[\boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^H]$. Assume the sources are uncorrelated, then $\mathbf{R}_\alpha$ is diagonal. The vectorized version of $\mathbf{R}_y$ is given by

$$\mathbf{r}_y = (\mathbf{A}_r^* \odot \mathbf{A}_r)\boldsymbol{\lambda} + vec(\sigma_n^2 \mathbf{I}_N), \tag{12}$$

where $\boldsymbol{\lambda}$ is an $L \times 1$ vector consisting of the diagonal elements of $\mathbf{R}_\alpha$.

It turns out that $\mathbf{A}_r^* \odot \mathbf{A}_r$ contains an antenna response matrix corresponding to an enlarged ULA of size $2M_r + 1$. Let $\mathbf{C} = \mathbf{A}_r^* \odot \mathbf{A}_r$ and its i-th column vector be $\mathbf{c}_i$, then $\mathbf{c}_i$ is given by $\mathbf{c}_i = \mathbf{a}_{r,i}^* \odot \mathbf{a}_{r,i}$, a vector of size $N^2 \times 1$. As the elements

of $\mathbf{a}_{r,i}$ are $e^{jm_\ell u_i}$ for $\ell = 0, 1, \cdots, N-1$, the elements of $\mathbf{c}_i$ are $e^{j(m_\ell - m_n)u_i}$. Note that $(m_\ell - m_n)$ are the elements in $\mathcal{D}$. For example, when $\mathcal{S} = \{0, 1, 3\}$ the numbers $(m_\ell - m_n)$ $0 \leq \ell, n \leq N-1$ are

$$0, \quad 1, \quad 3, \quad -1, \quad 0, \quad 2, \quad -3, \quad -2, \quad 0$$

There are some repeated elements and thus $\mathbf{C}$ has some repeated rows. We can select distinct rows of $\mathbf{C}$ that correspond to the consecutive integers in $\mathcal{D}$ using a $(2M_r + 1) \times N^2$ selection matrix. For the above example $\mathcal{D} = \{-3, -2, \cdots, 3\}$ and DoF is 7. We can use the 7×9 selection matrix

$$\mathbf{J}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Define $\tilde{\mathbf{A}}_r$ as

$$\tilde{\mathbf{A}}_r = \mathbf{J}_r (\mathbf{A}_r^* \odot \mathbf{A}_r), \qquad (13)$$

which is an antenna response matrix corresponding to a $(2M_r + 1)$-element ULA with antennas located at $-M_r d, -(M_r - 1)d, \cdots, M_r d$. Let $\tilde{\mathbf{r}}_y = \mathbf{J}_r \mathbf{r}_y$, then

$$\tilde{\mathbf{r}}_y = \tilde{\mathbf{A}}_r \boldsymbol{\lambda} + \sigma_n^2 \mathbf{J}_r vec(\mathbf{I}_N), \qquad (14)$$

where the ith column vector of $\tilde{\mathbf{A}}_r$ is given by $\tilde{\mathbf{a}}_{r,i} = \begin{bmatrix} e^{-jM_r u_i} & e^{-j(M_r - 1)u_i} & \cdots & e^{jM_r u_i} \end{bmatrix}^T$. We observe that $\tilde{\mathbf{r}}_y$ is equivalent to the output of a virtual ULA of size $2M_r + 1$ with a single transmit vector $\boldsymbol{\lambda}$. By applying spatial smoothing on $\tilde{\mathbf{r}}_y$ as in [7], we can obtain a $(M_r + 1) \times (M_r + 1)$ square matrix $\tilde{\mathbf{R}}_y$. Thanks to the ULA structure in $\tilde{\mathbf{r}}_y$, the signal subspace can be extracted from $\tilde{\mathbf{R}}_y$ and the angles estimated using useful subspace methods like MUSIC or ESPRIT. Up to M_r sources can be identified. Various sparse arrays have been proposed to design the spacing of the antennas so that M_r is in the order of N^2 e.g., minimum redundancy arrays [26], nested arrays [6]. Then the array size is enlarged from N to $O(N^2)$. The Cramér-Rao bound (CRB) for sparse arrays have been analyzed in [8], [9], [29].

There are some differences between the problem of direction finding for mmWave channels and that typically considered in sparse linear arrays for radar applications. For the mmWave channel in (1), we need to estimate AoAs as well as AoDs, not just AoAs. Furthermore, the signals corresponding to different paths, can be correlated. We will see that by using random SBT precoding, we can decorrelate the paths so that sparse array results can be applied. The decorrelation allows us to construct virtual arrays of augmented sizes as in sparse linear arrays.

B. VIRTUAL ARRAY AUGMENTATION USING SPARSE ARRAY SPACING

We use SBT precoder $\mathbf{F}_k$ and combiner $\mathbf{G}_k$ as in Lemma 1. Suppose the sets of integers used to construct $\mathbf{F}_k$ and $\mathbf{G}_k$ are given, respectively, by

$$\mathcal{S}_t = \{n_0, n_1, \cdots, n_{\bar{N}_t - 1}\}, \text{ and } \mathcal{S}_r = \{m_0, m_1, \cdots, m_{\bar{N}_r - 1}\}. \qquad (15)$$

For $\mathcal{S}_t$, the corresponding difference coarray is $\mathcal{D}_t = \{n_i - n_j, 0 \leq i, j < \bar{N}_t\}$. Suppose $\mathcal{D}_t$ contains consecutive integers $-M_t, -M_t + 1, \cdots, M_t$ and the DoF$= 2M_t + 1$. Similarly for the receiver side, the difference coarray $\mathcal{D}_r = \{m_i - m_j, 0 \leq i, j < \bar{N}_r\}$ contains consecutive integers $-M_r, -M_r + 1, \cdots, M_r$ and DoF$= 2M_r + 1$. To illustrate how sparse array results can be used in direction finding for mmWave systems, we first consider the special case when there is a single RF chain at the transmitter, i.e., $\bar{N}_t = 1$. The autocorrelation matrix $\mathbf{R}_y$ in (10) reduces to

$$\mathbf{R}_y = \bar{\mathbf{A}}_r \mathbf{R}_d \bar{\mathbf{A}}_r^H + \mathbf{R}_n. \qquad (16)$$

When the equivalent path gain vector $\mathbf{d}_k$ is uncorrelated, $\mathbf{R}_d$ is a diagonal matrix. How the equivalent path gains can be decorrelated will be addressed later in the design of precoders and combiners. Assuming $\mathbf{R}_d$ is diagonal and vectorizing $\mathbf{R}_y$, we have

$$\mathbf{r}_y = (\bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r)\boldsymbol{\lambda} + \mathbf{r}_n, \qquad (17)$$

where $\boldsymbol{\lambda}$ is an $L \times 1$ vector that consists of the diagonal elements of $\mathbf{R}_d$ and $\mathbf{r}_n = vec(\mathbf{R}_n)$ is the vectorized version of the noise autocorrelation matrix. Observe that the signal part in (17) has the same form as that in (12). Suppose the set $\mathcal{S}_r$ used to construct the SBT combiner is designed according to a sparse array and the associated DoF is $2M_r + 1$. Just like in Sec. 4.1, we can apply a $(2M_r + 1) \times \bar{N}_r^2$ selection matrix $\mathbf{J}_r$ on $\mathbf{r}_y$ so that $\mathbf{r}_y' = \mathbf{J}_r \mathbf{r}_y$ corresponds the output of a virtual array of size $2M_r + 1$. That is

$$\tilde{\mathbf{r}}_y = \tilde{\mathbf{A}}_r \boldsymbol{\lambda} + \mathbf{J}_r \mathbf{r}_n,$$

where $\tilde{\mathbf{A}}_r$ is as given in (13). Then the angles can be estimated as in sparse arrays.

Now let us address the more general case $\bar{N}_t \geq 1$. Vectorizing the autocorrelation matrix in (10), we have,

$$\mathbf{r}_y = \left(\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r\right)\boldsymbol{\lambda} + \mathbf{r}_n, \qquad (18)$$

a vector of size $(\bar{N}_t \bar{N}_r)^2 \times 1$.

Theorem 1: Suppose the precoder and combiner are SBT with associated sets of integers as in (15). Assume the equivalent paths are uncorrelated and $\mathbf{R}_d$ is diagonal. Then the autocorrelation matrix $\mathbf{R}_y$ of the received vector can be vectorized as $\mathbf{r}_y$ in (18). We can apply a selection matrix $\mathbf{J}$ on $\mathbf{r}_y$ so that $\tilde{\mathbf{r}}_y = \mathbf{J}\mathbf{r}_y$, a vector of size $(2M_r + 1)(2M_t + 1) \times 1$, is of the form

$$\tilde{\mathbf{r}}_y = (\tilde{\mathbf{A}}_t \odot \tilde{\mathbf{A}}_r)\boldsymbol{\lambda} + \mathbf{q}, \qquad (19)$$

where $\tilde{\mathbf{A}}_r$ is a $(2M_r + 1) \times L$ ULA antenna response matrix with the i-th column vector $\tilde{\mathbf{a}}_{r,i} = [e^{-jM_r u_i}\ e^{-j(M_r-1)u_i}\ \cdots\ e^{jM_r u_i}]^T$. The $(2M_t + 1) \times L$ matrix $\tilde{\mathbf{A}}_t$ is also a ULA antenna response matrix with the i-th column vector $\tilde{\mathbf{a}}_{t,i} = [e^{-jM_t p_i}\ e^{-j(M_t-1)p_i}\ \cdots\ e^{jM_t p_i}]^T$. The noise vector $\mathbf{q}$ is given by $\mathbf{q} = \mathbf{J}\mathbf{r}_n$.

A proof is given in Appendix A and a construction of the selection matrix $\mathbf{J}$ is given therein. Now let us consider the noise term $\mathbf{q}$ in (19). Suppose $\mathbf{G}_k$ is statistically semi-unitary $E[\mathbf{G}_k^H \mathbf{G}_k] = \mathbf{I}_{\bar{N}_r}$ and $E[\mathbf{n}_k \mathbf{n}_k^H] = \sigma_n^2 \mathbf{I}_{\bar{N}_t \bar{N}_r}$. In this case, the noise vector $\mathbf{q}$ has a close-form expression as given in the following lemma (a proof given in Appendix B).

Lemma 2: When the combiner $\mathbf{G}_k$ is statistically semi-unitary with $E[\mathbf{G}_k^H \mathbf{G}_k] = \mathbf{I}_{\bar{N}_r}$, the noise vector $\mathbf{q} = \mathbf{J}\mathbf{r}_n$ in Theorem 1 has only one nonzero element, right in the middle of the vector, and it is equal to the channel noise variance σ_n^2.

Observe that $\tilde{\mathbf{r}}_y$ in Theorem 1 is equivalent to the received vector when the receiver is equipped with a $(2M_r + 1)$-antenna ULA and the transmitter a $(2M_t + 1)$-antenna ULA. The signal and noise subspaces can be obtained from the vector $\tilde{\mathbf{r}}_y$ by using 2D matrix pencil [27] or 2D spatial smoothing [31], [37]. The spatially smoothed matrix contains both signal and noise parts. The signal part can be derived as in [27], [37]. But the noise part is different as the autocorrelation matrix has been rearranged as in Theorem 1; it is not the usual channel noise term and additional derivation is needed. To explain the behavior of noise after spatial smoothing is applied, we define the 1D selection matrix $\mathbf{J}_i^x = [\mathbf{0}_{(M_x+1) \times (M_x-i)}\ \mathbf{I}_{M_x+1}\ \mathbf{0}_{(M_x+1) \times i}]$ for $x \in \{t, r\}$ and $i = 0, 1, \cdots M_x$, and also the 2D selection matrix

$$\mathbf{J}_{i,j} = \mathbf{J}_i^t \otimes \mathbf{J}_j^r, \tag{20}$$

for $i = 0, 1, \cdots M_t$ and $j = 0, 1, \cdots M_r$. The 2D spatially smoothed matrix is defined as [37]

$$\mathbf{Y}_{ss} = [\mathbf{x}_{0,0}\quad \mathbf{x}_{0,1}\quad \cdots\quad \mathbf{x}_{0,M_r}\quad \mathbf{x}_{1,0}\quad \cdots\quad \mathbf{x}_{M_t,M_r}], \tag{21}$$

where $\mathbf{x}_{i,j} = \mathbf{J}_{i,j}\tilde{\mathbf{r}}_y$ and $\tilde{\mathbf{r}}_y$ is as given in (19). It is a square matrix of size $(M_t + 1)(M_r + 1) \times (M_t + 1)(M_r + 1)$. When spatial smoothing is applied on $\tilde{\mathbf{r}}_y$, the noise part is spatially smoothed along with the signal part. We have the following expression of $\mathbf{Y}_{ss}$.

Theorem 2: Suppose the combiner $\mathbf{G}_k$ is statistically semi-unitary as in Lemma 2. When we apply spatial smoothing on the vector $\tilde{\mathbf{r}}_y$ in (19), the spatially smoothed matrix $\mathbf{Y}_{ss}$ in (21) is given by

$$\mathbf{Y}_{ss} = (\tilde{\tilde{\mathbf{A}}}_t \odot \tilde{\tilde{\mathbf{A}}}_r)\mathbf{R}_d(\tilde{\tilde{\mathbf{A}}}_t \odot \tilde{\tilde{\mathbf{A}}}_r)^H + \sigma_n^2 \mathbf{I}_{(M_t+1)(M_r+1)}, \tag{22}$$

where $\tilde{\tilde{\mathbf{A}}}_t$ is an $(M_t + 1) \times L$ ULA antenna response matrix with the i-th column given by $\tilde{\tilde{\mathbf{a}}}_{t,i} = [1\ e^{jp_i}\ \cdots\ e^{jM_t p_i}]^T$ and $\tilde{\tilde{\mathbf{A}}}_r$ an $(M_r + 1) \times L$ ULA antenna response matrix with the i-th column given by $\tilde{\tilde{\mathbf{a}}}_{r,i} = [1\ e^{ju_i}\ \cdots\ e^{jM_r u_i}]^T$.

A proof is given in Appendix C. Theorem 2 means that $\mathbf{Y}_{ss}$ is equivalent to the autocorrelation matrix of a system with an $(M_r + 1)$-element ULA at the receiver and $(M_t + 1)$-element ULA at the transmitter. The signal subspace, i.e, the column space of $\mathbf{Y}_{ss}$, can be obtained by applying eigen value decomposition on $\mathbf{Y}_{ss}$ as in (11). Due to the embedded ULA structure in the signal subspace, 2D subspace methods such as MUSIC and ESPRIT can be used for angle estimation. The number of identifiable paths is up to $(M_t+1)(M_r+1)-1$ when 2D MUSIC [32] is applied and up to $\max(M_t(M_r + 1), M_r(M_t+1))$ when 2D ESPRIT is applied [33]. In the case that sparse array spacing is used to construct the precoder and combiner, M_t is in the order of $\bar{N}_t^2$ and M_r in the order of $\bar{N}_r^2$. After estimates of the angles are obtained, the path gains can be estimated using least squares [27]. Notice that the signal part of $\mathbf{Y}_{ss}$ in (22) is similar to that of $\mathbf{R}_y$ in (10) when the integers used to construct the precoder and combiner are uniform. The difference is that $\bar{\mathbf{A}}_r$ and $\bar{\mathbf{A}}_t$ in (10) correspond to ULA of sizes $\bar{N}_r$ and $\bar{N}_t$, respectively. But in (22), the two matrices $\tilde{\tilde{\mathbf{A}}}_r$ and $\tilde{\tilde{\mathbf{A}}}_t$ correspond to ULA of sizes $M_r + 1$ and M_t+1, respectively. By using sparse array spacing, the size of the ULA at the transmitter is enlarged from $\bar{N}_t$ to $O(\bar{N}_t^2)$, and the size of the ULA at the receiver enlarged from $\bar{N}_r$ to $O(\bar{N}_r^2)$.

In the above discussion we assume that the signals of different paths are uncorrelated and $\mathbf{R}_d$ is a diagonal matrix so that sparse array results can be used. The following lemma states that $\mathbf{R}_d$ is a diagonal matrix if the path gains are uncorrelated or when the prototypes are properly designed so that the equivalent path gains becomes uncorrelated.

Lemma 3: The autocorrelation matrix $\mathbf{R}_d = E[\mathbf{d}_k \mathbf{d}_k^H]$ is an invertible diagonal matrix if the following conditions are satisfied: (1) the beamforming and combining prototypes $\mathbf{f}_k$ and $\mathbf{g}_k$ are such that $E[|F_k(\omega)|^2] \neq 0$ and $E[|G_k(\omega)|^2] \neq 0$ for all ω, and (2) $E[\alpha_{k,i}\alpha_{k,j}^*] = 0$, for $i \neq j$ or $E[F_k(p_i)F_k^*(p_j)] = 0$ for $p_i \neq p_j$ or $E[G_k^*(u_i)G_k(u_j)] = 0$ for $u_i \neq u_j$.

Proof: For the i-th diagonal element, we have

$$[\mathbf{R}_d]_{i,i} = \sigma_s^2 E[|\alpha_{k,i}|^2]E[|G_k(u_i)|^2]E[|F_k(p_i)|^2],$$

which is positive if the first condition is satisfied. For the off-diagonal elements, note that

$$[\mathbf{R}_d]_{i,j} = \sigma_s^2 E[\alpha_{k,i}\alpha_{k,j}^*]E[F_k(p_i)F_k^*(p_j)]E[G_k^*(u_i)G_k(u_j)],$$

which is equal to zero if condition (2) holds. △△△

The second condition can be satisfied if the path gains $\alpha_{k,i}$ are uncorrelated random variables with zero mean. It happens when the channel exhibits fading during training, i.e., the separation of training blocks larger than the channel coherence times [23]. Then $\mathbf{R}_d$ becomes a diagonal invertible matrix and the paths are uncorrelated automatically even for deterministic precoder and combiner. On the other hand, if the channel coherence time is larger than the training period, i.e., time-invariant channels, the path gains stay the same during training and $E[\alpha_{k,i}\alpha_{k,j}^*] \neq 0$. In this case, the second condition in Lemma 3, i.e, path decorrelation, can be satisfied

by using random precoding and combining (to be discussed next). Then the equivalent paths are uncorrelated, irrespective of the channel coherence time.

Design of prototypes: For SBT precoders and combiners, only the prototypes $\mathbf{f}_k$ and $\mathbf{g}_k$ need to be designed. From (4), we see that the precoder prototype $\mathbf{f}_k$ has $N_t - n_{\bar{N}_t-1}$ coefficients and the combiner prototype $\mathbf{g}_k$ has $N_r - m_{\bar{N}_r-1}$ coefficients. Lemma 3 shows that the path gains can be successfully decorrelated if the prototype $F_k(\omega)$ or $G_k(\omega)$ is uncorrelated in frequency, i.e., $E[F_k(p_i)F_k^*(p_j)] = 0$ for $p_i \neq p_j$ or $E[G_k^*(u_i)G_k(u_j)] = 0$ for $u_i \neq u_j$. This can be satisfied approximately by designing the prototypes as follows. Let us choose the coefficients of the prototype to be binary random phases. That is,

$$[\mathbf{f}_k]_i = \gamma \beta_{k,i}, \text{ for } i = 0, 1, \cdots, N_t - n_{\bar{N}_t-1} - 1, \quad (23)$$

where $\beta_{k,i}$ are independent and $\beta_{k,i} = 1$ or -1 with equal probability, and $\gamma = 1/\sqrt{(N_t - n_{\bar{N}_t-1})}$ is a normalizing factor so that the prototype has unit norm. In this case, it can be shown that (a proof given in Appendix D)

$$|E[F_k(p_i)F_k^*(p_j)]| = \frac{1}{(N_t - n_{\bar{N}_t-1})}$$
$$\times \frac{|\sin(\frac{1}{2}(p_i - p_j)(N_t - n_{\bar{N}_t-1}))|}{|\sin(\frac{1}{2}(p_i - p_j))|}, \quad (24)$$

which is approximately zero when $(N_t - n_{\bar{N}_t-1})$ is large. The above equation also implies that

$$E[|F_k(\omega)|^2] = 1, \text{ for all } \omega. \quad (25)$$

That is, condition (1) of Lemma 3 is satisfied. When the prototype combiner $\mathbf{g}_k$ is designed in the same manner, we can verify that the statistically semi-unitary property $E[\mathbf{G}_k^H \mathbf{G}_k] = \mathbf{I}_{\bar{N}_r}$ used in Lemma 2 and Theorem 2 is satisfied. From (24), we see that there is significant correlation between $F_k(p_i)$ and $F_k(p_j)$ if two AoDs are close such that $|p_i - p_j| < 2\pi/(N_t - n_{\bar{N}_t-1})$. In this case, the path gains can still be decorrelated if the AoAs are sufficiently separated, i.e., $|u_i - u_j| > 2\pi/(N_r - n_{\bar{N}_r-1})$. As long as the AoAs and AoDs are not close at the same time, we can use random beamforming and combiner to decorrelate the path gains. When two paths have close AoAs and AoDs, they can be merged as one path and there is no need of decorrelating the two. With the design in (23), the coefficients of the precoder are $1, -1$, or 0, which can be implemented using phase shifters and switches [35].

Proposed Algorithm: The proposed angle estimation using random SBT precoding is as follows. For the k-th training block, we use random precoder $\mathbf{F}_k$ and random combiner $\mathbf{G}_k$ with prototypes designed according to (23). Given N_b received training blocks $\mathbf{Y}_1, \mathbf{Y}_2, \cdots \mathbf{Y}_{N_b}$, we vectorize them to $\mathbf{y}_1, \mathbf{y}_2, \cdots \mathbf{y}_{N_b}$. The empirical autocorrelation matrix is computed as

$$\mathbf{R}_{y,e} = \frac{1}{N_b} \sum_{k=1}^{N_b} \mathbf{y}_k \mathbf{y}_k^H$$
$$= (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) \mathbf{R}_{d,e} (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r)^H + \mathbf{R}_{n,e}, \quad (26)$$

where $\mathbf{R}_{d,e} = \frac{1}{N_b} \sum_{k=1}^{N_b} \mathbf{d}_k \mathbf{d}_k^H$ and $\mathbf{R}_{n,e} = \frac{1}{N_b} \sum_{k=1}^{N_b} \mathbf{n}_k \mathbf{n}_k^H + \mathbf{n}_k \mathbf{d}_k^H (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r)^H + (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) \mathbf{d}_k \mathbf{n}_k^H$. The ith equivalent path gain is given by

$$[\mathbf{R}_{d,e}]_{i,i} = \sigma_s^2 \frac{1}{N_b} \sum_{k=1}^{N_b} |\alpha_{k,i}|^2 |G_k(u_i)|^2 |F_k(p_i)|^2.$$

We can vectorize $\mathbf{R}_{y,e}$, and apply selection matrix as in Theorem 1 and spatial smoothing as in (21) to obtain $\mathbf{Y}_{ss,e}$, the empirical version of the spatially smoothed matrix $\mathbf{Y}_{ss}$ in (22). Based on $\mathbf{Y}_{ss,e}$, 2D subspace based methods like MUSIC or ESPRIT can be applied to estimate the angles. The above procedure is summarized in Steps 1-5 of Algorithm 1.

Notice that, with finite number of training vectors, $\mathbf{R}_{d,e}$ is not diagonal and the off-diagonal entries will interfere with the estimation of angles later. The problem can be greatly alleviated if estimates of the angles are available so that the interference can be estimated and canceled. We observe that the vectorized $\mathbf{R}_{y,e}$ is given by

$$\mathbf{r}_{y,e} = (\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^*) \otimes (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) \mathbf{r}_{d,e} + \mathbf{r}_{n,e},$$

where $\mathbf{r}_{d,e} = vec(\mathbf{R}_{d,e})$. In the above expression, one of the Khatri-Rao product in (18) is replaced with Kronecker product as $\mathbf{R}_{d,e}$ is not diagonal. We can rewrite $\mathbf{r}_{y,e}$ as

$$\mathbf{r}_{y,e} = (\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) \lambda_e$$
$$+ (\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^*) \otimes (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) \lambda_{off} + \mathbf{r}_{n,e}, \quad (27)$$

where λ_e is an $L \times 1$ vector that consists of the diagonal elements of $\mathbf{R}_{d,e}$, $\lambda_{off} = vec(\mathbf{R}'_{d,e})$, and $\mathbf{R}'_{d,e}$ is the same as $\mathbf{R}_{d,e}$ except that the diagonal elements are replaced by zeros. The first term is the desired signal part, while the second is the interfering term due to the off-diagonal entries of $\mathbf{R}_{d,e}$. Using $\mathbf{r}_{y,e}$, we can obtain estimates of the angles as in Steps 1-5 of Algorithm 1. Suppose the estimates of u_i and p_i are, respectively, $\hat{u}_i$ and $\hat{p}_i$. Based on these estimates, we form the antenna response matrices $\bar{\mathbf{A}}_{r,e}$ and $\bar{\mathbf{A}}_{t,e}$. Then the least-square estimate of $\mathbf{r}_{d,e}$ can be obtained by solving the problem

$$\hat{\mathbf{r}}_{d,e} = \arg \min_{\mathbf{r}_{d,e}} ||\mathbf{r}_{y,e} - (\bar{\mathbf{A}}_{t,e} \odot \bar{\mathbf{A}}_{r,e}^*) \otimes (\bar{\mathbf{A}}_{t,e}^* \odot \bar{\mathbf{A}}_{r,e}) \mathbf{r}_{d,e}||^2.$$
$$(28)$$

We unvectorize $\hat{\mathbf{r}}_{d,e}$ back to a $L \times L$ matrix $\hat{\mathbf{R}}_{d,e}$, whose off-diagonal elements $\hat{\mathbf{R}}_{d,e}$ are now the undesired part that will cause interference. Let $\hat{\lambda}_{off} = vec(\hat{\mathbf{R}}'_{d,e})$, where $\hat{\mathbf{R}}'_{d,e}$ is the same as $\hat{\mathbf{R}}_{d,e}$ except that the diagonal elements are replaced by zeros. We can cancel the interference due to the off-diagonal entries of $\hat{\mathbf{R}}_{d,e}$ using

$$\hat{\mathbf{r}}_{y,e} = \mathbf{r}_{y,e} - (\bar{\mathbf{A}}_{t,e} \odot \bar{\mathbf{A}}_{r,e}^*) \otimes (\bar{\mathbf{A}}_{t,e}^* \odot \bar{\mathbf{A}}_{r,e}) \hat{\lambda}_{off}. \quad (29)$$

The cancellation procedure is given in Steps 6-8 of Algorithm 1. After removing the interference, we can use $\hat{\mathbf{r}}_{y,e}$ in place of $\mathbf{r}_{y,e}$ and proceed with Steps 3-5 of Algorithm 1. Due to the removal of interference, we can obtain better angles estimates. The new estimates can also be used for another iteration of interference removal, after which angle

TABLE 1. Complexity of Algorithm 1.

Operation	Complexity
Step 2	$S_2 : O(\bar{N}_t^2 \bar{N}_r^2 N_b)$
Step 5 (ESPRIT)	$S_5 : O((M_t + 1)(M_r + 1)(2L^2 + L)t) + 2L^2 (M_r(M_t + 1) + M_t(M_r + 1)) + \frac{22}{3}L^3)$
	(For signal subspace extraction in ESPRIT, orthogonal iteration with t iterations [34] [39] is used.)
Step 7	$S_7 : O(L^4 \bar{N}_t^2 \bar{N}_r^2 + L^6/3)$
Step 8	$S_8 : O(\bar{N}_t^2 \bar{N}_r^2 (L^2 - L))$
Total	$S_2 + S_5(K + 1) + K(S_7 + S_8)$, for K iterations of interference cancellation

Algorithm 1 Angle Estimation Using SBT Precoding and Combining

1. *Collect N_b received block of vectors*

$$\mathbf{Y}_k = \mathbf{G}_k^H \mathbf{A}_r \mathbf{D}_{\alpha,k} \mathbf{A}_t^H \mathbf{F}_k + \mathbf{G}_k^H \mathbf{N}_k,$$

   and form the vectorized versions $\mathbf{y}_k = vec(\mathbf{Y}_k)$ for $k = 1, 2, \cdots, N_b$.
2. *Compute the empirical autocorrelation matrix*

$$\mathbf{R}_{y,e} = \frac{1}{N_b} \sum_{k=1}^{N_b} \mathbf{y}_k \mathbf{y}_k^H$$

   and its vectorized version $\mathbf{r}_{y,e}$.
3. Apply a selection matrix $\mathbf{J}$ on $\mathbf{r}_{y,e}$ as in Theorem 1 to obtain $\tilde{\mathbf{r}}_{y,e} = \mathbf{J}\mathbf{r}_{y,e}$.
4. Apply spatial smoothing on $\tilde{\mathbf{r}}_{y,e}$ as in (21) to obtain $\mathbf{Y}_{ss,e}$.
5. Obtain estimates $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_L$ of AoAs and estimates $\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_L$ of AoDs from $\mathbf{Y}_{ss,e}$ using 2D estimation methods like MUSIC or ESPRIT.
   ————————interference cancellation————————
6. Use the estimates obtained in Step. 5 to form antenna response matrices $\bar{\mathbf{A}}_{t,e}$ and $\bar{\mathbf{A}}_{r,e}$. The i-th column of $\bar{\mathbf{A}}_{t,e}$ is given by $\begin{bmatrix} 1 & e^{j\hat{p}_i n_1} & \cdots & e^{j\hat{p}_i n_{\bar{N}_t - 1}} \end{bmatrix}^T$ and the i-th column of $\bar{\mathbf{A}}_{r,e}$ by $\begin{bmatrix} 1 & e^{j\hat{u}_i m_1} & \cdots & e^{j\hat{u}_i m_{\bar{N}_t - 1}} \end{bmatrix}^T$.
7. Compute the least square estimate $\hat{\mathbf{r}}_{d,e} = (\mathbf{T}^H \mathbf{T})^{-1} \mathbf{T}^H \mathbf{r}_{y,e}$, where $\mathbf{T} = (\bar{\mathbf{A}}_{t,e} \odot \bar{\mathbf{A}}_{r,e}^*) \otimes (\bar{\mathbf{A}}_{t,e}^* \odot \bar{\mathbf{A}}_{r,e})$.
8. Cancel the interference due to the off-diagonal entries of $\hat{\mathbf{R}}_{d,e}$ using

$$\hat{\mathbf{r}}_{y,e} = \mathbf{r}_{y,e} - (\bar{\mathbf{A}}_{t,e} \odot \bar{\mathbf{A}}_{r,e}^*) \otimes (\bar{\mathbf{A}}_{t,e}^* \odot \bar{\mathbf{A}}_{r,e}) \hat{\boldsymbol{\lambda}}_{off}.$$

9. Set $\mathbf{r}_{y,e} = \hat{\mathbf{r}}_{y,e}$ and repeat Steps 3-5.

estimation can be performed again. Table 1 gives the order of complexity for the steps in Algorithm 1 that require computations. In Table 1, ESPRIT is used for Step 5 and the order of complexity is obtained from [34], [39].

Single RF Chain: The proposed method can not be directly applied if the transmitter or receiver has only one RF chain.

For example, when there is only one transmit RF chain, AoD can not be estimated. One way around this is to use extra training time to compensate for a single RF chain. For example, suppose there is only on transmit RF chain. We would like the transmitter to behave as if there were $\bar{N}_t$ RF chains and the hypothetical precoder $\mathbf{F}_k$ is $N_t \times \bar{N}_t$. With the training input block $\mathbf{S} = \mathbf{I}_{\bar{N}_t}$, the ith transmitter output vector is precisely the ith column vector of $\mathbf{F}_k$. For the first training vector of the kth block, the first column vector $\mathbf{f}_{k,0}$ of $\mathbf{F}_k$ is used as the actual beamformer. For the next training vector, the next column vector $\mathbf{f}_{k,1}$ is used as the actual beamformer. Preceding in this manner $\bar{N}_t$ times, we have $\bar{N}_t$ transmit output vectors that are equal to the output vectors of the hypothetical transmitter $\mathbf{F}_k$ with $\bar{N}_t$ RF chains. Thus, by using extra $\bar{N}_t - 1$ training vectors, we can obtain a transmitter that acts as if there were $\bar{N}_t$ RF chains. We can use the same trick too when the receiver has only one RF chain.

V. SIMULATIONS

Consider the channel model in (1) with antennas spaced by half wavelength. We use 10^4 channel realizations in the Monte Carlo evaluation. The channel is time-varying in Example 2 and time-invariant in the rest of the examples. The following setting is used, unless otherwise indicated. The numbers of transmit and receive antennas are $N_t = N_r = 64$, the numbers of RF chain are $\bar{N}_t = \bar{N}_r = 4$, and 100 training blocks are used in the simulation. The AoA and AoD are uniformly distributed over $[-\pi, \pi]$. In Algorithm 1, we use 2D ESPRIT to estimate the angles and four iterations of interference cancellation are applied.

Example 1: With the use of random beamforming, the beam pattern is different for every training block. Upon collecting N_b training blocks, we can compute the equivalent path gains using the empirical autocorrelation matrix given in (26). Suppose the actual path gains are time-invariant $\alpha_{k,i} = \alpha_i$ and the combining prototype is fixed for all training blocks. The ith equivalent path gain is given by $\sigma_s^2 |\alpha_i|^2 |G_1(u_i)|^2 F_{ave}(p_i)$, where

$$F_{ave}(\omega) = 1/N_b \sum_{k=1}^{N_b} |F_k(\omega)|^2$$

is the average beam pattern at the transmitter and $F_k(\omega)$ denotes the discrete-time Fourier transform of the prototype
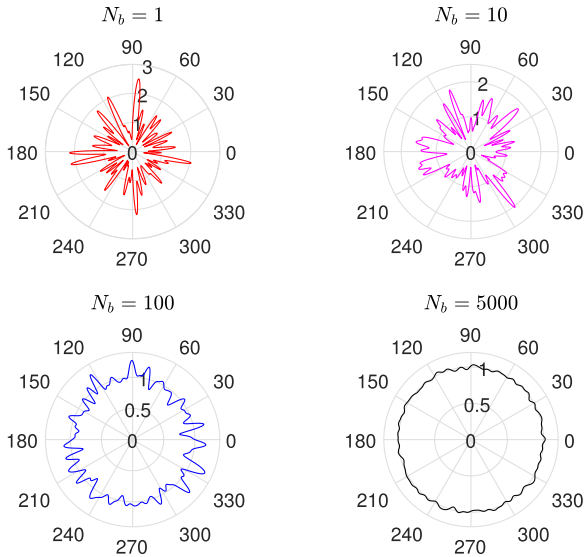
FIGURE 3. Average beam pattern of the transmit beamformer.

(a)

(b)

FIGURE 4. MSE as a function of the number of training blocks N_b for (a) AoA and (b) AoD.

for the kth training block. Fig. 3 plots the average beam pattern $F_{ave}(\omega)$ when the number of training blocks N_b is 1, 10, 100, and 5000. The average beam pattern approaches a constant as N_b increases. This is consistent with the fact that with binary random coefficients, the prototype satisfies $E[|F_k(\omega)|^2] = 1$ for all ω, as given in (25).

Example 2: In this example, we examine the case when the number of training blocks N_b is finite. For a time-varying channel, Fig. 4 plots the mean squared error as a function of N_b for different SNR $\sigma_s^2/\sigma_n^2 = -10$ dB, 0 dB, and 10 dB. The mean squared error $\frac{1}{L}\sum_{i=1}^{L} E[(\hat{p}_i - p_i)^2]$ for AoD and $\frac{1}{L}\sum_{i=1}^{L} E[(\hat{u}_i - u_i)^2]$ for AoA are computed for $L = 2$ paths, where $p_i = \pi\cos\theta_i^t$ and $u_i = \pi\cos\theta_i^r$. We have used $\theta_1^t = 20°$, $\theta_2^t = 40°$, and $\theta_1^r = -20°$, and $\theta_2^r = -40°$. We see that the error decreases steadily with N_b. Good estimates can be obtained even when there is only a moderate number of training vectors. For example, when $N_b = 100$, the error is around -40dB for SNR = 0 dB. As a benchmark, the CRB [8] of the estimation error has been plotted. The estimation error is close to the CRB except when the SNR is low and the number of training blocks is small. Therefore the CRB provides an accurate characterization of the estimation error.

Example 3: For time-invariant channels, the equivalent path gains can be approximately decorrelated by properly designing the random prototype when N_t is large. In this example we examine the performance as a function of N_t when the beamforming prototype have random coefficients as in (23) while the combining prototype is a deterministic vector $\mathbf{g}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ with $N_r = 64$. Fig. 5(a) shows the normalized mean squared error (NMSE) of the channel and SNR $\sigma_s^2/\sigma_n^2 = 5$ dB. The NMSE of the channel is defined as $E[||\hat{\mathbf{H}} - \mathbf{H}||_F^2/||\mathbf{H}||_F^2]$, where $\hat{\mathbf{H}}$ is the estimated channel matrix. The NMSE is shown for $\bar{N}_r = \bar{N}_t = 4, 5,$ and 6.
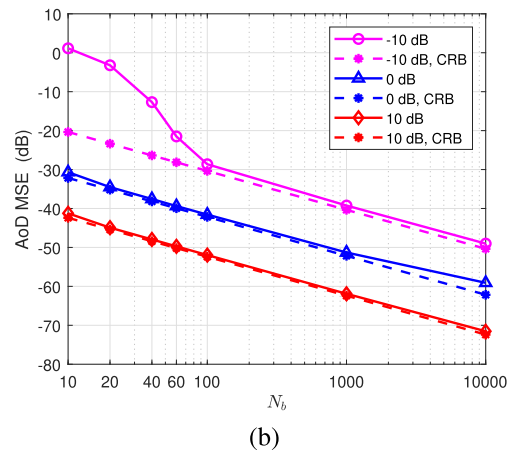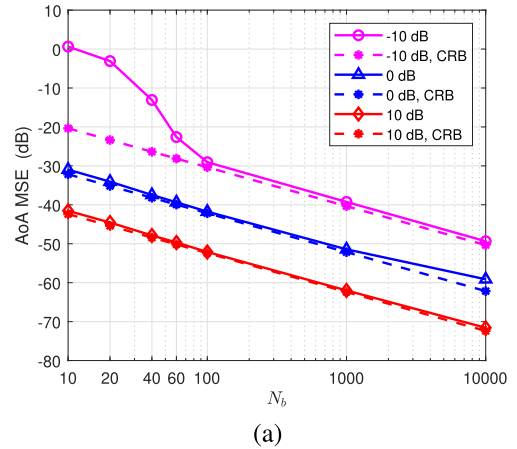
The set $\mathcal{S}_t$, taken from the table for minimum redundancy array in [36], are $\{0, 1, 4, 6\}$, $\{0, 1, 4, 7, 9\}$, and $\{0, 1, 4, 5, 11, 13\}$ for $\bar{N}_t = 4, 5$ and 6, respectively. We can see that the error reduces with N_t and $\bar{N}_t$. Without interference cancellation (IC), the performance for different $\bar{N}_t$ can be close for small N_t. For example, when $N_t = 16$, there is no difference between $\bar{N}_t = 5$ and $\bar{N}_t = 6$. This can be understood as follows. The prototype has $N_t - n_{\bar{N}_t-1}$ coefficients. When $N_t = 16$, the number of nonzero coefficients are 7 and 3 for $\bar{N}_t = 5$ and $\bar{N}_t = 6$, respectively. When the prototype has few nonzero coefficients, the gain from enlarged virtual arrays can be offset by the loss due to a lack of design freedom for decorrelating the paths. But when interference cancellation is applied, increasing $\bar{N}_t$ still lowers the NMSE for a small N_t. This is because interference cancellation reduces the interfering off-diagonal terms of $\mathbf{R}_{d,e}$, i.e., decorrelating the equivalent path gains.

Example 4: For a finite number of training vectors, the empirical autocorrelation matrix of the effective path gains $\mathbf{R}_{d,e}$ is not diagonal, which causes interference in angle estimation. In Fig. 6, we examine the usefulness of interference cancellation when the precoder and combiner prototypes are random, as designed in (23). The NMSE
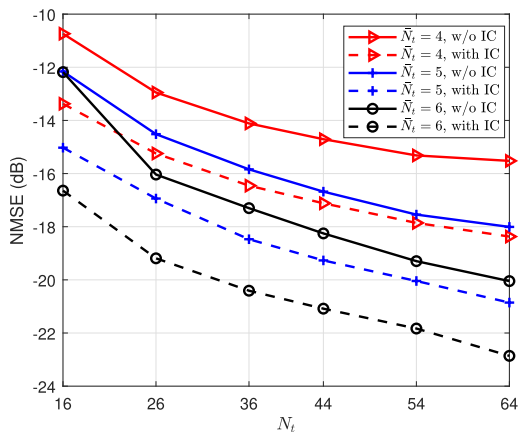
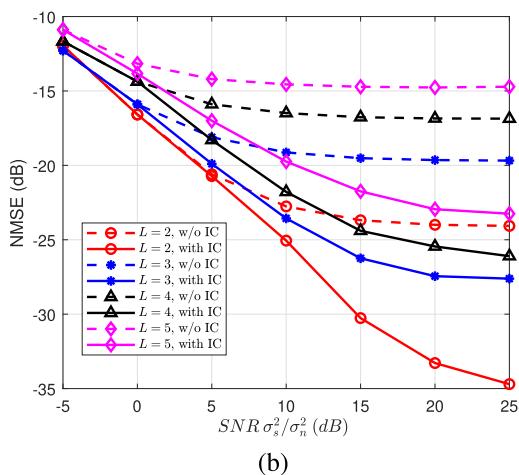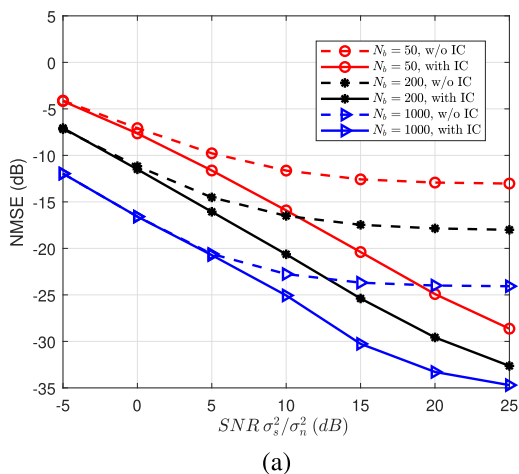**FIGURE 5.** NMSE of the channel as a function of N_t.

(a)

(b)

FIGURE 6. NMSE with and without interference cancellation (IC), (a) for different N_b when $L = 2$ and (b) for different L when $N_b = 1000$.

is plotted with and without interference cancellation for $L = 2$ in Fig. 6(a). The effect of interference cancellation is particularly significant for high SNR range when the number of training blocks is small, e.g., $N_b = 50$. For high SNR, the noise is small and the remaining path correlation

plays a major role in angle estimation. With cancellation, the interference is largely removed and the estimation accuracy greatly improved. Fig. 6(b) shows the NMSE for different number of paths L when $N_b = 1000$. As L increases, the $L \times L$ empirical autocorrelation matrix $\mathbf{R}_{d,e}$ has more off-diagonal terms and thus more interfering terms in angle estimation. A satisfactory NMSE can be achieved even for a larger L when interference cancellation is applied.

Example 5: When two angles are close, they may not be resolved correctly. Consider two AoAs, $\theta_1^r = 30° - \Delta\theta$ and $\theta_2^r = 30° + \Delta\theta$. The two AoAs are considered successfully identified if $|\hat{\theta}_i^r - \theta_i^r| < \Delta\theta$, for $i = 1, 2$ [9]. Fig. 7 shows the probability of resolution, i.e., the probability when the two AoAs are successfully identified for SNR = 5 dB. The AoDs are chosen as $\theta_1^t = 15°$, and $\theta_2^t = 45°$. We plot the probability as a function of $\Delta\theta$. When $\Delta\theta$ is 1°, the two AoAs can be successfully identified with high probability, around .86, .93, and .95, respectively for $N_b = 128, 256$, and 1000.
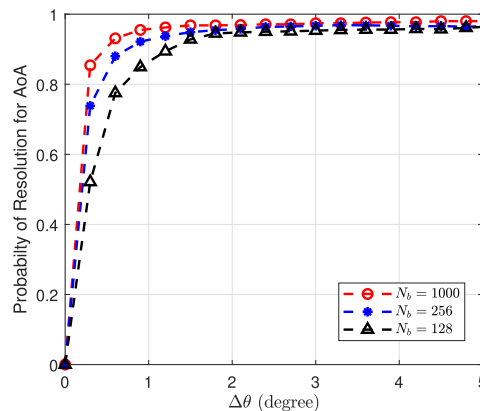

FIGURE 7. Probability of Resolution for $N_b = 100, 256, 1000$.

Example 6: In the data transmission to be followed after channel estimation, the precoder and combiner are designed according to the estimated channel, which results in a degradation. Fig. 8(a) shows the achievable rate, which is given by

$$\mathcal{R} = \log_2(\det(\mathbf{I}_{\bar{N}_r} + \rho \mathbf{R}_g^{-1} \mathbf{G}^H \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{G})),$$

where $\rho = \sigma_s^2/\sigma_n^2$ and $\mathbf{R}_g = \mathbf{G}^H \mathbf{G}$. The result is compared with [15], [19] and [20] for $L = 3$ paths. In [15], a multi-resolution training codebook of beamforming vectors is used for multi-stage closed loop angle estimation; feedback is required before the training of the next stage. In Fig. 8 four stages, i.e., three feedbacks, are used and in each stage the resolution is increased by a factor of four, resulting in a total of 432 training blocks. For [19], [20], the angles are estimated using compressive sensing, OMP in [19] and generalized block OMP in [20]. For both compressive sensing curves, the dictionary size is 100 for AoA and AoD. The resolution increases with the dictionary sizes, and so does the complexity, which increases with the product of dictionary sizes. The number of training blocks used is 256 for [19], [20]

and the proposed method. The proposed 'SBT' is very close to 'Perfect CSI' in middle and high SNR range. It enjoys a much lower complexity compared to compressive sensing based methods. In terms of complexity, it is in the order of $1.8 * 10^5$ for 'SBT' (computed from Table 1), $6.4 * 10^7$ for [15], and $4.1 * 10^7$ for both [19] and [20] (obtained from [38]). Fig. 8(b) shows the NMSE comparison of the systems considered in Fig. 8(a). We can see that a small estimation error can be achieved with 'SBT'.
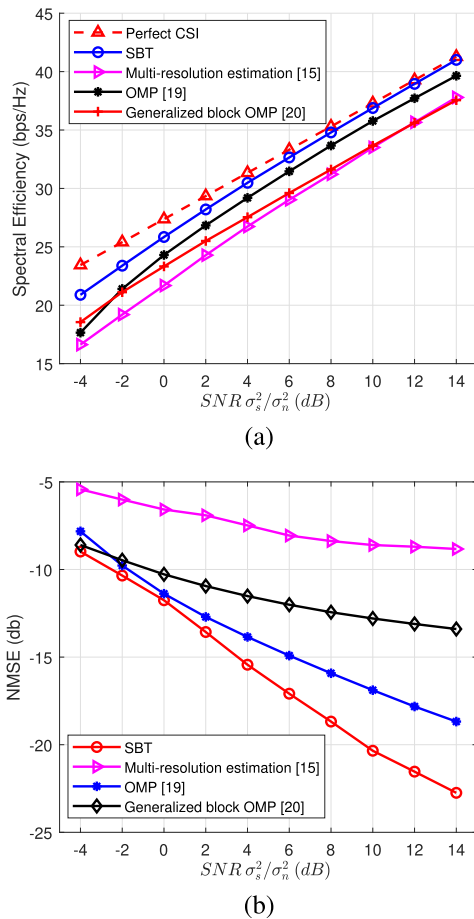

(a)

(b)

FIGURE 8. (a) Achievable rate performance and (b) NMSE performance.

VI. CONCLUSION

In this article we consider the estimation of AoA and AoD for mmWave massive MIMO communication systems using SBT precoders and combiners. By designing SBT precoder and combiner based on the spacing of sparse arrays, we can construct augmented virtual ULA and reduce estimation error. Due to random beamforming, the equivalent path gains can be decorrelated even if the channel is time-invariant. Complete decorrelation of different paths requires arrays of infinite sizes and infinitely many training vectors. However, with the aid of interference cancellation, simulations show that the proposed method achieves high estimation accuracy even for a finite number of training vectors and moderate antenna sizes.

APPENDIX A
PROOF OF THEOREM 1

The expression of $\mathbf{r}_y$ in the AoA-only case involves only the Khatri-Rao product of two antenna response matrices and we can apply a selection matrix to obtain a ULA matrix. In the general case, eq. (18) involves the Khatri-Rao product of four antenna response matrices $\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r$. We would like to switch the ordering so that $\bar{\mathbf{A}}_r$ and $\bar{\mathbf{A}}_r^*$ are next to each other, and $\bar{\mathbf{A}}_t$ and $\bar{\mathbf{A}}_t^*$ are also next to each other. This will allow us to exploit the results from sparse arrays directly. The order of the four matrices in the product can be changed with proper permutation. The following fact is useful for constructing such a permutation matrix.

Fact 1: Given two vectors $\mathbf{a}$ and $\mathbf{b}$, the Khatri-Rao products $\mathbf{a} \odot \mathbf{b}$ and $\mathbf{b} \odot \mathbf{a}$ are permutation equivalent [30]; one can be obtained from the other through a permutation. That is, $\mathbf{a} \odot \mathbf{b} = \mathbf{P}(\mathbf{b} \odot \mathbf{a})$, where $\mathbf{P}$ is a permutation matrix. Let $\mathbf{a}$ be $M \times 1$ and $\mathbf{b}$ be $N \times 1$. Then the $(M\ell + k)$-th row of $\mathbf{P}$ is $\mathbf{e}_{\ell+kN}^T$ for $\ell = 0, 1, \cdots, N-1$, and $k = 0, 1, \cdots, M-1$, where $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_{MN}$.

The noise vector $\mathbf{q}$ can be obtained by directly applying $\mathbf{J}$ on $\mathbf{r}_n$, so we consider only the signal term in the following. Using Fact 1, we can construct a permutation matrix $\bar{\mathbf{P}}_1$ such that $\bar{\mathbf{P}}_1(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r) = \bar{\mathbf{A}}_r \odot \bar{\mathbf{A}}_t^*$. Let $\mathbf{P}_1 = \mathbf{I}_{\bar{N}_t \bar{N}_r} \otimes \bar{\mathbf{P}}_1$ and $\mathbf{r}_1 = \mathbf{P}_1 \mathbf{r}_y$. Using the property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \odot \mathbf{D}) = (\mathbf{AC}) \odot (\mathbf{BD})$ and the expression of $\mathbf{r}_y$ in (18), the vector $\mathbf{r}_1$ can be written as $\mathbf{r}_1 = (\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r \odot \bar{\mathbf{A}}_t^*) \boldsymbol{\lambda}$. Now treating $\mathbf{r}_1$ as $((\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r) \odot \bar{\mathbf{A}}_t^*) \boldsymbol{\lambda}$, we can use Fact 1 again to construct a permutation matrix $\mathbf{P}_2$ so that $\mathbf{r}_2 = \mathbf{P}_2 \mathbf{r}_1$ is given by

$$\mathbf{r}_2 = (\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r) \boldsymbol{\lambda}. \qquad (30)$$

Let

$$\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2, \qquad (31)$$

then we have $\mathbf{r}_2 = \mathbf{P} \mathbf{r}_y$. Now the four matrices in the above expression of $\mathbf{r}_2$ are in the desired order.

As in the single transmit RF chain case, we can apply a selection matrix $\mathbf{J}_r$ on $\bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r$ so that $\tilde{\mathbf{A}}_r = \mathbf{J}_r(\mathbf{A}_r^* \odot \mathbf{A}_r)$ is an antenna response matrix corresponding to a $(2M_r + 1)$-element ULA. Similarly, we can apply a selection matrix $\mathbf{J}_t$ on $\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t$ so that $\tilde{\mathbf{A}}_t = \mathbf{J}_t(\mathbf{A}_t^* \odot \mathbf{A}_t)$ is an antenna response matrix corresponding to a $(2M_t + 1)$-element ULA. Let us apply the selection matrix $\mathbf{J}_t \otimes \mathbf{J}_r$ on $\mathbf{r}_2$, then

$$\begin{aligned}(\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{r}_2 &= (\mathbf{J}_t \otimes \mathbf{J}_r)(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r)\boldsymbol{\lambda} \\ &= (\mathbf{J}_t(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t)) \odot (\mathbf{J}_r(\bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r))\boldsymbol{\lambda} \\ &= (\tilde{\mathbf{A}}_t \odot \tilde{\mathbf{A}}_r)\boldsymbol{\lambda}.\end{aligned}$$

Let

$$\mathbf{J} = (\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{P}, \qquad (32)$$

where $\mathbf{P}$ is as in (31), then $\mathbf{J}$ is a selection matrix. Then, the signal part of $\tilde{\mathbf{r}}_y$ is

$$\mathbf{J}\mathbf{r}_y = (\tilde{\mathbf{A}}_t \odot \tilde{\mathbf{A}}_r)\boldsymbol{\lambda},$$

as given in Theorem 1.

APPENDIX B
PROOF OF LEMMA 2

As the noise $\mathbf{n}_k$ is uncorrelated, $\mathbf{r}_n = \sigma_n^2 \tilde{\mathbf{1}}$, where $\tilde{\mathbf{1}} = vec(\mathbf{I}_{\bar{N}_t \bar{N}_r})$. We have $\mathbf{q} = \mathbf{J}\mathbf{r}_n = \sigma_n^2 \mathbf{J}\tilde{\mathbf{1}}$. Observe that $\tilde{\mathbf{1}}$ contains only $\bar{N}_t \bar{N}_r$ identical nonzero coefficients. We can obtain $\mathbf{q}$ by tracking the nonzero coefficients upon the application of $\mathbf{J}$. The matrix $\mathbf{J}$ in (32) contains two parts, a selection matrix $\mathbf{J}_t \otimes \mathbf{J}_r$ and a permutation matrix $\mathbf{P}$, which in turns consists of permutation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. We consider the application of $\mathbf{J}_t \otimes \mathbf{J}_r$ and the two permutations separately.

Application of permutations. Let $\mathbf{r}_2 = \mathbf{P}\mathbf{r}_y$, we have

$$\mathbf{r}_2 = \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r\right)\boldsymbol{\lambda} + \sigma_n^2 \mathbf{P}\tilde{\mathbf{1}}. \quad (33)$$

Notice that $\tilde{\mathbf{1}}$ is the stacking of $\bar{N}_t \bar{N}_r$ standard vectors $\mathbf{e}_i$, for $i = 0, 1, \cdots \bar{N}_t \bar{N}_r - 1$. For each i, we can find unique $0 \le j_i \le \bar{N}_t - 1$ and $0 \le k_i \le \bar{N}_r - 1$ such that

$$i = j_i \bar{N}_r + k_i.$$

Furthermore there is a one-to-one correspondence between i and (j_i, k_i). Upon the first permutation, the location of the '1's move and we can verify that

$$\mathbf{P}_1 \tilde{\mathbf{1}} = \begin{bmatrix} \mathbf{e}_{\ell_0}^T & \mathbf{e}_{\ell_1}^T & \cdots & \mathbf{e}_{\ell_{\bar{N}_t \bar{N}_r - 1}}^T \end{bmatrix}^T, \quad (34)$$

where $\ell_i = k_i \bar{N}_t + j_i$. To consider the second permutation, let $\mathbf{q}_1 = \mathbf{P}_2 \mathbf{P}_1 \tilde{\mathbf{1}}$. It can be shown that the nonzero coefficient in $\mathbf{e}_{\ell_i}$ will move to the ℓ_i-th entry of $\mathbf{q}_1$, where $\ell_i = j_i \bar{N}_t \bar{N}_r^2 + i\bar{N}_r + k_i$. Using the expression $i = j_i N_r + k_i$, we have $\ell_i = \bar{N}_r^2 j_i (\bar{N}_t + 1) + k_i (\bar{N}_r + 1)$. The nonzero coefficients in $\mathbf{q}_1$ are located at $\bar{N}_r^2 m(\bar{N}_t + 1) + k(\bar{N}_r + 1)$ for $0 \le m \le \bar{N}_t - 1$ and $0 \le k \le \bar{N}_r - 1$.

Application of $\mathbf{J}_t \otimes \mathbf{J}_r$. The noise term in (19) is $\mathbf{q} = (\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{q}_1$. As the selection matrix $\mathbf{J}_t \otimes \mathbf{J}_r$ is designed according to the signal part, we turn our attention to the signal plus noise vector $\tilde{\mathbf{r}}_y$. Using the expression of $\mathbf{r}_2$ in (33) and $\tilde{\mathbf{r}}_y = (\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{r}_2$, we get

$$\tilde{\mathbf{r}}_y = (\mathbf{J}_t \otimes \mathbf{J}_r)\left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r\right)\boldsymbol{\lambda} + (\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{q}_1.$$

To consider the application of $\mathbf{J}_t \otimes \mathbf{J}_r$, let us unvectorize $\mathbf{r}_2$ and shape it to an $\bar{N}_r^2$ by $\bar{N}_t^2$ matrix,

$$\mathbf{R}_2 = \left(\bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r\right) \mathbf{R}_d \left(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t\right)^T + \mathbf{Q}_1, \quad (35)$$

where $\mathbf{Q}_1$ is an $\bar{N}_r^2$ by $\bar{N}_t^2$ matrix obtained by unvectorizing $\mathbf{q}_1$. Using $\tilde{\mathbf{r}}_y = (\mathbf{J}_t \otimes \mathbf{J}_r)\mathbf{r}_2$, we can unvectorize $\tilde{\mathbf{r}}_y$ and shape it to the $(2M_r + 1)$ by $(2M_t + 1)$ matrix $\tilde{\mathbf{R}}_y = \mathbf{J}_r \mathbf{R}_2 \mathbf{J}_t$. It follows that

$$\tilde{\mathbf{R}}_y = \mathbf{J}_r(\bar{\mathbf{A}}_r^* \odot \bar{\mathbf{A}}_r)\mathbf{R}_d \left(\mathbf{J}_t(\bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_t)\right)^T + \mathbf{Q}, \quad (36)$$

where $\mathbf{Q} = \mathbf{J}_r \mathbf{Q}_1 \mathbf{J}_t$. The noise vector $\mathbf{q}$ in (19) can be obtained by vectorizing $\mathbf{Q}$.

We can verify that when we unvectorize $\mathbf{q}_1$, the nonzero unity coefficients in $\mathbf{q}_1$ move to $(k(\bar{N}_r + 1), m(\bar{N}_t + 1))$-th element of $\mathbf{Q}_1$ for $k = 0, 1, \cdots, \bar{N}_r - 1$, and $m = 0, 1, \cdots, \bar{N}_t - 1$. Notice that the above row index $k(\bar{N}_r + 1)$

for $k = 0, 1, \cdots, \bar{N}_r - 1$ are also the indices of the rows of $\bar{\mathbf{A}}_r \odot \bar{\mathbf{A}}_r^*$ whose elements are all equal to unity. When we apply the selection matrix $\mathbf{J}_r$ on $\bar{\mathbf{A}}_r \odot \bar{\mathbf{A}}_r^*$ we keep only one row that corresponds to the unity rows of $\bar{\mathbf{A}}_r \odot \bar{\mathbf{A}}_r^*$. Similarly, the column index $m(\bar{N}_t + 1)$ in $(k(\bar{N}_r + 1), m(\bar{N}_t + 1))$ for $m = 0, 1, \cdots, \bar{N}_t - 1$ are also the indices of the rows of $\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_t^*$ whose elements are all equal to unity. When we apply the selection matrix $\mathbf{J}_t$ on $\bar{\mathbf{A}}_t \odot \bar{\mathbf{A}}_t^*$ we keep only one row that corresponds to the unity rows. Therefore we are left with only one nonzero coefficient in the noise term. Now we only need to determine where the nonzero coefficient is in the matrix $\mathbf{Q}$. As the unity row of $\tilde{\mathbf{A}}_t$ is in the middle of the matrix, i.e., M_t-th row, and the unity row of $\tilde{\mathbf{A}}_r$ is also in the middle of the matrix, i.e., M_r-th row, the nonzero coefficient is moved to the (M_r, M_t)-th entry. Unvectorizing $\mathbf{Q}$, we obtain $\mathbf{q}$ as stated in the Theorem.

APPENDIX C
PROOF OF THEOREM 2

The signal part of $\mathbf{Y}_{ss}$ can be obtained using the result from [27]. We will consider the noise part only. Let $\mathbf{q}_{i,j} = \mathbf{J}_{i,j}\mathbf{q}$, where $\mathbf{J}_{i,j}$ is as defined in (20). Then the noise part of $\mathbf{Y}_{ss}$ is

$$\begin{bmatrix} \mathbf{q}_{0,0} & \mathbf{q}_{0,1} & \cdots & \mathbf{q}_{0,M_r} & \mathbf{q}_{1,0} & \cdots & \mathbf{q}_{M_t,M_r} \end{bmatrix}. \quad (37)$$

We can obtain the expression in (22) if we can show that the column vectors in the above matrix are $\sigma_n^2 \mathbf{e}_0, \sigma_n^2 \mathbf{e}_1, \sigma_n^2 \mathbf{e}_2, \cdots$, where $\mathbf{e}_i$ is the i-th column of the identify matrix $\mathbf{I}_{(M_t+1)(M_r+1)}$. Notice that $\mathbf{q}_{i,j} = (\mathbf{J}_i^t \otimes \mathbf{J}_j^r)\mathbf{q}$ can also be expressed as

$$\mathbf{q}_{i,j} = vec(\mathbf{Q}_{i,j}),$$

where $\mathbf{Q}_{i,j} = \mathbf{J}_j^r \mathbf{Q}(\mathbf{J}_j^t)^T$ and $\mathbf{Q}$ is the $(2M_r + 1) \times (2M_t + 1)$ unvectorized version of $\mathbf{q}$ that is given in (36). As shown in the proof of Lemma 2, the matrix $\mathbf{Q}$ has only one nonzero coefficient, right at the center of the matrix. The expression $\mathbf{Q}_{i,j} = \mathbf{J}_j^r \mathbf{Q}(\mathbf{J}_j^t)^T$ means that $\mathbf{Q}_{i,j}$ is a submatrix of $\mathbf{Q}$. An example of $\mathbf{Q}$ for $M_t = M_r = 2$ is as shown in Fig. 9. The vector $\mathbf{q}_{0,0}$ is the vectorized version of $\mathbf{Q}_{0,0}$, which is shown in the figure, and $\mathbf{q}_{0,1}$ is the vectorized version of $\mathbf{Q}_{0,1}$, also shown. Each $\mathbf{Q}_{i,j}$ contains the nonzero entry of $\mathbf{Q}$ and thus each $\mathbf{q}_{i,j}$ is a standard vector scaled by σ_n^2. In particular, $\mathbf{q}_{0,0} = \sigma_n^2 \mathbf{e}_0$, and $\mathbf{q}_{0,1} = \sigma_n^2 \mathbf{e}_1$. Notice that upon vectorization
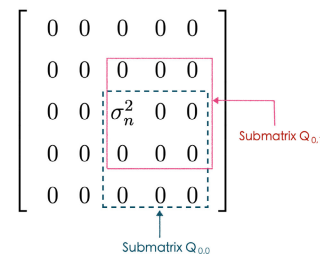

FIGURE 9. An example of $\mathbf{Q}$ for $M_t = M_r = 2$.

of $\mathbf{Q}_{i,j}$, the location of the nonzero entry in $\mathbf{q}_{i,j}$ moves down by one when j increases by one and the location of the nonzero entry moves down by $M_r + 1$ when i increases by one. Therefore the matrix in (37) is equal to $\sigma_n^2 \mathbf{I}_{(M_r+1)(M_t+1)}$.

APPENDIX D
PROOF OF (24)

Using the beamforming prototype in (23), we have $F_k(\omega) = \gamma \sum_{n=0}^{N_t-1-n_{\bar{N}_t-1}} \beta_{k,n} e^{-jn\omega}$. Then

$$
\begin{aligned}
& E[F_k(p_i)F_k^*(p_j)] \\
&= E\left[\gamma^2 \sum_{n=0}^{N_t-1-n_{\bar{N}_t-1}} \beta_{k,n} e^{-jnp_i} \sum_{m=0}^{N_t-1-n_{\bar{N}_t-1}} \beta_{k,m} e^{jmp_j} \right] \\
&= \gamma^2 \sum_{n=0}^{N_t-1-n_{\bar{N}_t-1}} \sum_{m=0}^{N_t-1-n_{\bar{N}_t-1}} E[\beta_{k,n}\beta_{k,m}] e^{-j(np_i-mp_j)}.
\end{aligned}
$$

As $E[\beta_{k,n}\beta_{k,m}]$ is equal to 1 for $n = m$ and it is equal to 0 for $n \neq m$, we have

$$
E[F_k(p_i)F_k^*(p_j)] = \gamma^2 \sum_{n=0}^{N_t-1-n_{\bar{N}_t-1}} e^{-jn(p_i-p_j)}.
$$

Using the above expression, we can obtain $|E[F_k(p_i)F_k^*(p_j)]|$ as given in (24).

REFERENCES

[1] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[3] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, Eds., *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[4] 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.

[5] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[6] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[7] C.-L. Liu and P. P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1438–1442, Sep. 2015.

[8] C.-L. Liu and P. P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digit. Signal Process.*, vol. 61, pp. 43–61, Feb. 2017.

[9] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2017.

[10] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[11] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[12] A. Liao, Z. Gao, Y. Wu, H. Wang, and M.-S. Alouini, "2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding," *IEEE Access*, vol. 5, pp. 24747–24757, Nov. 2017.

[13] 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.

[14] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

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

[16] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, Jul. 2017.

[17] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.

[18] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[19] 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.

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

[21] J. Zhang and M. Haardt, "Channel estimation for hybrid multi-carrier mmWave MIMO systems using three-dimensional unitary ESPRIT in DFT beamspace," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, Dec. 2017, pp. 1–5.

[22] D. Zhang, Y. Wang, Z. Su, and W. Xiang, "Millimeter wave channel estimation based on subspace fitting," *IEEE Access*, vol. 6, pp. 76126–76139, 2018.

[23] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based On 2-D Beamspace MUSIC method, *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.

[24] 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.

[25] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[26] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[27] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, Sep. 1992.

[28] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[29] A. Koochakzadeh and P. Pal, "Cramér–Rao bounds for underdetermined source localization," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 919–923, Jul. 2016.

[30] H. V. Henderson and S. R. Searle, "The vec-permutation matrix, the VEC operator and Kronecker products: A review," *Linear Multilinear Algebra*, vol. 9, no. 4, pp. 271–288, Jan. 1981.

[31] A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.

[32] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 4, pp. 817–827, Aug. 1984.

[33] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb. 1996.

[34] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1989.

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

[36] H. L. Van Trees, *Optimum Array Processing: Part IV Detection, Estimation, Modulation Theory*. Hoboken, NJ, USA: Wiley, 2002.

[37] J. Steinwandt, F. Roemer, M. Haardt, and G. D. Galdo, "Performance analysis of multi-dimensional ESPRIT-type algorithms for arbitrary and strictly non-circular sources with spatial smoothing," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2262–2276, May 2017.

[38] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.

[39] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

YUAN-PEI LIN (Senior Member, IEEE) was born in Taipei, Taiwan, in 1970. She received the B.S. degree in control engineering from National Chiao Tung University, Taiwan, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, in 1993 and 1997, respectively.

She joined the Department of Electrical and Control Engineering, National Chiao Tung University, in 1997. Her research interests include wireless communications and signal processing for digital communications.

She has coauthored two books *Signal Processing and Optimization for Transceiver Systems* and *Filter Bank Transceivers for OFDM and DMT Systems* (Cambridge University Press, 2010). She was a recipient of the Ta-You Wu Memorial Award, in 2004. She has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, the *EURASIP Journal on Applied Signal Processing*, and *Multidimensional Systems and Signal Processing* (Academic Press). She was a Distinguished Lecturer of the IEEE Circuits and Systems Society, from 2006 to 2007.

TING-MING YANG (Graduate Student Member, IEEE) was born in Taoyuan, Taiwan, in 1995. He received the B.S. degree in communication engineering from National Central University, Taoyuan, in 2018. He is currently pursuing the M.S. degree with the Department of Electrical Engineering, National Chiao Tung University, Taiwan. His research interests include signal processing for digital communications and wireless communications.

● ● ●