Reliable Beamspace Channel Estimation for Millimeter-Wave Massive MIMO Systems with Lens Antenna Array

Xinyu Gao, *Student Member, IEEE*, Linglong Dai, *Senior Member, IEEE*, Shuangfeng Han, *Member, IEEE*, Chih-Lin I, *Senior Member, IEEE*, and Xiaodong Wang, *Fellow, IEEE*

Abstract—Millimeter-wave (mm-wave) massive MIMO with lens antenna array can considerably reduce the number of required radio-frequency (RF) chains by beam selection. However, beam selection requires the base station to acquire the accurate information of beamspace channel. This is a challenging task as the size of beamspace channel is large, while the number of RF chains is limited. In this paper, we investigate the beamspace channel estimation problem in mm-wave massive MIMO systems with lens antenna array. Specifically, we first design an adaptive selecting network for mm-wave massive MIMO systems with lens antenna array, and based on this network, we further formulate the beamspace channel estimation problem as a sparse signal recovery problem. Then, by fully utilizing the structural characteristics of the mm-wave beamspace channel, we propose a support detection (SD)-based channel estimation scheme with reliable performance and low pilot overhead. Finally, the performance and complexity analyses are provided to prove that the proposed SD-based channel estimation scheme can estimate the support of sparse beamspace channel with comparable or higher accuracy than conventional schemes. Simulation results verify that the proposed SD-based channel estimation scheme outperforms conventional schemes and enjoys satisfying accuracy even in the low SNR region as the structural characteristics of beamspace channel can be exploited.

Index Terms—Massive MIMO, millimeter-wave communications, lens antenna array, beamspace channel estimation.

I. INTRODUCTION

MILLIMETER-WAVE (mm-wave) massive multiple-input multiple-output (MIMO) has been considered as a key technique for future 5G wireless communications [1], since it can achieve significant increase in data rates due to its wider bandwidth [2] and higher spectral efficiency [3]. However, realizing mm-wave massive MIMO in practice is not a trivial task. One key challenging problem is that each antenna in MIMO systems usually requires one dedicated radio-frequency (RF) chain (including digital-to-analog converter, up converter, etc.) [4]. This results in unaffordable hardware cost and energy consumption in mm-wave massive MIMO systems, as the number of antennas becomes huge (e.g., 256 antennas) [1], and the energy consumption of RF chain is high (e.g., about 250 mW per RF chain at mm-wave frequencies, compared with 30 mW per RF chain at cellular frequencies) [5]. If we consider the base station (BS) in a typical mm-wave massive MIMO system with 256 antennas, only the RF chains will consume 64 Watts, which is much higher than the energy consumption of current 4G micro-cell BS (several Watts) [1]. To reduce the number of required RF chains, mm-wave massive MIMO with lens antenna array has been recently proposed [6]. By employing lens antenna array (an electromagnetic lens with energy focusing capability and a matching antenna array with elements located on the focal surface of the lens), the spatial channel can be transformed into beamspace channel by concentrating the signals from different directions (beams) on different antennas [6]–[9]. Since the scattering at mm-wave frequencies is not rich, the number of effective prorogation paths in mm-wave communications is quite limited, occupying only a small number of beams. Therefore, the mm-wave beamspace channel is sparse [6], and we can select a small number of dominant beams according to the sparse beamspace channel to significantly reduce the dimension of MIMO system and the number of required RF chains [10]–[12]. As a result, mm-wave massive MIMO with lens antenna array can be considered as a promising solution to relieve the bottleneck of huge energy consumption [6].

Nevertheless, to achieve the capacity-approaching performance, beam selection requires the BS to acquire the information of beamspace channel of large size, which is difficult to realize, especially when the number of RF chains is limited. To solve this problem, some advanced schemes based on compressive sensing (CS) have been proposed very recently [13]–[17]. The key idea of these schemes is to efficiently utilize the sparsity of mm-wave channel in the angle domain. However, these schemes are designed for hybrid precoding systems [18]–[20], where the phase shifter network is realized by high-resolution phase shifters.

As a result, the analog precoders with much higher design freedom can be realized to improve the channel estimation accuracy. By contrast, in mm-wave massive MIMO systems with lens antenna array, although the phase shifter network can be replaced by lens antenna array to further reduce the hardware cost and energy consumption, the equivalently designed analog precoders will be restricted to discrete fourier transform (DFT) vectors. From the mathematical view, the analog precoders in mm-wave massive MIMO with lens antenna array have stronger constraint [9]. As a result, if we directly apply the channel estimation schemes designed for hybrid precoding systems, the performance will be not very satisfying. To estimate the beamspace channel in mm-wave massive MIMO systems with lens antenna array, sparsity mask detection (SMD)-based channel estimation is proposed very recently [21], [22]. The key idea is to first determine which beams with large power should be used (i.e., sparsity mask [10], [21], [23]) by a beam training procedure between the BS and users. Then, we can reduce the dimension of beamspace channel, and estimate the dimension-reduced channel by classical algorithms, such as least squares (LS). This scheme can efficiently estimate the beamspace channel with quite low computational complexity, and its pilot overhead is also low when the whole channel coherence time or beam coherence time as discussed in [21, Section V] are considered. However, the number of pilot symbols required to scan all the beams is proportional to the number of BS antennas, which is still large (e.g., 256 antennas). This can be reduced if we exploit the sparsity of beamspace channel by utilizing CS tools [17].

In this paper, we propose a support detection (SD)-based channel estimation scheme by utilizing the CS tools to estimate the beamspace channel with low pilot overhead.[1] The contributions of this paper can be summarized as follows:

1) We design an adaptive selecting network, which consists of a small number of 1-bit phase shifters, to replace the recently proposed selecting network in mm-wave massive MIMO with lens antenna array [6]. For data transmission, the proposed adaptive selecting network can select beams like the traditional one, while for channel estimation, it can perform as a combiner to obtain the efficient measurements of beamspace channel. Then, based on the proposed adaptive selecting network, we formulate the beamspace channel estimation problem as a sparse signal recovery problem [24].

2) We propose a SD-based channel estimation scheme. The basic idea is to decompose the total channel estimation problem into a series of sub-problems, each of which only considers one sparse channel component (a vector containing the information of a specific propagation direction) [25]. For each channel component, we first detect its support (i.e., the index set of nonzero elements in a sparse vector) by exploiting the structural characteristics of mm-wave beamspace channel. Then, the influence of this channel component is removed, and the support of the next channel component is detected in a similar method. After the supports of all channel components
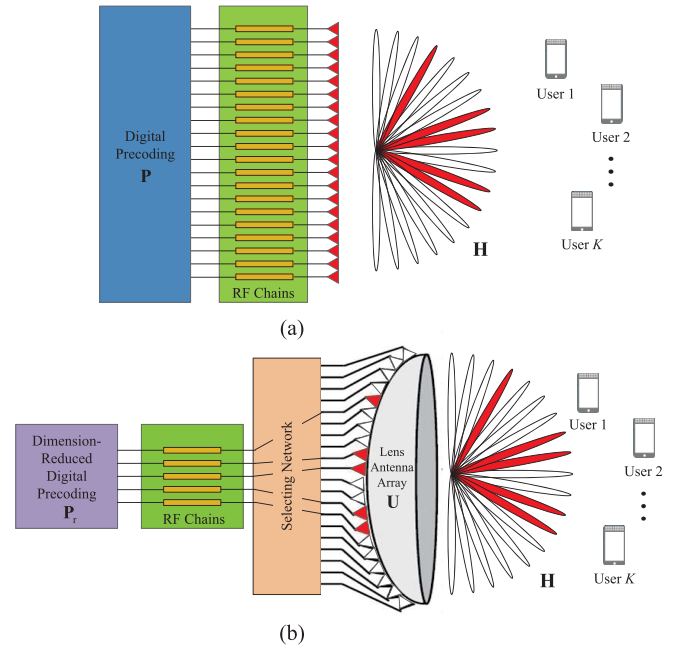
Fig. 1. Comparison of system architectures: (a) traditional mm-wave massive MIMO; (b) mm-wave massive MIMO with lens antenna array.

have been detected, the beamspace channel of large size can be estimated with low pilot overhead.

3) We prove that the proposed SD-based channel estimation scheme can detect the supports of channel components more accurately than the classical CS algorithms [24]. Complexity analysis shows that SD-based channel estimation also enjoys low complexity. Simulation results verify that SD-based channel estimation enjoys satisfying accuracy and low pilot overhead, even in the low signal-to-noise ratio (SNR) region.

The rest of the paper is organized as follows. In Section II, the system model of mm-wave massive MIMO with lens antenna array is described. In Section III, we specify the proposed SD-based channel estimation scheme and the analyses of its performance. Finally, the simulation results are provided in Section IV, and conclusions are drawn in Section V.

Notation: Lower-case and upper-case boldface letters $\mathbf{a}$ and $\mathbf{A}$ denote a vector and a matrix, respectively; $\mathbf{A}^H$, $\mathbf{A}^{-1}$, and $\text{tr}(\mathbf{A})$ denote the conjugate transpose, inversion, and trace of matrix $\mathbf{A}$, respectively; $\|\mathbf{A}\|_F$ denotes the Frobenius norm of matrix $\mathbf{A}$; $\|\mathbf{a}\|_2$ denotes the l_2-norm of vector $\mathbf{a}$; $|a|$ denotes the amplitude of scalar a; Card ($\mathcal{A}$) denotes the cardinality of set $\mathcal{A}$; Finally, $\mathbf{I}_K$ is the $K \times K$ identity matrix.

II. SYSTEM MODEL

We consider a time division duplexing (TDD) mm-wave massive MIMO system, where the BS employs N antennas and N_{RF} RF chains to simultaneously serve K single-antenna users [10]–[12]. In this section, we focus on the downlink model to explain the basic principle of mm-wave massive MIMO with lens antenna array, while in Section III, the uplink model will be considered for channel estimation, which is just a transposition of the downlink one according to the TDD channel reciprocity.

A. Traditional mm-wave Massive MIMO

Fig. 1 (a) shows the traditional mm-wave massive MIMO. For a narrowband system, the $K \times 1$ received signal vector

$\mathbf{y}^{\mathrm{DL}}$ for all K users in the downlink can be presented by

$$\mathbf{y}^{\mathrm{DL}} = \mathbf{H}^H \mathbf{P} \mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}^H \in \mathbb{C}^{K \times N}$ is the downlink channel matrix, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]$ is the uplink channel matrix according to the channel reciprocity [15], $\mathbf{h}_k$ of size $N \times 1$ is the channel vector between the BS and the kth user as will be discussed in details later, $\mathbf{s}$ of size $K \times 1$ is the original signal vector for all K users with normalized power $\mathbb{E}\left(\mathbf{s}\mathbf{s}^H\right) = \mathbf{I}_K$, $\mathbf{P}$ of size $N \times K$ is the precoding matrix satisfying $\mathrm{tr}\left(\mathbf{P}\mathbf{P}^H\right) \leq \rho$, where ρ is downlink transmit power. Finally, $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_{\mathrm{DL}}^2 \mathbf{I}_K\right)$ is the $K \times 1$ additive white Gaussian noise (AWGN) vector, where σ_{DL}^2 is the downlink noise power. It is obvious from Fig. 1 (a) that the number of required RF chains in traditional MIMO systems is equal to the number of antennas, i.e., $N_{\mathrm{RF}} = N$, which is usually large for mm-wave massive MIMO systems, e.g., $N_{\mathrm{RF}} = N = 256$ [1].

Next, we will introduce the channel vector $\mathbf{h}_k$ of the kth user. In this paper, we adopt the widely used Saleh-Valenzuela channel model for mm-wave communications as [13]

$$\mathbf{h}_k = \sqrt{\frac{N}{L_k+1}} \sum_{i=0}^{L_k} \beta_k^{(i)} \mathbf{a}\left(\psi_k^{(i)}\right) = \sqrt{\frac{N}{L_k+1}} \sum_{i=0}^{L_k} \mathbf{c}_{k,i}, \tag{2}$$

where $\mathbf{c}_{k,0} = \beta_k^{(0)} \mathbf{a}\left(\psi_k^{(0)}\right)$ is the line-of-sight (LoS) component of $\mathbf{h}_k$ with $\beta_k^{(0)}$ presenting the complex gain and $\psi_k^{(0)}$ denoting the spatial direction, $\mathbf{c}_{k,i} = \beta_k^{(i)} \mathbf{a}\left(\psi_k^{(i)}\right)$ for $1 \leq i \leq L_k$ is the ith non-line-of-sight (NLoS) component of $\mathbf{h}_k$, and L_k is the number of NLoS components, which can be usually obtained by channel measurement [26], $\mathbf{a}(\psi)$ is the $N \times 1$ array steering vector. For the typical uniform linear array (ULA) with N antennas, we have

$$\mathbf{a}(\psi) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi\psi m}\right]_{m \in I(N)}, \tag{3}$$

where $I(N) = \{p - (N-1)/2, \, p = 0, 1, \cdots, N-1\}$ is a symmetric set of indices centered around zero. The spatial direction is defined as $\psi \triangleq \frac{d}{\lambda}\sin\theta$ [6], where θ is the physical direction, λ is the wavelength of carrier, and d is the antenna spacing which usually satisfies $d = \lambda/2$ in mm-wave communications [1].

B. mm-wave Massive MIMO With Lens Antenna Array

The conventional channel (2) in the spatial domain can be transformed to the beamspace channel by employing a carefully designed lens antenna array [6] as shown in Fig. 1 (b). Essentially, such lens antenna array plays the role of a spatial DFT matrix $\mathbf{U}$ of size $N \times N$, which contains the array steering vectors of N orthogonal directions (beams) covering the entire space as

$$\mathbf{U} = \left[\mathbf{a}\left(\bar{\psi}_1\right), \mathbf{a}\left(\bar{\psi}_2\right), \cdots, \mathbf{a}\left(\bar{\psi}_N\right)\right]^H, \tag{4}$$

where $\bar{\psi}_n = \frac{1}{N}\left(n - \frac{N+1}{2}\right)$ for $n = 1, 2, \cdots, N$ are the spatial directions pre-defined by lens antenna array. Then, according
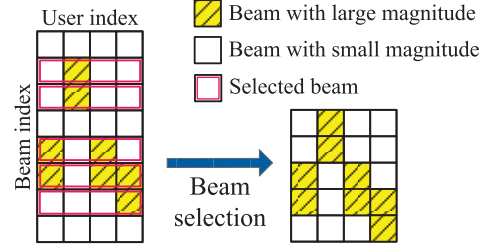

Fig. 2. Illustration of beam selection [10]–[12].

to Fig. 1 (b), the system model of mm-wave massive MIMO with lens antenna array can be represented by

$$\tilde{\mathbf{y}}^{\mathrm{DL}} = \mathbf{H}^H \mathbf{U}^H \mathbf{P} \mathbf{s} + \mathbf{n} = \tilde{\mathbf{H}}^H \mathbf{P} \mathbf{s} + \mathbf{n}, \tag{5}$$

where $\tilde{\mathbf{y}}^{\mathrm{DL}}$ is the received downlink signal vector in the beamspace, and the beamspace channel $\tilde{\mathbf{H}}$ is defined as

$$\tilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_K\right] = \mathbf{U}\mathbf{H} = [\mathbf{U}\mathbf{h}_1, \mathbf{U}\mathbf{h}_2, \cdots, \mathbf{U}\mathbf{h}_K], \tag{6}$$

where $\tilde{\mathbf{h}}_k$ is the beamspace channel vector between the BS and the kth user. It is worth pointing out that the beamspace channel $\tilde{\mathbf{H}}$ ($\tilde{\mathbf{h}}_k$) has a sparse structure [6], [23] due to the limited number of dominant scatterers in the mm-wave prorogation environments [13]. Therefore, as shown in Fig. 2, we can select only a small number of appropriate beams according to the sparse beamspace channel to reduce the dimension of MIMO system as

$$\tilde{\mathbf{y}}^{\mathrm{DL}} \approx \tilde{\mathbf{H}}_{\mathrm{r}}^H \mathbf{P}_{\mathrm{r}} \mathbf{s} + \mathbf{n}, \tag{7}$$

where $\tilde{\mathbf{H}}_{\mathrm{r}} = \tilde{\mathbf{H}}(b,:)_{b \in \mathcal{B}}$ with $\mathcal{B}$ denoting the sparsity mask (the beam set contains the indices of selected beams) [21], and $\mathbf{P}_{\mathrm{r}}$ is the corresponding dimension-reduced digital precoding matrix, which also satisfies the transmit power constraint as $\mathrm{tr}\left(\mathbf{P}_{\mathrm{r}}\mathbf{P}_{\mathrm{r}}^H\right) \leq \rho$. Due to the sparse nature of the beamspace channel $\tilde{\mathbf{H}}$, mm-wave massive MIMO with lens antenna array can significantly reduce the number of required RF chains without obvious performance loss as shown in Fig. 1 (b) [23], [27], [28]. Note that the smallest number of required RF chains should be $N_{\mathrm{RF}} = K$ to guarantee the spatial multiplexing gains of K users. Therefore, we consider $N_{\mathrm{RF}} = K$ without loss of generality in this paper.

To achieve the capacity-approaching performance, beam selection requires the information of beamspace channel $\tilde{\mathbf{H}}$. Fortunately, since the beamforming gains always exist in mm-wave massive MIMO systems with lens antenna array, the beamspace channel estimation will not suffer from the serious SNR loss [9]. However, as the number of RF chains is limited while the size of beamspace channel is large, estimating the beamspace channel with low pilot overhead is still a challenging problem.

III. BEAMSPACE CHANNEL ESTIMATION

In this section, based on the specific architecture of mm-wave massive MIMO with lens antenna array, we first introduce a pilot transmission strategy. Then, an adaptive selecting

network is designed to obtain the efficient measurements of beamspace channel for channel estimation. After that, a SD-based channel estimation scheme is proposed to estimate the beamspace channel with reliable performance and low pilot overhead. Finally, the performance and complexity analyses are provided to show the advantages of our scheme.

A. Pilot Transmission

To estimate the beamspace channel, in the uplink of TDD systems, all users need to transmit the known pilot sequences to the BS over Q instants (each user transmits one pilot symbol in each instant) for channel estimation, and we assume that the beamspace channel remains unchanged within such channel coherence time (i.e., Q instants) [29]. The impact of channel coherence time on the pilot overhead can be found in [21] and [30]. It is worth pointing out that for mm-wave communications, although the channel coherence time is usually small due to the high carrier frequency, it still contains quite a large number of symbols thanks to the large mm-wave bandwidth [21]. For example, when the carrier frequency is 28 GHz and the bandwidth is 1 GHz, a maximum speed of 30 m/s results in the small channel coherence time of 0.36 ms. However, the symbol duration is in the order of 1 ns, which means that the small channel coherence time still contains 400,000 symbols. In this paper, we consider the pilot transmission strategy, where Q instants are divided into M blocks and each block consists of K instants, i.e., $Q = MK^2$ (the pilot transmission strategy adopted in [21] and [30] can be also applied here). For the mth block, we define Ψ_m of size $K \times K$ as the pilot matrix, which contains K mutually orthogonal pilot sequences transmitted by K users over K instants [29], [32]. To normalize the uplink pilot power to unit, we have $\Psi_m \Psi_m^H = I_K$ and $\Psi_m^H \Psi_m = I_K$. A simple example of Ψ_m when $K = 4$ is

$$\Psi_m = \frac{1}{2} \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}. \quad (8)$$

Then, according to Fig. 1 (b) and the channel reciprocity [15] in TDD systems, the received uplink signal matrix $\tilde{Y}_m^{UL}$ of size $N \times K$ at the BS in the mth block can be presented as

$$\tilde{Y}_m^{UL} = UH\Psi_m + N_m = \tilde{H}\Psi_m + N_m, \quad m = 1, 2, \cdots, M, \quad (9)$$

where N_m is the $N \times K$ noise matrix in the mth block, whose entries are independent and identically distributed (i.i.d.) complex Gaussian random variables with mean zero and variance σ_{UL}^2 (the uplink noise power). As the uplink pilot power is normalized to 1, $1/\sigma_{UL}^2$ can be regarded as the uplink SNR.

B. Adaptive Selecting Network

We consider the mth block without loss of generality. During the pilot transmission, the BS should employ a combiner

---

[2]Note that in practical systems, the number of users may change but the maximum number of users that can be simultaneously served is usually fixed (e.g., at most 4 users in LTE-Advanced systems [31]). To this end, the total number of pilot symbols Q can be designed according to the maximum number. When the number of users exceeds the maximum one, we can resort to user scheduling [3].
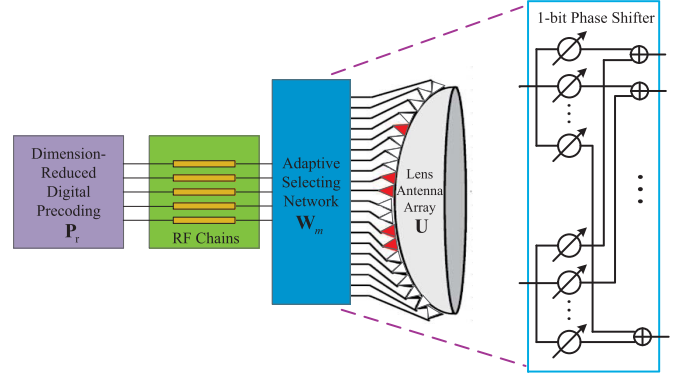

Fig. 3. Proposed adaptive selecting network for mm-wave massive MIMO with lens antenna array.

W_m of size $K \times N$ to combine the received uplink signal matrix $\tilde{Y}_m^{UL}$ (9). Then, we can obtain R_m of size $K \times K$ in the baseband sampled by $N_{RF} = K$ RF chains as

$$R_m = W_m \tilde{Y}_m^{UL} = W_m \tilde{H}\Psi_m + W_m N_m. \quad (10)$$

After that, by multiplying the known pilot matrix Ψ_m^H on the right side of (10), the $K \times K$ measurement matrix Z_m of the beamspace channel $\tilde{H}$ can be obtained by

$$Z_m = R_m \Psi_m^H = W_m \tilde{H} + N_m^{eff}, \quad (11)$$

where $N_m^{eff} = W_m N_m \Psi_m^H$ is the effective noise matrix.

Note that here we focus on estimating the beamspace channel $\tilde{h}_k$ of the kth user without loss of generality, and the similar method can be directly applied to other users to obtain the complete beamspace channel $\tilde{H}$. Then, after M blocks for the pilot transmission, we can obtain an $Q \times 1$ measurement vector $\bar{z}_k$ for $\tilde{h}_k$ as

$$\bar{z}_k = \begin{bmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{M,k} \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_M \end{bmatrix} \tilde{h}_k + \begin{bmatrix} n_{1,k}^{eff} \\ n_{2,k}^{eff} \\ \vdots \\ n_{M,k}^{eff} \end{bmatrix} \triangleq \bar{W}\tilde{h}_k + \bar{n}_k, \quad (12)$$

where $z_{m,k}$, $\tilde{h}_k$, and $n_{m,k}^{eff}$ are the kth column of Z_m, $\tilde{H}$, and N_m^{eff} in (11), respectively. $\bar{z}_k$, $\bar{W}$, and $\bar{n}_k$ are of size $Q \times 1$, $Q \times N$, and $Q \times 1$, respectively. Our target is to reliably reconstruct $\tilde{h}_k$ based on $\bar{z}_k$ with the number of pilot symbols Q as low as possible. However, if we directly utilize the recently proposed selecting network [11], [21] as shown in Fig. 1 (b) to design $\bar{W}$ (or equivalently W_m for $m = 1, 2, \cdots, M$), each row of $\bar{W}$ will have one and only one nonzero element. Consequently, to guarantee that the measurement vector $\bar{z}_k$ contains the complete information of the beamspace channel $\tilde{h}_k$, the number of pilot symbols Q should be at least larger than N, which is still high in mm-wave massive MIMO systems as mentioned above.

To this end, we propose an adaptive selecting network for mm-wave massive MIMO systems with lens antenna array as shown in Fig. 3, where the selecting network with switches in Fig. 1 (b) is replaced by 1-bit phase shifters. During the data transmission, the proposed adaptive selecting network

can be configured to realize the traditional function of beam selection.[3] Furthermore, during the beamspace channel estimation, this adaptive selecting network can be also adaptively used as an analog combiner $\mathbf{W}_m$ [20] to combine the uplink signals. Note that $\tilde{\mathbf{h}}_k$ is a sparse vector, as the number of dominant scatterers in the mm-wave prorogation environments is limited [10]. Therefore, by utilizing the proposed adaptive selecting network, we can guarantee that $\bar{\mathbf{z}}_k$ (12) has the complete information of $\tilde{\mathbf{h}}_k$ even if $Q < N$. Then, (12) can be formulated as a typical sparse signal recovery problem [17]. The motivation of such formulation is that we can significantly reduce the number of pilot symbols (i.e., Q can be much less than N) at the cost of some SNR loss [17].

Our next target is to design the analog combiner $\bar{\mathbf{W}}$. Under the framework of CS, to achieve the satisfying recovery accuracy, $\bar{\mathbf{W}}$ should be designed to make the mutual coherence

$$\mu \triangleq \max_{i \neq j} \left| \bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_j \right| \quad (13)$$

as small as possible, where $\bar{\mathbf{w}}_i$ is the ith column of $\bar{\mathbf{W}}$. There are already some matrices that have been proved to enjoy small μ, such as the i.i.d. Gaussian random matrix and Bernoulli random matrix [17]. In our paper, we select the Bernoulli random matrix as the combiner $\bar{\mathbf{W}}$, i.e., each element of $\bar{\mathbf{W}}$ is randomly selected from $\frac{1}{\sqrt{Q}}\{-1, +1\}$ with equal probability. This is due to the facts that: i) all elements of $\bar{\mathbf{W}}$ share the same normalized amplitude, which can be realized by phase shifters; ii) the resolution of phase shifter can be only 1 bit, since we only need to shift the phase by 0 or π. This means that the cost and energy consumption of the phase shifter network can be significantly reduced[4] (1-bit phase shifter is simple and consumes the energy quite similar to that of switch [35]).

C. SD-Based Channel Estimation

After $\bar{\mathbf{W}}$ has been designed by the proposed adaptive selecting network, (12) can be solved by the classical CS algorithms, such as orthogonal matching pursuit (OMP) and compressive sampling matching pursuit (CoSaMP) [36]. However, due to the limited transmit power of users, we usually expect that the channel can be estimated with low SNR [13], [37], [38]. In this case, $\tilde{\mathbf{h}}_k$ is overwhelmed by noise, and the support of $\tilde{\mathbf{h}}_k$ detected by the classical CS algorithms is inaccurate, leading to the deteriorated performance [17], [39]. In this paper, by utilizing the structural characteristics of mm-wave beamspace channel, we propose a SD-based channel estimation scheme, which can detect the support more accurately and achieve better performance than the classical CS algorithms,

[3]Specifically, we can turn off some phase shifters to realize "unselect" [33] and set some phase shifters to shift the phase 0 degree to realize "select" in beam selection. Alternatively, we can directly utilize the proposed adaptive selecting network to design an analog precoder for data transmission, which can further improve the performance. One possible way is to extend the simple conjugate analog precoder [34] to the scenarios where only 1-bit phase shifters are used. More efficient schemes will be left for our further work.

[4]Note that in practice, realizing the phase shifter network with 1-bit phase shifters as shown in Fig. 3 may be also a little complicated. If a certain performance loss is permitted, we can replace the phase shifters by switches, and the proposed scheme can be directly employed without modification.
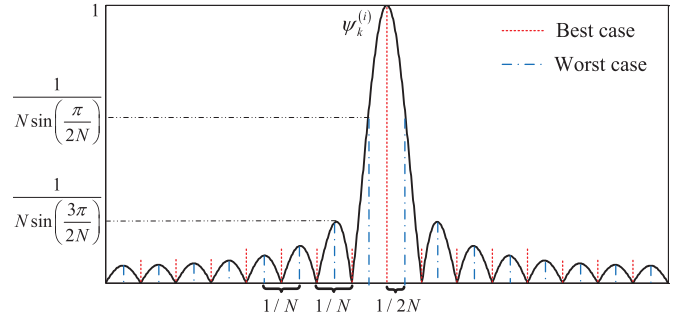

Fig. 4. The normalized amplitude distribution of the elements in $\tilde{\mathbf{c}}_{k,i}$.

especially in the low SNR region. In the following **Lemma 1**, we will first prove a special property of mm-wave beamspace channel, which is one of the two bases of the proposed SD-based channel estimation scheme.

Lemma 1: Represent the beamspace channel $\tilde{\mathbf{h}}_k$ as $\tilde{\mathbf{h}}_k = \sqrt{N/(L_k + 1)} \sum_{i=0}^{L_k} \tilde{\mathbf{c}}_{k,i}$, where $\tilde{\mathbf{c}}_{k,i} = \mathbf{U}\mathbf{c}_{k,i}$ is the ith channel component of $\tilde{\mathbf{h}}_k$ in the beamspace. Then, any two channel components $\tilde{\mathbf{c}}_{k,i}$ and $\tilde{\mathbf{c}}_{k,j}$ in the beamspace are asymptotically orthogonal when the number of antennas N in mm-wave massive MIMO systems tends to infinity [40], i.e.,

$$\lim_{N \to \infty} \left| \tilde{\mathbf{c}}_{k,i}^H \tilde{\mathbf{c}}_{k,j} \right| = 0, \quad \forall i, j = 0, 1, \cdots, L_k, \quad i \neq j. \quad (14)$$

Proof: The detailed proof can be found in [40]. ∎

Lemma 1 implies that for the mm-wave massive MIMO system with a large lens antenna array (e.g., $N = 256$), all channel components of $\tilde{\mathbf{h}}_k$ in the beamspace are approximately orthogonal to each other. As a result, we can decompose the total channel estimation problem into a series of independent sub-problems, each of which only considers one specific channel component. Specifically, we can first estimate the strongest channel component. After that, we can remove the influence of this component from the total estimation problem, and then the channel component with the second strongest power can be estimated. Such procedure will be repeated until all $(L_k + 1)$ channel components have been estimated. Next, in the following **Lemma 2**, we will prove another special structural characteristic of mm-wave beamspace channel to show how to estimate each channel component in the beamspace.

Lemma 2: Consider the ith channel component $\tilde{\mathbf{c}}_{k,i}$ in the beamspace, and assume V is an even integer without loss of generality. The ratio between the power P_V of V strongest elements of $\tilde{\mathbf{c}}_{k,i}$ and the total power P_T of $\tilde{\mathbf{c}}_{k,i}$ can be lower-bounded by

$$\frac{P_V}{P_T} \geq \frac{2}{N^2} \sum_{i=1}^{V/2} \frac{1}{\sin^2 \left(\frac{(2i-1)\pi}{2N} \right)}. \quad (15)$$

Moreover, once the position n_i^* of the strongest element of $\tilde{\mathbf{c}}_{k,i}$ is determined, the other $V - 1$ strongest elements will uniformly locate around it.

Proof: Based on (2)-(4), the ith channel component $\tilde{\mathbf{c}}_{k,i}$ in the beamspace can be presented as

$$\tilde{\mathbf{c}}_{k,i} = \beta_k^{(i)} \left[\Upsilon \left(\bar{\psi}_1 - \psi_k^{(i)} \right), \cdots, \Upsilon \left(\bar{\psi}_N - \psi_k^{(i)} \right) \right]^H, \quad (16)$$

where $\Upsilon(x) \triangleq \frac{\sin N\pi x}{N \sin \pi x}$. Fig. 4 shows the normalized amplitude distribution of the elements in $\tilde{\mathbf{c}}_{k,i}$, where the set of red dash lines (or blue dot dash lines) presents the set of spatial directions $\bar{\psi}_n = \frac{1}{N}\left(n - \frac{N+1}{2}\right)$ for $n = 1, 2, \cdots, N$ in (4). From Fig. 4, we can observe that when the practical spatial direction $\psi_k^{(i)}$ exactly equals one of the pre-defined spatial directions, there is only one strongest element containing all the power of $\tilde{\mathbf{c}}_{k,i}$, which is the best case. In contrast, the worst case will happen when the distance between $\psi_k^{(i)}$ and one of the pre-defined spatial directions is equal to $1/2N$. In this case, the power P_V of V strongest elements of $\tilde{\mathbf{c}}_{k,i}$ is

$$P_V = \frac{2\left(\beta_k^{(i)}\right)^2}{N^2} \sum_{i=1}^{V/2} \frac{1}{\sin^2\left(\frac{(2i-1)\pi}{2N}\right)}. \quad (17)$$

Besides, according to (16), the total power P_T of $\tilde{\mathbf{c}}_{k,i}$ can be calculated as

$$P_T = \tilde{\mathbf{c}}_{k,i}^H \tilde{\mathbf{c}}_{k,i} = \left(\beta_k^{(i)}\right)^2. \quad (18)$$

Based on (17) and (18), we can conclude that P_V/P_T is lower-bounded by

$$\frac{P_V}{P_T} \geq \frac{2}{N^2} \sum_{i=1}^{V/2} \frac{1}{\sin^2\left(\frac{(2i-1)\pi}{2N}\right)}. \quad (19)$$

Moreover, as shown in Fig. 4, once the position n_i^* of the strongest element of $\tilde{\mathbf{c}}_{k,i}$ is determined, the other $V-1$ strongest elements will uniformly locate around it. ∎

Note that by considering the channel with only LoS component, [10] has analyzed the power ratio between the total channel and the approximated channel with only two strongest beams. The method used in [10, Appendix] can be also modified and extended to prove **Lemma 2** in this paper. From **Lemma 2**, we can derive two important conclusions. The first one is that $\tilde{\mathbf{c}}_{k,i}$ can be considered as a sparse vector, since the most power of $\tilde{\mathbf{c}}_{k,i}$ is focused on a small number of dominant elements. For example, when $N = 256$ and $V = 8$, the lower-bound of P_V/P_T is about 95%. This means that we can retain only a small number (e.g., $V = 8$) of elements of $\tilde{\mathbf{c}}_{k,i}$ with strong power, and regard other elements as zero without obvious performance loss. The second one is that the support of the sparse channel component $\tilde{\mathbf{c}}_{k,i}$ can be uniquely determined by n_i^* as[5]

$$\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right) = \text{mod}_N\left\{n_i^* - \frac{V}{2}, \cdots, n_i^* + \frac{V-2}{2}\right\}, \quad (20)$$

where $\text{Card}\left(\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)\right) = V$, and $\text{mod}_N(\cdot)$ is the modulo operation with respect to N, which guarantees that all indices in $\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)$ belong to $\{1, 2, \cdots, N\}$. After the support of $\tilde{\mathbf{c}}_{k,i}$ has been detected, we can extract V columns from $\bar{\mathbf{W}}$ (12) according to $\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)$, and use the classical LS algorithm to estimate the nonzero elements of $\tilde{\mathbf{c}}_{k,i}$.

Based on the discussion above, the pseudo-code of the proposed SD-based channel estimation can be summarized in **Algorithm 1**, which can be explained as follows. During the

---

Algorithm 1 Proposed SD-Based Channel Estimation

Input:
  Measurement vector: $\bar{\mathbf{z}}_k$ in (12);
  Combining matrix: $\bar{\mathbf{W}}$ in (12);
  Total number of channel components: $L_k + 1$;
  Retained number of elements for each component: V.
Initialization: $\tilde{\mathbf{c}}_{k,i}^e = \mathbf{0}_{N\times1}$ for $0 \leq i \leq L_k$, $\bar{\mathbf{z}}_k^{(0)} = \bar{\mathbf{z}}_k$.
for $0 \leq i \leq L_k$
1. Detect the position of the strongest element of $\tilde{\mathbf{c}}_{k,i}$ as
  $n_i^* = \arg\max_{1\leq n\leq N} \left|\bar{\mathbf{w}}_n^H \bar{\mathbf{z}}_k^{(i)}\right|$, $\bar{\mathbf{w}}_n$ is the nth column of $\bar{\mathbf{W}}$;
2. Detect $\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)$ according to (20);
3. LS estimation of the nonzero elements of $\tilde{\mathbf{c}}_{k,i}$ as
  $\mathbf{f}_i = \left(\bar{\mathbf{W}}_i^H \bar{\mathbf{W}}_i\right)^{-1} \bar{\mathbf{W}}_i^H \bar{\mathbf{z}}_k^{(i)}$, $\bar{\mathbf{W}}_i = \bar{\mathbf{W}}(:,b)_{b\in\text{supp}(\tilde{\mathbf{c}}_{k,i})}$;
4. Form the estimated $\tilde{\mathbf{c}}_{k,i}^e$ as $\tilde{\mathbf{c}}_{k,i}^e\left(\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)\right) = \mathbf{f}_i$;
5. Remove the influence of $\tilde{\mathbf{c}}_{k,i}$ by $\bar{\mathbf{z}}_k^{(i+1)} = \bar{\mathbf{z}}_k^{(i)} - \bar{\mathbf{W}}\mathbf{Q}_{k,i}^e$
6. $i = i + 1$;
end for
7. $\mathcal{S}_T = \bigcup_{0\leq i\leq L_k} \text{supp}\left(\tilde{\mathbf{c}}_i\right)$;
8. $\mathbf{f}_T = \left(\bar{\mathbf{W}}_T^H \bar{\mathbf{W}}_T\right)^{-1} \bar{\mathbf{W}}_T^H \bar{\mathbf{z}}_k$, $\bar{\mathbf{W}}_T = \bar{\mathbf{W}}(:,b)_{b\in\mathcal{S}_T}$;
9. $\tilde{\mathbf{h}}_k^e = \mathbf{0}_{N\times1}$, $\tilde{\mathbf{h}}_k^e\left(\mathcal{S}_T\right) = \mathbf{f}_T$;
Output: Estimated beamspace channel for user k: $\tilde{\mathbf{h}}_k^e$.

---

ith iteration, we first detect the position n_i^* of the strongest element of $\tilde{\mathbf{c}}_{k,i}$ in step 1 by utilizing the low mutual coherence property of $\bar{\mathbf{W}}$ (12), which is the same as the classical CS algorithms [36]. Then in step 2, utilizing the structural characteristic of mm-wave beamspace channel as analyzed before, we can directly obtain $\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right)$ according to (20), which is the key step of the proposed SD-based channel estimation scheme as will be discussed later. After that, the nonzero elements of $\tilde{\mathbf{c}}_{k,i}$ are estimated by LS algorithm in step 3, and the influence of this channel component is removed in steps 4 and 5. Such procedure will be repeated ($i = i + 1$ in step 6) until all $(L_k + 1)$ channel components have been considered. It is worth pointing out that for the proposed SD-based channel estimation scheme, we do not directly estimate the beamspace channel as $\tilde{\mathbf{h}}_k^e = \sqrt{\frac{N}{L_k+1}} \sum_{i=0}^{L_k} \tilde{\mathbf{c}}_{k,i}^e$. This is because that most of the elements with small power are regarded as zero, which will lead to error propagation in the influence removal (step 5), especially when i is large. As a result, $\bar{\mathbf{z}}_k^{(i)}$ will be more and more inaccurate to estimate the nonzero elements in step 3. To this end, we only utilize $\bar{\mathbf{z}}_k^{(i)}$ to estimate the position n_i^* of the strongest element of $\tilde{\mathbf{c}}_{k,i}$ in step 1, which can still guarantee a satisfying recovery probability even if $\bar{\mathbf{z}}_k^{(i)}$ is a little inaccurate. Then, after the iterative procedure, we can obtain the total support $\mathcal{S}_T$ of $\tilde{\mathbf{h}}_k$ in step 7.[6] Using $\mathcal{S}_T$ and $\bar{\mathbf{z}}_k$, we can alleviate the impact of

---

[5]Correspondingly, when V is odd, the support of $\tilde{\mathbf{c}}_{k,i}$ should be $\text{supp}\left(\tilde{\mathbf{c}}_{k,i}\right) = \text{mod}_N\left\{n_i^* - \frac{V-1}{2}, \cdots, n_i^* + \frac{V-1}{2}\right\}$.

[6]In the proposed SD-based channel estimation scheme, we estimate $\text{Card}\left(\mathcal{S}_T\right)$ nonzero elements for $\tilde{\mathbf{h}}_k$. Although some of these nonzero elements may be deleted after beam selection, they are still useful to determine the set of selected beams to eliminate the multi-user interferences, just like the beam selection schemes proposed in [11] and [12]. Moreover, it is worth pointing out that estimating a little more nonzero elements does not require more pilot symbols. The only cost is the slightly increased computational complexity, as will be analyzed in Section III-E.

error propagation and estimate the beamspace channel more accurately in steps 8 and 9.

The key difference between **Algorithm 1** and the classical CS algorithms [36] is the support detection in step 2. In the classical CS algorithms, all the positions of nonzero elements are estimated one by one in an iterative procedure, which may be inaccurate, especially for the element whose power is not strong enough. By contrast, in our algorithm, we only estimate the position of the strongest element. Then, by utilizing the structural characteristics of mm-wave beamspace channel, we can directly obtain the support with higher accuracy. These conclusions analyzed above will be proved in the following subsection.

D. Performance Analysis of SD-Based Channel Estimation

In this subsection, we will prove that the proposed SD-based channel estimation can detect the support more accurately than the classical CS algorithms.

From **Algorithm 1**, we can observe that the accuracy of the detected support is determined by step 1, i.e., the estimation of the position of the strongest element. Therefore, to evaluate the accuracy of the detected support, we need to analyze the probability of the event "the position of the strongest element is correctly estimated" as shown in the following **Lemma 3**.

Lemma 3: Consider the LoS scenario, i.e., $\bar{\mathbf{h}}_k = \sqrt{N}\tilde{\mathbf{c}}_{k,0}$ and suppose that the strongest element $\tilde{h}_{k,n^*}$ of $\tilde{\mathbf{h}}_k$ satisfies[7]

$$\left|\tilde{h}_{k,n^*}\right| \geq \frac{\sqrt{8\sigma_{\mathrm{UL}}^2\,(1+\alpha)\ln N}}{(1-\mu)\,(1-\kappa)-2\mu\eta}, \tag{21}$$

where $\alpha > 0$ is a constant, and we define that

$$\eta \triangleq \frac{\sum\limits_{n=1}^{N}\left|\frac{1}{\sin((2n-1)\pi/2N)}\right| - \left|\frac{1}{\sin(\pi/2N)}\right|}{\left|\frac{1}{\sin(\pi/2N)}\right|}, \tag{22}$$

$$\kappa \triangleq \left|\frac{\sin(\pi/2N)}{\sin(3\pi/2N)}\right|. \tag{23}$$

Then, the probability Pr 1 *that the position of the strongest element is correctly estimated is lower-bounded by*

$$\mathrm{Pr}\,1 \geq \left(1 - \frac{1}{N^{\alpha+1}\sqrt{\pi\,(1+\alpha)\ln N}}\right)^N. \tag{24}$$

Proof: See Appendix A. ∎

In (21), η and κ are constants, and μ is fixed when the combiner $\bar{\mathbf{W}}$ has been designed. Therefore, the constant α can be interpreted as the effective SNR in general. Given the amplitude of the strongest element $\left|\tilde{h}_{k,n^*}\right|$, for the large uplink noise power σ_{UL}^2, the constant α should be small. Consequently, it will be more difficult to correctly estimate the position of the strongest element according to (24). In addition, for a large number of pilot symbols Q, μ in (21) is more likely to be small [36, Lemma 8]. As a result, α should be

---

[7]Here we only consider the LoS scenario for the expression simplicity. It is worth pointing out that the conclusions in **Lemma 3** can be also extended to the scenario with NLoS components, where the only changes are the expressions of η (22) and κ (23) as verified in Appendix A.
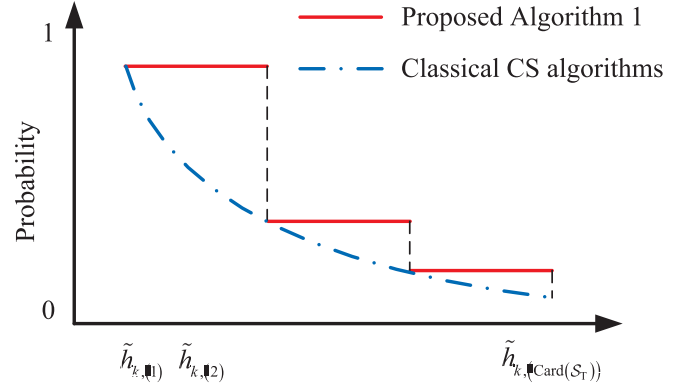

Fig. 5. An illustration of the probability comparison.

large, leading to a higher probability to correctly estimate the position of the strongest element.

From **Lemma 3**, we can conclude that the proposed SD-based channel estimation scheme can obtain the support with higher accuracy than the classical CS algorithms, which can be explained as follows: For SD-based channel estimation, once the position of the strongest element $\tilde{h}_{k,n^*}$ has been estimated, the positions of the rest of elements can be directly obtained from (20), which means that the probability to correctly estimate the positions of all nonzero elements can be lower-bounded by (24). In contrast, for the classical CS algorithms, the positions of all elements are estimated one by one in an iterative procedure. Take the second strongest element $\tilde{h}_{k,n^{**}}$ for example. In this case, $\left|\tilde{h}_{k,n^*}\right|$ in (21) should be replaced by $\left|\tilde{h}_{k,n^{**}}\right|$ (smaller than $\left|\tilde{h}_{n^*}\right|$), and the estimate error of the strongest element will equivalently lead to the increase of σ_{UL}^2. As a result, α will be small and the probability Pr 1 will decrease. This indicates that for the classical CS algorithms, the probability to correctly estimate the position of one element drops as the amplitude of this element decreases.

Such conclusion can be further intuitively illustrated by Fig. 5, where $\tilde{h}_{k,(n)}$ is the nth strongest nonzero element of $\tilde{\mathbf{h}}_k$ (i.e., $\left|\tilde{h}_{k,(1)}\right| \geq \left|\tilde{h}_{k,(2)}\right| \geq \cdots \geq \left|\tilde{h}_{k,(\mathrm{Card}(\mathcal{S}_T))}\right|$). The curve of **Algorithm 1** in Fig. 5 is divided into several segments due to the fact that we consider each channel component independently. From Fig. 5, we can observe that our algorithm is expected to detect the support more accurately, which will be also verified by simulation results in Section IV.

E. Complexity Analysis of SD-Based Channel Estimation

In this section, the complexity of SD-based channel estimation is discussed.

From **Algorithm 1**, we can observe that the complexity of the proposed SD-based channel estimation scheme mainly comes from steps 1, 3, 5, and 8. In step 1, we need to compute N inner products between two $Q \times 1$ vectors $\bar{\mathbf{w}}_n$ and $\bar{\mathbf{z}}_k^{(i)}$, whose complexity is $O(QN)$. In step 3, since $\bar{\mathbf{W}}_i$ is of size $Q \times V$, the LS algorithm can be employed with the complexity $O(V^2Q)$. In step 5, the influence of the ith channel component is removed by calculating the multiplication between

the $Q \times N$ matrix $\bar{\mathbf{W}}$ and the $N \times 1$ vector $\tilde{\mathbf{c}}_{k,i}^{e}$, which has the complexity $O(QN)$. Finally, in step 8, LS algorithm is used again with the complexity $O\left(\mathrm{Card}^2\left(\mathcal{S}_{\mathrm{T}}\right) Q\right)$ as $\bar{\mathbf{W}}_{\mathrm{T}}$ is of size $Q \times \mathrm{Card}\left(\mathcal{S}_{\mathrm{T}}\right)$.

To sum up, the complexity of the proposed SD-based channel estimation scheme can be presented as

$$O\left(L_k V^2 Q\right) + O\left(L_k N Q\right) + O\left(\mathrm{Card}^2\left(\mathcal{S}_{\mathrm{T}}\right) Q\right), \quad (25)$$

where we have $\mathrm{Card}\left(\mathcal{S}_{\mathrm{T}}\right) \leq V L_k$ according to step 7 in **Algorithm 1**. Since L_k and V are usually small as we have discussed before, we can conclude that the complexity of SD-based channel estimation is quite low, as it is comparable with that of LS algorithm. Finally, it should be also pointed out that the SMD-based channel estimation proposed in [21] enjoys much lower complexity compared with our scheme, as the dimension-reduced beamspace channel can be directly obtained by scanning all the beams. In addition, it also enjoys low pilot overhead when the whole channel/beam coherence time is considered, since the symbol duration is quite small due to the large bandwidth and there are a large number of symbols in the channel/beam coherence time [21, Table II]. However, to scan all the beams, the number of required pilot symbols in the SMD-based channel estimation scheme is still proportional to N, which is a large number (e.g., $N = 256$). By contrast, our scheme can further reduce the number of pilot symbols by utilizing the CS tools as will be shown in Section IV.

IV. SIMULATION RESULTS

In this section, we consider a typical mm-wave massive MIMO system, where the BS equips a lens antenna array with $N = 256$ antennas and $N_{\mathrm{RF}} = 16$ RF chains to serve $K = 16$ users. For the kth user, the spatial channel is generated as follows [10]: 1) one LoS component and $L_k = 2$ NLoS components; 2) $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$, and $\beta_k^{(i)} \sim \mathcal{CN}\left(0, 10^{-0.5}\right)$ for $i = 1, 2$; 3) $\psi_k^{(0)}$ and $\psi_k^{(i)}$ follow the i.i.d. uniform distribution within $[-0.5, 0.5]$. Finally, the uplink and downlink SNR are defined as $1/\sigma_{\mathrm{UL}}^2$ and $\rho/\sigma_{\mathrm{DL}}^2$, respectively.

Fig. 6 shows the normalized mean square error (NMSE) performance comparison among the SMD-based [21], OMP-based [41], and the SD-based channel estimation schemes. For SD-based channel estimation, we retain $V = 8$ strongest elements as analyzed above for each channel component. For OMP-based channel estimation [41], we assume that the sparsity level of the beamspace channel is equal to $V\left(L_k + 1\right) = 24$, and we also make a small modification on the designed sensing matrix to enable it to be applied in mm-wave massive MIMO systems with lens antenna array. Note that there are at most $V\left(L_k + 1\right)$ nonzero elements of $\tilde{\mathbf{h}}_k$ to be estimated. Therefore, the total number of instants Q for pilot transmission should be at least lager than $V\left(\max_k\left(L_k\right) + 1\right) = 24$. In this section, we assume that both SD-based and OMP-based channel estimation schemes employ $Q = 96$ instants (i.e., $M = 6$ blocks), and the SMD-based channel estimation scheme employ $Q = N = 256$ instants for pilot transmission.
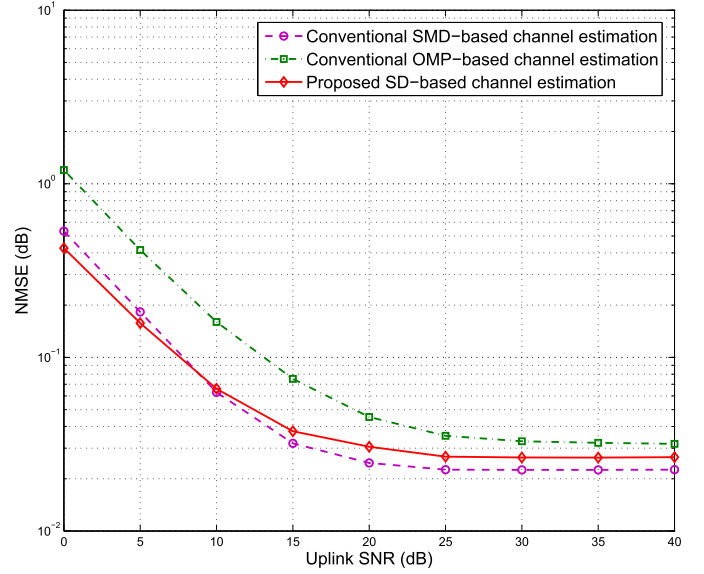

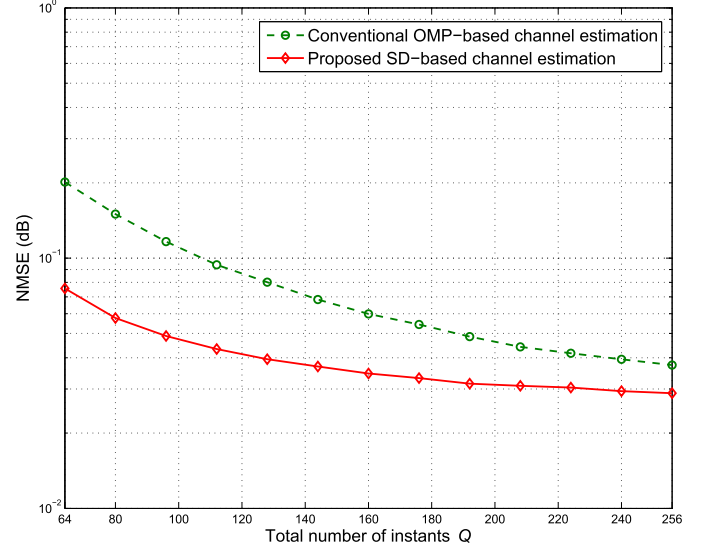Fig. 6. NMSE performance comparison among SMD-based, OMP-based, and SD-based channel estimation.

Fig. 7. NMSE performance comparison against the total number of instants Q for pilot transmission.

From Fig. 6, we first observe that the proposed SD-based channel estimation with $Q = 96$ instants can achieve the NMSE performance close to the SMD-based channel estimation in [21] with $Q = N = 256$ instants. In addition, we also observe that SD-based channel estimation enjoys higher accuracy than OMP-based channel estimation when the uplink SNR is low (e.g., less than 15 dB). When the uplink SNR is high, the performance gap becomes smaller. This can be explained by the fact that when the uplink SNR is high and the total number of instants Q is large enough, OMP-based channel estimation can also accurately estimate the support of the beamspace channel. However, the uplink SNR is usually low due to the limited transmit power of users [13]. Therefore, we can conclude that the proposed SD-based channel estimation scheme is attractive. Finally, Fig. 6 shows that when the uplink SNR is high enough, the NMSE performance of all considered channel estimation
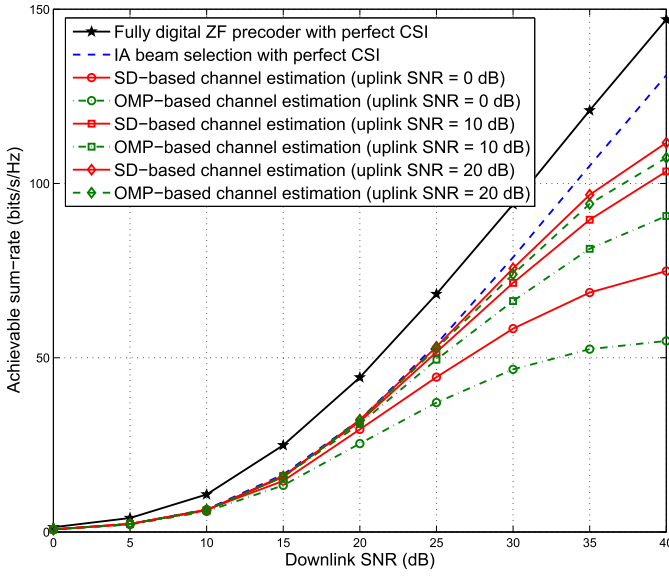
Fig. 8. Sum-rate comparison between IA beam selection with SD-based channel estimation and the one with OMP-based channel estimation.

Fig. 9. Sum-rate comparison between IA beam selection with SD-based channel estimation and the one with SMD-based channel estimation.

schemes will saturate. This is due to the fact that although the nonzero elements of beamspace channel can be estimated accurately with sufficiently high uplink SNR, the error induced by regarding the elements with small power as zero does not vanish.

Fig. 7 shows the NMSE performance comparison against the total number of instants Q, where the uplink SNR is set as 10 dB. From Fig. 7, we can observe that to achieve the same accuracy, the total number of instants Q required by SD-based channel estimation is much lower than OMP-based channel estimation. For example, to achieve the NMSE of 5×10^{-2}, the total number of instants required by OMP-based channel estimation is $Q = 190$, while the proposed SD-based channel estimation scheme only requires $Q = 120$ instants. Besides, since the number of instants Q required in SMD-based channel estimation is larger than N (e.g., $Q \geq N = 256$), we can also conclude that the proposed SD-based channel estimation scheme can achieve satisfying performance with further reduced pilot overhead.

Next, we will evaluate the impact of different beamspace channel estimation schemes on beam selection. Here, we adopt the interference-aware (IA) beam selection scheme proposed in our previous work [12], as it can guarantee that different users select different beams and support scenarios where $N_{RF} = K$. Its key idea is to classify all users into two user groups according to the potential inter-beam interferences. Specifically, if the beam with the largest power of a user is different from those of other users, this user will be regarded as a non-interference user (NIU). Otherwise, this user will be considered as an interference user (IU). For NIUs, the beams with largest power are directly selected, just like the traditional magnitude maximization beam selection scheme [10]. For IUs, a low-complexity incremental algorithm is employed. It selects the beams one by one in an incremental order, and in each step the beam with the greatest contribution to the achievable sum-rate is selected.
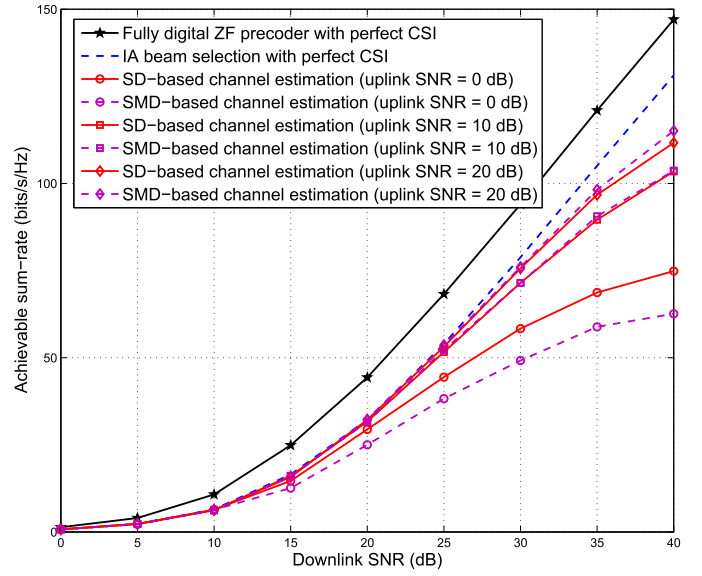
Fig. 8 provides the sum-rate comparison between IA beam selection with SD-based channel estimation and the one with OMP-based channel estimation. Note that the IA beam selection scheme can select the beams with low interference based on the beamspace channel estimated by any scheme, such as the SD-based, OMP-based, and SMD-based channel estimation. We can observe that by utilizing the proposed SD-based channel estimation instead of the OMP-based one, IA beam selection can achieve better performance, especially when the uplink SNR is low. Moreover, when the uplink SNR is moderate (e.g., 10 dB), IA beam selection with SD-based channel estimation can achieve the sum-rate performance close to the one with perfect channel state information (CSI). Finally, it is worth pointing out that when the downlink SNR is sufficiently high (e.g., 40 dB), the sum-rate performance of IA beam selection with all channel estimation schemes will saturate. This is due to the fact that with high downlink SNR, the sum-rate performance is dominated by the channel estimation error instead of noise.

Fig. 9 provides the sum-rate comparison between IA beam selection with SD-based channel estimation and the one with SMD-based channel estimation [21]. From Fig. 9, we observe that when the uplink SNR is high (e.g., 20 dB), IA beam selection with SMD-based channel estimation can achieve higher sum-rate than the one with SD-based channel estimation. This is due to the fact that in SMD-based channel estimation, by scanning all the beams with low noise power, the weak beams can be also accurately estimated to efficiently eliminate interferences in IA beam selection. However, when the uplink SNR is not high (e.g., 0 dB), IA beam selection with SD-based channel estimation can slightly outperform the one with SMD-based channel estimation. This can be explained by the fact that in SMD-based channel estimation, the N pilot symbols are respectively used to scan all N beams without further process to suppress the noise, which makes the selected strong beams suboptimal in low uplink SNR region. This further verifies that the proposed SD-based channel estimation

scheme can not only reduce the number of pilot symbols, but also achieve satisfying performance, especially when the uplink SNR is low.

V. CONCLUSIONS

This paper investigates the channel estimation problem for mm-wave massive MIMO systems with lens antenna array. Specifically, we first design an adaptive selecting network for mm-wave massive MIMO systems with lens antenna array to formulate the beamspace channel estimation problem as a sparse signal recovery problem. Then, by utilizing the special structural characteristics of mm-wave beamspace channel, we propose a SD-based channel estimation scheme with low pilot overhead. The performance analysis proves that the proposed SD-based channel estimation scheme can detect the support of sparse beamspace channel with higher accuracy than the classical CS algorithms. The complexity analysis further shows that SD-based channel estimation enjoys low complexity, which is comparable with that of LS algorithm. Simulation results verify that the proposed SD-based channel estimation scheme can achieve much better NMSE performance than conventional schemes, even in the low SNR region. This makes it more attractive for mm-wave massive MIMO systems with lens antenna array. In our future work, we will extend the proposed SD-based channel estimation scheme to the scenarios where users employ multiple antennas.

APPENDIX A
PROOF OF LEMMA 3

To prove **Lemma 3**, we need to prove the the following **Lemma 4** at first.

Lemma 4: Assume that the uplink noise vector $\bar{\mathbf{n}}_k$ in (12) follows the distribution $\mathcal{CN}\left(0, \sigma_{\mathrm{UL}}^2 \mathbf{I}\right)$, and define $\delta \triangleq \sqrt{2\sigma_{\mathrm{UL}}^2 (1 + \alpha) \ln N}$, where $\alpha > 0$ is a constant. Then, we have

$$\Pr\left\{\max_{1 \le n \le N}\left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\right| < \delta\right\} \ge \left(1 - \frac{1}{N^\alpha \sqrt{\pi (1 + \alpha) \ln N}}\right)^N, \tag{26}$$

where $\bar{\mathbf{w}}_n$ is the nth column of the combiner $\bar{\mathbf{W}}$ in (12).

Proof: Since $Q < N$, the N columns $\{\bar{\mathbf{w}}_n\}_{n=1}^N$ of $\bar{\mathbf{W}}$ are correlated. Therefore, we can conclude that $\{\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\}_{n=1}^N$ are jointly Gaussian, leading to the following result

$$\Pr\left\{\max_{1 \le n \le N}\left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\right| < \delta\right\} \ge \prod_{n=1}^N \Pr\left\{\left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\right| < \delta\right\}. \tag{27}$$

Since $\{\|\bar{\mathbf{w}}_n\|_2^2\}_{n=1}^N = 1$, each random Gaussian variable $\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k$ has mean zero and variance σ_{UL}^2. Then, we have

$$\Pr\left\{\left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\right| < \delta\right\} = 1 - \sqrt{\frac{2}{\pi \sigma_{\mathrm{UL}}^2}} \int_\delta^{+\infty} e^{-\frac{x^2}{2\sigma_{\mathrm{UL}}^2}} dx$$

$$\overset{(a)}{\ge} 1 - \sqrt{\frac{2}{\pi}} \frac{\sigma_{\mathrm{UL}}}{\delta} e^{-\frac{\delta^2}{2\sigma_{\mathrm{UL}}^2}}$$

$$\overset{(b)}{=} 1 - \frac{1}{N^{\alpha+1}\sqrt{\pi (1 + \alpha) \ln N}}, \tag{28}$$

where (a) is true due to the fact that [42]

$$\frac{1}{\sqrt{2\pi}} \int_y^{+\infty} e^{-\frac{x^2}{2}} dx \le \frac{1}{\sqrt{2\pi}\, y} e^{-\frac{y^2}{2}}, \tag{29}$$

and (b) is true because $\delta \triangleq \sqrt{2\sigma_{\mathrm{UL}}^2 (1 + \alpha) \ln N}$. Note that $\alpha > 0$ and N is the number of antennas. Therefore, it can be guaranteed that $1/\left(N^{\alpha+1}\sqrt{\pi (1 + \alpha) \ln N}\right) < 1$. Then, substituting (28) into (27), we can obtain the conclusion (26) of **Lemma 4**. ∎

By utilizing **Lemma 4**, we now can prove the conclusions in **Lemma 3**. Specifically, when the position of the strongest element $\tilde{h}_{k,n^*}$ is correctly estimated, we have

$$\left|\bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{z}}_k\right| > \max_{\substack{1 \le n \le N \\ n \ne n^*}} \left|\bar{\mathbf{w}}_n^H \bar{\mathbf{z}}_k\right|. \tag{30}$$

According to (12), the left side of (30) can be lower-bounded by

$$\left|\bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{z}}_k\right| = \left|\tilde{h}_{k,n^*} + \sum_{n \ne n^*} \bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{w}}_n \tilde{h}_{k,n} + \bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{n}}_k\right|$$

$$\overset{(a)}{\ge} \left|\tilde{h}_{k,n^*}\right| - \left|\sum_{n \ne n^*} \bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{w}}_n \tilde{h}_{k,n} + \bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{n}}_k\right|$$

$$\overset{(b)}{\ge} \left|\tilde{h}_{k,n^*}\right| - \delta - \left|\sum_{n \ne n^*} \bar{\mathbf{w}}_{n^*}^H \bar{\mathbf{w}}_n \tilde{h}_{k,n}\right|$$

$$\ge \left|\tilde{h}_{k,n^*}\right| - \delta - \mu \sum_{n \ne n^*} \left|\tilde{h}_{k,n}\right|$$

$$\overset{(c)}{\ge} \left|\tilde{h}_{k,n^*}\right| - \delta - \mu \eta \left|\tilde{h}_{k,n^*}\right|, \tag{31}$$

where (a) is true due to the triangle inequality, (b) is valid because of the assumption $\max_{1 \le n \le N}\left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k\right| < \delta$, and (c) is obtained by utilizing the structural characteristic of mm-wave beamspace channel as proved in **Lemma 2**, where we have

$$\sum_{n \ne n^*} \left|\tilde{h}_{k,n}\right| \le \frac{\sum_{n=1}^N \left|\frac{1}{\sin((2n-1)\pi/2N)}\right| - \left|\frac{1}{\sin(\pi/2N)}\right|}{\left|\frac{1}{\sin(\pi/2N)}\right|} \left|\tilde{h}_{k,n^*}\right|$$

$$= \eta \left|\tilde{h}_{k,n^*}\right|. \tag{32}$$

On the other hand, the right side of (30) can be upper-bounded by

$$\max_{\substack{1 \le n \le N \\ n \ne n^*}} \left|\bar{\mathbf{w}}_n^H \bar{\mathbf{z}}_k\right|$$

$$= \max_{\substack{1 \le n \le N \\ n \ne n^*}} \left|\bar{\mathbf{w}}_n^H \bar{\mathbf{n}}_k + \sum_{j=1}^N \bar{\mathbf{w}}_n^H \bar{\mathbf{w}}_j \tilde{h}_{k,j}\right|$$

$$\overset{(a)}{\le} \delta + \left|\tilde{h}_{k,n^{**}}\right| + \max_{\substack{1 \le n \le N \\ n \ne n^*}} \left|\sum_{j=1, j \ne n^{**}}^N \bar{\mathbf{w}}_n^H \bar{\mathbf{w}}_j \tilde{h}_{k,j}\right|$$

$$\le \delta + \left|\tilde{h}_{k,n^{**}}\right| + \mu \left|\tilde{h}_{k,n^*}\right| + \mu \eta \left|\tilde{h}_{k,n^*}\right| - \mu \left|\tilde{h}_{k,n^{**}}\right|$$

$$\overset{(b)}{\le} \delta + \mu \left|\tilde{h}_{k,n^*}\right| + \mu \eta \left|\tilde{h}_{k,n^*}\right| + (1 - \mu) \kappa \left|\tilde{h}_{k,n^*}\right|, \tag{33}$$

where $\tilde{h}_{k,n^{**}}$ is the second strongest element of $\tilde{\mathbf{h}}_k$, (a) is true due to the fact that n may be equal to n^{**}, leading to $\bar{\mathbf{w}}_{n^{**}}^H \bar{\mathbf{w}}_{n^{**}} = 1$, and (b) can be also obtained from **Lemma 2**, where we have

$$\left|\tilde{h}_{k,n^{**}}\right| \leq \frac{\sin(\pi/2N)}{\sin(3\pi/2N)}\left|\tilde{h}_{k,n^*}\right| = \kappa\left|\tilde{h}_{k,n^*}\right|. \qquad (34)$$

Combine (31) and (33) will yield

$$\left|\tilde{h}_{k,n^*}\right| \geq \frac{2\delta}{(1-\mu)(1-\kappa)-2\mu\eta}. \qquad (35)$$

Then, by utilizing **Lemma 4**, we can conclude that when

$$\left|\tilde{h}_{k,n^*}\right| \geq \frac{\sqrt{8\sigma_{\mathrm{UL}}^2(1+\alpha)\ln N}}{(1-\mu)(1-\kappa)-2\mu\eta}, \qquad (36)$$

the probability that the position of the strongest element is correctly estimated is lower-bounded by (24). $\blacksquare$

Finally, it is worth pointing out that the conclusions in **Lemma 3** can be also extended to the scenario with NLoS components, where the only changes are the expressions of η and κ. This is due to fact that all channel components of $\tilde{\mathbf{h}}_k$ are approximately orthogonal to each other as proved in **Lemma 1**. Therefore we can always find the upper-bounds of $\sum_{n \neq n^*}\left|\tilde{h}_{k,n}\right|/\left|\tilde{h}_{k,n^*}\right|$ and $\left|\tilde{h}_{k,n^{**}}\right|/\left|\tilde{h}_{k,n^*}\right|$ by utilizing the structural characteristic of mm-wave beamspace channel as proved in **Lemma 2**.

REFERENCES

[1] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[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] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[6] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[7] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[8] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1194–1206, Jun. 2014.

[9] N. Behdad and A. Sayeed, "Continuous aperture phased MIMO: Basic theory and applications," in *Proc. Allerton Conf.*, Sep./Oct. 2010, pp. 1196–1203.

[10] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3679–3684.

[11] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2222, Jun. 2015.

[12] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

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

[14] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[15] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE SPAWC Workshops*, Jun./Jul. 2015, pp. 146–150.

[16] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[17] 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.

[18] O. El 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.

[19] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5731–5737, Jul. 2016.

[20] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[21] J. Hogan and A. Sayeed, "Beam selection for performance-complexity optimization in high-dimensional MIMO systems," in *Proc. CISS*, Mar. 2016, pp. 337–342.

[22] L. Yang, Y. Zeng, and R. Zhang, "Efficient channel estimation for millimeter wave MIMO with limited RF chains," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[23] G. H. Song, J. Brady, and A. M. Sayeed, "Beamspace MIMO transceivers for low-complexity and near-optimal communication at mm-wave frequencies," in *Proc. IEEE ICASSP*, Apr. 2013, pp. 4394–4398.

[24] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[25] L. Dai, X. Gao, S. Han, I. Chih-Lin, and X. Wang, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *Proc. IEEE/CIC ICCC*, Jul. 2016, pp. 1–6.

[26] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[27] J. Brady, P. Thomas, D. Virgilio, and A. Sayeed, "Beamspace MIMO prototype for low-complexity Gigabit/s wireless communication," in *Proc. IEEE 15th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 135–139.

[28] J. Brady, J. Hogan, and A. Sayeed, "Multi-beam MIMO prototype for real-time multiuser communication at 28 GHz," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2016, pp. 1–6.

[29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[30] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[31] D. Bai *et al.*, "LTE-Advanced modem design: Challenges and perspectives," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 178–186, Feb. 2012.

[32] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[33] P. Y. Chen, C. Argyropoulos, and A. Alù, "Terahertz antenna phase shifters using integrally-gated graphene transmission-lines," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1528–1537, Apr. 2013.

[34] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[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, Jan. 2016.

[36] 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, Dec. 2007.

[37] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[38] Z. Gao, L. Dai, C. Hu, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[39] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.

[40] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[41] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. IEEE ITA Workshops*, Feb. 2015, pp. 90–97.

[42] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

Xinyu Gao (S'14) received the B.E. degree in communication engineering from the Harbin Institute of Technology, Heilongjiang, China, in 2014. He is currently pursuing the Ph.D. degree in electronic engineering with Tsinghua University, Beijing, China. His current research interests include massive MIMO and mm-wave communications, with the emphasis on signal detection and precoding. He has authored several journal and conference papers in the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE ICC, and the IEEE GLOBECOM. He received the National Scholarship in 2015 and the IEEE WCSP Best Paper Award in 2016.

Linglong Dai (M'11–SM'14) received the B.S. degree from Zhejiang University in 2003, the M.S. degree (Hons.) from the China Academy of Telecommunications Technology in 2006, and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. He co-authored the book *mmWave Massive MIMO: A Paradigm for 5G* (Academic Press, Elsevier, 2016). He has published over 50 IEEE journal papers and over 30 IEEE conference papers. He also holds 13 granted patents. His current research interests include massive MIMO, millimeter-wave communications, multiple access, and sparse signal processing. He received four conference Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the WCSP 2016, and the IEEE ICC 2017, and the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015. He currently serves as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE COMMUNICATIONS LETTERS.

Shuangfeng Han received the M.S. and Ph.D. degrees in electrical engineering from Tsinghua University in 2002 and 2006, respectively. He joined Samsung Electronics as a Senior Engineer in 2006, where he was involved MIMO, MultiBS, and MIMO. Since 2012, he has been a Senior Project Manager with the Green Communication Research Center, China Mobile Research Institute. His research interests are green 5G, massive MIMO, full duplex, NOMA, and EE-SE co-design. He is currently the Vice-Chair of the Wireless Technology Work Group of China's IMT-2020 (5G) Promotion Group.

Chih-Lin I received the Ph.D. degree in electrical engineering from Stanford University. She has been involved in multiple world-class companies and research institutes leading the research and development, including AT&T Bell Labs, AT&T HQ, ITRI of Taiwan, and ASTRI of Hong Kong. She is currently the China Mobile's Chief Scientist of Wireless Technologies and has established the Green Communications Research Center, spearheading major initiatives including key 5G technology research and development high energy efficiency system architectures, technologies and devices, green energy, and C-RAN and soft base stations. Her research interests are green communications, C-RAN, network convergence, bandwidth reframing, EE-SE co-design, massive MIMO, and active antenna arrays. She was an Elected Board Member of the IEEE Communications Society, the Chair of the Communications Society Meetings and Conferences Board, and the Founding Chair of the IEEE WCNC Steering Committee. She is currently an Executive Board Member of the GreenTouch and a Network Operator Council Member of the ETSI NFV. She received the IEEE TRANSACTIONS ON COMMUNICATION Stephen Rice Best Paper Award and is a winner of the CCCP National 1000 Talent Program.

Xiaodong Wang (S'98–M'98–SM'04–F'08) received the Ph.D. degree in electrical engineering from Princeton University. He is currently a Professor of electrical engineering with Columbia University, New York, NY, USA. His research interests fall in the general areas of computing, signal processing and communications, and has authored extensively in these areas. Among his publications is the book *Wireless Communication Systems: Advanced Techniques for Signal Reception*, (Prentice Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing. He received the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He has served as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY. He is listed as an ISI Highly-cited Author.