

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

Linglong Dai\*, Xinyu Gao\*, Shuangfeng Han<sup>†</sup>, Chih-Lin I<sup>†</sup>, and Xiaodong Wang<sup>‡</sup>

\*Tsinghua National Laboratory for Information Science and Technology (TNList),

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

<sup>†</sup>Green Communication Research Center, China Mobile Research Institute, Beijing 100053, China

<sup>‡</sup>Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

Email: daill@tsinghua.edu.cn

**Abstract**—By employing the lens antenna array, beamspace MIMO can utilize beam selection to reduce the number of required RF chains in mmWave massive MIMO systems without obvious performance loss. However, to achieve the capacity-approaching performance, beam selection requires the accurate information of beamspace channel of large size, which is challenging, especially when the number of RF chains is limited. To solve this problem, in this paper we propose a reliable support detection (SD)-based channel estimation scheme. Specifically, we propose to decompose the total beamspace channel estimation problem into a series of sub-problems, each of which only considers one sparse channel component. For each channel component, we first reliably detect its support by utilizing the structural characteristics of mmWave beamspace channel. Then, the influence of this channel component is removed from the total beamspace channel estimation problem. After the supports of all channel components have been detected, the nonzero elements of the sparse beamspace channel can be estimated with low pilot overhead. Simulation results show that the proposed SD-based channel estimation outperforms conventional schemes and enjoys satisfying accuracy, even in the low SNR region.

## I. INTRODUCTION

The integration of millimeter-wave (mmWave) and 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 and higher spectral efficiency.

However, realizing mmWave 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.). This results in unaffordable hardware complexity and energy consumption in mmWave massive MIMO systems, as the number of antennas becomes huge [1] and the energy consumption of RF chain is high at mmWave frequencies [2]. To reduce the number of required RF chains, the concept of beamspace MIMO has been recently proposed in the pioneering work [3]. By employing the lens antenna array instead of the conventional electromagnetic antenna array, beamspace MIMO can transform the conventional spatial channel into beamspace channel by concentrating the signals from different directions (beams) on different antennas [3]. Since the scattering in mmWave communications is not rich, the number of effective propagation paths is quite limited [1],

occupying only a small number of beams. As a result, the mmWave beamspace channel is sparse [3], and we can select a small number of dominant beams to significantly reduce the dimension of MIMO system and the number of required RF chains without obvious performance loss [4].

Nevertheless, beam selection requires the base station (BS) to acquire the information of beamspace channel of large size, which is challenging, 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 [5]–[7]. The key idea of these schemes is to utilize the sparsity of mmWave channels in the angle domain to efficiently estimate the mmWave massive MIMO channel of large size. However, these schemes are designed for hybrid precoding systems [8], where the phase shifter network can generate beams with sufficiently high angle resolution to improve the channel estimation accuracy. By contrast, in beamspace MIMO systems, although the phase shifter network can be replaced by lens antenna array to further reduce the hardware cost and energy consumption, the generated beams are predefined with a fixed yet limited angle resolution. If we directly apply the existing channel estimation schemes to beamspace MIMO systems with lens antenna array, the performance will be not very satisfying [3]. To the best of our knowledge, the channel estimation problem for beamspace MIMO systems has not been well addressed in the literature.

In this paper, by fully utilizing the structural characteristics of mmWave beamspace channel, we propose a reliable support detection (SD)-based channel estimation scheme. The basic idea is to decompose the total beamspace 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). For each channel component, we first detect its support (i.e., the index set of nonzero elements in a sparse vector) according to the estimated position of the strongest element. Then, the influence of this channel component is removed from the total beamspace channel estimation problem, and the support of the next channel component is detected in a similar method. After the supports of all channel components have been detected, the nonzero elements of the sparse beamspace channel can be estimated with low pilot overhead. Simulation results show

that the proposed SD-based channel estimation outperforms conventional schemes, especially in the low signal-to-noise ratio (SNR) region, which is more attractive for mmWave massive MIMO systems where low SNR is the typical case before beamforming [5].

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

## II. SYSTEM MODEL

In this paper, we consider a typical mmWave massive MIMO system working in time division duplexing (TDD) model, where the BS employs  $N$  antennas and  $N_{\text{RF}}$  RF chains to simultaneously serve  $K$  single-antenna users [3], [4]. As shown in Fig. 1 (a), for conventional MIMO systems in the spatial domain, the  $K \times 1$  received signal vector  $\mathbf{y}^{\text{DL}}$  for all  $K$  users in the downlink can be presented by

$$\mathbf{y}^{\text{DL}} = \mathbf{H}^H \mathbf{P} \mathbf{s} + \mathbf{n}, \quad (1)$$

where  $\mathbf{H}^H \in \mathbb{C}^{K \times N}$  is the downlink channel matrix,  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]$  is the uplink channel matrix according to the channel reciprocity [7],  $\mathbf{h}_k$  of size  $N \times 1$  is the channel vector between the BS and the  $k$ th user,  $\mathbf{s}$  of size  $K \times 1$  is the signal vector for all  $K$  users with normalized power  $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = \mathbf{I}_K$ ,  $\mathbf{P}$  of size  $N \times K$  is the precoding matrix satisfying the total transmit power constraint  $\rho$  as  $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq \rho$ . Finally,  $\mathbf{n} \sim \mathcal{CN}(0, \sigma_{\text{DL}}^2 \mathbf{I}_K)$  is the  $K \times 1$  noise vector, where  $\sigma_{\text{DL}}^2$  is the downlink noise power.

### A. MmWave channel model

In this paper, we adopt the widely used Saleh-Valenzuela channel model to embody the low rank and spatial correlation characteristics of mmWave communications as [3]–[7]

$$\mathbf{h}_k = \sqrt{\frac{N}{L+1}} \sum_{i=0}^L \beta_k^{(i)} \mathbf{a}(\psi_k^{(i)}) = \sqrt{\frac{N}{L+1}} \sum_{i=0}^L \mathbf{c}_i, \quad (2)$$

where  $\mathbf{c}_0 = \beta_k^{(0)} \mathbf{a}(\psi_k^{(0)})$  is the line-of-sight (LoS) component of the  $k$ th user with  $\beta_k^{(0)}$  presenting the complex gain and  $\psi_k^{(0)}$  denoting the spatial direction,  $\mathbf{c}_i = \beta_k^{(i)} \mathbf{a}(\psi_k^{(i)})$  for  $1 \leq i \leq L$  is the  $i$ th non-line-of-sight (NLoS) component of the  $k$ th user, and  $L$  is the total number of NLoS components which can be usually obtained by channel measurement [9],  $\mathbf{a}(\psi)$  is the  $N \times 1$  array steering vector. For the typical uniform linear array (ULA) with  $N$  antennas, we have

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

where  $\mathcal{I}(N) = \{l - (N-1)/2, l = 0, 1, \dots, N-1\}$  is a symmetric set of indices centered around zero. The spatial direction is defined as  $\psi \triangleq \frac{d}{\lambda} \sin \theta$  [3], where  $\theta$  is the physical direction,  $\lambda$  is the signal wavelength, and  $d$  is the antenna spacing which usually satisfies  $d = \lambda/2$  at mmWave frequencies.

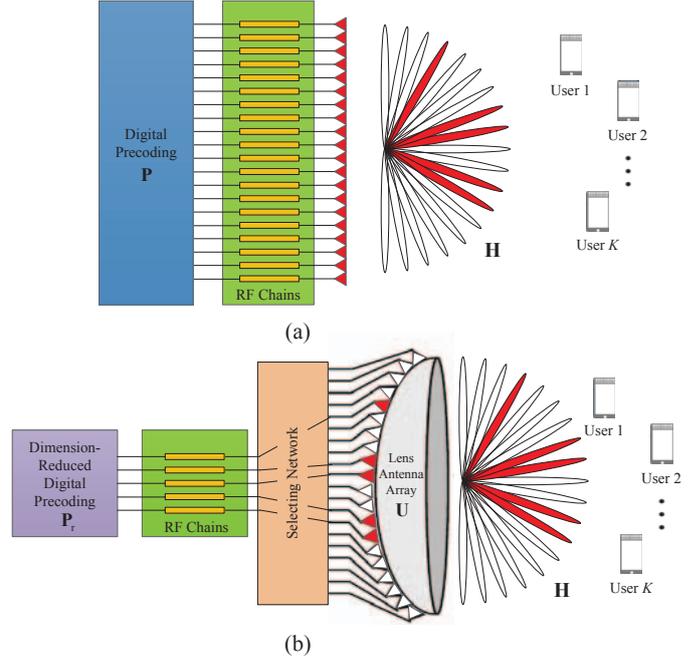


Fig. 1. System architectures: (a) conventional MIMO; (b) beamspace MIMO.

### B. Beamspace MIMO

The conventional channel (2) in the spatial domain can be transformed to the beamspace channel by employing a carefully designed lens antenna array [3] as shown in Fig. 1 (b). Specifically, such lens antenna array plays the role of a spatial discrete fourier transform (DFT) matrix  $\mathbf{U}$ , which contains the array steering vectors of  $N$  orthogonal directions (beams) covering the entire angle space as [3]

$$\mathbf{U} = [\mathbf{a}(\bar{\psi}_1), \mathbf{a}(\bar{\psi}_2), \dots, \mathbf{a}(\bar{\psi}_N)]^H, \quad (4)$$

where  $\bar{\psi}_n = \frac{1}{N} (n - \frac{N+1}{2})$  for  $n = 1, 2, \dots, N$ . Then, by assuming  $N_{\text{RF}} = K$  without loss of generality, the system model of beamspace MIMO can be represented by

$$\tilde{\mathbf{y}}^{\text{DL}} = \mathbf{H}^H \mathbf{U}^H \mathbf{B} \mathbf{P}_r \mathbf{s} + \mathbf{n} = \tilde{\mathbf{H}}^H \mathbf{B} \mathbf{P}_r \mathbf{s} + \mathbf{n}, \quad (5)$$

where  $\tilde{\mathbf{y}}^{\text{DL}}$  is the received downlink signal vector in the beamspace,  $\tilde{\mathbf{H}}^H = \mathbf{H}^H \mathbf{U}^H = (\mathbf{U}\mathbf{H})^H$  is defined as the downlink beamspace channel matrix whose  $N$  columns correspond to  $N$  orthogonal beams,  $\mathbf{B}$  of size  $N \times K$  is the selecting matrix whose entries belong to  $\{0, 1\}$ . For example, if the  $n$ th beam is selected by the  $k$ th user, the element of  $\mathbf{B}$  at the  $n$ th row and the  $k$ th column is 1. Finally,  $\mathbf{P}_r$  of size  $K \times K$  is the dimension-reduced digital precoding matrix. It is worth pointing out that the beamspace channel  $\tilde{\mathbf{H}}^H$  (or equivalently  $\tilde{\mathbf{H}}$ ) has a sparse structure [3], [4] due to the limited number of dominant scatters in the mmWave propagation environments [1]. Therefore, we can select only a small number of appropriate beams according to the sparse beamspace channel to significantly reduce the effective channel dimension without obvious performance loss. Consequently, only a small-size digital precoder  $\mathbf{P}_r$  is required, leading to a small number of required RF chains. Unfortunately, acquiring the beamspace channel of large size in practice is challenging, especially when the number of RF chains is limited.

### III. BEAMSPACE CHANNEL ESTIMATION

In this section, based on the beamspace MIMO architecture, we first introduce a pilot transmission strategy. After that, an adaptive selecting network is designed to obtain the measurements of the beamspace channel. Finally, a SD-based channel estimation is proposed to estimate the beamspace channel with limited number of RF chains and low pilot overhead.

#### 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) [10]. 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.,  $Q = MK$ . For the  $m$ th block, we define  $\Psi_m$  of size  $K \times K$  as the pilot matrix, which contains  $K$  mutually orthogonal pilot sequences transmitted by  $K$  users over  $K$  instants [10]. Obviously, we have  $\Psi_m \Psi_m^H = \mathbf{I}_K$  and  $\Psi_m^H \Psi_m = \mathbf{I}_K$ .

Then, according to Fig. 1 (b) and the channel reciprocity [7] in TDD systems, the received uplink signal matrix  $\tilde{\mathbf{Y}}_m^{\text{UL}}$  of size  $N \times K$  at the BS in the  $m$ th block can be presented as

$$\tilde{\mathbf{Y}}_m^{\text{UL}} = \mathbf{U}\mathbf{H}\Psi_m + \mathbf{N}_m = \tilde{\mathbf{H}}\Psi_m + \mathbf{N}_m, \quad m=1, 2, \dots, M, \quad (6)$$

where  $\mathbf{N}_m$  is the  $N \times K$  noise matrix in the  $m$ th block, whose entries are independent and identically distributed (i.i.d.) complex Gaussian random variables with mean zero and variance  $\sigma_{\text{UL}}^2$  (the uplink noise power).

#### B. Measurements of the beamspace channel

We consider the  $m$ th block without loss of generality. During the pilot transmission, the BS should employ a combiner  $\mathbf{W}_m$  of size  $K \times N$  to combine the received uplink signal matrix  $\tilde{\mathbf{Y}}_m^{\text{UL}}$  (6). Then, we can obtain  $\mathbf{R}_m$  of size  $K \times K$  in the baseband sampled by  $N_{\text{RF}} = K$  RF chains as

$$\mathbf{R}_m = \mathbf{W}_m \tilde{\mathbf{Y}}_m^{\text{UL}} = \mathbf{W}_m \tilde{\mathbf{H}} \Psi_m + \mathbf{W}_m \mathbf{N}_m. \quad (7)$$

After that, by multiplying the known pilot matrix  $\Psi_m^H$  on the right side of (7), the  $K \times K$  measurement matrix  $\mathbf{Z}_m$  of the beamspace channel  $\tilde{\mathbf{H}}$  can be obtained by

$$\mathbf{Z}_m = \mathbf{R}_m \Psi_m^H = \mathbf{W}_m \tilde{\mathbf{H}} + \mathbf{N}_m^{\text{eff}}, \quad (8)$$

where  $\mathbf{N}_m^{\text{eff}} = \mathbf{W}_m \mathbf{N}_m \Psi_m^H$  is the effective noise matrix.

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

$$\tilde{\mathbf{z}}_k = \begin{bmatrix} \mathbf{z}_{1,k} \\ \mathbf{z}_{2,k} \\ \vdots \\ \mathbf{z}_{M,k} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_M \end{bmatrix} \tilde{\mathbf{h}}_k + \begin{bmatrix} \mathbf{n}_{1,k}^{\text{eff}} \\ \mathbf{n}_{2,k}^{\text{eff}} \\ \vdots \\ \mathbf{n}_{M,k}^{\text{eff}} \end{bmatrix} \triangleq \bar{\mathbf{W}} \tilde{\mathbf{h}}_k + \bar{\mathbf{n}}_k, \quad (9)$$

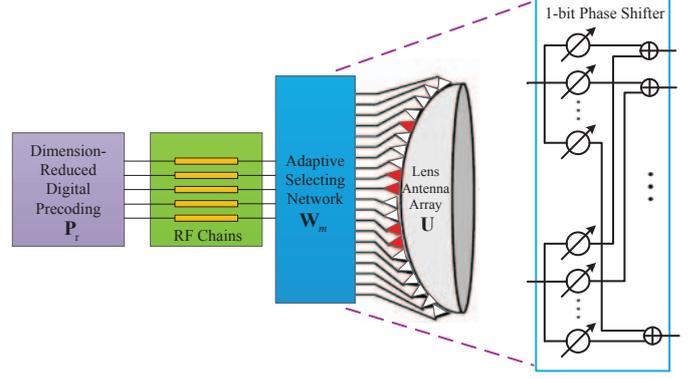


Fig. 2. Proposed adaptive selecting network in beamspace MIMO systems.

where  $\mathbf{z}_{m,k}$ ,  $\tilde{\mathbf{h}}_k$ , and  $\mathbf{n}_{m,k}^{\text{eff}}$  are the  $k$ th column of  $\mathbf{Z}_m$ ,  $\tilde{\mathbf{H}}$ , and  $\mathbf{N}_m^{\text{eff}}$  in (8), respectively.  $\tilde{\mathbf{z}}_k$ ,  $\bar{\mathbf{W}}$ , and  $\bar{\mathbf{n}}_k$  are of size  $Q \times 1$ ,  $Q \times N$ , and  $Q \times 1$ , respectively. Our target is to reliably reconstruct  $\tilde{\mathbf{h}}_k$  based on  $\tilde{\mathbf{z}}_k$  with the pilot overhead  $Q$  as low as possible. However, if we directly utilize the traditional selecting network in beamspace MIMO systems as shown in Fig. 1 (b) to design  $\bar{\mathbf{W}}$  (or equivalently  $\mathbf{W}_m$  for  $m = 1, 2, \dots, M$ ), each row of  $\bar{\mathbf{W}}$  will have one and only one nonzero element [11]. Consequently, to guarantee that the measurement vector  $\tilde{\mathbf{z}}_k$  contains the complete information of the beamspace channel  $\tilde{\mathbf{h}}_k$ , the pilot overhead  $Q$  should be at least larger than  $N$ , which is unaffordable since the number of antennas  $N$  is usually huge in mmWave massive MIMO systems as mentioned above.

To this end, we propose an adaptive selecting network for mmWave massive MIMO systems with lens antenna array as shown in Fig. 2, where the selecting network in Fig. 1 (b) is replaced by a phase shifter network. During the data transmission, the proposed adaptive selecting network can be configured to realize the traditional function of beam selection<sup>1</sup>. Furthermore, during the beamspace channel estimation, this adaptive selecting network can be also adaptively used as an analog combiner  $\mathbf{W}_m$  [8] to combine the uplink signals. With the help of the proposed adaptive selecting network, we can guarantee that  $\tilde{\mathbf{z}}_k$  (9) has the complete information of  $\tilde{\mathbf{h}}_k$ , even if  $Q < N$ . Moreover, due to the limited number of dominant scatters in the mmWave propagation environments [4],  $\tilde{\mathbf{h}}_k$  is a sparse vector. Therefore, (9) can be formulated as a typical sparse signal recovery problem [12].

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  $\mu \triangleq \max_{i \neq j} |\bar{\mathbf{w}}_i^H \bar{\mathbf{w}}_j|$  as small as possible, where  $\bar{\mathbf{w}}_i$  is the  $i$ th column of  $\bar{\mathbf{W}}$ . There are already some matrices that have been proved to enjoy small  $\mu$ , such as the i.i.d. Gaussian random matrix and Bernoulli random matrix [12]. 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

<sup>1</sup>Specifically, we can turn off some phase shifters to realize “unselect” and set some phase shifters to shift the phase 0 degree to realize “select” for beam selection.

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 phases 0 degree and  $\pi$  degree. This means that the cost and energy consumption of the phase shifter network can be significantly reduced [13].

### C. SD-based channel estimation

After  $\bar{\mathbf{W}}$  has been designed by the proposed adaptive selecting network, (9) can be solved by classical CS algorithms, such as orthogonal matching pursuit (OMP) and compressive sampling matching pursuit (CoSaMP) [14]. However, when the uplink SNR is low, which is the typical case in mmWave massive MIMO systems due to the lack of beamforming gain and the low transmit power of users [5],  $\tilde{\mathbf{h}}_k$  will be overwhelmed by noise. As a result, the support of  $\tilde{\mathbf{h}}_k$  detected by classical CS algorithms is usually inaccurate, leading to the deteriorated performance. In this paper, by utilizing the structural characteristics of mmWave beamspace channel, we propose a SD-based channel estimation, which can detect the support more accurately and achieve better performance than classical CS algorithms, especially in the low SNR region.

In the following **Lemma 1**, we will first prove a special property of mmWave beamspace channel, which is one of the two bases of the proposed SD-based channel estimation.

**Lemma 1.** Represent the beamspace channel  $\tilde{\mathbf{h}}_k$  as  $\tilde{\mathbf{h}}_k = \sqrt{N/(L+1)} \sum_{i=0}^L \tilde{\mathbf{c}}_i$ , where  $\tilde{\mathbf{c}}_i = \mathbf{U}\mathbf{c}_i$  is the  $i$ th channel component of  $\tilde{\mathbf{h}}_k$  in the beamspace. Then, any two channel components  $\tilde{\mathbf{c}}_i$  and  $\tilde{\mathbf{c}}_j$  are asymptotically orthogonal when the number of antennas  $N$  in mmWave massive MIMO systems with lens antenna array tends to infinity, i.e.,

$$\lim_{N \rightarrow \infty} |\tilde{\mathbf{c}}_i^H \tilde{\mathbf{c}}_j| = 0, \quad \forall i, j = 0, 1, \dots, L, \quad i \neq j. \quad (10)$$

*Proof:* Based on (2)-(4), we have

$$\begin{aligned} \tilde{\mathbf{c}}_i^H \tilde{\mathbf{c}}_j &= \beta_k^{(i)*} \beta_k^{(j)} \mathbf{a}^H(\psi_k^{(i)}) \mathbf{U}^H \mathbf{U} \mathbf{a}(\psi_k^{(j)}) \\ &= \beta_k^{(i)*} \beta_k^{(j)} \Upsilon(\psi_k^{(i)} - \psi_k^{(j)}), \end{aligned} \quad (11)$$

where  $\Upsilon(x) \triangleq \frac{\sin N\pi x}{N \sin \pi x}$ . Note that  $\psi_k^{(i)}$  and  $\psi_k^{(j)}$  belong to  $[-0.5, 0.5]$  according to the definitions in (3). Therefore, as long as  $(\psi_k^{(i)} - \psi_k^{(j)}) \neq 0$  or 1, which can be guaranteed by  $i \neq j$ , we have

$$\left| \Upsilon(\psi_k^{(i)} - \psi_k^{(j)}) \right| \leq \frac{1}{N} \left| \frac{1}{\sin \pi (\psi_k^{(i)} - \psi_k^{(j)})} \right|. \quad (12)$$

Based on (12), we can conclude that

$$0 \leq \lim_{N \rightarrow \infty} |\tilde{\mathbf{c}}_i^H \tilde{\mathbf{c}}_j| \leq \lim_{N \rightarrow \infty} \frac{1}{N} \left| \frac{\beta_k^{(i)*} \beta_k^{(j)}}{\sin \pi (\psi_k^{(i)} - \psi_k^{(j)})} \right| = 0, \quad (13)$$

which verifies the conclusion (10).  $\blacksquare$

**Lemma 1** implies that we can decompose the total beamspace channel estimation problem into a series of independent sub-problems, each of which only considers one

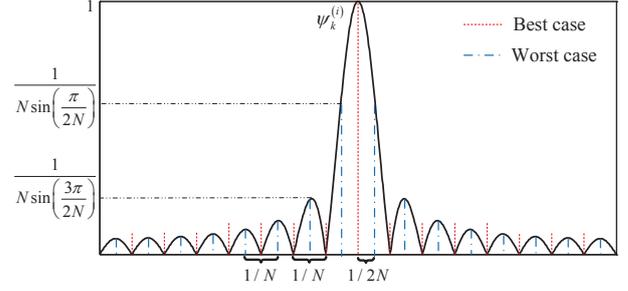


Fig. 3. The normalized amplitude distribution of the elements in  $\tilde{\mathbf{c}}_i$ .

specific channel component approximately orthogonal to the others. 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  $(L+1)$  channel components have been estimated. Next, in the following **Lemma 2**, we will prove another special structural characteristic of mmWave beamspace channel to show how to estimate each channel component in the beamspace.

**Lemma 2.** Consider the  $i$ th channel component  $\tilde{\mathbf{c}}_i$  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  $\tilde{\mathbf{c}}_i$  and the total power  $P_T$  of  $\tilde{\mathbf{c}}_i$  can be lower-bounded by

$$\frac{P_V}{P_T} \geq \frac{2}{N^2} \sum_{i=1}^{V/2} \frac{1}{\sin^2 \left( \frac{(2i-1)\pi}{2N} \right)}. \quad (14)$$

Moreover, once the position  $n_i^*$  of the strongest element of  $\tilde{\mathbf{c}}_i$  is determined, the other  $V-1$  strongest elements will uniformly locate around it with the interval  $1/N$ .

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

$$\tilde{\mathbf{c}}_i = \beta_k^{(i)} \left[ \Upsilon(\bar{\psi}_1 - \psi_k^{(i)}), \dots, \Upsilon(\bar{\psi}_N - \psi_k^{(i)}) \right]^H. \quad (15)$$

Fig. 3 shows the normalized amplitude distribution of the elements in  $\tilde{\mathbf{c}}_i$ , where the set of red dash lines (or blue dot dash lines) presents the set of spatial directions  $\bar{\psi}_n = \frac{1}{N} (n - \frac{N+1}{2})$  for  $n = 1, 2, \dots, N$  in (4) pre-defined by lens antenna array. From Fig. 3, we can observe that when the practical spatial direction  $\psi_k^{(i)}$  exactly equals one pre-defined spatial direction, there is only one strongest element containing all the power of  $\tilde{\mathbf{c}}_i$ , which is the best case. In contrast, the worst case will happen when the distance between  $\psi_k^{(i)}$  and one pre-defined spatial direction is equal to  $1/2N$ . In this case, the power  $P_V$  of  $V$  strongest elements of  $\tilde{\mathbf{c}}_i$  is

$$P_V = \frac{2 \left( \beta_k^{(i)} \right)^2}{N^2} \sum_{i=1}^{V/2} \frac{1}{\sin^2 \left( \frac{(2i-1)\pi}{2N} \right)}. \quad (16)$$

Besides, according to (15), the total power  $P_T$  of  $\tilde{\mathbf{c}}_i$  can be calculated as  $P_T = \tilde{\mathbf{c}}_i^H \tilde{\mathbf{c}}_i = \left( \beta_k^{(i)} \right)^2$ . Therefore, we can conclude that  $P_V/P_T$  is lower-bounded by (14). Moreover, as

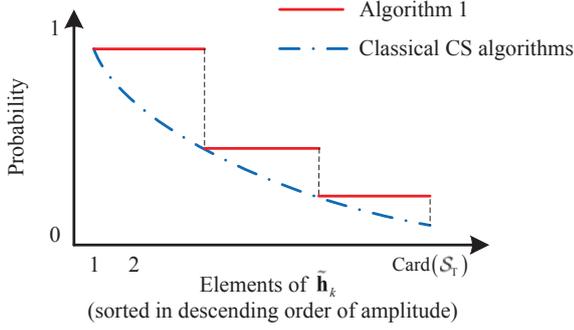


Fig. 4. An illustration of the probability comparison.

shown in Fig. 3, once the position  $n_i^*$  of the strongest element of  $\tilde{\mathbf{c}}_i$  is determined, the other  $V - 1$  strongest elements will uniformly locate around it with the interval  $1/N$ . ■

From **Lemma 2**, we can derive two important conclusions. The first one is that  $\tilde{\mathbf{c}}_i$  can be considered as a sparse vector, since the most power of  $\tilde{\mathbf{c}}_i$  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  $\tilde{\mathbf{c}}_i$  with strong power and regard other elements as zero without obvious performance loss. The second one is that the support of sparse vector  $\tilde{\mathbf{c}}_i$  can be uniquely determined by  $n_i^*$  as<sup>2</sup>

$$\text{supp}(\tilde{\mathbf{c}}_i) = \text{mod}_N \left\{ n_i^* - \frac{V}{2}, \dots, n_i^* + \frac{V-2}{2} \right\}, \quad (17)$$

where  $\text{Card}(\text{supp}(\tilde{\mathbf{c}}_i)) = V$ , and  $\text{mod}_N(\cdot)$  is the modulo operation with respect to  $N$ , which guarantees that all indices in  $\text{supp}(\tilde{\mathbf{c}}_i)$  belong to  $\{1, 2, \dots, N\}$ . After the support of  $\tilde{\mathbf{c}}_i$  has been detected, we can extract  $V$  columns from  $\tilde{\mathbf{W}}$  (9) according to  $\text{supp}(\tilde{\mathbf{c}}_i)$ , and use the classical LS algorithm to estimate the nonzero elements of  $\tilde{\mathbf{c}}_i$ .

Based on the discussion so far, the pseudo-code of the proposed SD-based channel estimation can be summarized in **Algorithm 1**<sup>3</sup>, which can be explained as follows. During the  $i$ th iteration, we first detect the position  $p_i$  of the strongest element of  $\tilde{\mathbf{c}}_i$  in step 1. Then in step 2, utilizing the structural characteristics of beamspace channel as analyzed above, we can directly obtain  $\text{supp}(\tilde{\mathbf{c}}_i)$  according to (17). After that, the nonzero elements of  $\tilde{\mathbf{c}}_i$  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 the last channel component is considered. Note that for the proposed SD-based channel estimation, we do not directly estimate the beamspace channel as  $\tilde{\mathbf{h}}_k^e = \sqrt{\frac{N}{L+1}} \sum_{i=0}^L \tilde{\mathbf{c}}_i^e$ . 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, especially when  $i$

<sup>2</sup>Correspondingly, when  $V$  is odd, the support of  $\tilde{\mathbf{c}}_i$  should be  $\text{supp}(\tilde{\mathbf{c}}_i) = \text{mod}_N \left\{ n_i^* - \frac{V-1}{2}, \dots, n_i^* + \frac{V-1}{2} \right\}$ .

<sup>3</sup>Note that the proposed SD-based channel estimation can be easily extended to the scenarios where users employ multiple antennas. In this case, the beamspace channel is a block sparse matrix instead of a sparse vector [3], and we can use **Algorithm 1** for every column of this 2D sparse matrix.

**Input:**

- Measurement vector:  $\bar{\mathbf{z}}_k$  in (9);
- Combining matrix:  $\tilde{\mathbf{W}}$  in (9);
- Total number of channel components:  $L + 1$ ;
- Retained number of elements for each component:  $V$ .

**Initialization:**  $\tilde{\mathbf{c}}_i^e = \mathbf{0}_{N \times 1}$  for  $0 \leq i \leq L$ ,  $\bar{\mathbf{z}}_k^{(0)} = \bar{\mathbf{z}}_k$ .

**for**  $0 \leq i \leq L$

1. Detect the position of the strongest element in  $\tilde{\mathbf{c}}_i$  as  $p_i = \arg \max_{1 \leq n \leq N} |\bar{\mathbf{w}}_n \bar{\mathbf{z}}_k^{(i)}|$ ,  $\bar{\mathbf{w}}_n$  is the  $n$ th row of  $\tilde{\mathbf{W}}$ ;
2. Detect  $\text{supp}(\tilde{\mathbf{c}}_i)$  according to (17);
3. LS estimation of the nonzero elements of  $\tilde{\mathbf{c}}_i$  as  $\mathbf{f}_i = (\tilde{\mathbf{W}}_i \tilde{\mathbf{W}}_i^H)^{-1} \tilde{\mathbf{W}}_i \bar{\mathbf{z}}_k^{(i)}$ ,  $\tilde{\mathbf{W}}_i = \tilde{\mathbf{W}}(l, \cdot)_{l \in \text{supp}(\tilde{\mathbf{c}}_i)}$ ;
4. Form the estimated  $\tilde{\mathbf{c}}_i^e$  as  $\tilde{\mathbf{c}}_i^e(\text{supp}(\tilde{\mathbf{c}}_i)) = \mathbf{f}_i$ ;
5. Remove the influence of  $\tilde{\mathbf{c}}_i$  as  $\bar{\mathbf{z}}_k^{(i+1)} = \bar{\mathbf{z}}_k^{(i)} - \tilde{\mathbf{W}}^H \tilde{\mathbf{c}}_i^e$ ;
6.  $i = i + 1$ ;

**end for**

7.  $\mathcal{S}_T = \bigcup_{0 \leq i \leq L} \text{supp}(\tilde{\mathbf{c}}_i)$ ;

8.  $\mathbf{f}_T = (\tilde{\mathbf{W}}_T \tilde{\mathbf{W}}_T^H)^{-1} \tilde{\mathbf{W}}_T \bar{\mathbf{z}}_k$ ,  $\tilde{\mathbf{W}}_T = \tilde{\mathbf{W}}(l, \cdot)_{l \in \mathcal{S}_T}$ ;

9.  $\tilde{\mathbf{h}}_k^e = \mathbf{0}_{N \times 1}$ ,  $\tilde{\mathbf{h}}_k^e(\mathcal{S}_T) = \mathbf{f}_T$ ;

**Output:** Estimated beamspace channel for user  $k$ :  $\tilde{\mathbf{h}}_k^e$ .

**Algorithm 1:** Proposed SD-based channel estimation.

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  $p_i$  in step 1, which can still guarantee a high recovery probability even if  $\bar{\mathbf{z}}_k^{(i)}$  is inaccurate [14]. Then, after the iterative procedure, we can obtain the total support  $\mathcal{S}_T$  of  $\tilde{\mathbf{h}}_k$  in step 7. Using  $\mathcal{S}_T$  and  $\bar{\mathbf{z}}_k$ , we can alleviate the impact of error propagation and estimate the beamspace channel more accurately in steps 8 and 9.

The key difference between **Algorithm 1** and classical CS algorithms [14] is the step of support detection. In classical CS algorithms, all the positions of nonzero elements are estimated 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 mmWave beamspace channel, we can directly obtain the accurate support with higher probability as illustrated in Fig. 4. Moreover, we can also observe that the most complicated part of the proposed SD-based channel estimation is the LS algorithm, i.e., step 3 and step 8. Therefore, the computational complexity of SD-based channel estimation is  $\mathcal{O}(LV^2M) + \mathcal{O}(\text{Card}^2(\mathcal{S}_T)M)$  ( $\text{Card}(\mathcal{S}_T) \leq VL$ ), which is comparable with that of LS algorithm, since  $L$  and  $V$  are usually small as discussed above.

#### IV. SIMULATION RESULTS

In this section, we consider a typical mmWave massive MIMO system, where the BS equips a lens antenna array with  $N = 256$  antennas and  $N_{\text{RF}} = 16$  RF chains to simultaneously serve  $K = 16$  users. For the  $k$ th user, the spatial channel is generated as follows [3], [4]: 1) one LoS component and  $L = 2$  NLoS components; 2)  $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$ , and

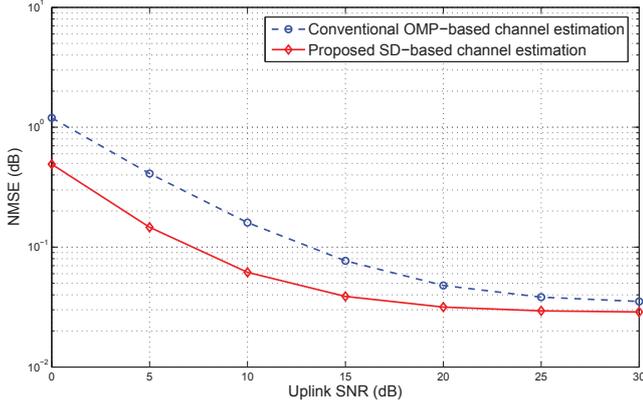


Fig. 5. NMSE performance comparison.

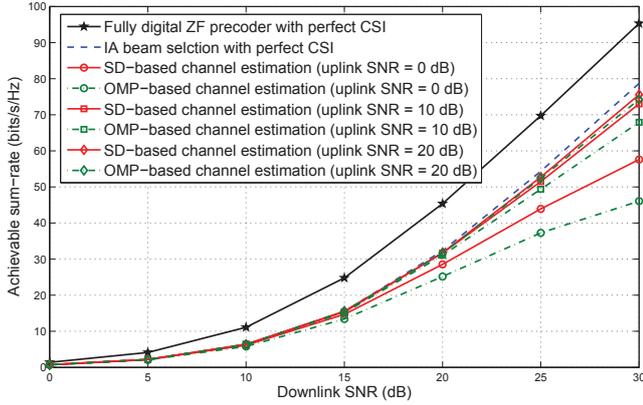


Fig. 6. Sum-rate performance comparison.

$\beta_k^{(i)} \sim \mathcal{CN}(0, 10^{-2})$  for  $i = 1, 2, 3$ ;  $\psi_k^{(0)}$  and  $\psi_k^{(i)}$  follow the i.i.d. uniform distribution within  $[-0.5, 0.5]$ .

Fig. 5 shows the normalized mean square error (NMSE) performance comparison between the proposed SD-based channel estimation and the conventional OMP-based channel estimation (i.e., using OMP to solve (9)), where the total number of instants  $Q$  for pilot transmission is  $Q = 96$  (i.e.,  $M = 6$  blocks). For SD-based channel estimation, we retain  $V = 8$  strongest elements as analyzed above for each channel component, while for OMP, we assume that the sparsity level of the beamspace channel is equal to  $V(L+1) = 24$ . From Fig. 5, we can observe that SD-based channel estimation enjoys much better NMSE performance than OMP-based channel estimation, especially when the uplink SNR is low (e.g., less than 15 dB). Since low SNR is the typical case in mmWave communications before beamforming [5], we can conclude that the proposed SD-based channel estimation is more attractive for mmWave massive MIMO systems.

Next, we evaluate the impact of different beamspace channel estimation schemes on beam selection. We adopt the interference-aware (IA) beam selection proposed in [15] as it can support the  $N_{\text{RF}} = K$  scenario, and the dimension-reduced digital precoder  $\mathbf{P}_r$  in (5) is selected as the zero-forcing (ZF) precoder. Fig. 6 provides the sum-rate performance of IA beam selection with different channels. We can observe that by utilizing the proposed SD-based channel estimation, IA beam selection can achieve better performance,

especially when the uplink SNR is low. More importantly, when the uplink SNR is moderate (e.g., 10 dB), IA beam selection with SD-based channel estimation, which only requires 16 RF chains, can achieve the sum-rate performance not far away from the fully digital ZF precoder with 256 RF chains and perfect channel state information (CSI).

## V. CONCLUSIONS

This paper investigates the beamspace channel estimation problem for mmWave massive MIMO systems with lens antenna array. Specifically, we first propose an adaptive selecting network with low cost to obtain the efficient measurements of beamspace channel. Then, we propose a SD-based channel estimation, where the key idea is to utilize the structural characteristics of mmWave beamspace channel to reliably detect the channel support. Analysis shows that the computational complexity of the proposed scheme is comparable with the classical LS algorithm. Simulation results verify that the proposed SD-based channel estimation can achieve much better NMSE performance than the conventional OMP-based channel estimation, especially in the low SNR region.

## 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] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, 2014.
- [3] J. Brady, N. Behdad, and A. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Ant. and Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [4] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3679–3684.
- [5] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [6] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE ICASSP*, Apr. 2015, pp. 2909–2913.
- [7] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE SPAWC Workshops*, Jun. 2015, pp. 146–150.
- [8] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "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.
- [9] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "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.
- [10] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [11] P. 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] 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.
- [13] C. A. Balanis, *Antenna theory: analysis and design*. John Wiley & Sons, 2012.
- [14] 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.
- [15] 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.