

# Wideband BeamSpace Channel Estimation for Millimeter-Wave MIMO Systems Relying on Lens Antenna Arrays

Xinyu Gao, *Student Member, IEEE*, Linglong Dai, *Senior Member, IEEE*, Shidong Zhou, *Senior Member, IEEE*, Akbar M. Sayeed, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

**Abstract**—Beamspace channel estimation is indispensable for millimeter-wave (mmWave) MIMO systems relying on lens antenna arrays for achieving substantially increased data rates, despite using a small number of radio-frequency (RF) chains. However, most of the existing beamspace channel estimation schemes have been designed for narrowband systems, while the rather scarce wideband solutions tend to assume that the sparse beamspace channel exhibits a common support in the frequency domain, which has a limited validity owing to the effect of beam squint caused by the wide bandwidth in practice. In this paper, we investigate the wideband beamspace channel estimation problem without the common support assumption. Specifically, by exploiting the effect of beam squint, we first prove that each path component of the wideband beamspace channel exhibits a unique frequency-dependent sparse structure. Inspired by this structure, we then propose a successive support detection (SSD)-based beamspace channel estimation scheme, which successively estimates all the sparse path components following the classical idea of successive interference cancellation (SIC). For each path component, its support at different frequencies is jointly estimated to improve the accuracy by utilizing the proved sparse structure, and its influence is removed to estimate the remaining path components. The performance analysis shows that the proposed SSD-based scheme can accurately estimate the wideband beamspace channel at a low complexity. Simulation results verify that the proposed SSD-based scheme enjoys a reduced pilot overhead, and yet achieves an improved channel estimation accuracy.

**Index Terms**—MIMO, millimeter-wave, lens antenna array, wideband beamspace channel estimation.

## I. INTRODUCTION

Millimeter-wave (mmWave) multiple-input multiple-output (MIMO) working at 30-300 GHz has been

X. Gao, L. Dai, and S. Zhou are with the Beijing National Research Center for Information Science and Technology (BNRIST), Department of Electronic Engineering, Beijing 100084, China (e-mails: xy-gao14@mails.tsinghua.edu.cn, daill@tsinghua.edu.cn, zhousd@tsinghua.edu.cn).

A. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706, USA (email: akbar@engr.wisc.edu).

L. Hanzo is with the Department of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK (email: lh@ecs.soton.ac.uk).

This work was supported by the National Natural Science Foundation of China for Outstanding Young Scholars (Grant No. 61722109), the National Science and Technology Major Project of China (Grant No. 2018ZX03001004-003), the Royal Academy of Engineering under the UK-China Industry Academia Partnership Programme Scheme (Grant No. UK-CAAPP/49), and the US National Science Foundation (Grant Nos. 1629713, 1703389, and 1548996).

recently recognized as a promising technique to substantially increase the data rates of wireless communications [1], since it can provide a very wide bandwidth (e.g., 2-5 GHz) [2]. However, in the conventional MIMO architecture working at sub-6 GHz cellular frequencies, each antenna requires a dedicated radio-frequency (RF) chain (including the digital-to-analog/analog-to-digital converter, mixer, and so on) [3], [4]. Employing this architecture in mmWave MIMO will lead to unaffordable hardware cost and power consumption due to the following two reasons [5]: 1) the number of antennas is usually very large to compensate for the severe path loss (e.g., 256 antennas may be used at mmWave frequencies instead of 8 antennas at cellular frequencies) [6]; 2) the power consumption of the RF chain is high due to the increased sampling rate (e.g., 250 mW/RF chain at mmWave frequencies, compared to 30 mW/RF chain at cellular frequencies) [7]. To solve this problem, mmWave MIMO relying on lens antenna array has been proposed [8]. By employing the lens antenna array (an electromagnetic lens with power focusing capability and a matching antenna array with elements located on the focal surface of the lens [9]), we can focus the signal power arriving from different directions on different antennas [10], and transform the mmWave MIMO channel from the spatial domain to its sparse beamspace representation (i.e., beamspace channel) [11]. This allows us to select a small number of power-focused beams for significantly reducing the effective MIMO dimension and the associated number of RF chains. Consequently, the high power consumption and hardware cost of mmWave MIMO systems can be mitigated [12]–[14].

To select the power-focused beams, a high-dimensional beamspace channel is required at the base station (BS). However, this is not a trivial task in mmWave MIMO systems relying on lens antenna arrays, since the number of RF chains is much smaller than the number of antennas so that we cannot directly observe the complete channel in the baseband [15]. To circumvent this problem, some beamspace channel estimation schemes have been proposed in [16]–[20]. For example, in [16], a training-based scheme is proposed. It first scans all the beams and only retains a few strong beams. Then, the least squares (LS) algorithm is employed for estimating the reduced-dimensional beamspace channel. In [17], a modified version of [16] is proposed, where the overhead of beam training is reduced by simultaneously scanning several beams with the help of power splitters at the BS. In [18], a support detection based scheme is proposed for further reducing the

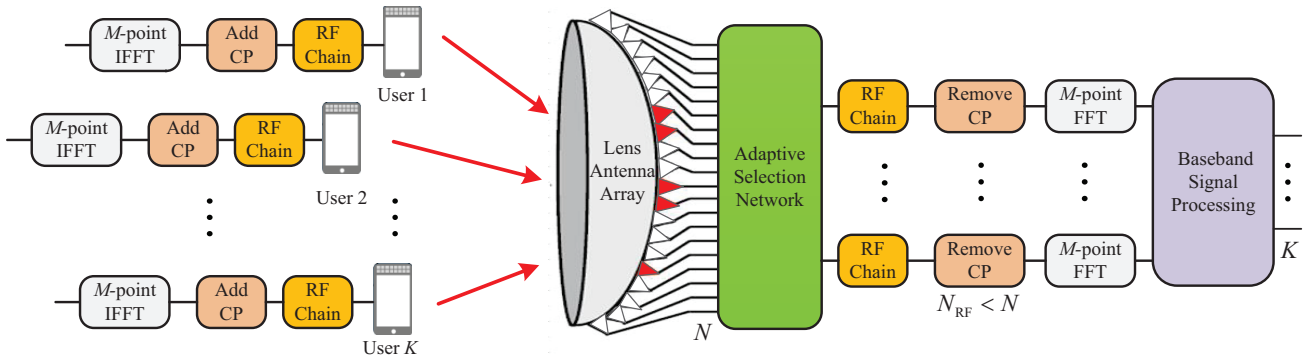


Fig. 1. Architecture of wideband mmWave MIMO-OFDM system relying on lens antenna array.

pilot overhead. It exploits the sparsity of the beamspace channel to directly estimate the channel support (i.e., the index set of nonzero elements in a sparse vector). However, all of these schemes have been designed for narrowband systems, while realistic mmWave MIMO systems are more likely to be of wideband nature for achieving high data rates. For wideband systems, there are only a few recent contributions. In [19], a simultaneous orthogonal matching pursuit (SOMP)-based scheme is proposed. It first regards the wideband beamspace channel estimation problem as a multiple measurement vector (MMV) problem associated with a common support (i.e., the channel support at different frequencies is assumed to be the same), and then solves it by the SOMP algorithm. In [20], an orthogonal matching pursuit (OMP)-based scheme is proposed. It first estimates the support of the wideband beamspace channel at some frequencies independently by the OMP algorithm. Then, it combines them into the common support at all frequencies. Unfortunately, the common support assumption in [19], [20] has limited validity in the practical wideband mmWave MIMO systems. As discussed in [21], the combination of a wide bandwidth and a large number of antennas will make the channel spreading factor defined in [21] larger than one, and the effect of “beam squint” becomes more obvious, where “beam squint” is used to imply that the indices of the power-focused beams are frequency-dependent [22]. As a result, the support of wideband beamspace channels also tends to be frequency-dependent, and the existing wideband solutions [19], [20] relying on the common support assumption will suffer from an obvious performance loss in practice.

In this paper, inspired by the classical successive interference cancellation (SIC) conceived for multi-user signal detection [23], we propose a successive support detection (SSD)-based wideband beamspace channel estimation scheme without the common support assumption. Specifically, the contributions of this paper can be summarized as follows:

1) By exploiting the effect of beam squint, we first prove that each path component of the wideband beamspace channel exhibits a unique frequency-dependent sparse structure. Specifically, for each sparse path component, we demonstrate that: i) its frequency-dependent support is uniquely determined by its spatial direction at the carrier frequency; ii) this spatial direction can be estimated by tentatively generating several beamspace windows (BWins) to capture the path power.

2) Inspired by the idea of SIC, we propose to decompose the wideband beamspace channel estimation problem into a series of sub-problems, each of which only considers a single path component. For each path component, its support observed at different frequencies is estimated jointly to improve the accuracy by utilizing the proved sparse structure, and then its influence is removed to estimate the remaining path components. The performance analysis shows that the proposed scheme can accurately estimate the wideband beamspace channel at a low complexity.

3) We provide extensive simulation results to verify the advantages of the proposed SSD-based scheme. We demonstrate that our scheme achieves a satisfactory channel estimation accuracy at a lower pilot overhead than the existing schemes. We also show that our wideband scheme performs well in narrowband systems.

The rest of the paper is organized as follows. In Section II, the system model of wideband mmWave MIMO-OFDM relying on lens antenna array is introduced, and the problem of wideband beamspace channel estimation is formulated when single-antenna users are considered. In Section III, the proposed SSD-based scheme is specified, together with its performance analysis. In Section IV, the proposed SSD-based scheme is extended to the scenario with multiple-antenna users. In Section V, our simulation results are provided to verify the advantages of the proposed SSD-based scheme. Finally, our conclusions are drawn in Section VI.

*Notation:* Lower-case and upper-case boldface letters  $\mathbf{a}$  and  $\mathbf{A}$  denote a vector and a matrix, respectively;  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^{-1}$ , and  $\mathbf{A}^\dagger$  denote the transpose, conjugate transpose, inverse, and pseudo inverse of matrix  $\mathbf{A}$ , respectively;  $\|\mathbf{A}\|_2$  and  $\|\mathbf{A}\|_F$  denote the spectral norm and Frobenius norm of matrix  $\mathbf{A}$ , respectively;  $\|\mathbf{a}\|_2$  denotes the  $l_2$ -norm of vector  $\mathbf{a}$ ;  $|a|$  denotes the amplitude of scalar  $a$ ;  $|\mathcal{S}|$  denotes the cardinality of set  $\mathcal{S}$ ;  $\mathbf{A}(\mathcal{S}, :)$  and  $\mathbf{A}(:, \mathcal{S})$  denote the sub-matrices of  $\mathbf{A}$  consisting of the rows and columns indexed by  $\mathcal{S}$ , respectively;  $\mathbf{a}(\mathcal{S})$  denotes the sub-vector of  $\mathbf{a}$  indexed by  $\mathcal{S}$ . Finally,  $\mathbf{I}_N$  is the identity matrix of size  $N \times N$ .

## II. SYSTEM MODEL

As shown in Fig. 1, we consider an uplink time division duplexing (TDD) wideband mmWave MIMO-OFDM system with  $M$  sub-carriers. The BS employs an  $N$ -element lens

antenna array and  $N_{\text{RF}}$  RF chains to simultaneously serve  $K$  users. In this section, we assume that each user employs single antenna, while in Section IV, multiple-antenna users will be considered. Next, we will first introduce the wideband beamspace channel. Then, the wideband beamspace channel estimation problem will be formulated.

#### A. Wideband beamspace channel

We commence with the wideband mmWave MIMO channel in the conventional spatial domain. To characterize the dispersive mmWave MIMO channel [24], we adopt the widely used Saleh-Valenzuela multipath channel model presented in the frequency domain. The  $N \times 1$  spatial channel  $\mathbf{h}_m$  of a certain user at sub-carrier  $m$  ( $m = 1, 2, \dots, M$ ) can be presented as [4], [21], [25]

$$\mathbf{h}_m = \sqrt{\frac{N}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \mathbf{a}(\varphi_{l,m}), \quad (1)$$

where  $L$  is the number of resolvable paths,  $\beta_l$  and  $\tau_l$  are the complex gain and the time delay of the  $l$ -th path, respectively,  $\varphi_{l,m}$  is the spatial direction at sub-carrier  $m$  defined as

$$\varphi_{l,m} = \frac{f_m}{c} d \sin \theta_l, \quad (2)$$

where  $f_m = f_c + \frac{f_s}{M} (m - 1 - \frac{M-1}{2})$  is the frequency of sub-carrier  $m$  with  $f_c$  and  $f_s$  representing the carrier frequency and the bandwidth (sampling rate), respectively,  $c$  is the speed of light,  $\theta_l$  is the physical direction, and  $d$  is the antenna spacing, which is usually designed according to the carrier frequency as  $d = c/2f_c$  [4]. Note that in narrowband mmWave systems with  $f_s \ll f_c$ , we have  $f_m \approx f_c$ , and  $\varphi_{l,m} \approx \frac{1}{2} \sin \theta_l$  is frequency-independent. However, in wideband mmWave systems,  $f_m \neq f_c$ , and  $\varphi_{l,m}$  is frequency-dependent. Finally,  $\mathbf{a}(\varphi_{l,m})$  is the array response vector of  $\varphi_{l,m}$ . For the typical  $N$ -element uniform linear array (ULA), we have  $\mathbf{a}(\varphi_{l,m}) = \frac{1}{\sqrt{N}} e^{-j2\pi\varphi_{l,m} \mathbf{p}_a}$ , where  $\mathbf{p}_a = [-\frac{N-1}{2}, -\frac{N+1}{2}, \dots, \frac{N-1}{2}]^T$  [4].

The spatial channel  $\mathbf{h}_m$  can be transformed to its beamspace representation by employing the lens antenna array, as shown in Fig. 1. Essentially, this lens antenna array plays the role of an  $N \times N$ -element spatial discrete Fourier transform (DFT) matrix  $\mathbf{U}_a^1$ , which contains the array response vectors of  $N$  orthogonal directions (beams) covering the entire space as [8]

$$\mathbf{U}_a = [\mathbf{a}(\bar{\varphi}_1), \mathbf{a}(\bar{\varphi}_2), \dots, \mathbf{a}(\bar{\varphi}_N)], \quad (3)$$

where  $\bar{\varphi}_n = \frac{1}{N} (n - \frac{N+1}{2})$  for  $n = 1, 2, \dots, N$  are the spatial directions pre-defined by the lens antenna array. Accordingly,

<sup>1</sup>The reason why the lens antenna array realizes the spatial DFT can be found in [10, Lemma 1]. Explicitly, it is shown that the power-focusing capability of the lens relies on the spatial phase shifters on the lens' aperture, which usually cannot be adjusted according to different frequencies. As a result, the response of the lens antenna array cannot be frequency-dependent as in (1). However, we would like to surmise that it may be possible but rather challenging to conceive a frequency-dependent lens antenna array capable of compensating for the effect of beam squint. In this case, the proposed SSD-based scheme can be further simplified to its narrowband version as we have proposed in [18], since the beamspace channel at different sub-carriers will have the common support.

the wideband beamspace channel  $\tilde{\mathbf{h}}_m$  at sub-carrier  $m$  can be presented by

$$\tilde{\mathbf{h}}_m = \mathbf{U}_a^H \mathbf{h}_m = \sqrt{\frac{N}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \tilde{\mathbf{c}}_{l,m}, \quad (4)$$

where  $\tilde{\mathbf{c}}_{l,m}$  denotes the  $l$ -th path component at sub-carrier  $m$  in the beamspace, and  $\tilde{\mathbf{c}}_{l,m}$  is determined by  $\varphi_{l,m}$  as

$$\begin{aligned} \tilde{\mathbf{c}}_{l,m} &= \mathbf{U}_a^H \mathbf{a}(\varphi_{l,m}) \\ &= [\Xi(\varphi_{l,m} - \bar{\varphi}_1), \Xi(\varphi_{l,m} - \bar{\varphi}_2), \dots, \Xi(\varphi_{l,m} - \bar{\varphi}_N)]^T, \end{aligned} \quad (5)$$

where  $\Xi(x) = \frac{\sin N\pi x}{\sin \pi x}$  is the Dirichlet sinc function [12].

Based on the power-focusing capability of  $\Xi(x)$  [12], [18], we know that most of the power of  $\tilde{\mathbf{c}}_{l,m}$  is focused on only a small number of elements. Additionally, due to the limited scattering in mmWave systems,  $L$  is also small [24], [26]. Therefore,  $\tilde{\mathbf{h}}_m$  should be a sparse vector [27]. However, since  $\varphi_{l,m}$  in (5) is frequency-dependent in wideband mmWave systems (i.e.,  $f_m \neq f_c$ ), the beam power distribution of the  $l$ -th path component should be different at different sub-carriers, i.e.,  $\tilde{\mathbf{c}}_{l,m_1} \neq \tilde{\mathbf{c}}_{l,m_2}$  for  $m_1 \neq m_2$ . This effect is termed as beam squint [21], which is a key difference between wideband and narrowband systems. For example, when we consider a narrowband system with  $\theta_l = -\pi/4$ ,  $N = 32$ , and  $f_c = 28$  GHz, the beam power distribution of the  $l$ -th path component is shown by the black line in Fig. 2, which is fixed. By contrast, when we extend this system to a wideband one with  $M = 128$  and  $f_s = 4$  GHz, the beam power distributions of  $\tilde{\mathbf{c}}_{l,1}$  and  $\tilde{\mathbf{c}}_{l,M}$  are shown by the blue line and red line in Fig. 2, respectively. We observe that  $\tilde{\mathbf{c}}_{l,1}$  only has a single strong beam  $\tilde{\mathbf{c}}_{l,1}$  (6), while  $\tilde{\mathbf{c}}_{l,M}$  has 2 strong beams, namely  $\tilde{\mathbf{c}}_{l,M}$  (4) and  $\tilde{\mathbf{c}}_{l,M}$  (5), which are different. Fig. 3 shows the effect of beam squint from another perspective, where the parameters are the same as in Fig. 2, and the curve indexed by  $n$  represents the power variation of the  $n$ th element (beam) of  $\tilde{\mathbf{c}}_{l,m}$  over frequency. We observe from Fig. 3 that in contrast to the narrowband systems where the power of each beam is frequency-independent [18], the power of each beam in wideband systems varies significantly over frequency. Due to beam squint and the fact that the beamspace channel is the summation of several resolvable path components, we can conclude that the support of the beamspace channel should be frequency-dependent, which is different from the common support assumption considered in the existing beamspace channel estimation schemes<sup>2</sup>.

#### B. Problem formulation

In TDD systems, the users are required to transmit pilot sequences to the BS for uplink channel estimation, and the channel is assumed to remain unchanged during this period [29], [30]. In this paper, we adopt the widely used orthogonal pilot transmission strategy, and therefore the channel estimation invoked for each user is independent [31]. Let us consider a specific user without loss of generality, and define  $s_{m,q}$  as its transmitted pilot at sub-carrier  $m$  and instant  $q$  (each

<sup>2</sup>It is worth pointing out that beam squint also exists in wideband mmWave MIMO systems using the conventional phased arrays [28]. The proposed channel estimation scheme in this paper can be also used in such systems.

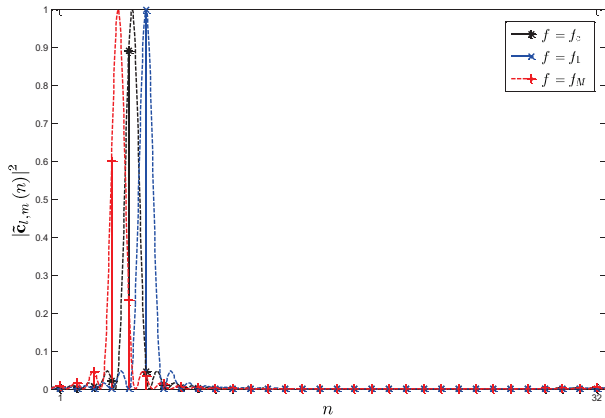


Fig. 2. Beam power distributions of the  $l$ -th path component in a narrowband system and in a wideband system.

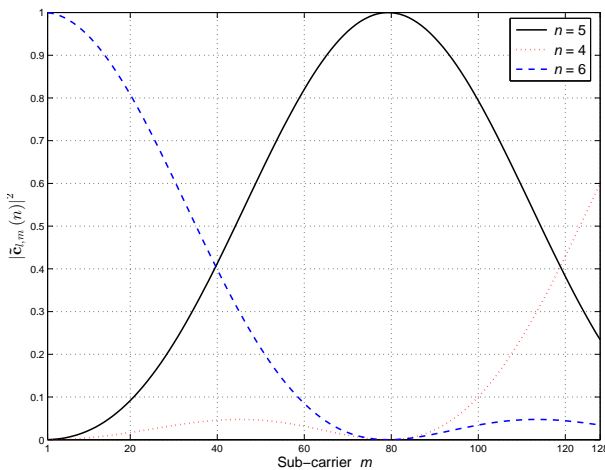


Fig. 3. Beam power variation over frequency.

user transmits one pilot per instant) before the  $M$ -point IFFT and cyclic prefix (CP) adding [20]. Then, as shown in Fig. 1, the  $N_{\text{RF}} \times 1$  received pilot vector  $\mathbf{y}_{m,q}$  at the BS after receiver combining (realized by the adaptive selection network [18]), CP removal, and  $M$ -point FFT can be presented as [20]

$$\mathbf{y}_{m,q} = \mathbf{W}_q \tilde{\mathbf{h}}_m s_{m,q} + \mathbf{W}_q \mathbf{n}_{m,q}, \quad m = 1, 2, \dots, M, \quad (6)$$

where  $\mathbf{W}_q$  of size  $N_{\text{RF}} \times N$  is the receiver combining matrix (fixed at different sub-carriers due to the analog hardware limitation [20]) and  $\mathbf{n}_{m,q} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$  of size  $N \times 1$  is the noise vector with  $\sigma^2$  representing the noise power. After  $Q$  instants of pilot transmission, we can obtain the overall measurement vector  $\tilde{\mathbf{y}}_m = [\mathbf{y}_{m,1}^T, \mathbf{y}_{m,2}^T, \dots, \mathbf{y}_{m,Q}^T]^T$  as

$$\tilde{\mathbf{y}}_m = \tilde{\mathbf{W}} \tilde{\mathbf{h}}_m + \mathbf{n}_m^{\text{eff}}, \quad m = 1, 2, \dots, M, \quad (7)$$

where we assume  $s_{m,q} = 1$  for  $q = 1, 2, \dots, Q$  without loss of generality [32], and define  $\mathbf{n}_m^{\text{eff}}$  as the effective noise vector. Furthermore, we define  $\tilde{\mathbf{W}} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_Q^T]^T$  of size  $QN_{\text{RF}} \times N$  as the overall combining matrix, which is designed according to the hardware realization of the adaptive selection network. For example, if the adaptive selection network is realized by low-cost 1-bit phase shifters

as in [18]<sup>3</sup>, the elements of  $\tilde{\mathbf{W}}$  can be randomly selected from the set  $\frac{1}{\sqrt{QN_{\text{RF}}}} \{-1, +1\}$  with equal probability. Here the normalization factor  $\frac{1}{\sqrt{QN_{\text{RF}}}}$  is used for guaranteeing that  $\tilde{\mathbf{W}}$  has unit-norm columns [34]. The reason we adopt a randomly selected matrix is that it has been shown to have a low mutual-column coherence, and therefore can be expected to achieve a high recovery accuracy according to well-established compressive sensing theory [35]. Finally, it should be noted that hardware impairments are indeed imposed on the adaptive selection network, leading to an element-wise gain/phase offset in  $\mathbf{W}_q$ , which cannot be fully captured in the estimated channel. This is a common problem inherent in most of the popular channel estimation schemes conceived for hybrid analog and digital architectures [4], since the analog modules (e.g., phase shifter network) are usually involved in the channel estimation procedure. Fortunately, since the gain/phase offsets are usually not serious in practice, the channel estimation accuracy degradation caused by hardware impairments will not be significant.

According to (7), we now can recover  $\tilde{\mathbf{h}}_m$  given  $\tilde{\mathbf{y}}_m$  and  $\tilde{\mathbf{W}}$ . Since  $\tilde{\mathbf{h}}_m$  is sparse, this problem can be solved relying on compressive sensing (CS) algorithms with a significantly reduced number of instants for pilot transmission (i.e.,  $Q \ll (N/N_{\text{RF}})$ ) [27], [36]. However, most of the existing schemes using CS algorithms have been designed for narrowband systems [16]–[18], while mmWave MIMO systems are more likely to be of wideband nature for achieving high data rates. For wideband systems, only the SOMP-based scheme [19] and the OMP-based scheme [20] have been proposed, but they assume that  $\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M$  share a common support, which is not strictly valid in practice due to the effect of beam squint, as shown in Fig. 2 and Fig. 3 [21].

### III. WIDEBAND BEAMSPACE CHANNEL ESTIMATION

In this section, we first explicitly demonstrate that the wideband beamspace channel exhibits a sparse structure. Then, we propose an efficient SSD-based scheme. Finally, the associated performance analysis is provided to quantify the advantages of our scheme.

#### A. Sparse structure of wideband beamspace channel

As shown in Fig. 2 and Fig. 3, the common support assumption is not strictly valid in practice due to the effect of beam squint. Fortunately, the wideband beamspace channel still exhibits a unique frequency-dependent sparse structure. This will be proved by the following lemmas, which constitute the basics of the proposed SSD-based scheme.

**Lemma 1.** Consider the  $l$ -th path component of the wideband beamspace channel. The frequency-dependent support  $\mathcal{T}_{l,m}$  of  $\tilde{\mathbf{c}}_{l,m}$  for  $m = 1, 2, \dots, M$  is uniquely determined by the

<sup>3</sup>It is worth pointing out that during data transmission, an adaptive selection network relying on 1-bit phase shifters can also be configured to realize conventional beam selection [12]. To achieve this, we can turn off some phase shifters to realize “unselect” [33] and set some phase shifters to shift the phase  $0^\circ$  to realize “select”.

spatial direction  $\varphi_{l,c}$  of the  $l$ -th path at the carrier frequency  $f_c$ , which is defined as  $\varphi_{l,c} = (f_c/c) d \sin \theta_l = (1/2) \sin \theta_l$ .

*Proof:* Based on the analysis in [18], the index of the strongest element  $n_{l,m}^*$  of  $\tilde{\mathbf{c}}_{l,m}$  is determined by  $\varphi_{l,m}$  as

$$n_{l,m}^* = \arg \min_n |\varphi_{l,m} - \bar{\varphi}_n|, \quad (8)$$

where  $\bar{\varphi}_n$  is defined in (3). Then, the support of  $\tilde{\mathbf{c}}_{l,m}$  can be obtained by

$$\mathcal{T}_{l,m} = \Theta_N \{n_{l,m}^* - \Omega, \dots, n_{l,m}^* + \Omega\}, \quad (9)$$

where  $\Theta_N(x) = \text{mod}_N(x-1) + 1$  is the mod function guaranteeing that all elements in  $\mathcal{T}_{l,m}$  belong to  $\{1, 2, \dots, N\}$ , and  $\Omega$  determines how much power can be preserved by assuming that  $\tilde{\mathbf{c}}_{l,m}$  is a sparse vector with support  $\mathcal{T}_{l,m}$ . For example, when  $N = 256$  and  $\Omega = 4$ , at least 96% of the power can be preserved [18]. The reasonable nature of (9) can be explained as follows. In practice,  $\varphi_{l,m}$  is arbitrary, which is usually different from the pre-defined beam directions  $\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N$ . In this case, the power of  $\tilde{\mathbf{c}}_{l,m}$  will be distributed across several beams. According to the properties of  $\Xi(x)$  in  $\tilde{\mathbf{c}}_{l,m}$ ,  $\Xi(x)$  is larger when  $x$  is closer to 0, and we know that the indices of these power-focused beams should be adjacent. The detailed proof can be found in [18, **Lemma 2**].

On the other hand, based on (2) and the definition of  $\varphi_{l,c}$ ,  $\varphi_{l,m}$  can be rewritten following [21] as

$$\varphi_{l,m} = \left\{ 1 + \frac{f_s}{M f_c} \left( m - 1 - \frac{M-1}{2} \right) \right\} \varphi_{l,c}, \quad (10)$$

which is only determined by  $\varphi_{l,c}$  ( $M, f_c, f_s$  are given system parameters). As a result, once  $\varphi_{l,c}$  is known, the support  $\mathcal{T}_{l,m}$  of  $\tilde{\mathbf{c}}_{l,m}$  for  $m = 1, 2, \dots, M$  can be obtained based on (8) and (9). ■

**Lemma 1** implies that  $\varphi_{l,c}$  is a crucial parameter for determining  $\mathcal{T}_{l,m}$  for  $m = 1, 2, \dots, M$ . In the following **Lemma 2**, we will provide some insights about how to estimate  $\varphi_{l,c}$ .

**Lemma 2.** Let us define  $\mathbf{C}_n = [\tilde{\mathbf{c}}_{1,1}, \tilde{\mathbf{c}}_{1,2}, \dots, \tilde{\mathbf{c}}_{1,M}]$ , where we assume  $\varphi_{l,c} = \bar{\varphi}_n$ . Then, the power of the  $s$ -th row  $\mathbf{C}_n(s, :)$  of  $\mathbf{C}_n$  can be calculated as

$$\|\mathbf{C}_n(s, :)\|_2^2 = \frac{M}{\alpha_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{n-s}{N} + \Delta\varphi \right) d\Delta\varphi, \quad (11)$$

where  $\alpha_n = f_s \bar{\varphi}_n / f_c$ . Moreover, if we define a beamspace window (BWin)  $\Upsilon_n = \Theta_N \{n - \Delta_n, \dots, n + \Delta_n\}$  centered around  $n$ , the ratio  $\gamma$  between the power of the sub-matrix  $\mathbf{C}_n(\Upsilon_n, :)$  and the power of  $\mathbf{C}_n$  can be presented as

$$\gamma = \frac{\|\mathbf{C}_n(\Upsilon_n, :)\|_F^2}{\|\mathbf{C}_n\|_F^2} = \frac{1}{\alpha_n} \sum_{i=-\Delta_n}^{\Delta_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{i}{N} + \Delta\varphi \right) d\Delta\varphi. \quad (12)$$

*Proof:* Based on (5), the power of the  $s$ -th row  $\mathbf{C}_n(s, :)$  of  $\mathbf{C}_n$  can be calculated as

$$\|\mathbf{C}_n(s, :)\|_2^2 = \sum_{m=1}^M \Xi^2(\varphi_{l,m} - \bar{\varphi}_s). \quad (13)$$

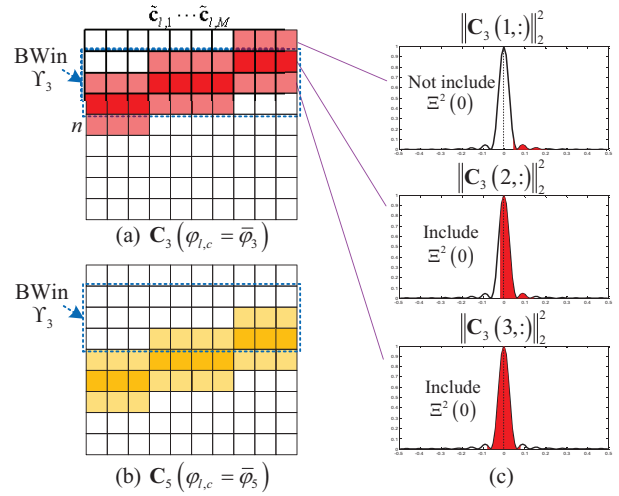


Fig. 4. Illustration of the power distribution: a)  $\mathbf{C}_3$ ; b)  $\mathbf{C}_5$ ; c)  $\mathbf{C}_3(1, :)$ ,  $\mathbf{C}_3(2, :)$ , and  $\mathbf{C}_3(3, :)$ .

Defining  $\Delta\varphi_m = \frac{f_s \varphi_{l,c}}{M f_c} (m - 1 - \frac{M-1}{2})$ , we can rewrite (13) based on (10) as

$$\begin{aligned} \|\mathbf{C}_n(s, :)\|_2^2 &= \sum_{m=1}^M \Xi^2(\varphi_{l,c} + \Delta\varphi_m - \bar{\varphi}_s) \\ &\stackrel{(a)}{=} \sum_{m=1}^M \Xi^2 \left( \frac{n-s}{N} + \Delta\varphi_m \right), \end{aligned} \quad (14)$$

where (a) is valid since  $\varphi_{l,c} = \bar{\varphi}_n$ . Note that  $M$  is usually a large number (e.g.,  $M = 512$ ). Therefore,  $\Delta\varphi_m$  is small and the summation in (14) can be well-approximated by its integral form as

$$\|\mathbf{C}_n(s, :)\|_2^2 = \frac{M}{\alpha_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{n-s}{N} + \Delta\varphi \right) d\Delta\varphi, \quad (15)$$

where the integral interval is determined by  $\Delta\varphi_1$  and  $\Delta\varphi_M$  with  $\frac{M-1}{M} \approx 1$ . Furthermore, based on (15), the power of  $\mathbf{C}_n(\Upsilon_n, :)$  can be written as

$$\begin{aligned} \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 &= \frac{M}{\alpha_n} \sum_{i \in \Upsilon_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{n-i}{N} + \Delta\varphi \right) d\Delta\varphi \\ &\stackrel{(a)}{=} \frac{M}{\alpha_n} \sum_{i=-\Delta_n}^{\Delta_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{i}{N} + \Delta\varphi \right) d\Delta\varphi, \end{aligned} \quad (16)$$

where (a) is due to the fact that  $\Upsilon_n$  is centered around  $n$ . On the other hand, since  $\sum_{s=1}^N \Xi^2(\varphi_{l,m} - \bar{\varphi}_s) = 1$ , the total power of  $\|\mathbf{C}_n\|_F^2$  is  $M$ . Then, the conclusions can be derived. ■

According to **Lemma 2**, we observe that if  $\varphi_{l,c} = \bar{\varphi}_n$ , the most power of  $\mathbf{C}_n$  can be captured by a carefully designed BWin  $\Upsilon_n$  centered around  $n$ . For example, given  $N = 256$ ,  $f_c = 28$  GHz,  $f_s = 4$  GHz,  $\varphi_{l,c} = \bar{\varphi}_1$ , we can capture  $\gamma \approx 92\%$  of the power of  $\mathbf{C}_1$  by using the BWin  $\Upsilon_1 = \Theta_{256} \{1 - 8, \dots, 1 + 8\}$  ( $\Delta_1 = 8$ ). On the other hand, if  $\varphi_{l,c} \neq \bar{\varphi}_1$ , e.g.,  $\varphi_{l,c} = \bar{\varphi}_{10}$ , using  $\Upsilon_1$  to capture the power of  $\mathbf{C}_{10}$  will lead to serious power leakage, where we only have  $\gamma \approx 47\%$ . This observation is further illustrated in Fig. 4 (a) and (b). Therefore, we can conclude that the BWin  $\Upsilon_n$  centered around  $n$  can be considered as a feature specialized for  $\varphi_{l,c} = \bar{\varphi}_n$ , which can be exploited for estimating  $\varphi_{l,c}$ .



The next problem is how to design  $\Delta_n$  in the Bwin  $\Upsilon_n$ . Note that our target is to estimate  $\varphi_{l,c}$  by using different BWins to capture the power of the  $l$ -th path component<sup>4</sup>, and we assume that  $\varphi_{l,c}$  belongs to the set  $\{\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N\}$  pre-defined by the lens antenna array (the corresponding quantization error is negligible when  $N$  is large, e.g.,  $N = 256$  [8]). For the specific case where  $\varphi_{l,c} = \bar{\varphi}_n$  (i.e.,  $\mathbf{C}_n = [\tilde{\mathbf{c}}_{l,1}, \tilde{\mathbf{c}}_{l,2}, \dots, \tilde{\mathbf{c}}_{l,M}]$ ),  $\varphi_{l,c}$  can only be correctly estimated when the following condition is satisfied

$$\|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 > \max_{n' \neq n} \left( \|\mathbf{C}_n(\Upsilon_{n'}, :)\|_F^2 \right). \quad (17)$$

In practice,  $\mathbf{C}_n$  may be corrupted by interference or noise. To overcome this problem,  $\Upsilon_n$  should be designed to capture the power  $\mathbf{C}_n$  as much as possible, which is formulated as

$$\Upsilon_n^* = \arg \max_{\Upsilon_n} \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2, \quad (18)$$

while  $\Upsilon_{n'}$  for  $n' \neq n$  should be designed to capture the power of  $\mathbf{C}_n$  as little as possible, leading to

$$\Upsilon_{n'}^* = \arg \min_{\Upsilon_{n'}} \|\mathbf{C}_n(\Upsilon_{n'}, :)\|_F^2. \quad (19)$$

Upon considering all the cases  $\varphi_{l,c} = \bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N$  and collecting the optimization problems related to  $\Upsilon_n$ , we conclude that  $\Upsilon_n$  should be designed to optimize the following problem

$$\max_{\Delta_n} \left( \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 - \max_{n' \neq n} \|\mathbf{C}_n(\Upsilon_{n'}, :)\|_F^2 \right), \quad (20)$$

where  $\mathbf{C}_{n'}$  is constructed with  $\varphi_{l,c} = \bar{\varphi}_{n'}$ , and we replace the optimization variable  $\Upsilon_n$  by  $\Delta_n$ , since designing  $\Upsilon_n$  is equivalent to designing  $\Delta_n$ . Due to the power-focusing capability of  $\Xi(x)$ , we know that  $\|\mathbf{C}_{n'}(\Upsilon_n, :)\|_F^2$  will be larger, if  $n'$  is closer to  $n$ . Therefore, the inner maximization in (20) can be presented as<sup>5</sup>

$$\max_{n' \neq n} \|\mathbf{C}_{n'}(\Upsilon_n, :)\|_F^2 = \|\mathbf{C}_{\Theta_N(n+1)}(\Upsilon_n, :)\|_F^2. \quad (21)$$

Based on (21) and **Lemma 2**, the target to maximize in (20) can be rewritten as

$$\begin{aligned} & \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 - \|\mathbf{C}_{\Theta_N(n+1)}(\Upsilon_n, :)\|_F^2 \quad (22) \\ & \stackrel{(a)}{\approx} \frac{M}{\alpha_n} \sum_{s=-\Delta_n}^{\Delta_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \left( \Xi^2\left(\frac{s}{N} + \Delta\varphi\right) - \Xi^2\left(\frac{s+1}{N} + \Delta\varphi\right) \right) d\Delta\varphi \\ & \stackrel{(b)}{=} \frac{M}{\alpha_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \left( \Xi^2\left(\frac{\Delta_n}{N} + \Delta\varphi\right) - \Xi^2\left(\frac{\Delta_n+1}{N} + \Delta\varphi\right) \right) d\Delta\varphi \\ & \stackrel{(c)}{=} \frac{M}{\alpha_n} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \left( \Xi^2\left(\frac{\Delta_n}{N} + \Delta\varphi\right) - \Xi^2\left(\frac{\Delta_n+1}{N} + \Delta\varphi\right) \right) d\Delta\varphi, \end{aligned}$$

where (a) is reasonable since  $\bar{\varphi}_{\Theta_N(n+1)} \approx \bar{\varphi}_n$  and  $\alpha_{\Theta_N(n+1)} \approx \alpha_n$  with large  $N$ , (b) is obtained by exchanging the orders of integral and summation, and (c) is true due to the fact that  $\Xi^2(x) = \Xi^2(-x)$ .

<sup>4</sup>Note that the method described above is heuristic. In Section V, we will verify that this method is simple and efficient. Designing the optimal method to estimate  $\varphi_{l,c}$  is also interesting, which will be left for our future work.

<sup>5</sup>It can be also written as  $\max_{n' \neq n} \|\mathbf{C}_{n'}(\Upsilon_n, :)\|_F^2 = \|\mathbf{C}_{\Theta_N(n-1)}(\Upsilon_n, :)\|_F^2$ , and the final results will be the same.

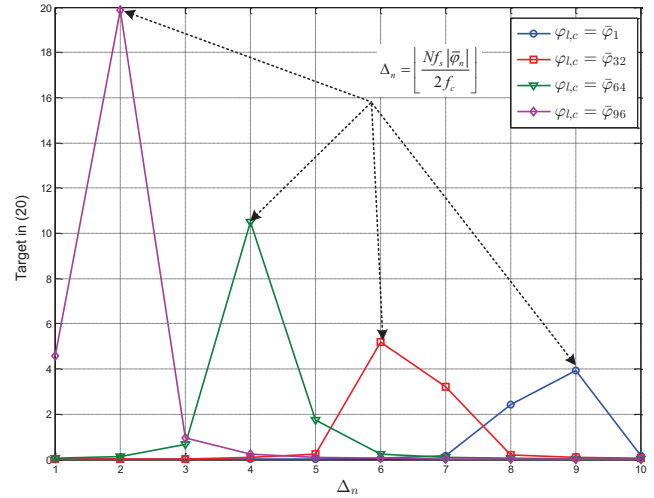


Fig. 5. Relationship between the target in (20) and  $\Delta_n$  when  $N = 256$ ,  $M = 128$ ,  $f_c = 28$  GHz, and  $f_s = 4$  GHz.

From (22), we know that the optimal  $\Delta_n$  should make the degradation from  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2\left(\frac{\Delta_n}{N} + \Delta\varphi\right) d\Delta\varphi$  to  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2\left(\frac{\Delta_n+1}{N} + \Delta\varphi\right) d\Delta\varphi$  the largest. Based on the power-focusing capability of  $\Xi(x)$ , we can conclude that if the integral includes  $\Xi^2(0)$ , it will be a large value. Otherwise, it should be small. An example is shown in Fig. 4 (c). Since  $\|\mathbf{C}_3(2, :)\|_2^2 = \frac{M}{\alpha_3} \int_{-\frac{\alpha_3}{2}}^{\frac{\alpha_3}{2}} \Xi^2\left(\frac{1}{N} + \Delta\varphi\right) d\Delta\varphi$  includes  $\Xi^2(0)$ , while  $\|\mathbf{C}_3(1, :)\|_2^2 = \frac{M}{\alpha_3} \int_{-\frac{\alpha_3}{2}}^{\frac{\alpha_3}{2}} \Xi^2\left(\frac{2}{N} + \Delta\varphi\right) d\Delta\varphi$  does not,  $\|\mathbf{C}_3(2, :)\|_2^2$  is much larger than  $\|\mathbf{C}_3(1, :)\|_2^2$ . This means that the largest degradation will happen when  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2\left(\frac{\Delta_n}{N} + \Delta\varphi\right) d\Delta\varphi$  includes  $\Xi^2(0)$ , while  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2\left(\frac{\Delta_n+1}{N} + \Delta\varphi\right) d\Delta\varphi$  does not. In other words,  $\Delta_n$  should satisfy

$$\left\lfloor \frac{\Delta_n}{N} - \frac{|\alpha_n|}{2} \right\rfloor = \left\lfloor \frac{\Delta_n}{N} - \frac{f_s |\bar{\varphi}_n|}{2f_c} \right\rfloor = 0, \quad (23)$$

where  $\lfloor x \rfloor$  returns the largest integer smaller than  $x$ . It indicates that the optimal  $\Delta_n$  is

$$\Delta_n = \left\lfloor \frac{N f_s |\bar{\varphi}_n|}{2f_c} \right\rfloor. \quad (24)$$

Note that the optimal  $\Delta_n$  has the similar form to the channel spreading factor defined in [21], which is utilized to quantify the severity of beam squint. However, they are actually two different parameters designed for different purposes and derived in different ways.

In Fig. 5, we plot the relationship between the target in (20) and  $\Delta_n$  when  $N = 256$ ,  $M = 128$ ,  $f_c = 28$  GHz, and  $f_s = 4$  GHz. We observe from Fig. 5 that the  $\Delta_n$  in (24) satisfies the requirement above and can be expected to achieve a good performance. Moreover, Fig. 5 also shows that the derived  $\Delta_n$  varies with different  $\varphi_{l,c}$  values (leading to different  $\mathbf{C}_n$  as defined in **Lemma 2**). The reason for this is that different  $\varphi_{l,c}$  values incur different degrees of power leakage due to beam squint. Therefore, we have to change  $\Delta_n$  to make sure that we can always capture most of the power of  $\mathbf{C}_n$ .

**Input:**Measurement matrix:  $\bar{\mathbf{Y}}$ Combining matrix:  $\bar{\mathbf{W}}$ Total number of channel paths:  $L$ **Initialization:**  $\mathbf{R} = \bar{\mathbf{Y}}$ **for**  $1 \leq l \leq L$ 1.  $\Upsilon_n = \Theta_N \{n - \Delta_n, \dots, n + \Delta_n\}$ ,  $\Delta_n = \left\lfloor \frac{N f_s |\bar{\varphi}_n|}{2 f_c} \right\rfloor$ 2.  $\mathbf{A}_l = \bar{\mathbf{W}}^H \mathbf{R}$ 3.  $n_{l,c}^* = \arg \max_n \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F^2}{|\Upsilon_n|}$ 4.  $\varphi_{l,c} = \bar{\varphi}_{n_{l,c}^*}$   
**for**  $1 \leq m \leq M$ 5.  $\varphi_{l,m} = \left\{ 1 + \frac{f_s}{M f_c} (m - 1 - \frac{M-1}{2}) \right\} \varphi_{l,c}$ 6.  $n_{l,m}^* = \arg \min_n |\varphi_{l,m} - \bar{\varphi}_n|$ 7.  $\mathcal{T}_{l,m} = \Theta_N \{n_{l,m}^* - \Omega, \dots, n_{l,m}^* + \Omega\}$ 8.  $\tilde{\mathbf{c}}_{l,m} = \mathbf{0}_{N \times 1}$ ,  $\tilde{\mathbf{c}}_{l,m}(\mathcal{T}_{l,m}) = \bar{\mathbf{W}}^\dagger(:, \mathcal{T}_{l,m}) \mathbf{r}_m$ 9.  $\mathbf{r}_m = \mathbf{r}_m - \bar{\mathbf{W}}(:, \mathcal{T}_{l,m}) \tilde{\mathbf{c}}_{l,m}(\mathcal{T}_{l,m})$ **end for****end for****for**  $1 \leq m \leq M$ 10.  $\tilde{\mathcal{T}}_m = \mathcal{T}_{1,m} \cup \mathcal{T}_{2,m} \cup \dots \cup \mathcal{T}_{L,m}$ 11.  $\tilde{\mathbf{h}}_m = \mathbf{0}_{N \times 1}$ ,  $\tilde{\mathbf{h}}_m(\tilde{\mathcal{T}}_m) = \bar{\mathbf{W}}^\dagger(:, \tilde{\mathcal{T}}_m) \bar{\mathbf{y}}_m$ **end for****Output:**Estimated beamspace channel:  $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M]$ **Algorithm 1:** SSD-based wideband beamspace channel estimation scheme.

### B. Proposed SSD-based scheme

Based on the sparse structure proved above, we propose an efficient SSD-based scheme to estimate the wideband beamspace channel. Its key idea is to decompose the total channel estimation problem into a series of sub-problems, each of which only considers a single path component. We first estimate the support of the strongest path component at all sub-carriers jointly. Then, its influence is removed for estimating the second strongest path component. This procedure is repeated until all path components have been considered.

To realize it, we first rewrite (7) as

$$\bar{\mathbf{Y}} = \bar{\mathbf{W}} \tilde{\mathbf{H}} + \mathbf{N}, \quad (25)$$

where  $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_M]$ ,  $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M]$ , and  $\mathbf{N} = [\mathbf{n}_1^{\text{eff}}, \mathbf{n}_2^{\text{eff}}, \dots, \mathbf{n}_M^{\text{eff}}]$ . Then, the pseudo-code of the proposed SSD-based scheme can be summarized in **Algorithm 1** and discussed as follows.

For the initialization, we set  $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M] = \bar{\mathbf{Y}}$ , where  $\mathbf{r}_m$  denotes the residual at sub-carrier  $m$ .

For the  $l$ -th path component, we first estimate  $\varphi_{l,c}$  based on **Lemma 2**. Specifically, in step 1, we generate  $N$  BWins  $\Upsilon_n = \Theta_N \{n - \Delta_n, \dots, n + \Delta_n\}$  with  $\Delta_n = \left\lfloor \frac{N f_s |\bar{\varphi}_n|}{2 f_c} \right\rfloor$  for  $n = 1, 2, \dots, N$ . Then, in step 2, we calculate the correlation matrix  $\mathbf{A}_l$  between  $\bar{\mathbf{W}}$  and  $\mathbf{R}$  as  $\mathbf{A}_l = \bar{\mathbf{W}}^H \mathbf{R}$ . Based on the low mutual coherence property of  $\bar{\mathbf{W}}$  (i.e.,  $\bar{\mathbf{W}}^H \bar{\mathbf{W}} \approx \mathbf{I}_N$ ) as in the classical OMP or SOMP algorithms [35], in step 3, we can utilize the  $N$  BWins to capture the power of  $\mathbf{A}_l$ , and

obtain the index  $n_{l,c}^*$  of the spatial direction of the  $l$ -th path component at the carrier frequency  $f_c$  as

$$n_{l,c}^* = \arg \max_n \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F^2}{|\Upsilon_n|}, \quad (26)$$

where we divide  $\|\mathbf{A}_l(\Upsilon_n, :)\|_F^2$  by  $|\Upsilon_n| = 2\Delta_n + 1$  to avoid that the large BWin captures more noise power. Finally, in step 4,  $\varphi_{l,c}$  is estimated as  $\varphi_{l,c} = \bar{\varphi}_{n_{l,c}^*}$ .

After  $\varphi_{l,c}$  has been estimated, the frequency-dependent support  $\mathcal{T}_{l,m}$  of the  $l$ -th path component for  $m = 1, 2, \dots, M$  can be obtained by **Lemma 1**. Specifically, in steps 5 and 6, we compute the spatial direction  $\varphi_{l,m}$  and the index  $n_{l,m}^*$  of the strongest element at sub-carrier  $m$  based on (10) and (8). Then, in step 7,  $\mathcal{T}_{l,m}$  can be obtained based on (9).

After the support estimation, we remove the influence of the  $l$ -th path component to estimate the remaining path components. Specifically, in step 8, based on  $\mathcal{T}_{l,m}$ , we estimate the nonzero elements of the  $l$ -th path component  $\tilde{\mathbf{c}}_{l,m}$  at sub-carrier  $m$  by the LS algorithm. Then, its influence is removed in step 9 by

$$\mathbf{r}_m = \mathbf{r}_m - \bar{\mathbf{W}}(:, \mathcal{T}_{l,m}) \tilde{\mathbf{c}}_{l,m}(\mathcal{T}_{l,m}). \quad (27)$$

The procedure above will then be repeated until the supports of all path components have been estimated. In the end, we estimate  $\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M$  independently. Specifically, in step 10, we formulate the complete support  $\tilde{\mathcal{T}}_m$  of  $\tilde{\mathbf{h}}_m$  as

$$\tilde{\mathcal{T}}_m = \mathcal{T}_{1,m} \cup \mathcal{T}_{2,m} \cup \dots \cup \mathcal{T}_{L,m}. \quad (28)$$

Then, in step 11, the nonzero elements of  $\tilde{\mathbf{h}}_m$  are estimated by the LS algorithm.

Note that the key difference between our scheme and the conventional schemes is the support detection. For example, for the OMP-based scheme, the support of wideband beamspace channel at different sub-carriers is estimated independently [20], which is vulnerable to noise. As a result, the detected support may be inaccurate, especially in the low SNR region [35]. For the SOMP-based scheme, the support at different sub-carriers is estimated jointly, but it assumes the common support [19]. Due to the effect of beam squint, this assumption will lead to serious performance loss, especially in the high SNR region. By contrast, in our scheme, we jointly recover the support without the common support assumption. By fully exploiting the frequency-dependent sparse structure of wideband beamspace channel, our scheme can be expected to achieve a higher accuracy. These conclusions will be further verified in Section V by simulation results.

In the end of this sub-section, we would like to point out that in the proposed SSD-based scheme, we assume that the number of resolvable paths  $L$  is known in advance. A suggested  $L$  can be obtained in advance by channel measurements [24]. For example, a measurement campaign carried out in New York City has shown that the average number of resolvable paths in a 28 GHz propagation environment is 6.8 with a standard deviation of 2.2. Therefore we can set  $L = 9$  in this case ignoring the low-probability cases of having  $L > 9$ . In practice, the actual number of resolvable paths should be a little lower than  $L$ , but this will not significantly affect the

performance of the proposed SSD-based scheme. Moreover, it is worth pointing out that the prior knowledge of  $L$  is not a necessary condition for our scheme. When  $L$  cannot be obtained in advance, we can borrow the idea of the classical OMP and SOMP algorithms, and run the proposed SSD-based scheme several times [36]. Specifically, during the  $t$ -th run ( $t = 1, 2, \dots$ ), we set  $L = t$ , and define the channel estimation update as  $\frac{1}{MN} \left\| \tilde{\mathbf{H}}^{(t)} - \tilde{\mathbf{H}}^{(t-1)} \right\|_F^2$ , where  $\tilde{\mathbf{H}}^{(t-1)}$  and  $\tilde{\mathbf{H}}^{(t)}$  represent the previous estimated channel at the  $(t-1)$ -th run and the current estimated channel at the  $t$ -th run, respectively. If the update is smaller than a threshold  $\zeta$  (e.g.,  $\zeta = 0.1$ ), then we will terminate the procedure. In Section V, we will verify that by utilizing this method, the proposed SSD-based scheme can still achieve a satisfactory accuracy without the knowledge of  $L$ .

### C. Performance analysis

In this sub-section, we will prove that the proposed SSD-based scheme is capable of correctly estimating the key parameters  $\varphi_{l,c}$  for  $l = 1, 2, \dots, L$  with a certain probability.

To do this, we first rewrite  $\tilde{\mathbf{H}}$  in (25) as  $\tilde{\mathbf{H}} = \mathbf{T}\mathbf{Z}$ . Here, we define  $\mathbf{Z}$  of size  $\sum_{n=1}^N |\Upsilon_n| \times M$  as an enlarged version of  $\tilde{\mathbf{H}}$ , which can be presented as  $\mathbf{Z} = [\mathbf{Z}_1^H, \mathbf{Z}_2^H, \dots, \mathbf{Z}_N^H]^H$  with

$$\mathbf{Z}_n = \begin{cases} \tilde{\mathbf{H}}(\Upsilon_n, :), & \text{if } n \in \{n_1, n_2, \dots, n_L\}, \\ \mathbf{0}_{|\Upsilon_n| \times M}, & \text{if } n \notin \{n_1, n_2, \dots, n_L\}, \end{cases} \quad (29)$$

and  $\tilde{\mathbf{H}}(\Upsilon_n, :) = \beta_l \mathbf{C}_{n_l}(\Upsilon_n, :)$ . Here, for the  $l$ -th path component, we assume that  $\varphi_{l,c} = \bar{\varphi}_{n_l}$  and that all its power can be captured by  $\mathbf{C}_{n_l}(\Upsilon_n, :)$  (this assumption only leads to a negligible performance loss, as we have proved in **Lemma 2**). Correspondingly,  $\mathbf{T}$  of size  $N \times \sum_{n=1}^N |\Upsilon_n|$  is the transformation matrix. More specifically,  $\mathbf{T}$  can be presented as  $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_N]$ , where  $\mathbf{T}_n$  is of size  $N \times |\Upsilon_n|$  and its  $i$ -th column ( $i = 1, 2, \dots, |\Upsilon_n|$ ) only has a single nonzero element 1 at the location  $\Upsilon_n(i)$  with  $\Upsilon_n(i)$  representing the  $i$ -th element selected from the set  $\Upsilon_n$ . By utilizing this transformation, we can transfer each path component in  $\tilde{\mathbf{H}}$  to a specific block in  $\mathbf{Z}$ , as illustrated in Fig. 6, where different blocks are non-overlapping.

According to the definitions of  $\mathbf{Z}$  and  $\mathbf{T}$ , we can rewrite (25) as

$$\tilde{\mathbf{Y}} = [\bar{\mathbf{W}}(:, \Upsilon_1), \bar{\mathbf{W}}(:, \Upsilon_2), \dots, \bar{\mathbf{W}}(:, \Upsilon_N)] \mathbf{Z} + \mathbf{N}. \quad (30)$$

Then, the key matrix  $\mathbf{A}_l(\Upsilon_n, :)$  used for estimating  $\varphi_{l,c}$  (i.e., step 2 of **Algorithm 1**) can be presented as

$$\begin{aligned} \mathbf{A}_l(\Upsilon_n, :) & \\ &= \bar{\mathbf{W}}^H(:, \Upsilon_n) \tilde{\mathbf{Y}} \\ &= \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N} + \sum_{i=1}^N \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}^H(:, \Upsilon_i) \mathbf{Z}_i. \end{aligned} \quad (31)$$

Next, we define a pair of auxiliary parameters  $\mu$  and  $\mu_B$  as

$$\mu \triangleq \max_{1 \leq n \leq N} \max_{i, j \in \Upsilon_n, i \neq j} |\bar{\mathbf{W}}^H(:, i) \bar{\mathbf{W}}(:, j)|, \quad (32)$$

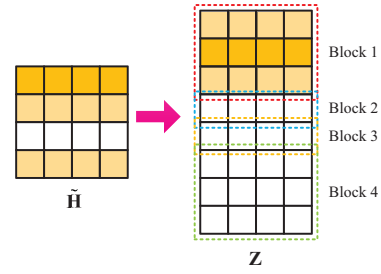


Fig. 6. Transformation from  $\tilde{\mathbf{H}}$  to  $\mathbf{Z}$ .

and

$$\mu_B \triangleq \max_{1 \leq i, j \leq N, i \neq j} \frac{1}{\sqrt{|\Upsilon_i| |\Upsilon_j|}} \|\bar{\mathbf{W}}^H(:, \Upsilon_i) \bar{\mathbf{W}}(:, \Upsilon_j)\|_2, \quad (33)$$

respectively. Note that  $\mu$  is exactly the same as the sub-coherence of the dictionary  $\bar{\mathbf{W}}$  in compress sensing theory [37], while  $\mu_B$  can be regarded as a generalized version of the block coherence introduced in [37].

Then, based on the discussion above, we have the following **Lemma 3**.

**Lemma 3.** For the  $l$ -th path component, assume that  $\varphi_{l,c} = \bar{\varphi}_{n_l}$  and that

$$\begin{aligned} & \frac{(1 - (|\Upsilon_{n_l}| - 1)\mu - \mu_B |\Upsilon_{n_l}|) |\beta_l| \|\mathbf{C}_{n_l}(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \\ & \geq 2\sigma\sqrt{\alpha M} + 2\mu_B \sum_{n_i \in \mathcal{L} \setminus n_l} \sqrt{|\Upsilon_{n_i}|} |\beta_i| \|\mathbf{C}_{n_i}(\Upsilon_{n_i}, :)\|_F \end{aligned} \quad (34)$$

for some constant  $\alpha$ , where  $\mathcal{L} = \{n_1, n_2, \dots, n_L\}$ . Then, with a probability exceeding

$$\prod_{n=1}^N \left(1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}\right), \quad (35)$$

the proposed SSD-based scheme can correctly estimate  $\varphi_{l,c}$ .

*Proof:* To prove **Lemma 3**, we first list two useful lemmas, which have been proved in [38].

**Lemma 4.** Let  $\mathbf{u}$  be a Gaussian random vector of size  $d \times 1$ . Assuming that  $\mathbf{u}$  has a mean of  $\mathbf{0}$  and a covariance of  $\mathbf{I}_d$ , we have

$$\Pr \left\{ \|\mathbf{u}\|_2^2 \geq t^2 \right\} \leq 0.8dt^{d-2}e^{-t^2/2}. \quad (36)$$

*Proof:* Please refer to [38, **Lemma 4**]. ■

**Lemma 5.** Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$  be  $M$  jointly Gaussian random vectors. Let us assume that  $\mathbb{E}(\mathbf{v}_m) = \mathbf{0}$  for  $m = 1, 2, \dots, M$ , but that the covariances of the vectors are unspecified and that the vectors are not necessarily independent. Then, we have

$$\begin{aligned} & \Pr \left\{ \|\mathbf{v}_1\|_2 \leq c_1, \|\mathbf{v}_2\|_2 \leq c_2, \dots, \|\mathbf{v}_M\|_2 \leq c_M \right\} \\ & \geq \prod_{m=1}^M \Pr \left\{ \|\mathbf{v}_m\|_2 \leq c_m \right\}. \end{aligned} \quad (37)$$

*Proof:* Please refer to [38, **Lemma 3**]. ■

Based on **Lemma 4** and **Lemma 5**, we have the following **Lemma 6**.

**Lemma 6.** Let us assume that each column  $\mathbf{n}_m^{\text{eff}}$  of  $\mathbf{N}$  in (25) is a Gaussian random vector of size  $QN_{\text{RF}} \times 1$  with a mean



of  $\mathbf{0}$  and a covariance of  $\sigma^2 \mathbf{I}_{Q_{N_{\text{RF}}}}$ . Then, we have

$$\Pr \left( \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right) \geq 1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}, \quad (38)$$

where we define

$$\tau_n^2 = \sigma^2 (1 + (|\Upsilon_n| - 1) \mu) \alpha, \quad (39)$$

and  $\alpha$  is a constant value introduced to guarantee that  $0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2} \leq 1$  for  $n = 1, 2, \dots, N$ .

*Proof:* Note that  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}}$  is also a Gaussian random vector with a mean of  $\mathbf{0}$  and a covariance of  $\sigma^2 \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$ . Let us now define an auxiliary vector  $\mathbf{u}$  as

$$\mathbf{u} \triangleq \frac{1}{\sigma} (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n))^{-\frac{1}{2}} \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}}. \quad (40)$$

We know that  $\mathbf{u}$  should be a Gaussian random vector of size  $|\Upsilon_n| \times 1$  with a mean of  $\mathbf{0}$  and a covariance of  $\mathbf{I}_{|\Upsilon_n|}$ . As a result, we have

$$\begin{aligned} & \Pr \left\{ \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right\} \\ &= \Pr \left\{ \sigma^2 \left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n))^{\frac{1}{2}} \mathbf{u} \right\|_2^2 \leq \tau_n^2 \right\} \\ &\stackrel{(a)}{\geq} \Pr \left\{ \sigma^2 \left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)) \right\|_2 \|\mathbf{u}\|_2^2 \leq \tau_n^2 \right\}, \end{aligned} \quad (41)$$

where (a) is due to the spectral norm property of a matrix [39]. Note that all the diagonal elements of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  are equal to 1, while all the off-diagonal elements of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  have amplitudes smaller than  $\mu$  according to the definition in (32). Therefore, by the Gershgorin circle theorem [39], we know that the largest singular value of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  should be upper-bounded by  $1 + (|\Upsilon_n| - 1) \mu$ , which means that

$$\left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)) \right\|_2 \leq 1 + (|\Upsilon_n| - 1) \mu. \quad (42)$$

Correspondingly, we have

$$\begin{aligned} & \Pr \left\{ \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right\} \\ &\geq \Pr \left\{ \|\mathbf{u}\|_2^2 \leq \frac{\tau_n^2}{\sigma^2 (1 + (|\Upsilon_n| - 1) \mu)} \right\}. \end{aligned} \quad (43)$$

According to the definition of  $\tau_n^2$  in (39), the right side of (43) can be rewritten as

$$\Pr \left\{ \|\mathbf{u}\|_2^2 \leq \frac{\tau_n^2}{\sigma^2 (1 + (|\Upsilon_n| - 1) \mu)} \right\} = \Pr \left\{ \|\mathbf{u}\|_2^2 \leq \alpha \right\}. \quad (44)$$

Based on **Lemma 4**, we can conclude that

$$\begin{aligned} & \Pr \left( \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right) \\ &\geq 1 - \Pr \left\{ \|\mathbf{u}\|_2^2 \geq \alpha \right\} \\ &= 1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}. \end{aligned} \quad (45)$$

which completes the proof.  $\blacksquare$

Next, we continue the proof of **Lemma 3**. For the  $l$ -th path component,  $\varphi_{l,c}$  can only be correctly estimated if

$$\frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}}, \quad (46)$$

where  $\mathbf{A}_l(\Upsilon_n, :)$  is given by (31). Let us consider a specific event

$$\text{B} = \left\{ \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2, n = 1, 2, \dots, N \right\}. \quad (47)$$

Based on **Lemma 5** and **Lemma 6**, we know that event B will occur with a probability exceeding (35). When it occurs, the right side of (46) can be upper-bounded by

$$\begin{aligned} & \max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}} \\ &= \max_{n \notin \mathcal{L}} \frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N} + \sum_{n_i \in \mathcal{L}} \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F}{\sqrt{|\Upsilon_n|}} \\ &\leq \max_{n \notin \mathcal{L}} \frac{\|\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N}\|_F}{\sqrt{|\Upsilon_n|}} \\ &\quad + \max_{n \notin \mathcal{L}} \frac{\sum_{n_i \in \mathcal{L}} \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \right\|_2 \|\mathbf{Z}_{n_i}\|_F}{\sqrt{|\Upsilon_n|}}. \end{aligned} \quad (48)$$

Since  $\mathbf{N}$  has  $M$  Gaussian random columns and event B occurs, we know that

$$\|\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N}\|_F \leq \sqrt{M} \tau_n \stackrel{(a)}{\leq} \sigma \sqrt{\alpha M |\Upsilon_n|}, \quad (49)$$

where (a) is true due to the fact that  $\mu \leq 1$ . Combining this result and the definition of  $\mu_B$  in (33), (48) can be further upper-bounded by

$$\max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}} \leq \sigma \sqrt{\alpha M} + \mu_B \sum_{n_i \in \mathcal{L}} \sqrt{|\Upsilon_{n_i}|} \|\mathbf{Z}_{n_i}\|_F. \quad (50)$$

On the other hand, the left side of (46) can be lower-bounded by

$$\begin{aligned} & \frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \\ &= \frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \mathbf{N} + \sum_{n_i \in \mathcal{L}} \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F}{\sqrt{|\Upsilon_{n_l}|}} \\ &\geq \frac{1}{\sqrt{|\Upsilon_{n_l}|}} \left( \left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_l}) \mathbf{Z}_{n_l} \right\|_F - \left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \mathbf{N} \right\|_F \right. \\ &\quad \left. - \left\| \sum_{n_i \in \mathcal{L} \setminus n_l} \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F \right). \end{aligned} \quad (51)$$

Also according to the Gershgorin circle theorem as we have used in **Lemma 6**, the first term of (51) can be lower-bounded by

$$\frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_l}) \mathbf{Z}_{n_l} \right\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \frac{(1 - (|\Upsilon_{n_l}| - 1) \mu)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F. \quad (52)$$

Since event B occurs, the second term of (51) is lower-bounded by  $-\sigma \sqrt{\alpha M}$ . Finally, similar to the manipulation in (50), the third term of (51) can be lower-bounded by

$$-\mu_B \left( \sum_{n_i \in \mathcal{L}} \sqrt{|\Upsilon_{n_i}|} \|\mathbf{Z}_{n_i}\|_F - \sqrt{|\Upsilon_{n_l}|} \|\mathbf{Z}_{n_l}\|_F \right). \quad (53)$$

Upon combining these results, we have

$$\frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \frac{(1 - (|\Upsilon_{n_l}| - 1)\mu)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F - \sigma\sqrt{\alpha M} - \mu_B \left( \sum_{n_i \in \mathcal{L}} \sqrt{|\Upsilon_{n_i}|} \|\mathbf{Z}_{n_i}\|_F - \sqrt{|\Upsilon_{n_l}|} \|\mathbf{Z}_{n_l}\|_F \right). \quad (54)$$

Considering (50) and (54) together, we can conclude that when

$$\frac{(1 - (|\Upsilon_{n_l}| - 1)\mu - \mu_B |\Upsilon_{n_l}|)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F \geq 2\sigma\sqrt{\alpha M} + 2\mu_B \sum_{n_i \in \mathcal{L} \setminus n_l} \sqrt{|\Upsilon_{n_i}|} \|\mathbf{Z}_{n_i}\|_F, \quad (55)$$

$\varphi_{l,c}$  can be correctly estimated with a probability exceeding (35). Substituting the fact  $\|\mathbf{Z}_{n_l}\|_F = |\beta_l| \|\mathbf{C}_{n_l}(\Upsilon_{n_l}, :)\|_F$  in (29) to (55), we can finally complete this proof. ■

Next, we give some insights of **Lemma 3**. From (35), we observe that when  $\alpha$  belongs to the feasible region which makes  $0 \leq \prod_{n=1}^N (1 - 0.8|\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}) \leq 1$ , the probability monotonically increases with the increased  $\alpha$ . In addition, when  $\alpha$  is large enough, the term  $\alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}$  approaches 0, and the probability is close to 1. Based on these facts, we can conclude from **Lemma 3** that when the noise power  $\sigma^2$  is large, the allowed  $\alpha$  should be small when  $\mu$ ,  $\mu_B$ , and the power of each path component  $|\beta_i|^2 \|\mathbf{C}_{n_i}(\Upsilon_{n_i}, :)\|_F^2$  ( $i = 1, 2, \dots, L$ ) are given. Therefore, the probability of correctly estimating  $\varphi_{l,c}$  is low. Moreover, when the number of instants  $Q$  for pilot transmission is large,  $\bar{\mathbf{W}}$  of size  $QN_{\text{RF}} \times N$  can be expected to have lower  $\mu$  and  $\mu_B$  [35], leading to a larger allowed  $\alpha$ . As a result, the probability of correctly estimating  $\varphi_{l,c}$  should be high.

The conclusions above are further verified by Fig. 7. The simulation parameters of Fig. 7 are set as follows. We consider a wideband mmWave MIMO-OFDM system with  $N = 256$ ,  $f_c = 28$  GHz,  $f_s = 4$  GHz, and  $M = 512$ . The channel is assumed to have  $L = 3$  paths with  $\varphi_{1,c} = \bar{\varphi}_{192} = 0.2480$  ( $\theta_1 \approx \pi/6$ ),  $\varphi_{2,c} = \bar{\varphi}_{38} = -0.3535$  ( $\theta_2 \approx -\pi/4$ ), and  $\varphi_{3,c} = \bar{\varphi}_{239} = 0.4316$  ( $\theta_3 \approx -\pi/3$ ), and all path components have the same normalized power  $|\beta_i|^2 \|\mathbf{C}_{n_i}(\Upsilon_{n_i}, :)\|_F^2 = M$  ( $i = 1, 2, 3$ ). Finally, we generate two overall combining matrices  $\bar{\mathbf{W}}$ . The first one with  $\mu = 0.0606$  and  $\mu_B = 0.0247$  is obtained when  $Q = 16$ , while the second one with  $\mu = 0.0736$  and  $\mu_B = 0.0271$  is obtained when  $Q = 12$ . Fig. 7 shows the probability of correctly estimating  $\varphi_{1,c}$ , where we observe the trends consistent with the conclusions of **Lemma 3**. Moreover, Fig. 7 also verifies that the derived lower-bound in (35) is tight, especially when the SNR is high.

In the end, we would like to point out that when the pilot overhead  $Q$  is small (i.e.,  $\mu$  and  $\mu_B$  in (34) are large), the inequality in (34) may be meaningless ( $\sqrt{\alpha} < 0$ ), and the lower-bound in (35) is unavailable. Although  $Q$  being small is not the typical case for channel estimation, deriving the universal result is still of great interest. We will try to solve this challenging problem in our future work.

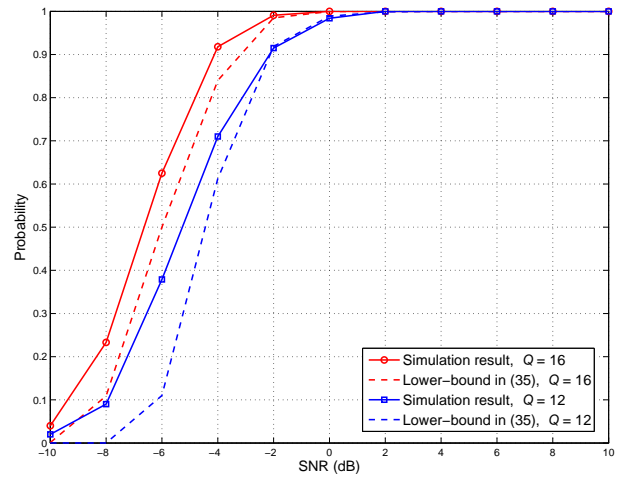


Fig. 7. The probability of correctly estimating  $\varphi_{1,c}$ .

#### D. Complexity analysis

In this subsection, we evaluate the complexity of the proposed SSD-based scheme in terms of the number of complex multiplications. According to **Algorithm 1**, we observe that the complexity is dominated by steps 2, 3, 8, 9, 11.

In step 2, we need to compute the multiplication between  $\bar{\mathbf{W}}^H$  of size  $N \times QN_{\text{RF}}$  and  $\mathbf{R}$  of size  $QN_{\text{RF}} \times M$ , which has a complexity in order of  $\mathcal{O}(NN_{\text{RF}}MQ)$ . In step 3, the power of  $N$  sub-matrices  $\mathbf{A}_l(\Upsilon_n, :)$  for  $n = 1, 2, \dots, N$  is calculated. This can be solved by calculating the power of  $N$  rows of  $\mathbf{A}_l$  in advance, where the complexity is in order of  $\mathcal{O}(NM)$ . In step 8, the pseudo inverse of  $\bar{\mathbf{W}}(:, \mathcal{T}_{l,m})$  of size  $QN_{\text{RF}} \times (2\Omega + 1)$ , and the multiplication between  $\bar{\mathbf{W}}^\dagger(:, \mathcal{T}_{l,m})$  of size  $(2\Omega + 1) \times QN_{\text{RF}}$  and  $\mathbf{r}_m$  of size  $QN_{\text{RF}} \times 1$  are required. Therefore, this step involves a complexity in order of  $\mathcal{O}(N_{\text{RF}}Q\Omega^2)$ . In step 9, we compute the multiplication between  $\bar{\mathbf{W}}(:, \mathcal{T}_{l,m})$  and  $\tilde{\mathbf{c}}_{l,m}(\mathcal{T}_{l,m})$  of size  $(2\Omega + 1) \times 1$  at a complexity in order of  $\mathcal{O}(N_{\text{RF}}Q\Omega)$ . Finally, in step 11, the LS algorithm is used again like in step 8, where  $\bar{\mathbf{W}}(:, \tilde{\mathcal{T}}_m)$  is of size  $QN_{\text{RF}} \times |\tilde{\mathcal{T}}_m|$  and  $\tilde{\mathbf{y}}_m$  is of size  $QN_{\text{RF}} \times 1$ . As a result, this step has the complexity in order of  $\mathcal{O}(N_{\text{RF}}Q|\tilde{\mathcal{T}}_m|^2)$ , where  $|\tilde{\mathcal{T}}_m| \leq L(2\Omega + 1)$ .

Note that steps 2, 3, 8, 9 are executed  $L$  times, and step 11 is executed only once. Therefore, the overall complexity of the proposed SSD-based scheme can be summarized as

$$\mathcal{O}(NML) + \mathcal{O}(MN_{\text{RF}}QL\Omega^2) + \mathcal{O}(MN_{\text{RF}}QL^2\Omega^2). \quad (56)$$

By contrast, the complexity of both the OMP-based and SOMP-based schemes can be presented as  $\mathcal{O}(MN_{\text{RF}}QL^3\Omega^3) + \mathcal{O}(NMN_{\text{RF}}QL\Omega)$  [19], [20]. Note that  $\Omega$  is usually much smaller than  $N$  (e.g.,  $\Omega = 4 \ll N = 256$ ) as proved in [18], we can conclude that the complexity of the proposed SSD-based scheme is lower than that of the conventional OMP-based and SOMP-based schemes.

#### IV. EXTENSION TO MULTIPLE-ANTENNA USERS

In this section, we will discuss how to extend the proposed SSD-based scheme to the scenario where each user employs an  $U$ -element lens antenna array like the BS.

### A. Problem formulation

In this case, the  $N \times U$  spatial channel  $\mathbf{H}_m$  between the BS and a certain user at sub-carrier  $m$  can be presented as [4]

$$\mathbf{H}_m = \sqrt{\frac{NU}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \mathbf{a}(\varphi_{l,m}) \mathbf{b}^H(\psi_{l,m}), \quad (57)$$

where  $\psi_{l,m} = \left\{1 + \frac{f_s}{Mf_c} (m-1 - \frac{M-1}{2})\right\} \psi_{l,c}$  is the spatial direction at the user side similar to  $\varphi_{l,m}$  in (2),  $\psi_{l,c} = \frac{1}{2} \sin \vartheta_l$  is the spatial direction at the carrier frequency  $f_c$  with  $\vartheta_l$  representing the corresponding physical direction, and  $\mathbf{b}(\psi_{l,m})$  is the  $U \times 1$  array response vector of  $\psi_{l,m}$ . For ULA, we have  $\mathbf{b}(\psi_{l,m}) = \frac{1}{\sqrt{U}} e^{-j2\pi\psi_{l,m} \mathbf{p}_b}$ , where  $\mathbf{p}_b = [-\frac{U-1}{2}, -\frac{U+1}{2}, \dots, \frac{U-1}{2}]^T$ .

Accordingly, the wideband beamspace channel  $\tilde{\mathbf{H}}_m$  to be estimated can be presented by [40]

$$\tilde{\mathbf{H}}_m = \mathbf{U}_a^H \mathbf{H}_m \mathbf{U}_b, \quad (58)$$

where  $\mathbf{U}_b = [\mathbf{b}(\bar{\psi}_1), \mathbf{b}(\bar{\psi}_2), \dots, \mathbf{b}(\bar{\psi}_U)]$  is the  $U \times U$  spatial DFT matrix realized by the lens antenna array at the user side and  $\bar{\psi}_u = \frac{1}{U} (u - \frac{U+1}{2})$  for  $u = 1, 2, \dots, U$  are the corresponding pre-defined spatial directions.

To estimate  $\tilde{\mathbf{H}}_m$ , we assume that each user uses only a single RF chain (user is likely to use cheaper hardware with lower power consumption than the BS [41]) to transmit orthogonal pilot sequences in the uplink, and adopts the adaptive selection network as shown in Fig. 1 for precoding the pilot sequences. Then, similar to (6), the received pilot vector  $\mathbf{y}_{m,q}$  for a certain user at sub-carrier  $m$  and instant  $q$  can be written as

$$\mathbf{y}_{m,q} = \mathbf{W}_q \tilde{\mathbf{H}}_m \mathbf{f}_q s_{m,q} + \mathbf{W}_q \mathbf{n}_{m,q}, \quad m = 1, 2, \dots, M, \quad (59)$$

where  $\mathbf{f}_q$  of size  $U \times 1$  is the precoding vector. Like  $\mathbf{W}_q$ ,  $\mathbf{f}_q$  is also fixed at different sub-carriers, and its elements can be randomly selected from the set  $\{-1, +1\}$  with equal probability if they are realized by low-cost 1-bit phase shifters as in [18]. By assuming  $s_{m,q} = 1$  and exploiting the relationship  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$  [15], we can rewrite (59) as

$$\mathbf{y}_{m,q} = (\mathbf{f}_q^T \otimes \mathbf{W}_q) \tilde{\tilde{\mathbf{h}}}_m + \mathbf{W}_q \mathbf{n}_{m,q}, \quad m = 1, 2, \dots, M, \quad (60)$$

where  $\otimes$  denotes the Kronecker product, and  $\tilde{\tilde{\mathbf{h}}}_m$  of size  $NU \times 1$  is defined as  $\tilde{\tilde{\mathbf{h}}}_m = \text{vec}(\tilde{\mathbf{H}}_m)$ .

After  $Q$  instants of pilot transmission, the overall measurement vector  $\bar{\mathbf{y}}_m = [\mathbf{y}_{m,1}^T, \mathbf{y}_{m,2}^T, \dots, \mathbf{y}_{m,Q}^T]^T$  similar to (7) can be obtained as

$$\bar{\mathbf{y}}_m = \bar{\Phi} \tilde{\tilde{\mathbf{h}}}_m + \mathbf{n}_m^{\text{eff}}, \quad m = 1, 2, \dots, M, \quad (61)$$

where  $\bar{\Phi}$  of size  $QN_{\text{RF}} \times NU$  is defined as

$$\bar{\Phi} = [(\mathbf{f}_1^T \otimes \mathbf{W}_1)^T, (\mathbf{f}_2^T \otimes \mathbf{W}_2)^T, \dots, (\mathbf{f}_Q^T \otimes \mathbf{W}_Q)^T]^T. \quad (62)$$

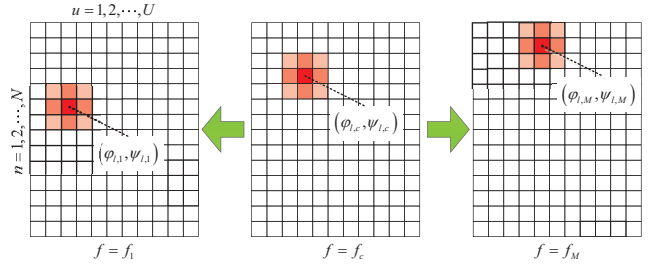


Fig. 8. Power distribution of the  $l$ -th path component  $\tilde{\mathbf{c}}_{l,m} \tilde{\mathbf{d}}_{l,m}^H$  in the frequency domain.

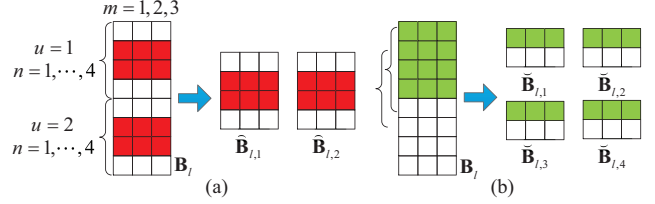


Fig. 9. Illustration of dividing  $\mathbf{B}_l$  into blocks when  $N = 4$ ,  $U = 2$ , and  $M = 3$ : a)  $U$  blocks  $\tilde{\mathbf{B}}_{l,u}$  for  $u = 1, 2$ ; b)  $N$  blocks  $\tilde{\mathbf{B}}_{l,n}$  for  $n = 1, 2, 3, 4$ .

### B. Extension of the proposed SSD-based scheme

To extend the proposed SSD-based scheme to the scenario with multiple-antenna users, we first rewrite (58) based on (57) as

$$\begin{aligned} \tilde{\mathbf{H}}_m &= \sqrt{\frac{NU}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \mathbf{U}_a^H \mathbf{a}(\varphi_{l,m}) \mathbf{b}^H(\psi_{l,m}) \mathbf{U}_b \\ &= \sqrt{\frac{NU}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \tilde{\mathbf{c}}_{l,m} \tilde{\mathbf{d}}_{l,m}^H, \end{aligned} \quad (63)$$

where  $\tilde{\mathbf{d}}_{l,m}$  is determined by  $\psi_{l,m}$  as

$$\begin{aligned} \tilde{\mathbf{d}}_{l,m} &= \mathbf{U}_b^H \mathbf{b}(\psi_{l,m}) \\ &= [\Xi(\psi_{l,m} - \bar{\psi}_1), \Xi(\psi_{l,m} - \bar{\psi}_2), \dots, \Xi(\psi_{l,m} - \bar{\psi}_U)]^T. \end{aligned} \quad (64)$$

From (63), we observe that the  $l$ -th path component  $\tilde{\mathbf{c}}_{l,m} \tilde{\mathbf{d}}_{l,m}^H$  of  $\tilde{\mathbf{H}}_m$  now should be a matrix exhibiting 2D sparsity as shown in Fig. 8, whose power is focused on a few rows and columns. Moreover, since  $\tilde{\mathbf{d}}_{l,m}$  shares the same structure as  $\tilde{\mathbf{c}}_{l,m}$ , we can also derive the following conclusions by extending **Lemma 1** and **Lemma 2**: 1) the row support  $\mathcal{T}_{l,m}$  (the index set of power-focused rows) and column support  $\mathcal{S}_{l,m}$  (the index set of power-focused columns) of  $\tilde{\mathbf{c}}_{l,m} \tilde{\mathbf{d}}_{l,m}^H$  can be uniquely determined by  $\varphi_{l,c}$  and  $\psi_{l,c}$ , respectively; 2)  $\varphi_{l,c}$  can be estimated by utilizing  $N$  row BWins  $\Upsilon_n = \Theta_N \{n - \Delta_n, \dots, n + \Delta_n\}$  with  $\Delta_n = \lfloor Nf_s |\bar{\varphi}_n| / 2f_c \rfloor$  for  $n = 1, 2, \dots, N$ , while  $\psi_{l,c}$  can be estimated by  $U$  column BWins  $X_u = \Theta_U \{u - \Delta_u, \dots, u + \Delta_u\}$  with  $\Delta_u = \lfloor Uf_s |\bar{\psi}_u| / 2f_c \rfloor$  for  $u = 1, 2, \dots, U$ .

Based on the analysis above, we can extend the proposed SSD-based scheme to estimate  $\tilde{\mathbf{H}}_m$ . To do this, we first rewrite (61) as

$$\bar{\mathbf{Y}} = \bar{\Phi} \tilde{\tilde{\mathbf{H}}} + \mathbf{N}, \quad (65)$$

where  $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_M]$ ,  $\tilde{\tilde{\mathbf{H}}} = [\tilde{\tilde{\mathbf{h}}}_1, \tilde{\tilde{\mathbf{h}}}_2, \dots, \tilde{\tilde{\mathbf{h}}}_M]$ , and  $\mathbf{N} = [\mathbf{n}_1^{\text{eff}}, \mathbf{n}_2^{\text{eff}}, \dots, \mathbf{n}_M^{\text{eff}}]$ . Then, we replace  $\bar{\mathbf{W}}$ ,  $\tilde{\mathbf{H}}$ , and  $\mathbf{A}_l$  in **Algorithm 1** by  $\bar{\Phi}$ ,  $\tilde{\tilde{\mathbf{H}}}$ , and  $\mathbf{B}_l = \bar{\Phi}^H \mathbf{R}$ , respectively.

After that, we estimate  $\varphi_{l,c}$  and  $\psi_{l,c}$  based on  $\Upsilon_n$  and  $X_u$  like steps 1-4. Specifically, based on the low mutual coherence property of  $\tilde{\Phi}$  (i.e.,  $\tilde{\Phi}^H \tilde{\Phi} \approx \mathbf{I}_{NU}$ ) as in the classical OMP or SOMP algorithms [35] and the relationship  $\tilde{\mathbf{h}}_m(n + (u-1)N) = \tilde{\mathbf{H}}_m(n, u)$ , we can divide  $\mathbf{B}_l$  into  $U$  blocks as shown in Fig. 9 (a). The  $u$ -th block  $\hat{\mathbf{B}}_{l,u}$  of size  $N \times M$  contains the rows belonging to the set  $\{1+(u-1)N, 2+(u-1)N, \dots, N+(u-1)N\}$  of  $\mathbf{B}_l$ . Then,  $\varphi_{l,c}$  can be estimated as

$$\varphi_{l,c} = \bar{\varphi}_{n_{l,c}^*}, n_{l,c}^* = \arg \max_n \sum_{u=1}^U \frac{\|\hat{\mathbf{B}}_{l,u}(\Upsilon_n, :)\|_F^2}{|\Upsilon_n|}. \quad (66)$$

Alternatively, we can also divide  $\mathbf{B}_l$  into  $N$  blocks as shown in Fig. 9 (b). The  $n$ -th block  $\tilde{\mathbf{B}}_{l,n}$  of size  $U \times M$  contains the rows belonging to the set  $\{n, n+N, \dots, n+(U-1)N\}$  of  $\mathbf{B}_l$ . Then,  $\psi_{l,c}$  can be estimated as

$$\psi_{l,c} = \bar{\psi}_{u_{l,c}^*}, u_{l,c}^* = \arg \max_u \sum_{n=1}^N \frac{\|\tilde{\mathbf{B}}_{l,n}(X_u, :)\|_F^2}{|X_u|}. \quad (67)$$

After  $\varphi_{l,c}$  and  $\psi_{l,c}$  have been estimated, we can calculate  $\varphi_{l,m}$  and  $\psi_{l,m}$  based on their definitions, and the row support  $\mathcal{T}_{l,m}$  and column support  $\mathcal{S}_{l,m}$  of the  $l$ -th path component  $\tilde{\mathbf{c}}_{l,m} \tilde{\mathbf{d}}_{l,m}^H$  at sub-carrier  $m$  can be obtained like steps 5-7 as

$$\mathcal{T}_{l,m} = \Theta_N \{n_{l,m}^* - \Omega, \dots, n_{l,m}^* + \Omega\}, \quad (68)$$

$$\mathcal{S}_{l,m} = \Theta_U \{u_{l,m}^* - \Omega, \dots, u_{l,m}^* + \Omega\}, \quad (69)$$

respectively, where we define

$$n_{l,m}^* = \arg \min_n |\varphi_{l,m} - \bar{\varphi}_n|, \quad (70)$$

$$u_{l,m}^* = \arg \min_u |\psi_{l,m} - \bar{\psi}_u|. \quad (71)$$

Once  $\mathcal{T}_{l,m}$  and  $\mathcal{S}_{l,m}$  have been acquired, the support  $\mathcal{D}_{l,m}$  of  $l$ -th path component can be directly calculated by

$$\mathcal{D}_{l,m} = \{n + (u-1)N \mid n \in \mathcal{T}_{l,m}, u \in \mathcal{S}_{l,m}\}, \quad (72)$$

and the influence of this path component can be removed like steps 8 and 9. Repeating this procedure until all path components have been considered, we can finally obtain the overall support of  $\tilde{\mathbf{h}}_m$  for  $m = 1, 2, \dots, M$  and estimate the corresponding nonzero elements by the LS algorithm like steps 10 an 11.

In the end, we would like to point out that in practice, the users are more likely to employ a conventional antenna array, since the lens antenna array is usually bulky at the time of writing. In this case, the power of wideband beamspace channel at a specific sub-carrier will be focused on a small number of rows instead of a small number of low-dimensional sub-matrices. This property allows us to further simplify our scheme. Specifically, when the conventional antenna array is employed at the user side, we do not have to estimate the column support  $\mathcal{S}_{l,m}$  any more. After we have estimated the row support  $\mathcal{T}_{l,m}$  by (68) and (70), the support  $\mathcal{D}_{l,m}$  of the  $l$ -th path component can be directly obtained as  $\mathcal{D}_{l,m} = \{n + (u-1)N \mid n \in \mathcal{T}_{l,m}, u \in \{1, 2, \dots, U\}\}$ .

## V. SIMULATION RESULTS

In this section, we first consider a wideband mmWave MIMO-OFDM system, where the BS equips an  $N = 256$ -element lens antenna array and  $N_{\text{RF}} = 8$  RF chains to serve  $K = 8$  single-antenna users. The carrier frequency is  $f_c = 28$  GHz, the number of sub-carriers is  $M = 512$ , and the bandwidth is  $f_s = 4$  GHz<sup>6</sup>. The spatial channel of each user in (1) is generated as follows [19]: 1)  $L = 3$ ; 2)  $\beta_l \sim \mathcal{CN}(0, 1)$ ; 3)  $\theta_l \sim \mathcal{U}(-\pi/2, \pi/2)$ ; 4)  $\tau_l \sim \mathcal{U}(0, 20\text{ns})$  and  $\max_l \tau_l = 20\text{ns}$ . We define the SNR for channel estimation as  $1/\sigma^2$ . Finally, we use the normalized mean square error (NMSE) to quantify the accuracy of channel estimation for each user, which is mathematically defined as

$$\mathbb{E} \left( \frac{1}{M} \sum_{m=1}^M \frac{\|\tilde{\mathbf{h}}_m - \tilde{\mathbf{h}}_m^e\|_2^2}{\|\tilde{\mathbf{h}}_m\|_2^2} \right), \quad (73)$$

where  $\tilde{\mathbf{h}}_m^e$  is the estimated beamspace channel at sub-carrier  $m$ .

We first compare the proposed SSD-based scheme and the conventional wideband schemes. Fig. 10 shows the NMSE of channel estimation against the SNR, where for all schemes we use  $Q = 16$  instants per user for pilot transmission. For the SSD-based scheme, we set  $\Omega = 4$  following the suggestion in [18]. For both the OMP-based [20] and the SOMP-based [19] schemes, we assume that the sparsity level is  $L(2\Omega + 1) = 27 < N_{\text{RF}}Q = 128$ . We also consider the oracle LS scheme as our benchmark, where the support of the wideband beamspace channel at different sub-carriers is assumed to be perfectly known. Note that for all the schemes mentioned above, we regard the elements of wideband beamspace channel having indices outside the support as zeros.

We observe from Fig. 10 that the accuracy of the OMP-based scheme is not satisfactory when the SNR is low, since it ignores the potential sparse structure of the wideband beamspace channel which can be exploited to suppress the noise. On the other hand, the accuracy of the SOMP-based scheme deteriorates when the SNR is high. This is because that the common support assumption is not strictly valid in wideband systems due to the effect of beam squint. By contrast, the proposed SSD-based scheme enjoys a much higher accuracy than the OMP-based and the SOMP-based schemes in all considered SNR regions, since it can fully exploit the sparse structure of the wideband beamspace channel. Actually, the proposed SSD-based scheme has already achieved the NMSE quite close to that of the oracle LS scheme. Moreover, Fig. 10 also shows that when the SNR is high (e.g., from 20-30 dB), there is a NMSE floor for all schemes. This can be explained by the fact that although the nonzero elements of the wideband beamspace channel can be estimated accurately at the sufficiently high SNR, the error induced by regarding the elements with low power as zeros does not vanish. Finally,

<sup>6</sup>Note that for future wireless communications such as 5G, the sub-carrier spacing of OFDM can be adjusted from 15 KHz to 240 KHz [42]. Therefore,  $M = 512$  is capable to support a large bandwidth, e.g., 4 GHz. Moreover, it is worth pointing out that the impact of  $M$  on different beamspace channel estimation schemes is negligible, since the NMSE (or sum-rate) is calculated by averaging over  $M$ .

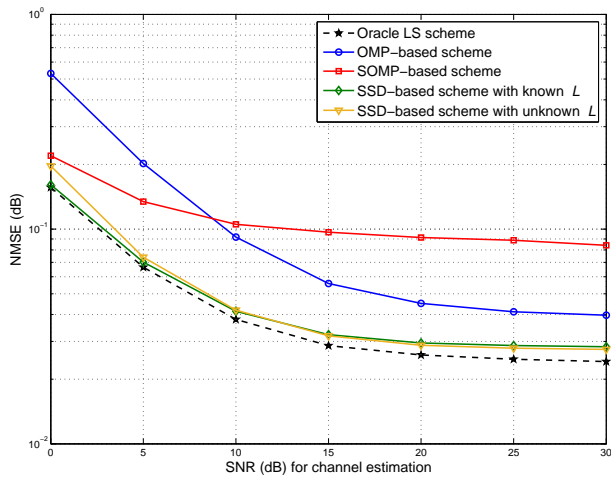
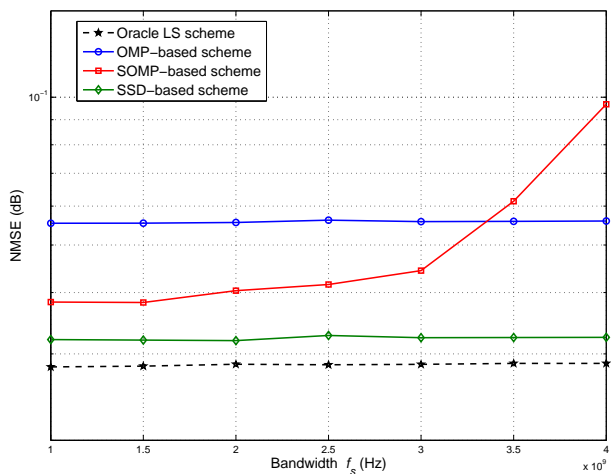


Fig. 10. NMSE comparison against the SNR for channel estimation.

Fig. 11. NMSE comparison against the bandwidth  $f_s$ .

we also observe from Fig. 10 that by utilizing the method described in the end of Section III-B, the proposed SSD-based scheme with unknown  $L$  (we empirically set  $\zeta = 0.1$ ) can achieve the accuracy quite close to the one with known  $L$ . This indicates that the prior knowledge of  $L$  is actually not necessary in the proposed SSD-based scheme.

Fig. 11 shows the NMSE comparison against the bandwidth  $f_s$ , where the SNR is set as 15 dB and the other simulation parameters are the same as those in Fig. 10. We observe from Fig. 11 that when  $f_s$  is low (e.g., 1 GHz), the effect of beam squint is less pronounced and the SOMP-based scheme can also achieve the satisfactory performance. However, as  $f_s$  increases, the SOMP-based scheme becomes more and more inaccurate. When  $f_s$  is high enough (e.g., 4 GHz), its performance becomes even worse than that of the OMP-based scheme. This is due to the fact that for large  $f_s$ , the support of the wideband beamspace channel at different sub-carriers will be more divergent, and the common support assumption leads to more serious accuracy degradation. By contrast, we observe that the proposed SSD-based scheme is robust to  $f_s$ . This indicates that our scheme works well even if the effect of beam squint is not pronounced.

Fig. 12 shows the achievable sum-rate of the wideband beam selection proposed in [21] along with different beamspace

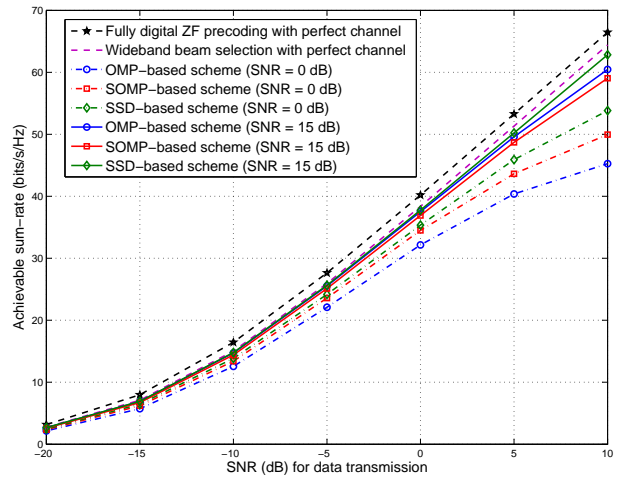
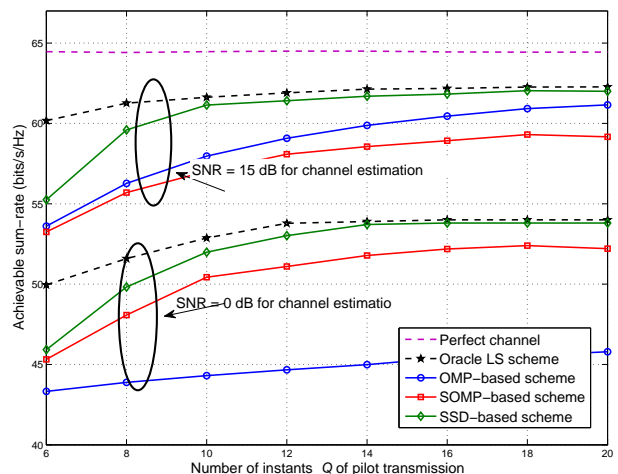


Fig. 12. Sum-rate comparison against the SNR for data transmission.

Fig. 13. Sum-rate comparison against the number of instants  $Q$  of pilot transmission.

channel estimation schemes, where the simulation parameters are the same as those in Fig. 10. We observe from Fig. 12 that by utilizing the proposed SSD-based scheme, the system achieves a higher sum-rate, especially when the SNR for channel estimation is low (e.g., 0 dB). Since the SNR for channel estimation is usually low in TDD systems due to the limited transmit power of users, we can conclude that our scheme is attractive in practice. Moreover, Fig. 12 also shows that when the SNR for channel estimation is moderate (e.g., 15 dB), the wideband beam selection using the SSD-based scheme achieves a sum-rate close to the one with perfect channel. Finally, we observe that the performance order of all schemes changes with the channel estimation SNR. This is due to the fact that when the SNR is high during data transmission, the sum-rate performance of wideband beam selection is dominated by channel estimation error.

Fig. 13 shows the impact of the number of instants  $Q$  of pilot transmission on different beamspace channel estimation schemes, where the SNR for data transmission is set as 10 dB. From Fig. 13, we observe that to achieve the same sum-rate, the number of instants  $Q$  required by the proposed SSD-based scheme is much lower than the conventional schemes both in the low and the moderate SNR regions. Therefore, we can



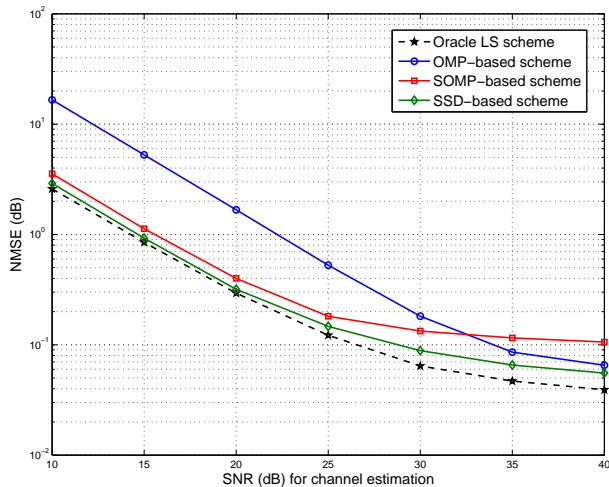


Fig. 14. NMSE comparison in the case with multiple-antenna users.

also conclude that the proposed SSD-based scheme achieves satisfactory performance at a low pilot overhead.

Finally, in Fig. 14 we evaluate the NMSE performance of the proposed SSD-based scheme in the case of multiple-antenna users. The simulation parameters are set as follows: 1) each user employs an  $U = 32$ -element lens antenna array; 2)  $\vartheta_l \sim \mathcal{U}(-\pi/2, \pi/2)$ ; 3)  $Q = 128$  instants per user (this value is larger than that in Fig. 10 since we need to estimate a much higher-dimensional channel  $\tilde{\mathbf{h}}_m = \text{vec}(\tilde{\mathbf{H}}_m)$  with more nonzero elements); 4) the other parameters are the same as those in Fig. 10. From Fig. 14, we observe the trends similar to those in Fig. 10, i.e., the SSD-based scheme enjoys a higher accuracy than the conventional schemes and achieves the NMSE close to the oracle LS scheme. This verifies that our scheme still performs well in the case of multiple-antenna users. Moreover, we would like to point out that in the case of multiple-antenna users, the SNR required for channel estimation to achieve a satisfactory NMSE is usually higher than in the case of single-antenna users. However, as we can see from Fig. 12, even if the SNR is not high enough (e.g., 0 dB or 15 dB) during channel estimation, wideband beam selection can still achieve a satisfactory sum-rate. The reason for this is that for data transmission only the reduced-dimensional beamspace channel having a much smaller size is effective. Although the NMSE performance may not be good enough for channel estimation at low SNRs, the reduced-dimensional beamspace channel's estimate is already accurate enough for data transmission. Therefore, in practice we do not have to estimate the beamspace channel so accurately at the cost of requiring a high SNR.

## VI. CONCLUSIONS

This paper investigated the wideband beamspace channel estimation problem for mmWave MIMO systems relying on lens antenna arrays. Specifically, we first proved that each path component of the wideband beamspace channel exhibits a unique frequency-dependent sparse structure. Then, by exploiting this sparse structure, we proposed an efficient SSD-based beamspace channel estimation scheme, where both single-antenna users and multiple-antenna users were considered. The

performance analysis showed that our scheme can accurately estimate the beamspace channel at a low complexity. The simulation results verified that: i) our scheme achieves a better NMSE performance than existing schemes in all considered SNR regions; ii) our scheme performs well even if the effect of beam squint is not pronounced; iii) our scheme considerably reduces the pilot overhead. In our future work, we will extend the proposed SSD-based scheme to 3D mmWave MIMO systems, where the elevation directions are also considered.

## REFERENCES

- [1] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.
- [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] 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.
- [4] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] N. Behdad and A. Sayeed, "Continuous aperture phased MIMO: Basic theory and applications," in *Proc. Allerton Conference*, Sep. 2010, pp. 1196–1203.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [16] J. Hogan and A. Sayeed, "Beam selection for performance-complexity optimization in high-dimension MIMO systems," in *Proc. CISS*, Mar. 2016, pp. 337–342.
- [17] 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.
- [18] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.

- [19] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [20] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [21] J. Brady and A. Sayeed, "Wideband communication with high-dimensional arrays: New results and transceiver architectures," in *Proc. IEEE ICC Workshops*, Jun. 2015, pp. 1042–1047.
- [22] X. Lin, S. Wu, L. Kuang, Z. Ni, X. Meng, and C. Jiang, "Estimation of sparse massive MIMO-OFDM channels with approximately common support," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1179–1182, Jan. 2017.
- [23] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [24] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [25] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.
- [26] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [27] W. U. Bajwa, J. Haupt, A. 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.
- [28] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [29] 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.
- [30] 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.
- [31] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [32] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. SPAWC Workshops*, Jun. 2015, pp. 146–150.
- [33] P. Chen, C. Argyropoulos, and A. Al, "Terahertz antenna phase shifters using integrally-gated graphene transmission-lines," *IEEE Trans. Antennas Propagat.*, vol. 61, no. 4, pp. 1528–1537, Apr. 2013.
- [34] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [35] 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.
- [36] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
- [37] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [38] Z. Ben-Haim, Y. C. Eldar *et al.*, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1032–1047, Sep. 2011.
- [39] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012.
- [40] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [41] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [42] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.