A Block Sparsity Based Estimator for mmWave Massive MIMO Channels With Beam Squint

Mingjin Wang, *Student Member, IEEE*, Feifei Gao, *Senior Member, IEEE*, Nir Shlezinger, *Member, IEEE*, Mark F. Flanagan, *Senior Member, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Multiple-input multiple-output (MIMO) millimeter wave (mmWave) communication is a key technology for next generation wireless networks. One of the consequences of utilizing a large number of antennas with an increased bandwidth is that array steering vectors vary among different subcarriers. Due to this effect, known as *beam squint*, the conventional channel model is no longer applicable for mmWave massive MIMO systems. In this paper, we study channel estimation under the resulting non-standard model. To that aim, we first analyze the beam squint effect from an array signal processing perspective, resulting in a model which sheds light on the *angle-delay* sparsity of mmWave transmission. We next design a compressive sensing based channel estimation algorithm which utilizes the shift-invariant block-sparsity of this channel model. The proposed algorithm jointly computes the *off-grid* angles, the *off-grid* delays, and the complex gains of the multi-path channel. We show that the newly proposed scheme reflects the mmWave channel more accurately and results in improved performance compared to traditional approaches. We then demonstrate how this approach can be applied to recover both the uplink as well as the downlink channel in frequency division duplex (FDD) systems, by exploiting the angle-delay reciprocity of mmWave channels.**

***Index Terms*—mmWave, massive MIMO, channel estimation, beam squint, block sparsity, angle-delay reciprocity.**

Manuscript received April 27, 2019; revised September 7, 2019 and October 25, 2019; accepted November 11, 2019. Date of publication November 28, 2019; date of current version December 17, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yimin D. Zhang. This work was supported in part by the National Natural Science Foundation of China under Grant 61831013 Grant 61771274, in part by the Beijing Municipal Natural Science Foundation under Grant 4182030 and Grant L182042, in part by the R&D Project of the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant JCYJ20180306170617062, in part by the National Science Foundation for Distinguished Young Scholars of China under Grant 61625106, in part by the National Natural Science Foundation of China under Grant 61531011, in part by the China Scholarship Council under Grant 201906210396, and in part by Futurewei Technologies. *(Corresponding author: Feifei Gao.)*

M. Wang is with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: wmj17@mails.tsinghua.edu.cn).

F. Gao is with the Department of Automation, Tsinghua University, Beijing 100084, China, with the Institute for Artificial Intelligence, Tsinghua University, Beijing 100084, China, with the Beijing National Research Center for Information Science and Technology, Beijing 100084, China, and also with the Key Laboratory of Digital TV System of Guangdong Province and Shenzhen City, Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China (e-mail: feifeigao@ieee.org).

N. Shlezinger and Y. C. Eldar are with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: nirshlezinger1@gmail.com; yonina@weizmann.ac.il).

M. F. Flanagan is with the School of Electrical and Electronic Engineering, University College Dublin, 4 Dublin, Ireland (e-mail: mark.flanagan@ieee.org).

Digital Object Identifier 10.1109/TSP.2019.2956677

I. Introduction

THE proliferation of wireless services such as multimedia, virtual reality, network video, vehicular networking, and the Internet of Things gives rise to continuously increasing demands on the transmission rate and quality of service. These demands include higher throughput, shorter delays, improved connectivity, denser networks, and better user experience [1]. To meet all these requirements, it is necessary to exploit higher frequencies, and in particular, the millimeter wave (mmWave) band, to overcome the spectral congestion of standard wireless frequency bands [2]. An additional method to increase the spectral efficiency and to improve spatial resolution is to equip the base station (BS) with a large-scale antenna array [3]–[5]. This technique is commonly referred to as *massive MIMO*. Due to the short wavelength of mmWave, utilizing large antenna arrays is also essential for successfully implementing mmWave communications. In particular, the increased number of antennas can be used to implement directed beamforming, thus overcoming the dominant path-loss induced at mmWave with line-of-sight [6], [7].

A large body of research has been devoted to understanding the potential and challenges associated with mmWave massive MIMO communications in recent years. Rappaport *et al.* proposed a model for mmWave channels based on an extensive measurement campaign and demonstrated that the mmWave band can effectively support high-speed data transmission [8]. The work [9] derived capacity bounds for mmWave massive MIMO communications based on the model of [8]. Additionally, [10] outlined the benefits, challenges, and potential solutions associated with cellular networks utilizing mmWave massive MIMO technology.

In order to achieve the potential benefits of mmWave massive MIMO communications, it is critical to have accurate channel state information (CSI). However, the channel characteristics in mmWave bands are quite different from their conventional sub-6 GHz counterparts. In particular: 1) experimental studies [11], [12] have shown that electromagnetic waves in mmWave bands suffer from severe path loss and have difficulty in bypassing obstacles; 2) mmWave channels have been shown to exhibit sparsity in the angle and delay domains, which is not encountered in microwave frequencies [13], [14]. Furthermore, due to the narrow angle spread of each cluster, the channel covariance matrices (CCMs) of the resulting channel models are typically low-rank. These properties indicate that any efficient channel estimator should be able to build upon this inherent structure.

Different low-complexity estimation algorithms have been designed to exploit the sparsity or low-rank property of the channel, including CCM based approaches [15], compressed sensing (CS) algorithms [16], [17], and angle domain based methods [18]. In particular, under the assumption of a finite scattering environment, the work [15] mathematically demonstrated the low-rank feature of the CCMs in mmWave communications and proposed a joint spatial division multiplexing algorithm to reduce the effective dimensions of the channel. In [16], the authors used the low-rank structure of the CCMs to cast the channel estimation task into a quadratic semidefinite programming (SDP) problem, which was solved using a polynomial SDP method. With a given unitary dictionary matrix known to the BS, [17] represented a virtual channel which has a common sparsity due to the fact that the users share the same local scatters. A joint orthogonal matching pursuit (OMP) recovery algorithm was then presented in [17] to estimate the channel and reduce feedback. To exploit the angle information for sparse channel estimation, [18] designed a fast discrete Fourier transform (DFT) based spatial rotation algorithm to concentrate most of the channel power on limited DFT grids and efficiently obtain the angle information for both frequency division duplex (FDD) as well as time division duplex (TDD) systems. In particular, [18] used an array signal processing aided channel estimation scheme, where the angle information of the user is exploited to simplify channel estimation. A detailed overview on signal processing methods, including array signal processing techniques, for mmWave massive MIMO communications can be found in [12].

An important drawback of the sparse channel estimation approaches mentioned above stems from the fact that they use *on-grid* estimation [19], [20] to solve the optimization problem, namely, they divide the continuous parameter space into a finite set of grid points. The sparse channel is then estimated assuming a discrete dictionary, resulting in increased estimation errors as the exact parameter does not necessarily lie on the discrete grid. Such grid mismatch introduces quantization error in addition to the channel recovery uncertainty, which may reduce the ability to accurately estimate the channel.

Another drawback of previously proposed channel estimators, e.g., [15]–[18], is that the massive MIMO model used is directly obtained from the conventional MIMO model. This model only observes the phase differences but does not capture the propagation delay of the same incident signal observed at different antennas. This effect is negligible in conventional MIMO setups with a relatively small number of antennas, however, it cannot be ignored when the antenna array grows larger and the bandwidth increases. This phenomenon is referred to as the *spatial-wideband effect* [21]. For orthogonal frequency division multiplexing (OFDM) systems, this effect makes the array response vary with frequency, causing the beams observed by the receiver to "deviate" as a function of frequency [22], which is also known as the *beam squint effect* (BSE) [23]. Since mmWave communications highly rely on the precise alignment of beams between the transmitters and the receivers, the BSE may result in severe performance degradation if not carefully treated.

The BSE was experimentally evaluated in [24], which measured the beam squinting range of 15 degrees with a 4-element patch array over 6 GHz bandwidth with central frequency of 60 GHz. The works [25], [26] proposed to mitigate the BSE by integrating interconnected slow-wave structures and metamaterial cells along the array feed network. In [27], the authors designed a beamforming codebook to compensate for the BSE by imposing an achievable rate constraint. Nevertheless, these efforts [25]–[27] aim only to compensate for the channel performance loss caused by the BSE, and do not provide a systematic channel estimation scheme under this effect.

In this paper, we develop a set of channel estimation algorithms for mmWave massive MIMO systems, in both the uplink channel as well as the downlink channel, accounting for the BSE. Our major contributions are summarized as follows:

- *Modeling block sparsity with BSE:* We decompose the channel vectors into angle, delay, and gain parameters, and establish a model for the BSE using these parameters. It is shown that in the presence of BSE, conventional channel models as well as the burst-sparsity structure proposed in [28] do not accurately represent mmWave massive MIMO systems. Then, we demonstrate that the task of estimating angle and delay parameters can each be expressed as a block-sparse signal recovery problem [29].
- *Off-grid channel parameter estimation:* We propose a pilot-based block-iterative gradient descent algorithm to estimate the angle and delay parameters for the uplink channel. The parameter estimation algorithm operates in an off-grid manner, namely, the estimated channel coefficients can take values in a continuous set. The proposed algorithm is inspired by grid refinement [30], while avoiding its drawbacks in terms of recovery performance and computational complexity.
- *Angle-delay pairing:* Based on the off-grid block-iterative estimation algorithm, we derive an efficient angle and delay pairing algorithm with low computing complexity.
- *Angle reciprocity for downlink channel estimation:* Angle reciprocity was previously studied in [18], [31], [32], where it was shown that the angular information obtained from uplink training can facilitate downlink channel estimation in FDD systems. However, the BSE is not studied in any of those three papers. By exploiting the frequency insensitivity of angle and delay, i.e., *angle-delay reciprocity*, we develop an effective low overhead algorithm for downlink channel estimation in FDD systems with BSE, thus tackling one of the major problems noted in the massive MIMO literature [33].

The rest of this paper is organized as follows: Section II introduces the BSE in the time and angle domains, and formulates the mmWave massive MIMO-OFDM system model. Section III illustrates the block sparsity based uplink angle/delay estimation algorithm which accounts for the BSE and shows how it can be used to reconstruct the uplink channel. Section IV designs a downlink channel estimation scheme for FDD massive MIMO systems with low complexity and low overhead based

on the guidelines used for deriving the uplink channel estimator. Numerical results are provided in Section V, and Section VI concludes this paper.

Notation: Throughout this paper, vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively; transpose, conjugate, Hermitian, inverse, and pseudo-inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$, respectively; $\|\mathbf{A}\|_F$ is the Frobenius norm of the matrix $\mathbf{A}$; $[\mathbf{A}]_{i,j}$ is the (i,j)th entry of $\mathbf{A}$; indices of vectors and matrices start at 0; $[\mathbf{A}]_{i,:}$ represents the ith row of the matrix $\mathbf{A}$; $[\mathbf{A}]_{:,j}$ represents the jth column of the matrix $\mathbf{A}$; $\mathrm{vec}(\mathbf{A})$ represents column-major vectorization of the matrix $\mathbf{A}$, i.e., the operation of stacking the columns of matrix $\mathbf{A}$ to form a vector; $\|\boldsymbol{h}\|_2$ denotes the Euclidean norm of the vector $\boldsymbol{h}$; $\otimes$ denotes the Kronecker product, and $\odot$ is the Hadamard product of matrices; $\mathbf{I}_N$ is the $N \times N$ identity matrix; $\mathbb{R}$ and $\mathbb{C}$ represent the sets of real and complex numbers, respectively; $\mathcal{R}(\cdot)$ is the real part of a complex number, vector or matrix; $\lfloor a \rfloor$ is the downward rounding operation of a real number a.

II. Channel Model

A. mmWave Massive MIMO Uplink Channel Model

We consider a mmWave massive MIMO system, focusing on the uplink transmission. The BS is equipped with a uniform linear array (ULA) consisting of M antennas, where M is a large integer, and the antenna spacing is $d = \lambda_c^{\mathrm{ul}}/2$. Here, $\lambda_c^{\mathrm{ul}} \triangleq c/f_c^{\mathrm{ul}}$, where f_c^{ul} is the uplink center frequency, λ_c^{ul} is the wavelength, and c is the speed of light. To present the channel model, highlighting the spatial wideband effect, we discuss the case of a single user with a single antenna and assume a noiseless setup in this subsection. The proposed model can be extended to multiple users with multiple antennas by properly adapting the arguments in the sequel.

To formulate the channel model, we use $\alpha[i]$ to denote the discrete-time baseband transmitted symbol, with sampling period T_s. The continuous-time baseband transmit signal $\bar{s}(t)$ can thus be expressed as

$$\bar{s}(t) = \sum_{i=-\infty}^{+\infty} \alpha[i] g(t - iT_s), \tag{1}$$

where $g(t)$ is the pulse shaping function. After modulating, the passband transmit signal $\tilde{s}(t)$ can be written as

$$\tilde{s}(t) = \mathcal{R}\left\{ \bar{s}(t) e^{j2\pi f_c^{\mathrm{ul}} t} \right\}. \tag{2}$$

Let P denote the number of paths between the user and the BS. Each path has direction of arrival (DOA) $\theta_p^{\mathrm{ul}} \in [-\pi/2, \pi/2)$ and passband gain $\beta_p^{\mathrm{ul}} \in \mathbb{R}^+$. Denote the pth path delay between the transmitter and the first receive antenna by τ_p^{ul}. Unlike conventional MIMO models, the large array aperture of massive MIMO receivers results in non-negligible delays among different antennas. Those delays are present even for received signal components corresponding to the same channel path, as illustrated in Fig. 1. The extra delay of the pth path from the mth receive antenna compared to the first receive antenna is given by

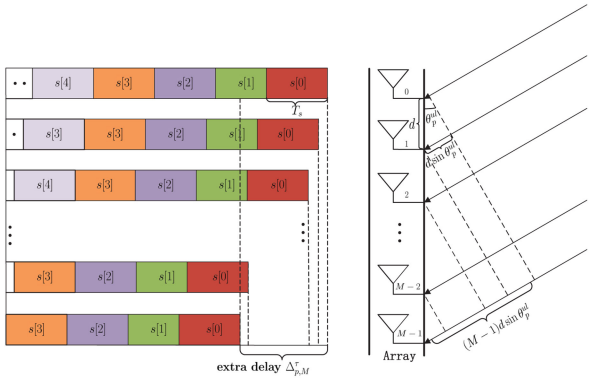

Fig. 1. Illustration of the non-negligible delays between the signals arriving at different antennas.

$$\Delta_{p,m}^{\tau} = \frac{md \cdot \sin\theta_p^{\mathrm{ul}}}{c} = \frac{md \cdot \sin\theta_p^{\mathrm{ul}}}{\lambda_c^{\mathrm{ul}} f_c^{\mathrm{ul}}}. \tag{3}$$

Consequently, the passband receive signal at the mth antenna can be written as

$$\begin{aligned} \tilde{y}_m^{\mathrm{ul}}(t) &= \sum_{p=1}^{P} \beta_p^{\mathrm{ul}} \tilde{s}(t - \tau_p^{\mathrm{ul}} - \Delta_{p,m}^{\tau}) \\ &= \sum_{p=1}^{P} \mathcal{R}\left\{ \beta_p^{\mathrm{ul}} \bar{s}(t - \tau_p^{\mathrm{ul}} - \Delta_{p,m}^{\tau}) e^{j2\pi f_c^{\mathrm{ul}}(t - \tau_p^{\mathrm{ul}} - \Delta_{p,m}^{\tau})} \right\}. \end{aligned} \tag{4}$$

For conventional MIMO systems, where M is small or the bandwidth is narrow, it typically holds that $\Delta_{p,m}^{\tau} \ll T_s$ for each $m \in \{0, 1, \ldots, M-1\}$. Thus the signals at different antennas satisfy $\bar{s}(t - \tau_p^{\mathrm{ul}} - \Delta_{p,m}^{\tau}) \approx \bar{s}(t - \tau_p^{\mathrm{ul}})$. Namely, different antennas at the BS effectively observe a synchronized signal. In such scenarios, the standard MIMO channel output model, i.e., a linear convolution between the individual channel impulse responses and the same source signal, faithfully represents the received signal.

However, for mmWave massive MIMO systems, $\Delta_{p,m}^{\tau}$ cannot be ignored for larger values of m, and the approximation $\bar{s}(t - \tau_p^{\mathrm{ul}} - \Delta_{p,m}^{\tau}) \approx \bar{s}(t - \tau_p^{\mathrm{ul}})$ no longer holds. In this case, the signal observed by the first antenna will include a different time shift compared to the signals observed by other antennas. This phenomenon is referred to as the *spatial wideband effect* [21]. As an illustrative example, consider a massive MIMO mmWave system with $M = 128$ BS antennas, $\theta_p^{\mathrm{ul}} = 60^\circ$, mmWave carrier frequency $f_c^{\mathrm{ul}} = 60$ GHz, and system bandwidth $f_s = \frac{1}{T_s} = 2$ GHz, which is chosen according to the 3GPP standards [34] for mmWave band channel models. Under this setting, the signal delay between the first antenna and the last one, computed via (3), is $1.85T_s$, which is clearly non-negligible.

In the presence of the spatial wideband effect, it is difficult to formulate a unified discrete-time MIMO channel model since the signals arrive at different antennas with relative delays which are fractions of the sampling period. To derive a convenient model which facilitates analysis, we observe the transmission in an antenna-by-antenna manner. After downconverting the

passband symbol in (4) (by multiplying $e^{-j2\pi f_c^{\text{ul}} t}$) to switch from passband to baseband representation, the continuous-time baseband receive signal at the mth antenna is given by

$$\bar{y}_m^{\text{ul}}(t) = \sum_{p=1}^{P} \beta_p^{\text{ul}} \bar{s}(t - \tau_p^{\text{ul}} - \Delta_{p,m}^{\tau}) e^{j2\pi f_c^{\text{ul}}(-\tau_p^{\text{ul}} - \Delta_{p,m}^{\tau})}$$

$$= \left(\sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} e^{-j2\pi f_c^{\text{ul}} \Delta_{p,m}^{\tau}} \delta(t - \tau_p^{\text{ul}} - \Delta_{p,m}^{\tau}) \right) * \bar{s}(t), \quad (5)$$

where $\bar{\beta}_p^{\text{ul}} \triangleq \beta_p^{\text{ul}} e^{-j2\pi f_c^{\text{ul}} \tau_p^{\text{ul}}}$ is the equivalent complex channel gain, and $*$ denotes the convolution operator. Taking the Fourier transform of (5), we obtain the frequency-domain representation of the received signal at the mth antenna as:

$$y_m^{\text{ul}}(f) = \int_{-\infty}^{+\infty} \bar{y}_m^{\text{ul}}(t) e^{-j2\pi f t} dt$$

$$= \left(\sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} e^{-j2\pi f_c^{\text{ul}} \Delta_{p,m}^{\tau}} e^{-j2\pi f(\tau_p^{\text{ul}} + \Delta_{p,m}^{\tau})} \right) s(f)$$

$$= \left(\sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} e^{-j(m-1)\phi_p^{\text{ul}}} e^{-j2\pi f \Delta_{p,m}^{\tau}} e^{-j2\pi f \tau_p^{\text{ul}}} \right) s(f), \quad (6)$$

where $\phi_p^{\text{ul}} \triangleq \frac{2\pi d \cdot \sin(\theta_p^{\text{ul}})}{\lambda_c^{\text{ul}}} \in [-\pi, \pi)$ is defined as the normalized DOA. For clarity, in the rest of the paper, the term "DOA" will refer to the normalized DOA.

By stacking the received signal $y_m^{\text{ul}}(f)$ over all M antennas into a single vector representation, we can write

$$\boldsymbol{y}^{\text{ul}}(f) = [y_1^{\text{ul}}(f), y_2^{\text{ul}}(f), \ldots, y_M^{\text{ul}}(f)]^T = \boldsymbol{h}^{\text{ul}}(f) s(f), \quad (7)$$

where

$$\boldsymbol{h}^{\text{ul}}(f) = \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \boldsymbol{a}(\phi_p^{\text{ul}}, f) e^{-j2\pi f \tau_p^{\text{ul}}}, \quad (8)$$

is the uplink channel frequency response, and

$$\boldsymbol{a}(\phi_p^{\text{ul}}, f) = \left[1, \ldots, e^{-j(M-1)\phi_p^{\text{ul}}} e^{-j2\pi f \Delta_{p,M}^{\tau}}\right]^T$$

$$= \left[1, \ldots, e^{-j(M-1)(1+\frac{f}{f_c^{\text{ul}}})\phi_p^{\text{ul}}}\right]^T, \quad (9)$$

is the wideband array steering vector. Note that unlike conventional array steering vectors, see, e.g., [16]–[18], the wideband array steering vector $\boldsymbol{a}(\phi_p^{\text{ul}}, f)$ is *frequency-dependent*.

B. Beam Squint Effect in OFDM Systems

We next focus on OFDM signaling with N subcarriers and subcarrier spacing f_0. We henceforth assume that the number of antennas M satisfies $M \leq 2N \frac{f_c^{\text{ul}}}{f_s}$. This assumption is reasonable for mmWave systems in which the bandwidth is very large, thus the number of subcarriers N is of the same order as M, and the carrier frequency f_c^{ul} is much larger than the system bandwidth f_s.

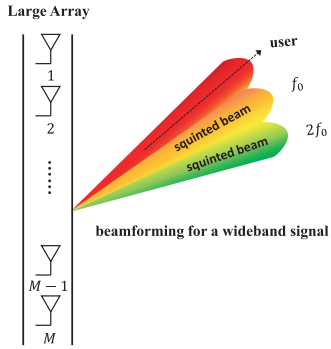

Fig. 2. Illustration of the beam squint effect directing different beam directions with the same angle.

Using (8), the uplink channel vector corresponding to the nth subcarrier can be written as

$$\boldsymbol{h}^{\text{ul}}(nf_0) = \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \boldsymbol{a}(\phi_p^{\text{ul}}, nf_0) e^{-j2\pi n f_0 \tau_p^{\text{ul}}}$$

$$= \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \left[\boldsymbol{a}(\phi_p^{\text{ul}}, 0) e^{-j2\pi n f_0 \tau_p^{\text{ul}}}\right] \odot [\boldsymbol{W}(\phi_p^{\text{ul}})]_{:,n}, \quad (10)$$

for each $n = \{0, 1, \ldots, N-1\}$, where $\boldsymbol{W}(\phi_p^{\text{ul}})$ is the $M \times N$ matrix of wideband factors whose (m, n)th entry is given by

$$[\boldsymbol{W}(\phi_p^{\text{ul}})]_{m,n} \triangleq \exp\left(-jm \frac{nf_0}{f_c^{\text{ul}}} \phi_p^{\text{ul}}\right). \quad (11)$$

The overall $M \times N$ uplink channel matrix for the OFDM system can then be written as

$$\mathbf{H}^{\text{ul}} = \left[\boldsymbol{h}^{\text{ul}}(0), \boldsymbol{h}^{\text{ul}}(f_0), \ldots, \boldsymbol{h}^{\text{ul}}((N-1)f_0)\right]. \quad (12)$$

In previous works, e.g., [16]–[18], which do not account for the spatial-wideband effect, the factor $\Delta_{p,m}^{\tau}$ in (9) is assumed to be zero. In this case, the array steering vector of any subcarrier reduces to $\boldsymbol{a}(\phi_p^{\text{ul}}, 0)$, namely, it is *frequency-independent*. When the frequency dependence of the array steering vector is not accounted for, the postulated beam direction at the nth subcarrier is assumed to be $\boldsymbol{a}(\phi_p^{\text{ul}}, 0)$, which has a phase offset of nf_0 compared to the true beam direction $\boldsymbol{a}(\phi_p^{\text{ul}}, nf_0)$. This causes a deviation between the spatially oriented beam and the user's true direction, commonly referred to as beam squint. This deviation can lead to significant performance degradation. Fig. 2 depicts an example of beam squint where the beams at different subcarriers point in different directions relative to the same user angle.

C. Channel Characteristics Due to the BSE

We now discuss the unique channel characteristics which arise in the presence of the BSE, focusing on OFDM signaling. It follows from (10) that $\boldsymbol{H}^{\text{ul}}$ can be modeled as the sum of the contributions from P paths via

$$\mathbf{H}^{\text{ul}} = \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \left[\mathbf{a}(\phi_p^{\text{ul}}, 0) \mathbf{b}^T(\tau_p^{\text{ul}})\right] \odot \mathbf{W}(\phi_p^{\text{ul}})$$

$$= \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \boldsymbol{\Xi}(\phi_p^{\text{ul}}, \tau_p^{\text{ul}}), \quad (13)$$

where $\boldsymbol{b}(\tau_p^{\text{ul}})$ is an $N \times 1$ vector defined as

$$\boldsymbol{b}(\tau_p^{\text{ul}}) \triangleq \left[1, e^{-j2\pi f_0 \tau_p^{\text{ul}}}, \ldots, e^{-j2\pi(N-1) f_0 \tau_p^{\text{ul}}}\right]^T, \tag{14}$$

and $\boldsymbol{\Xi}(\phi_p^{\text{ul}}, \tau_p^{\text{ul}}) \triangleq [\mathbf{a}(\phi_p^{\text{ul}}, 0)\mathbf{b}^T(\tau_p^{\text{ul}})] \odot \mathbf{W}(\phi_p^{\text{ul}}) \in \mathbb{C}^{M \times N}$.

Using these notations, we state the following asymptotic channel characteristic taken from [21], which holds when the number of antennas and the number of the subcarriers both grow arbitrarily large.

Theorem 1: If the conditions $\frac{d \cdot f_s}{\lambda_c^{\text{up}} \cdot f_c^{\text{ul}}} < 1$ and $\frac{M-1}{2N} \frac{f_s}{f_c^{\text{ul}}} < 1$ are both satisfied, then, as $M \to \infty$ and $N \to \infty$, the following property holds [21]

$$\lim_{M,N \to \infty} \frac{1}{MN} \text{vec}(\boldsymbol{\Xi}(\phi_p^{\text{ul}}, \tau_p^{\text{ul}})^H \text{vec}(\boldsymbol{\Xi}(\phi_s^{\text{ul}}, \tau_s^{\text{ul}})) = \begin{cases} 1, & \text{if } (\phi_p^{\text{ul}}, \tau_p^{\text{ul}}) = (\phi_s^{\text{ul}}, \tau_s^{\text{ul}}) \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

The first condition in Theorem 1 commonly holds, since the antenna spacing d is typically half of the wavelength λ, and the system bandwidth f_s, which is typically determined by the signal bandwidth, is far less than its carrier frequency in mmWave massive MIMO systems. The second condition holds under our system stated in the previous subsection.

Theorem 1 implies that in the massive MIMO regime with a sufficiently large number of subcarriers, paths with different angles or delays can be distinguished easily. We will exploit this property in Section III where we consider the reconstruction of the uplink channel.

For mmWave communications, the number of significant paths is typically much smaller compared to that encountered in standard sub 6-GHz systems [13], [14], and thus $\boldsymbol{H}^{\text{ul}}$ can be represented by a few steering vectors in the spatial and delay domains. This sparse property indicates that CS methods can be utilized to efficiently estimate the channel parameters, as we show in the following section.

III. Uplink Channel Estimation

In this section we propose algorithms for estimating uplink mmWave massive MIMO channels, accounting for the BSE. In particular, we assume that the BS serves K users using an orthogonal frequency division multiple access (OFDMA) protocol [35] with precise synchronization, in which channel estimation is carried out using dedicated pilots in an FDD manner. Non-overlapping subcarriers are assigned to different users. To model the observed signal used for channel estimation, we let $\mathcal{N}_k$ denote the set of subcarrier frequencies utilized by the kth user, and use $T \triangleq N/K$ to denote its cardinality.[1] Since all subcarriers are allocated among the users without spectral overlapping, it holds that $\mathcal{N}_k \bigcap \mathcal{N}_l = \varnothing, k \neq l$, and

$$\mathcal{N}_1 \bigcup \cdots \bigcup \mathcal{N}_K = \{0, 1, \ldots, N-1\}. \tag{16}$$

[1]Here, we only use one OFDM block to estimate the uplink channel parameters, and T is an integer. When N is not an integer multiple of K, we use $T = \lfloor \frac{N}{K} \rfloor$. If K is too large such that T is not long enough to estimate the multiple channels, one can utilize multiple OFDM blocks for channel estimation.

Despite the short channel coherence time caused by mmWave band carrier frequency, it still contains many symbols for channel estimation due to the large bandwidth. For example, when the carrier frequency is 60 GHz and the bandwidth is 2 GHz, a maximum speed of 60 m/s results in small channel coherence time of 0.083 ms. However, the sampling duration is in the order of 0.5 ns, and the receiver has enough symbols to utilize for channel estimation.

While in the previous section we focused on a single user, here we consider multiple users. Therefore, we henceforth use the notation $\boldsymbol{H}_k^{\text{ul}}$ to denote the channel of the kth user, similarly to (12). The a-priori known pilot sequence transmitted by the kth user is denoted by $\boldsymbol{s}_k^{\text{ul}} \in \mathbb{C}^{T \times 1}$, and is assumed to have non-zero entries. This assumption accommodates a broad range of pilot sequences used in practice, such as Zadoff-Chu (ZC) sequences [36]. The received pilots from the kth user at the BS, aggregated over the corresponding subcarriers assigned to the kth user, can be expressed as

$$\boldsymbol{Y}_{\mathcal{N}_k} = \boldsymbol{H}_{\mathcal{N}_k}^{\text{ul}} \boldsymbol{S}_k^{\text{ul}} + \boldsymbol{E}_{\mathcal{N}_k}, \tag{17}$$

where $\boldsymbol{H}_{\mathcal{N}_k}^{\text{ul}} \in \mathbb{C}^{M \times T}$ is the subset of columns of $\boldsymbol{H}_k^{\text{ul}}$ with column indices in $\mathcal{N}_k$, $\boldsymbol{S}_k^{\text{ul}} \triangleq \text{diag}\{\boldsymbol{s}_k^{\text{ul}}\} \in \mathbb{C}^{T \times T}$ is the diagonal kth user pilot matrix, and $\boldsymbol{E}_{\mathcal{N}_k}$ is additive noise with i.i.d. zero-mean unit variance proper complex Gaussian entries.

Our goal is to reconstruct the complete channel of the kth user, $\boldsymbol{H}_k^{\text{ul}}$, from the channel output $\boldsymbol{Y}_{\mathcal{N}_k}$. To that aim, we first identify the sparse characteristics of the unknown channel in Subsection III-A. Then, in Subsection III-B and III-C we exploit this sparse nature to efficiently estimate the unknown channel DOAs and delays, respectively. Finally, in Subsection III-D we show how these estimates can be combined to recover the unknown channel.

Since the pilot symbols of different users do not overlap in frequency, it follows from (17) that the channel estimation procedure can be carried out individually for each user. Therefore, for clarity, in the rest of Section III and in Section IV, we omit the user index k.

A. Sparse Representation

To model the sparse nature of the mmWave channel coefficients with the BSE, we denote the set of utilized subcarriers by $\mathcal{N} = \{n_1, n_2, \ldots, n_T\}$. It follows from (10) that the qth column of $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}$, $q \in \{1, 2, \ldots, T\}$, can be written as

$$[\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}]_{:,q} = \sum_{p=1}^{P} \bar{\beta}_p^{\text{ul}} \boldsymbol{a}(\phi_p^{\text{ul}}, n_q f_0) e^{-j2\pi n_q f_0 \tau_p^{\text{ul}}} = \sum_{p=1}^{P} \bar{z}_{q,p}^{\text{ul}} \boldsymbol{a}(\phi_p^{\text{ul}}, n_q f_0), \tag{18}$$

where $\bar{z}_{q,p}^{\text{ul}} = \bar{\beta}_p^{\text{ul}} e^{-j2\pi n_q f_0 \tau_p^{\text{ul}}}$.

We assume that the number of possible paths, denoted by L, is a relatively large number, and is much larger than the number of actual paths ($L \gg P$). Define

$$\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_q f_0) \triangleq [\boldsymbol{a}(\vartheta_1^{\text{ul}}, n_q f_0), \ldots, \boldsymbol{a}(\vartheta_L^{\text{ul}}, n_q f_0)], \tag{19}$$

as an *overcomplete sub-dictionary* based on (9), where $\boldsymbol{\vartheta}^{\text{ul}} = [\vartheta_1^{\text{ul}}, \vartheta_2^{\text{ul}}, \ldots, \vartheta_L^{\text{ul}}]$ and $\vartheta_i^{\text{ul}} \triangleq -\pi + \frac{2i\pi}{L}, i \in \{1, 2, \ldots, L\}$ divides the continuous angle space uniformly. Since L is large, the true DOA angles $\boldsymbol{\phi}^{\text{ul}} \triangleq [\phi_1^{\text{ul}}, \ldots, \phi_P^{\text{ul}}]$ can be approximated (with some quantization error) to be a subset of $\boldsymbol{\vartheta}^{\text{ul}}$. This indicates that we may use the overcomplete sub-dictionary $\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_q f_0)$ to represent $[\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}]_{:,q}$ as

$$[\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}]_{:,q} \approx \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_q f_0)\boldsymbol{z}_q, \tag{20}$$

where $\boldsymbol{z}_q$ is a $L \times 1$ sparse vector whose ith element is

$$z_{q,i} = \begin{cases} \bar{z}_{q,p}^{\text{ul}}, & i = \text{argmin}_{k\in\{1,\ldots,L\}} |\vartheta_k^{\text{ul}} - \phi_p^{\text{ul}}|, p = 1, \ldots, P \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

Specifically, for each $q \in \{1, 2, \ldots, T\}$, $\boldsymbol{z}_q$ has at most P nonzero values, i.e., it is P-*sparse*. It follows from (20) that $\mathbf{H}_{\mathcal{N}}^{\text{ul}}$ may be approximated as

$$\boldsymbol{H}_{\mathcal{N}}^{\text{ul}} \approx \left[\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_1 f_0)\boldsymbol{z}_1, \ldots, \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_T f_0)\boldsymbol{z}_T\right]. \tag{22}$$

We note that when the BSE is not present, the term $\boldsymbol{a}(\vartheta_i^{\text{ul}}, n_q f_0)$ reduces to $\boldsymbol{a}(\vartheta_i^{\text{ul}}, 0)$, which is *frequency-independent*. In this case, $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}$ becomes

$$\begin{aligned} \boldsymbol{H}_{\mathcal{N}}^{\text{ul}} &\approx \left[\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)\boldsymbol{z}_1, \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)\boldsymbol{z}_2, \ldots, \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)\boldsymbol{z}_T\right] \\ &= \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)[\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_T] = \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)\boldsymbol{Z}, \end{aligned} \tag{23}$$

where $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_T]$. From (21) it holds that the sparsity pattern of $\boldsymbol{z}_q$ is independent of q, i.e., all the vectors $\{\boldsymbol{z}_q\}_{q=1}^T$ have their non-zero elements in the same entries, so that the matrix $\boldsymbol{Z}$ has at most P nonzero rows occuring on a common index set. Consequently, rather than trying to estimate the channel parameters from each subcarrier independently, the parameters are jointly estimated by combining all the subcarriers, namely, by recasting the estimation of the mmWave channel as a *multiple measurement vector* (MMV) *problem* [29], [37]. To exploit the common support structure of $\boldsymbol{Z}$, one can stack the rows of $\boldsymbol{Z}$ and $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}$ into vectors. Then, (23) can be converted into a sparse recovery problem:

$$\text{vec}((\boldsymbol{H}_{\mathcal{N}}^{\text{ul}})^T) \approx (\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0) \otimes \boldsymbol{I}_T)\text{vec}(\boldsymbol{Z}^T). \tag{24}$$

In the presence of BSE, $\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_q f_0)$ varies from subcarrier to subcarrier. In this case, we cannot directly obtain an expression of the form (23). Nevertheless, given that $\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, n_q f_0)$ has a fixed phase offset $n_q f_0$ compared to $\boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)$, we write $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}} \approx \boldsymbol{A}(\boldsymbol{\vartheta}^{\text{ul}}, 0)\bar{\boldsymbol{Z}}$, where the nonzero columns of $\bar{\boldsymbol{Z}}$ have a regular 'shift' characteristic. This shift property is illustrated in Fig. 3, in which each square corresponds to a vector entry: black squares represent the nonzero elements while blank squares indicate zeros. Therefore, to ensure that each column of $\boldsymbol{Z}$ still has the same nonzero positions, we propose to design a shift-invariant transform such that the common sparse support of the transformed $\boldsymbol{Z}$ satisfies the same sparsity pattern behavior as in the absence of beam squint.

To that aim, we first define the $MT \times LT$ matrix

$$\boldsymbol{D}_{\text{bs}}(\boldsymbol{\vartheta}^{\text{ul}}) \triangleq \left[\boldsymbol{D}(\vartheta_1^{\text{ul}}), \boldsymbol{D}(\vartheta_2^{\text{ul}}), \ldots, \boldsymbol{D}(\vartheta_L^{\text{ul}})\right], \tag{25}$$

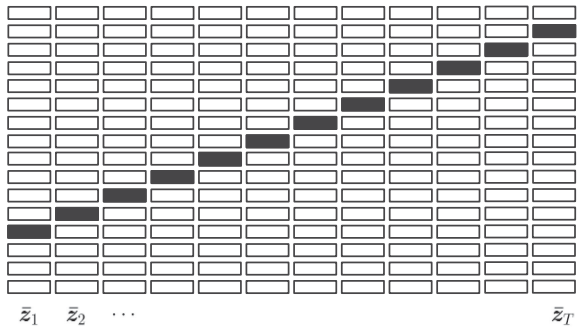

Fig. 3. Example of the structure of $\bar{\boldsymbol{Z}}$ for $P = 1$.

where the subscript "*bs*" stands for beam squint. Here, $\boldsymbol{D}(\vartheta_i^{\text{ul}})$ is an $MT \times T$ matrix given by

$$\begin{aligned} \boldsymbol{D}(\vartheta_i^{\text{ul}}) &= \begin{bmatrix} [\boldsymbol{a}(\vartheta_i^{\text{ul}}, 0)]_1 \cdot \boldsymbol{I}_T \\ [\boldsymbol{a}(\vartheta_i^{\text{ul}}, 0)]_2 \cdot \boldsymbol{I}_T \\ \vdots \\ [\boldsymbol{a}(\vartheta_i^{\text{ul}}, 0)]_M \cdot \boldsymbol{I}_T \end{bmatrix} \odot \begin{bmatrix} \boldsymbol{\Phi}_1(\vartheta_i^{\text{ul}}) \\ \boldsymbol{\Phi}_2(\vartheta_i^{\text{ul}}) \\ \vdots \\ \boldsymbol{\Phi}_M(\vartheta_i^{\text{ul}}) \end{bmatrix} \\ &= [\boldsymbol{a}(\vartheta_i^{\text{ul}}, 0) \otimes \boldsymbol{I}_T] \odot \left[\boldsymbol{\Phi}_1^T(\vartheta_i^{\text{ul}}), \ldots, \boldsymbol{\Phi}_M^T(\vartheta_i^{\text{ul}})\right]^T, \end{aligned} \tag{26}$$

and $\boldsymbol{\Phi}_m(\vartheta_i^{\text{ul}})$ is a $T \times T$ frequency rotation matrix with parameter ϑ_i^{ul} which can be expressed as

$$\begin{aligned} \boldsymbol{\Phi}_m(\vartheta_i^{\text{ul}}) \triangleq \text{diag}\Bigg\{ & e^{-j(m-1)\frac{n_1 f_0}{f_c^{\text{ul}}}\frac{2\pi d \sin(\vartheta_i^{\text{ul}})}{\lambda_c^{\text{ul}}}}, e^{-j(m-1)} \\ & \times e^{\frac{n_2 f_0}{f_c^{\text{ul}}}\frac{2\pi d \sin(\vartheta_i^{\text{ul}})}{\lambda_c^{\text{ul}}}}, \ldots, e^{-j(m-1)\frac{n_T f_0}{f_c^{\text{ul}}}\frac{2\pi d \sin(\vartheta_i^{\text{ul}})}{\lambda_c^{\text{ul}}}} \Bigg\}. \end{aligned} \tag{27}$$

The application of the rotation matrix allows to express $\text{vec}((\boldsymbol{H}_{\mathcal{N}}^{\text{ul}})^T)$ using the matrix $\boldsymbol{Z}$, similarly to (24). Specifically, it can be verified that $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}$ is transformed into an MMV sparse representation given by

$$\text{vec}((\boldsymbol{H}_{\mathcal{N}}^{\text{ul}})^T) \approx \boldsymbol{D}_{\text{bs}}(\boldsymbol{\vartheta}^{\text{ul}})\text{vec}(\boldsymbol{Z}^T), \tag{28}$$

where, as in (22)–(23), the approximation in (28) stems from the fact that the DOAs do not necessarily lie on the grid $\boldsymbol{\vartheta}^{\text{ul}}$.

We henceforth refer to $\boldsymbol{D}_{\text{bs}}(\boldsymbol{\vartheta}^{\text{ul}})$ as the *sensing matrix* [43], as it represents a linear dimension reduction of $\text{vec}(\boldsymbol{Z}^T)$. Since $\boldsymbol{Z}$ is P-row sparse, $\boldsymbol{Z}^T$ will be P-column sparse. This in turn implies that $\text{vec}(\boldsymbol{Z}^T)$ is P-*block sparse* [44], which facilitates its recovery using block sparsity methods, as detailed in the following subsection.

B. Off-Grid DOA Estimation Algorithm

We next study the recovery of the DOA vector from the channel output at the observed subcarriers $\boldsymbol{Y}_{\mathcal{N}}$, given in (17). Define $\boldsymbol{x} \triangleq \text{vec}((\boldsymbol{Z}\boldsymbol{S}^{\text{ul}})^T)$, recalling that $\boldsymbol{S}^{\text{ul}}$ is the pilot matrix with its diagonal elements being the priori known pilot sequence. Since the pilot elements are non-zero, $\boldsymbol{x}$ exhibits the same block-sparse structure as $\text{vec}(\boldsymbol{Z}^T)$. Also, since the pilot matrix $\boldsymbol{S}^{\text{ul}}$ is diagonal, we formulate the channel output $\boldsymbol{Y}_{\mathcal{N}}$ as an MMV

sparse representation via

$$\begin{aligned}\text{vec}((\boldsymbol{Y}_{\mathcal{N}})^T) &\approx \boldsymbol{D}_{\text{bs}}(\boldsymbol{\vartheta}^{\text{ul}})\text{vec}\left((\boldsymbol{Z}\boldsymbol{S}^{\text{ul}})^T\right) + \text{vec}(\boldsymbol{E}_{\mathcal{N}}^T) \\ &= \boldsymbol{D}_{\text{bs}}(\boldsymbol{\vartheta}^{\text{ul}})\boldsymbol{x} + \text{vec}(\boldsymbol{E}_{\mathcal{N}}^T). \end{aligned} \tag{29}$$

Previously proposed algorithms for channel estimation in massive MIMO systems [16], [17] assume that the actual DOA values, represented by the entries of the vector $\boldsymbol{\phi}^{\text{ul}}$, coincide with values in the grid vector $\boldsymbol{\vartheta}^{\text{ul}}$. Namely, the DOA values lie on the discrete grid, and there is a one-to-one correspondence between the non-zero indexes of $\boldsymbol{x}$ and $\boldsymbol{\phi}^{\text{ul}}$. Then, the DOA vector is recovered from the estimated $\boldsymbol{x}$. We henceforth refer to such methods as *on-grid* algorithms. However, since $\boldsymbol{\phi}^{\text{ul}}$ typically takes values in some continuous non-countable set, the resolution of on-grid DOA estimation is only $(\frac{2\pi}{L})$, which is also known as *grid mismatch*. This grid mismatch induces quantization error and degrades the estimation accuracy. Although the resolution of DOA estimation can be improved by increasing L, denser grids possibly imply non-feasible and higher computational complexity.

To circumvent the grid mismatch, *off-grid* solutions have been broadly studied [30], [38]–[42]. In off-grid estimation, the estimated DOAs are not restricted to a specific grid but rather take on value in the continuous parameter space. The main approaches for off-grid recovery proposed include:

1) *Taylor expansion.* In [38], the non-linear dependence in the DOA parameter is linearized via a first order Taylor series expansion, resulting in a formulation from which their values are recovered without discretization. However, this method heavily depends on the accuracy of the expansion.
2) *Atomic norm denoising.* In atomic norm denoising methods, the sparse signals are recovered by solving an atomic norm based objective function [39]. The function is then converted into a semi-definite program that is solved by off-the-shelf solvers in an off-grid manner. However, solving the atomic norm objective becomes computationally complex for large scale problems. This is especially difficult in our mmWave massive MIMO channel estimation problem, in which the dimensionality of the multivariate quantities tends to be very large.
3) *Grid refinement.* Grid refinement methods (also referred to as dynamic grid in [40]) were introduced by Malioutov *et al.* [30] to mitigate the effect of grid mismatch in DOA estimation. This approach adaptively refines the grid around candidate spatial locations with an updating process. The original formulation of grid refinement in [30] induces high computational complexity. This approach was extended in [41], which developed an iterative algorithm based on sparse Bayesian learning (SBL) for sparse vector recovery and off-grid refinement. However, [41] only considered i.i.d. sparsity (i.e., the entries of the sparse vectors are assumed to be i.i.d.). In practice, massive MIMO channels have more sophisticated structured sparsity. Recently, Mamandipoor *et al.* [42] proposed a Newtonized orthogonal matching pursuit (NOMP) algorithm that uses Newton refinements for the over-sampled grids to improve the grid refinement method. During each iteration of NOMP, the Newton refinements sidestep the potential basis mismatch from discretizing a continuous parameter space. However, such gradient-based algorithms have to calculate the gradient of the grids corresponding to the coarse paths at each iteration, resulting in increased computational complexity which may limit their application in massive MIMO systems.

Here, we propose an algorithm for recovering the DOAs which is inspired by grid refinement, while avoiding its drawbacks in terms of recovery performance and computational complexity. Our method starts with a fixed known dense grid, for which (29) represents a linear transformation of an unknown block sparse vector $\boldsymbol{x}$. The algorithm then alternates as follows: first, for a fixed grid, it recovers a block-sparse $\boldsymbol{x}$; then, for fixed block-sparse $\boldsymbol{x}$, we adjust the angle grid and accordingly the projection matrix $\boldsymbol{D}_{bs}$ to further minimize the cost. By repeating these steps iteratively, we are able to recover DOA angles which minimize the cost function without necessarily lying on the original grid.

Similar to grid refinement, the proposed approach changes the grid iteratively. However, while the aforementioned iterative methods locally refine parameters and refresh all the points corresponding to the estimated coarse paths, our method reduces the number of grid points iteratively in a continuous manner. The benefits of the proposed technique over conventional grid refinement are numerically demonstrated in Section V.

To explain the algorithm in detail, define $L_{\phi}^{(0)}$ as the initial guess of the number of unknown paths that will be gradually decreased and tuned during the estimation procedure. With a slight abuse of notation, we use $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_{L_{\phi}^{(0)}}]$ as the off-grid DOAs to be estimated. Since we do not know the number of paths in advance, $L_{\phi}^{(0)}$ is set initially to a relatively large number, and thus estimating $\boldsymbol{\phi}$ can be formulated as a sparse signal recovery problem with an unknown parametric dictionary $\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi})$. In this framework, the objective is not only to estimate the sparse signal, but also to optimize/refine the angle grid such that the parametric dictionary approaches the true sparsifying dictionary.

To proceed, we recall the definition of the block l_0-norm:

Definition 1: The T-block ℓ_o-norm of a $TL \times 1$ vector $\boldsymbol{x} = [\boldsymbol{x}^T[1], \ldots, \boldsymbol{x}^T[L]]^T$ is defined as [37]

$$\|\boldsymbol{x}\|_{0,T} \triangleq \sum_{i=1}^{L} \mathcal{I}(\|\boldsymbol{x}[i]\|_2 > 0), \tag{30}$$

where $\mathcal{I}(\|\boldsymbol{x}[i]\|_2 > 0)$ is an indicator function which equals 1 if $\|\boldsymbol{x}[i]\|_2 > 0$ and 0 otherwise, and $\boldsymbol{x}[i]$ is the ith block of $\boldsymbol{x}$ containing T consecutive elements. Note that the T-block ℓ_o-norm with $T = 1$ reduces to the conventional ℓ_o-norm.

Using (29), we formulate the DOA estimation problem for a fixed known grid $\boldsymbol{\phi}$ exploiting the prior knowledge of $\boldsymbol{x}$ being block sparse as

$$\begin{aligned} &\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{0,T} \\ &\text{s. t. } \|\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T) - \boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi})\boldsymbol{x}\|_2 \leq \xi, \end{aligned} \tag{31}$$

where ξ is an error tolerance parameter that is related to the noise statistics. To recover an off-grid estimate of the DOAs, we first recast problem (31) as an iterative reweighted least squares objective, as in, e.g., [45]. Then, we use the resulting objective to tune the grid vector $\boldsymbol{\phi}$, supporting an off-grid estimate in a computationally feasible fashion.

To formulate the iterative algorithm, let $\boldsymbol{x}^{(\omega)}$ and $\boldsymbol{\phi}^{(\omega)}$ be the estimations of $\boldsymbol{x}$ and $\boldsymbol{\phi}$ at the ωth iteration, respectively. At each iteration we form the following $L_\phi^{(\omega)}T \times L_\phi^{(\omega)}T$ matrix:

$$\boldsymbol{G}^{(\omega)} \triangleq \text{diag}\left[\frac{1}{(\|\boldsymbol{x}^{(\omega)}[1]\|_2^2)+\epsilon}\cdots,\frac{1}{(\|\boldsymbol{x}^{(\omega)}[L_\phi^{(\omega)}]\|_2^2)+\epsilon}\right] \otimes \boldsymbol{I}_T, \tag{32}$$

where $\epsilon > 0$ is a positive parameter ensuring that $\boldsymbol{G}^{(\omega)}$ is well-defined, and $L_\phi^{(\omega)}$ is the number of unknown angles at the ωth iteration. The block-sparsity problem (31) is then recast as a reweighed least squares objective problem:

$$\min_{\boldsymbol{x}} \boldsymbol{x}^H\boldsymbol{G}^{(\omega)}\boldsymbol{x} + \lambda^{(\omega)}\|\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T) - \boldsymbol{D}_{\text{bs}}\left(\boldsymbol{\phi}^{(\omega)}\right)\boldsymbol{x}\|_2^2. \tag{33}$$

The objective in (33) consists of two terms: the weighted norm $\boldsymbol{x}^H\boldsymbol{G}^{(\omega)}\boldsymbol{x}$, which controls the level of block sparsity of the recovered vector $\boldsymbol{x}$, and the term $\|\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T) - \boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}^{(\omega)})\boldsymbol{x}\|_2^2$, which represents the accuracy of the estimation. The balance between the two terms is controlled by the regularization parameter $\lambda^{(\omega)}$ which we set to

$$\lambda^{(\omega)} = \frac{MT}{\|\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T) - \boldsymbol{D}_{\text{bs}}\left(\boldsymbol{\phi}^{(\omega)}\right)\boldsymbol{x}^{(\omega)}\|_2^2}. \tag{34}$$

For a given $\boldsymbol{\phi}^{(\omega)} = \boldsymbol{\phi}$, the optimal value of $\boldsymbol{x}$ in (33) is

$$\boldsymbol{x}^{(\omega+1)}|\boldsymbol{\phi} = \left(\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi})\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}) + \left(\lambda^{(\omega)}\right)^{-1}\boldsymbol{G}^{(\omega)}\right)^{-1} \times \boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi})\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T). \tag{35}$$

Substituting $\boldsymbol{x}^{(\omega+1)}|\boldsymbol{\phi}$ back into (33), we optimize the grid $\boldsymbol{\phi}^{(\omega)}$ in light of the objective (33) by minimizing

$$v(\boldsymbol{\phi}^{(\omega)}) \triangleq -\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)^H\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi})\left(\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi}^{(\omega)})\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}^{(\omega)}) + \left(\lambda^{(\omega)}\right)^{-1}\boldsymbol{G}^{(\omega)}\right)^{-1}\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi}^{(\omega)})\text{vec}(\boldsymbol{Y}_{\mathcal{N}}^T). \tag{36}$$

Since directly minimizing (36) is computationally complex, we propose to gradually decrease it by selecting $\boldsymbol{\phi}^{(\omega+1)}$ that satisfies $v(\boldsymbol{\phi}^{(\omega+1)}) \leq v(\boldsymbol{\phi}^{(\omega)})$ for the next iteration. Since $v(\boldsymbol{\phi}^{(\omega)})$ is differentiable with respect to $\boldsymbol{\phi}$, the $(\omega+1)$th estimation can be obtained by the gradient descent method:

$$\boldsymbol{\phi}^{(\omega+1)} = \boldsymbol{\phi}^{(\omega)} - u\frac{\partial v(\boldsymbol{\phi}^{(\omega)})}{\partial \boldsymbol{\phi}}, \tag{37}$$

where u is a step size, and the derivative expression is given in closed-form in the Appendix. The update rule in (37) implies that, even if $\boldsymbol{\phi}^{\text{ul}}$ is initialized to a large grid, the iterative algorithm allows the updated DOA estimates to deviate from this initial grid, resulting in off-grid estimation.

Algorithm 1: DOA Estimation.

- **Initialize:** Input $\boldsymbol{Y}_{\mathcal{N}}$, $\boldsymbol{x}^{(0)} = \boldsymbol{0}$, $\boldsymbol{\phi}^{(0)} = \boldsymbol{0}$, $L_\phi^{(0)}$, and $\lambda^{(0)}$. Iteration index $\omega = 0$.
- **Step 1:** At iteration ω, based on the previous results $\boldsymbol{x}^{(\omega)}$, $\lambda^{(\omega)}$, refresh $\boldsymbol{G}^{(\omega)}$ in (32) and construct the function $v(\boldsymbol{\phi}^{(\omega)})$ via (36).
- **Step 2:** Update $\boldsymbol{\phi}^{(\omega+1)}$ via (37).
- **Step 3:** Compute $\boldsymbol{x}^{(\omega+1)}$ via (35) and $\lambda^{(\omega+1)}$ via (34).
- **Step 4:** If $\|\boldsymbol{x}[i]\|_2^2 < \mu$, then remove it, and delete the relevant angle $\phi_i^{(\omega+1)}$ via (38).
- **Step 5:** Update $L_\phi^{(\omega+1)}$ via (38).
- **Step 6:** If $\|\boldsymbol{x}^{(\omega+1)} - \boldsymbol{x}^{(\omega)}\|_2 \geq \eta$, then set $\omega \leftarrow \omega+1$ and go to step 1, otherwise stop.

In the proposed algorithm, the main complexity lies in calculating $\boldsymbol{x}^{(\omega+1)}|\boldsymbol{\phi}$ and the first derivative $\frac{\partial v(\boldsymbol{\phi}^{(\omega)})}{\partial \boldsymbol{\phi}}$. To reduce the computational complexity, a pruning method is introduced: for every $i = 1, 2, \ldots, L_\phi^{(\omega)}$, if $\|\boldsymbol{x}^{(\omega+1)}[i]\|_2^2$ is smaller than some fixed threshold μ, then we delete $\boldsymbol{x}^{(\omega+1)}[i]$ from the vector $\boldsymbol{x}^{(\omega+1)}$, and correspondingly delete the angle $\phi_i^{(\omega+1)}$ from the vector $\boldsymbol{\phi}^{(\omega+1)}$. We then set $L_\phi^{(\omega)}$ to be the length of the preserved $\boldsymbol{\phi}^{(\omega+1)}$:

$$\begin{cases} \boldsymbol{x}_*^{(\omega+1)} = \boldsymbol{x}^{(\omega+1)}(\|\boldsymbol{x}^{(\omega+1)}[i]\|_2^2 > \mu), \quad i = 1, \ldots, L_\phi^{(\omega)} \\ \boldsymbol{\phi}_*^{(\omega+1)} = \boldsymbol{\phi}^{(\omega+1)}(\|\boldsymbol{x}^{(\omega+1)}[i]\|_2^2 > \mu), \quad i = 1, \ldots, L_\phi^{(\omega)} \\ L_\phi^{(\omega+1)} = \text{Length}\left(\boldsymbol{\phi}_*^{(\omega+1)}\right), \end{cases} \tag{38}$$

where $(\cdot)_*$ represents the preserved vector at each iteration.

When the iterative algorithm satisfies its termination criterion, i.e., $\|\boldsymbol{x}^{(\omega+1)} - \boldsymbol{x}^{(\omega)}\|_2$ is less than some predefined η, the number of paths P is estimated using the number of non-zero block values of $\boldsymbol{x}$. The proposed algorithm is summarized as **Algorithm 1**.

C. Delay Estimation Algorithm

We next consider the recovery of the delays of each path. Note that (13) implies that the delay factor τ_p^{ul} has a Vandermonde vector $\boldsymbol{b}(\tau_p^{\text{ul}})$ appearing in the expression for each row of $\boldsymbol{H}_{\mathcal{N}}^{\text{ul}}$. Consequently, we can use the same block structure to estimate the delay by vectorizing $\boldsymbol{Y}_{\mathcal{N}}$. Similarly, we use $\boldsymbol{\tau} \triangleq [\tau_1, \tau_2, \ldots, \tau_{L_\tau^{(0)}}]$, as the off-grid delays to be estimated, where $\tau_i = \frac{i}{L_\tau^{(0)} N f_o}$, $i \in \{1, 2, \ldots, L_\tau^{(0)}\}$, and $L_\tau^{(0)}$ is the initial number of unknown delays. Thus, the sensing matrix for delay estimation can be formulated similarly to (24). To that aim, define

$$\begin{aligned}\boldsymbol{D}_t(\boldsymbol{\tau}) &\triangleq [\boldsymbol{b}(\tau_1)\otimes\boldsymbol{I}_M, \boldsymbol{b}(\tau_2)\otimes\boldsymbol{I}_M, \ldots, \boldsymbol{b}(\tau_{L_\tau}^{(0)})\otimes\boldsymbol{I}_M] \\ &= [\boldsymbol{d}_t(\tau_1), \boldsymbol{d}_t(\tau_2), \ldots, \boldsymbol{d}_t(\tau_{L_\tau^{(0)}})],\end{aligned} \tag{39}$$

where $\boldsymbol{d}_t(\tau_i) \triangleq \boldsymbol{b}(\tau_i)\otimes\boldsymbol{I}_M$. It can be readily checked that

$$\text{vec}(\boldsymbol{Y}_{\mathcal{N}}) \approx \boldsymbol{D}_t(\boldsymbol{\tau})\boldsymbol{x}_t + \text{vec}(\boldsymbol{E}_{\mathcal{N}}), \tag{40}$$

Algorithm 2: Delay Estimation.

- **Initialize:** Input $\boldsymbol{Y}_{\mathcal{N}}$, ${\boldsymbol{x}_t}^{(0)} = \boldsymbol{0}, \boldsymbol{\tau}^{(0)} = \boldsymbol{0}$, $L_\tau^{(0)}$, and $\lambda^{(0)}$. Iteration index $\omega_t = 0$.
- **Step 1:** At iteration ω_t, based on the previous results ${\boldsymbol{x}_t}^{(\omega_t)}, \lambda^{(\omega_t)}$, refresh $\boldsymbol{G_t}$ in (32) and calculate: $v_t(\boldsymbol{\tau}^{(\omega)}) = -\text{vec}(\boldsymbol{Y}_{\mathcal{N}})^H \boldsymbol{D}_t(\boldsymbol{\tau}^{(\omega_t)})(\boldsymbol{D}_t^H(\boldsymbol{\tau}^{(\omega_t)})$ $\boldsymbol{D}_t(\boldsymbol{\tau}^{(\omega_t)}) + (\lambda_t^{(\omega)})^{-1}\boldsymbol{G_t}^{(\omega_t)})^{-1}\boldsymbol{D}_t^H(\boldsymbol{\tau}^{(\omega_t)})\text{vec}(\boldsymbol{Y}_{\mathcal{N}})$.
- **Step 2:** Search $v_t(\boldsymbol{\tau}^{(\omega_t)})$ with gradient descent method for a new estimation of $\boldsymbol{\tau}^{(\omega_t+1)}$ via $\boldsymbol{\tau}^{(\omega_t+1)} = \boldsymbol{\tau}^{(\omega_t)} - u\frac{\partial v_t(\boldsymbol{\tau}^{(\omega_t)})}{\partial \boldsymbol{\tau}}$.
- **Step 3:** Compute ${\boldsymbol{x}_t}^{(\omega_t+1)} = (\boldsymbol{D}_t^H(\boldsymbol{\tau}^{(\omega_t)})\boldsymbol{D}_t(\boldsymbol{\tau}^{(\omega_t)}) + \lambda_t^{-1}\boldsymbol{G}_t^{(\omega_t)})^{-1}\boldsymbol{D}_t(\boldsymbol{\tau}^{(\omega_t)})^H\text{vec}(\boldsymbol{Y}_{\mathcal{N}})$ and $\lambda_t^{(\omega_t+1)}$ viae: (34).
- **Step 4:** If $\|\boldsymbol{x}_t^{(\omega_t+1)}[i]\| < \mu$, then remove it, and delete the relevant delay $\tau_i^{(\omega_t+1)}$.
- **Step 5:** Update $L_\tau^{(\omega_t+1)}$ based on the new length of $(\boldsymbol{\tau})^{(\omega_t+1)}$.
- **Step 6:** If $\|{\boldsymbol{x}_t}^{(\omega_t+1)} - {\boldsymbol{x}_t}^{(\omega_t)}\|_2 \geq \eta$, then set $\omega_t \leftarrow \omega_t + 1$ and go to step 1, otherwise stop.

where $\boldsymbol{x}_t$ is an $ML_\tau^{(0)} \times 1$ block sparse vector defined similarly to the vector $\boldsymbol{x}$ introduced in the previous subsection.

For a fixed delay grid $\boldsymbol{\tau}$, the delay estimation is then expressed as

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{x}_t\|_{0,M} \quad \text{s. t.} \quad \|\text{vec}(\boldsymbol{Y}_{\mathcal{N}}) - \boldsymbol{D}_t(\boldsymbol{\tau})\boldsymbol{x}_t\|_2 \leq \xi_t, \tag{41}$$

which can be solved in a similar manner as (31), namely, by iteratively updating the estimate of the vector $\boldsymbol{x}_t$ and the grid $\boldsymbol{\tau}$ using a block-sparsity boosting iterative reweighted least squares objective. The proposed delay estimation procedure and algorithm is summarized in **Algorithm 2**.

D. Uplink Channel Reconstruction

In the previous subsections we showed how the estimations of the DOAs and the delays, denoted $\hat{\boldsymbol{\phi}}^{\text{ul}}$ and $\hat{\boldsymbol{\tau}}^{\text{ul}}$ are separately obtained, along with the number of paths $\hat{P}$. Yet, one still has to match each DOA value to its corresponding delay, namely, to match the entries of $\hat{\boldsymbol{\phi}}^{\text{ul}}$ to their corresponding entries $\hat{\boldsymbol{\tau}}^{\text{ul}}$. A trivial approach is to search all of the $\hat{P}^{\hat{P}}$ possible pairings among $\hat{\boldsymbol{\phi}}^{\text{ul}}$ and $\hat{\boldsymbol{\tau}}^{\text{ul}}$, but this would incur a heavy computational burden. Fortunately, based on Theorem 1, if $\hat{\phi}_i^{\text{ul}}$ and $\hat{\tau}_j^{\text{ul}}$ belong to the same path, then the projection of their channel vector $\text{vec}(\boldsymbol{\Xi}(\hat{\phi}_i^{\text{ul}}, \hat{\tau}_j^{\text{ul}}))$ onto $\text{vec}(\boldsymbol{Y}_{\mathcal{N}})$ will be a larger value compared to other (mismatched) combination.

Consequently, we can try different combinations of $\text{vec}(\boldsymbol{\Xi}(\hat{\phi}_i^{\text{ul}}, \hat{\tau}_j^{\text{ul}}))$ to perform the inner product operation with $\text{vec}(\boldsymbol{Y}_{\mathcal{N}})$, and select the maximum value from each operation as the correct matching of the angle $\hat{\phi}^{\text{ul}}$ for delay $\hat{\tau}^{\text{ul}}$. Next, we delete the angle and delay that have been matched already, and perform the inner product for the remaining angles and delays. Under this pairing procedure, at most $\sum_{p=1}^{\hat{P}} p^2 = \hat{P}^2$ inner product operations are required.

The above sequential procedure, which facilitates the angle-delay pairing in terms of complexity, may induce error propagation in the presence of mismatch. Nonetheless, as also observed in our numerical study in Section V, such mismatches do not have a significant effect on the reconstruction performance. This can be explained by recalling that by *Theorem 1*, if one pairing of angle-delay is mismatched, the inner product of its channel vector with $\text{vec}(\boldsymbol{Y}_{\mathcal{N}})$ is too weak to be selected. Consequently, a mismatch commonly occurs when two incident paths have very similar angle, delay and energy information, simultaneously, and can thus be treated as a dominant cluster, resulting in a minimal degradation in reconstruction performance.

After the pairing process is concluded, one can recover the complex channel gains. To that aim, we stack the obtained channel vectors (with correct matching) $\text{vec}(\boldsymbol{\Xi}(\hat{\phi}_p^{\text{ul}}, \hat{\tau}_p^{\text{ul}}))$, where $1 \leq p \leq \hat{P}$, as columns to form a matrix:

$$\boldsymbol{B} \triangleq [\text{vec}(\boldsymbol{\Xi}(\hat{\phi}_1^{\text{ul}}, \hat{\tau}_1^{\text{ul}})), \ldots, \text{vec}(\boldsymbol{\Xi}(\hat{\phi}_{\hat{P}}^{\text{ul}}, \hat{\tau}_{\hat{P}}^{\text{ul}}))]. \tag{42}$$

Now, we note that if there is no error in the estimation of the DOAs and the delays, then the true channel gains can be computed via $\boldsymbol{\beta} = \boldsymbol{B}^\dagger \text{vec}(\boldsymbol{H}^{\text{ul}})$. However, as we do not have access to the true channel $\boldsymbol{H}^{\text{ul}}$, we estimate the gains by applying the same transformation to the least-squares estimation of $\boldsymbol{H}^{\text{ul}}$, i.e., $\boldsymbol{Y}_{\mathcal{N}}(\boldsymbol{S}^{\text{ul}})^{-1}$. Specifically, the uplink complex channel gains are estimated as

$$\hat{\boldsymbol{\beta}}^{\text{ul}} = \boldsymbol{B}^\dagger \text{vec}\left(\boldsymbol{Y}_{\mathcal{N}}(\boldsymbol{S}^{\text{ul}})^{-1}\right) = [\hat{\beta}_1^{\text{ul}}, \hat{\beta}_2^{\text{ul}}, \ldots, \hat{\beta}_{\hat{P}}^{\text{ul}}]^T. \tag{43}$$

The overall uplink channel at all N subcarriers of the kth user are then reconstructed from (13) as

$$\hat{\boldsymbol{H}}_k^{\text{ul}} = \sum_{p=1}^{\hat{P}} \hat{\beta}_p^{\text{ul}} \boldsymbol{\Xi}(\hat{\phi}_p^{\text{ul}}, \hat{\tau}_p^{\text{ul}}). \tag{44}$$

We summarize the overall proposed uplink channel reconstruction scheme as **Algorithm 3**.

IV. Downlink Channel Estimation

In the previous section we proposed an algorithm for estimating the uplink channel. However, in order to establish reliable bi-directional communications, the downlink channel must also be estimated. One of the major challenges in massive MIMO communications stems from the fact that in FDD systems, in which different bands are assigned to uplink and downlink transmissions, the downlink channel cannot be immediately deduced from the uplink channel. In the following we show how, for mmWave massive MIMO systems with the BSE, the estimation scheme designed for uplink channels in the previous section can be extended to downlink channels by exploiting a phenomenon called *angle-delay reciprocity*.

In particular, we consider pilot-aided downlink channel estimation in which the BS transmits a-priori known pilot sequence in a similar manner to the uplink channel estimation phase. To deal with the BSE, we design the downlink estimation strategy to use dedicated pilots for each path and using different

Algorithm 3: Uplink Channel Reconstruction Scheme.

Input: The BS receives matrix $\boldsymbol{Y}_{\mathcal{N}}$ via (17).

Output: The estimated channel matrix $\hat{\boldsymbol{H}}_k^{\text{ul}}$ over N subcarriers, Number of paths $\hat{P}$, DOAs $\hat{\boldsymbol{\phi}}^{\text{ul}}$, delays $\hat{\boldsymbol{\tau}}^{\text{ul}}$, and complex gain $\hat{\boldsymbol{\beta}}^{\text{ul}}$ for each user.

1: Calculate the DOAs $\phi_i, 1 \le i \le \hat{P}$, using **Algorithm 1** with input $\boldsymbol{Y}_{\mathcal{N}}$.
2: Calculate the delays $\tau_j, 1 \le j \le \hat{P}$, using **Algorithm 2** with input $\boldsymbol{Y}_{\mathcal{N}}$.
3: **while** $p < \hat{P}$ **do**
4: value $\leftarrow$ zeros$[\hat{P}-p+1, \hat{P}-p+1]$
5: **for** $j = p : \hat{P}$ **do**
6: value$[p,j] \leftarrow \text{vec}(\boldsymbol{Y}_{\mathcal{N}})^H \text{vec}(\boldsymbol{\Xi}(\phi_p, \tau_j))$
7: **end for**
8: $[pos_p, pos_j] \leftarrow \textbf{find}(\textbf{max}(\text{value}))$
9: $\hat{\phi}_p^{\text{ul}} = \phi_{pos_p}; \hat{\tau}_p^{\text{ul}} = \tau_{pos_j}$
10: $p \leftarrow p+1$
11: **end while**
12: Stack vec$(\boldsymbol{\Xi}(\hat{\phi}_p^{\text{ul}}, \hat{\tau}_p^{\text{ul}}))$ into the columns of the matrix: $\boldsymbol{B}_k = [\text{vec}(\boldsymbol{\Xi}(\hat{\phi}_1^{\text{ul}}, \hat{\tau}_1^{\text{ul}})), \ldots, \text{vec}(\boldsymbol{\Xi}(\hat{\phi}_{\hat{P}}^{\text{ul}}, \hat{\tau}_{\hat{P}}^{\text{ul}}))]$.
13: Calculate the complex gain as $\hat{\boldsymbol{\beta}} = \boldsymbol{B}^{\dagger} \text{vec}(\boldsymbol{Y}_{\mathcal{N}} (\boldsymbol{S}^{ul})^{-1})$.
14: Recover the channel matrix $\hat{\boldsymbol{H}}_k^{\text{ul}}$ via (44).
15: **return** $\hat{\boldsymbol{\phi}}^{\text{ul}}, \hat{\boldsymbol{\tau}}^{\text{ul}}, \hat{\boldsymbol{\beta}}^{\text{ul}}, \hat{\boldsymbol{H}}_k^{\text{ul}}$.

beamforming vectors for different subcarriers with respect to the given path. To present our scheme, we first elaborate on the structure of mmWave massive MIMO downlink channels in the presence of BSE in Subsection IV-A. Then, in Subsection IV-B, we discuss the angle-delay reciprocity, a property which we exploit in Subsection IV-C to generate the beamformed pilots and to formulate our downlink channel estimation algorithm.

A. Downlink Channel Structure

To formulate the downlink channel, we denote its center frequency f_c^{dl} and wavelength $\lambda^{\text{dl}} = c/f_c^{\text{dl}}$. The downlink array steering vector is expressed as

$$\boldsymbol{a}^{\text{dl}}(\phi^{\text{dl}}, f) = \left[1, e^{-j(1+\frac{f}{f_c^{\text{dl}}})\phi^{\text{dl}}}, \ldots, e^{-j(M-1)(1+\frac{f}{f_c^{\text{dl}}})\phi^{\text{dl}}}\right]^T, \tag{45}$$

where $\phi^{\text{dl}} \triangleq \frac{2\pi d \cdot \sin\theta^{\text{dl}}}{\lambda^{\text{dl}}}$ is the normalized direction of departure (DOD), and θ^{dl} is the downlink DOD. We henceforth use the term "DOD" to refer to the normalized DOD $\boldsymbol{\phi}^{\text{dl}}$.

Similar to the uplink case, at frequency f, the downlink channel observed by the kth user is written as:

$$\boldsymbol{h}_k^{\text{dl}}(f) = \sum_{p=1}^{P^{\text{dl}}} \bar{\beta}_p^{\text{dl}} \boldsymbol{a}^{\text{dl}}(\phi_p^{\text{dl}}, f) e^{-j2\pi f \tau_p^{\text{dl}}}, \tag{46}$$

where P^{dl} is the number of downlink paths, $\bar{\beta}_p^{\text{dl}}$ and τ_p^{dl} are the complex gain and multi-path delay of the pth downlink path, respectively. The channel parameters, P^{dl}, $\bar{\beta}_p^{\text{dl}}$, τ_p^{dl}, and ϕ_p^{dl} depend on the specific user index k, which is omitted for brevity, as in the previous section.

As in (11) and (13), we define :

$$\left[\boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}})\right]_{m,n} \triangleq \exp\left(-jm\frac{nf_0}{f_c^{\text{dl}}}\phi_p^{\text{dl}}\right);$$
$$\boldsymbol{\Xi}^{\text{dl}}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}) \triangleq \left[\boldsymbol{a}(\phi_p^{\text{dl}}, 0)\boldsymbol{b}^T(\tau_p^{\text{dl}})\right] \odot \boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}}). \tag{47}$$

Then, the downlink $1 \times MN$ frequency channel vector from the BS to the kth user is given by

$$\begin{aligned}\boldsymbol{H}_k^{\text{dl}} &= \left[\text{vec}\left[\boldsymbol{h}_k^{\text{dl}}(0), \boldsymbol{h}_k^{\text{dl}}(f_0), \ldots, \boldsymbol{h}_k^{\text{dl}}((N-1)f_0)\right]\right]^H \\ &= \sum_{p=1}^{P^{\text{dl}}} \left[\bar{\beta}_p^{\text{dl}} \text{vec}\left(\left[\boldsymbol{a}^{\text{dl}}(\phi_p^{\text{dl}}, 0)\boldsymbol{b}^T(\tau_p^{\text{dl}})\right] \odot \boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}})\right)\right]^H \\ &= (\boldsymbol{\beta}^{\text{dl}})^H \boldsymbol{\Omega}_k(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}})^H, \end{aligned} \tag{48}$$

where $\boldsymbol{\beta}^{\text{dl}} = [\bar{\beta}_1^{\text{dl}}, \bar{\beta}_2^{\text{dl}}, \ldots, \bar{\beta}_{P^{\text{dl}}}^{\text{dl}}]^T$ is the downlink channel complex gain vector, and

$$\boldsymbol{\Omega}_k(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}}) = [\text{vec}\left(\boldsymbol{\Xi}^{\text{dl}}(\phi_1^{\text{dl}}, \tau_1^{\text{dl}})\right), \ldots, \text{vec}\left(\boldsymbol{\Xi}^{\text{dl}}(\phi_{P^{\text{dl}}}^{\text{dl}}, \tau_{P^{\text{dl}}}^{\text{dl}})\right)]. \tag{49}$$

From (48), the downlink channel consists of a set of DODs, delays, and complex gains, and thus has a similar structure as the uplink case. Note that in FDD systems, channel reciprocity does not hold, and the channel must be estimated independently by each user. However, this estimation can be facilitated by accounting for the angle-delay reciprocity of mmWave channels, presented in the following subsection.

B. Angle-Delay Reciprocity

Unlike in TDD systems [46], FDD channels are not reciprocal, namely, downlink and uplink transmissions undergo different channels due to their different frequency bands. However, since the propagation paths of electromagnetic waves are reciprocal, only the signal wave that physically reverses the uplink path reaches users during downlink transmission. It has been shown in [47]–[49] that the conductivity and relative permittivity of most materials remain unchanged if the frequency of the electromagnetic wave does not vary much, say less than 1 GHz. Hence, the angle components of the uplink and downlink channels commonly are the same in FDD systems [18], [31], [32]. Moreover, since the downlink electromagnetic wave travels the same distance as the uplink, the delay components for the uplink and downlink channel are the same. This phenomenon is commonly referred to as *angle-delay reciprocity*, and it implies that

$$\theta_p^{\text{dl}} = \theta_p^{\text{ul}}, \quad \tau_p^{\text{dl}} = \tau_p^{\text{ul}}, \quad P^{\text{dl}} = P^{\text{ul}}. \tag{50}$$

From (50) it holds that the uplink and downlink channels have the same number of paths as well as the same angle and delay parameters, which can be estimated at the BS. Previous studies show that the channel complex gains are generally different between the uplink and downlink [50], [51] due to the phase difference. Therefore, to acquire accurate downlink channel in FDD systems, the users need to estimate the remaining downlink channel gain and feedback this gain to the BS. The resulting computational overhead, as we show in the sequel, can be made affordable.

C. Downlink Channel Estimation

We now show how the angle-delay reciprocity may be exploited to estimate the downlink channel. Here, the BS estimates (50), using the number of uplink paths along with their DOAs and delays obtained via **Algorithm 3**.

Recall that the BS transmits beamformed pilots to each user during downlink channel estimation. In particular, the BS sends a-priori known pilots in each estimated path. Let $\boldsymbol{s}_p^{\text{dl}} \in \mathbb{C}^{1\times T}$ denote the pilots targeting the pth path over T subcarriers with index set $\mathcal{N} = \{n_1, n_2, \ldots, n_T\}$. These pilots are orthogonal over the different paths, i.e., $\boldsymbol{s}_i^{\text{dl}}(\boldsymbol{s}_j^{\text{dl}})^H = \delta(i-j)$. To formulate how these pilots are beamformed prior to their transmission, we use the channel structure in (48) and formulate the columns of the matrix $\boldsymbol{\Xi}^{\mathcal{N}}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}})$ as follows: for the n_qth ($q = 1, 2, \ldots, T$) pilot subcarrier, the corresponding column of $\boldsymbol{\Xi}^{\mathcal{N}}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}})$ is given by

$$\boldsymbol{\Xi}^{\mathcal{N}}_{:,q}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}) = \left[\boldsymbol{a}(\phi_p^{\text{dl}}, 0)\boldsymbol{b}_{n_q}^T(\tau_p^{\text{dl}})\right] \odot [\boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}})]_{:,n_q}, \quad (51)$$

where $\boldsymbol{b}_{n_q}^T(\tau_p^{\text{dl}})$ is the n_qth element of $\boldsymbol{b}^T(\tau_p^{\text{dl}})$, and $[\boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}})]_{:,n_q}$ is the n_qth column of $\boldsymbol{W}^{\text{dl}}(\phi_p^{\text{dl}})$. By letting $\boldsymbol{F}_{n_q}^p(\phi_p^{\text{dl}}, \tau_p^{\text{dl}})$ be the beamforming vector for the kth user from the pth path on the n_qth pilot carrier, the corresponding channel output (prior to the addition of noise) can be written as

$$r_{n_q,p}^{\text{dl}} = \left[\bar{\beta}_p^{\text{dl}}\text{vec}(\boldsymbol{\Xi}^{\mathcal{N}}_{:,q}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}))\right]^H \boldsymbol{F}_{n_q}^p\left(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}\right)\boldsymbol{s}_p^{\text{dl}}(q). \quad (52)$$

In order to point the downlink beam to the pth path, we set the beamforming vector to be

$$\boldsymbol{F}_{n_q}^p\left(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}\right) = \frac{1}{M}\text{vec}\left(\boldsymbol{\Xi}^{\mathcal{N}}_{:,q}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}})\right). \quad (53)$$

The $1 \times T$ received vector from all T pilot carriers is now written as

$$\begin{aligned}\boldsymbol{y}_{\mathcal{N}}^{\text{dl}} &= \left[\sum_{p=1}^{P^{\text{dl}}} r_{n_1,p}^{\text{dl}}, \quad \sum_{p=1}^{P^{\text{dl}}} r_{n_2,p}^{\text{dl}} \quad , \ldots, \quad \sum_{p=1}^{P^{\text{dl}}} r_{n_T,p}^{\text{dl}}\right] + \boldsymbol{e}^{\text{dl}} \\ &= \boldsymbol{H}_{\mathcal{N}}^{\text{dl}}\boldsymbol{F}_{\mathcal{N}}(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}})\boldsymbol{S}^{\text{dl}} + \boldsymbol{e}^{\text{dl}}, \end{aligned} \quad (54)$$

where $\boldsymbol{H}_{\mathcal{N}}^{\text{dl}} \triangleq \sum_{p=1}^{P^{\text{dl}}}\left[\bar{\beta}_p^{\text{dl}}\text{vec}(\boldsymbol{\Xi}^{\mathcal{N}}(\phi_p^{\text{dl}}, \tau_p^{\text{dl}}))\right]^H$ is the $1 \times MT$ downlink channel over the T pilot carriers, $\boldsymbol{S}^{\text{dl}} = [(\boldsymbol{s}_1^{\text{dl}})^T, (\boldsymbol{s}_2^{\text{dl}})^T, \ldots, (\boldsymbol{s}_{P^{\text{dl}}}^{\text{dl}})^T]^T$ is the pilot matrix, and $\boldsymbol{F}_{\mathcal{N}}(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}})$ is the beamforming matrix given by

$$\begin{aligned}\boldsymbol{F}_{\mathcal{N}}(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}}) = \frac{1}{MT}[&\text{vec}(\boldsymbol{\Xi}^{\mathcal{N}}(\phi_1^{\text{dl}}, \tau_1^{\text{dl}})), \ldots, \\ &\text{vec}(\boldsymbol{\Xi}^{\mathcal{N}}(\phi_{P^{\text{dl}}}^{\text{dl}}, \tau_{P^{\text{dl}}}^{\text{dl}}))]. \end{aligned} \quad (55)$$

Due to the beamforming matrix in (55), the user can use the a-priori knowledge of $\boldsymbol{S}^{\text{dl}}$ to recover its downlink channel complex gain vector using simple least-squares estimation:

$$(\hat{\boldsymbol{\beta}}^{\text{dl}})^H = \boldsymbol{y}_{\mathcal{N}}^{\text{dl}}(\boldsymbol{S}^{\text{dl}})^{\dagger}. \quad (56)$$

To allow the BS to recover the complete downlink channel, each user feedbacks its estimated gain vector $\hat{\boldsymbol{\beta}}^{\text{dl}}$ to the BS, completing the downlink channel reconstruction via:

$$\hat{\boldsymbol{H}}_k^{\text{dl}} = (\hat{\boldsymbol{\beta}}^{\text{dl}})^H\boldsymbol{\Omega}_k(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}})^H. \quad (57)$$

Algorithm 4: Downlink Channel Reconstruction Scheme.

Input: The uplink channel estimation parameters $\hat{\boldsymbol{\phi}}^{\text{ul}}, \hat{\boldsymbol{\tau}}^{\text{ul}}$, $\hat{P}$ for the kth user; The downlink orthogonal pilots $\boldsymbol{S}^{\text{dl}}$ for the kth user.

Output: The reconstruction downlink channel matrix $\hat{\boldsymbol{H}}_k^{\text{dl}}$.

1: Uplink channel estimation: the kth user sends orthogonal pilots $\boldsymbol{S}^{\text{ul}}$ to BS. The BS uses **Algorithm 3** for estimating uplink channel, recovering the DOAs, delays, and multi-path numbers.
2: Downlink training: The BS generates the beamforming matrix $\boldsymbol{F}_{\mathcal{N}}(\boldsymbol{\phi}^{\text{dl}}, \boldsymbol{\tau}^{\text{dl}})$ in (55), and sends the pilots $\boldsymbol{S}^{\text{dl}}$ to the kth user. After receiving the pilots, the kth user recovers the complex gain $\hat{\boldsymbol{\beta}}^{\text{dl}}$ via (56).
3: Downlink channel reconstruction: the kth user feedbacks the downlink complex gain to the BS. Then the downlink channel is reconstructed via (57)
4: **return** $\hat{\boldsymbol{H}}_k^{\text{dl}}$.

We summarize the downlink channel reconstruction scheme as **Algorithm 4**.

V. Simulations

In this section, we demonstrate the effectiveness of the proposed algorithms for uplink and downlink channel estimation compared to conventional channel estimation algorithms. In particular, we show that our approach significantly outperforms previously proposed methods, which either restrict the solution set to a finite grid, as in [16], [17], or, alternatively, do not take into account the BSE. We also compare our off-grid recovery schemes to conventional grid refinement [30].

We consider a BS equipped with a ULA with element spacing $d = \lambda^{\text{ul}}/2$. All $K = 8$ users are randomly distributed in the service area and each has a single antenna. The pilots are uniformly distributed over all the $N = 64$ subcarriers, thus each user is assigned $T = 8$ pilot subcarriers. The transmit bandwidth is 1 GHz with uplink center frequency $f_c^{\text{ul}} = 60$ GHz and downlink center frequency $f_c^{\text{dl}} = 61$ GHz. For the on-grid approach, we take $L = 1024$ grid points. For the proposed off-grid approach, we correspondingly set $L_\phi = 1024$ as the initial resolution. The simulated mmWave channels are generated via (12) with $P = 6$, $\bar{\beta}_p \sim \mathcal{CN}(0, 1)$, $\phi_p \sim \mathcal{U}(-\pi, \pi)$, and $\tau_p \sim \mathcal{U}(0, \frac{1}{(N-1)f_0})$, where $\mathcal{CN}$ and $\mathcal{U}$ represent the complex normal and the uniform distribution, respectively, recalling that N and f_0 denote the number of subcarrier and the carrier spacing, respectively. The signal-to-noise ratio (SNR) is defined as σ_p^2/σ_n^2, where σ_p^2 is the pilot power. The performance of angle and delay estimation are measured by the corresponding mean-square error (MSE) values, defined as

$$\text{MSE}_{\boldsymbol{\phi}} \triangleq \frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J}\|\boldsymbol{\phi}_{k,j} - \hat{\boldsymbol{\phi}}_{k,j}\|_2^2, \quad (58)$$

$$\text{MSE}_{\boldsymbol{\tau}} \triangleq \frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J}\|\boldsymbol{\tau}_{k,j} - \hat{\boldsymbol{\tau}}_{k,j}\|_2^2, \quad (59)$$

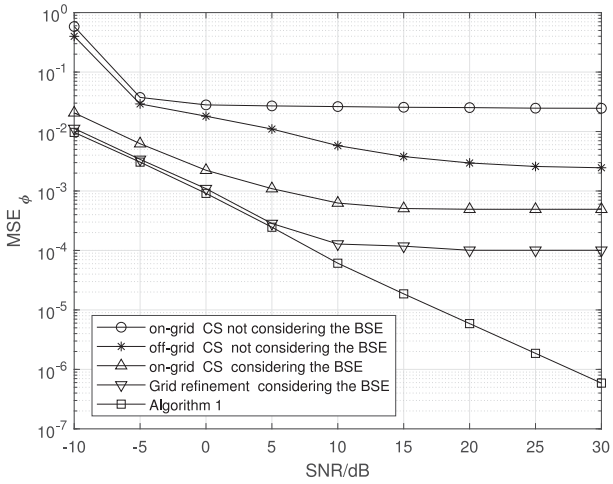

Fig. 4. MSE vs. SNR, DOA estimation, uplink channel.

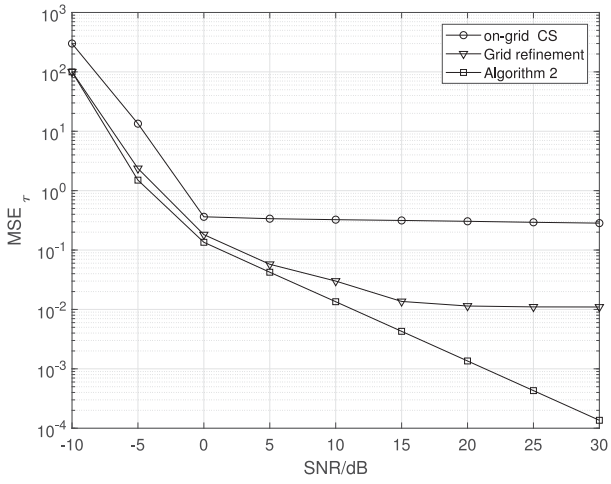

Fig. 5. MSE vs. SNR, delay estimation, uplink channel.

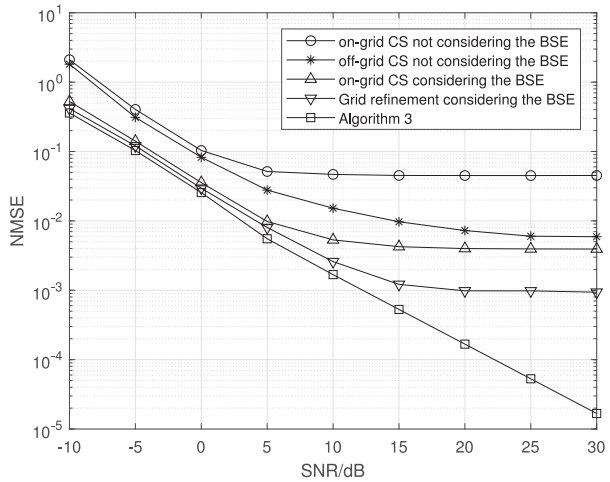

Fig. 6. NMSE vs. SNR, reconstructed uplink channel.

respectively. Here, $J = 1000$ is the number of Monte-Carlo trials. The channel estimation performance is measured in terms of the normalized mean square error (NMSE):

$$\text{NMSE} \triangleq \frac{1}{KJ} \sum_{k=1}^{K} \sum_{j=1}^{J} \frac{\|\boldsymbol{H}_{k,j} - \hat{\boldsymbol{H}}_{k,j}\|_F^2}{\|\boldsymbol{H}_{k,j}\|_F^2}. \tag{60}$$

Fig. 4 shows the uplink MSE$_{\boldsymbol{\phi}}$ versus SNR of the proposed algorithm compared with on-grid CS [17], grid refinement approach [30] and conventional channel modeling [45] that ignores the BSE. While the MMV structure is not considered in these algorithms [17], [30], [45], in the following we allow the competing algorithms to exploit this structure in order to maintain a fair comparison. Observing Fig. 4, we note that the curve for the MSE in recovering the DOAs of the proposed algorithm decreases linearly as SNR increases (indicating that MSE$_{\boldsymbol{\phi}}$ decays exponentially with increasing SNR), while all the other methods meet error floors at high SNR. This error floor exhibited by previously proposed estimators, which emphasizes the benefits of our algorithms in high SNR values, is a result of a model mismatch which can be attributed to: 1) For the on-grid algorithms, the grid mismatch restricts the resolution of DOA estimation to be $2\pi/L$; 2) For the grid refinement algorithm, the error floor occurs due to convergence to local optimum points for some of the parameters, as mentioned in Section III-B. 3) For the off-grid algorithm, neglecting the BSE significantly decreases channel estimation accuracy. It is also observed from the second and third curves in Fig. 4 that the degradation due to ignoring the BSE is more substantial compared to grid mismatch.

Fig. 5 depicts the MSE in recovering the delays for the same setup. Since the BSE does not influence the delay estimation, we only present the results for the on-grid, grid refinement, and the proposed off-grid approaches. Similarly to Fig. 4, the on-grid and grid refinement methods inevitably encounter an error floor at high SNRs, while the proposed off-grid method consistently improves performance.

Next, we compare the NMSE in recovering the uplink channel using **Algorithm 3**. The results are depicted in Fig. 6. It is observed that the NMSE curve for the proposed algorithm decreases linearly as the SNR increases, achieving significantly better performance than competing methods. Furthermore, the channel estimation accuracy of the other four techniques all meet error floors at high SNR, in correspondence with their MSE$_{\boldsymbol{\phi}}$ and MSE$_{\boldsymbol{\tau}}$ performance.

We next study the effect of bandwidth on the performance of the proposed estimators. The SNR is set to 10 dB and the pilots are uniformly distributed over all $N = 64$ subcarriers. Fig. 7 depicts the MSE in estimating the DOAs for uplink mmWave massive MIMO channels of the proposed algorithm compared to off-grid approaches [45] that either ignore the BSE or do not utilize the MMV structure under various transmission bandwidths. Note that, the propagation delays across the array of $f_s = 20$ MHz, $f_s = 100$ MHz, and $f_s = 1$ GHz are $0.0011T_s$, $0.053T_s$, and $0.53T_s$, respectively. It is observed in Fig. 7 that the proposed algorithm achieves the best estimation accuracy and that its superiority over previously proposed estimators is consistent for various bandwidths. When the bandwidth is as small as 20 MHz, it is noted that the off-grid approach considering the MMV structure but ignoring the BSE has the same performance as our method. This is because the BSE is not pronounced when

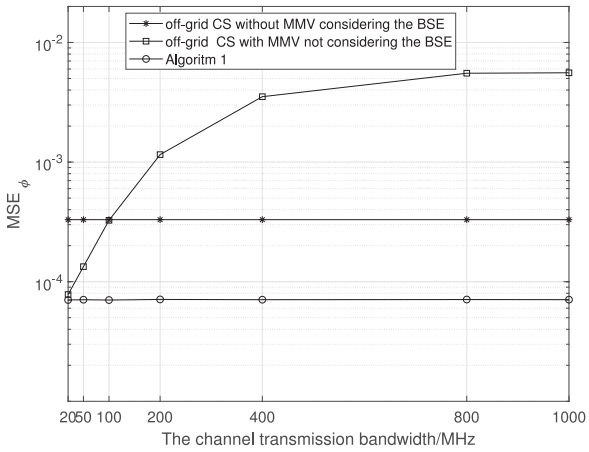

Fig. 7. MSE vs. bandwidth, DOA estimation, uplink channel.

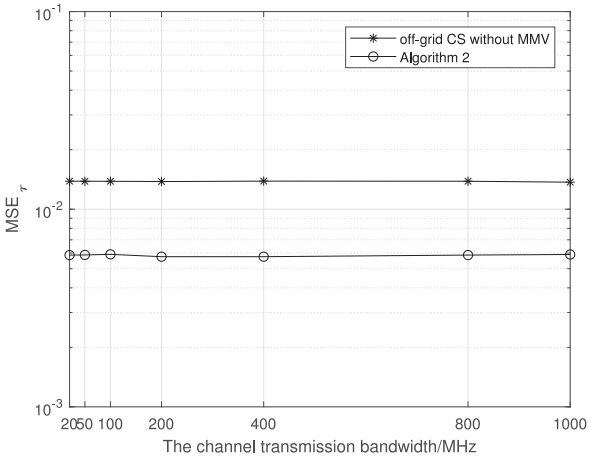

Fig. 8. MSE vs. bandwidth, delay estimation, uplink channel.

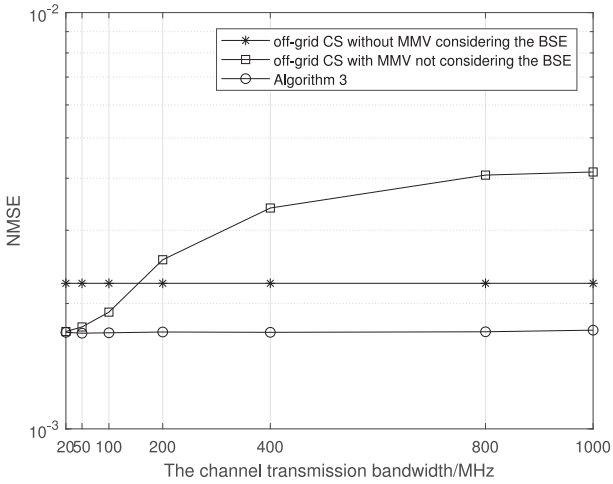

Fig. 9. NMSE vs. bandwidth, reconstructed uplink channel.

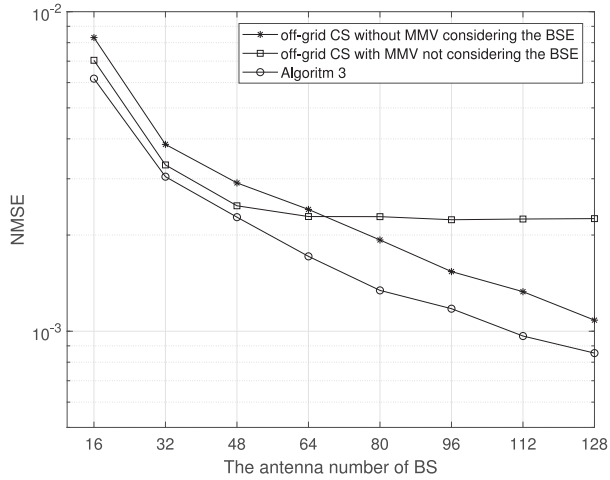

Fig. 10. NMSE vs. antenna number of BS, reconstructed uplink channel.

the bandwidth is small. However, as the bandwidth increases, the performance of the algorithms which ignore the BSE quickly deteriorates. Furthermore, the algorithm that does not utilize the MMV structure but considers the BSE performs worse than the proposed one, and the performance gap remains constant for different bandwidth values. This demonstrates that properly exploiting the block sparsity can improve the performance and this improvement is not affected by bandwidth.

Fig. 8 displays the uplink MSE in recovering the delays with the same system setup in Fig. 7. Because the BSE does not affect the delay estimation, the MSE in estimating the path delays for both algorithms remains constant as the bandwidth increases. Similarly to Fig. 7, the proposed algorithm still achieves superior performance for all bandwidth values.

Fig. 9 compares the NMSE of the reconstructed uplink channel via **Algorithm 3** with the competing approaches under various bandwidths. As the bandwidth increases, we observe that the NMSE of our algorithm remains constant and yields the best performance over all the techniques, as can be indicated from the results depicted in Figs. 7–8.

In Fig. 10, we show how the uplink channel estimation performance changes over the number of antennas in comparison with Fig. 9. The SNR is set to 10 dB, uplink carrier frequency $f_c^{\text{ul}} = 60$ GHz, the transmission bandwidth is kept as 100 MHz, and the pilots are uniformly distributed over all $N = 64$ subcarriers. It is observed in Fig. 10 that when the number of antennas M increases, satisfying $M > 48$, the BSE gradually becomes severe and the proposed algorithm (**Algorithm 3**) considering BSE becomes most beneficial. Moreover, when M increases to 64, the conventional method that does not consider BSE still meets an error floor as the antenna number increasing. This demonstrates the importance of properly accounting for the BSE in massive MIMO systems, in which the antenna number is typically large, compared to conventional MIMO systems where M is relatively small.

In the final example, we evaluate the NMSE of the reconstructed channel for both the uplink and downlink with different number of antennas, in Fig. 11. Observing Fig. 11, we note that the proposed algorithm is very effective in estimating the downlink channel and achieves a consistent improvement in performance as SNR increases. Nevertheless, the NMSE in recovering the downlink channel is always worse than that of the uplink channel because the angle and delay parameters of the downlink channel originated from their uplink counterparts and

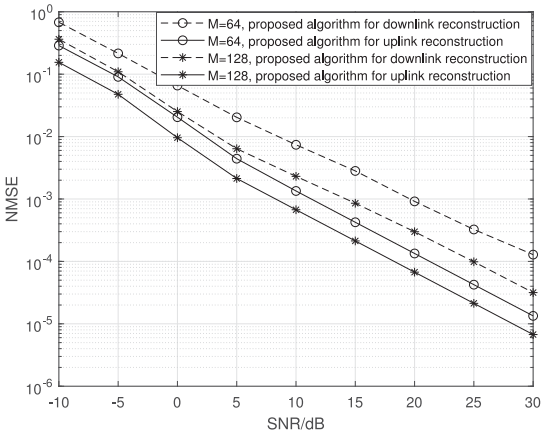

Fig. 11. NMSE vs. SNR of the reconstructed channel for both uplink and downlink channel under different antenna size.

may include estimation errors. These errors are combined with the estimation errors which arise due to the presence of noisy observations. Yet, as the number of antennas M increases, the channel recovery performance of both the uplink and downlink improves accordingly, indicating the potential benefits of our channel estimators for massive MIMO systems.

VI. Conclusion

In this paper, we designed channel estimation algorithms under a non-negligible beam squint effect in mmWave massive MIMO systems. We first showed that the recovery of the DOAs and delays of each channel path can be represented as an MMV problem using a shift invariant transformation, and developed an algorithm for recovering an off-grid estimate of these parameters. We then showed how these recovered values can be used to estimate the overall downlink channel. By exploiting the angle-delay reciprocity of mmWave channels, we extended the results derived for uplink channel estimation to a computationally efficient approach with low overhead for downlink channel estimation in FDD systems. Compared to previously proposed channel estimators, which either adopted an on grid approach or, alternatively, did not account for the beam squint modeling, the proposed algorithms provide significantly better performance. Numerical simulation results demonstrated the effectiveness of the proposed techniques, and have shown that properly taking the BSE into account is critical for mmWave massive MIMO systems.

Appendix

In this appendix, we compute the derivative of $v(\boldsymbol{\phi})$ in (37). The derivative of $v_t(\boldsymbol{\tau})$ in **Algorithm 2** is obtained in a similar fashion and is thus omitted for brevity.

From (36),

$$v(\boldsymbol{\phi}) = -\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)^H \boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}) \left(\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi})\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}) + \left(\lambda^{(\omega)}\right)^{-1}\boldsymbol{G}^{(\omega)}\right)^{-1} \boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi})\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T). \tag{61}$$

Define:

$$\boldsymbol{X} \triangleq \boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi})\left(\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi})\boldsymbol{D}_{\text{bs}}(\boldsymbol{\phi}) + \left(\lambda^{(\omega)}\right)^{-1}\boldsymbol{G}^{(\omega)}\right)^{-1}\boldsymbol{D}_{\text{bs}}^H(\boldsymbol{\phi}). \tag{62}$$

Based on the chain rule, the first derivative of $v(\boldsymbol{\phi})$ with respect to ϕ_i can be computed as

$$\frac{\partial v(\boldsymbol{\phi})}{\partial \phi_i} = \operatorname{tr}\left(\left(\frac{\partial v(\boldsymbol{\phi})}{\partial \boldsymbol{X}}\right)^T \frac{\partial \boldsymbol{X}}{\partial \phi_i}\right) + \operatorname{tr}\left(\left(\frac{\partial v(\boldsymbol{\phi})}{\partial \boldsymbol{X}^*}\right)^T \frac{\partial \boldsymbol{X}^*}{\partial \phi_i}\right), \tag{63}$$

where

$$\begin{aligned}\frac{\partial \boldsymbol{X}}{\partial \phi_i} &= \frac{\partial}{\partial \phi_i}\left(\boldsymbol{D}_{bs}(\boldsymbol{\phi})\left(\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \rho\boldsymbol{I}\right)^{-1}\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\right)\\ &= \frac{\partial \boldsymbol{D}_{bs}(\boldsymbol{\phi})}{\partial \phi_i}\left(\left(\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \rho\boldsymbol{I}\right)^{-1}\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\right)\\ &\quad + \left(\boldsymbol{D}_{bs}(\boldsymbol{\phi})\left(\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \rho\boldsymbol{I}\right)^{-1}\right)\frac{\partial \boldsymbol{D}_{bs}^H(\boldsymbol{\phi})}{\partial \phi_i}\\ &\quad - \boldsymbol{D}_{bs}(\boldsymbol{\phi})\left(\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \rho\boldsymbol{I}\right)^{-1}\\ &\quad \times\left(\frac{\partial \boldsymbol{D}_{bs}^H(\phi)}{\partial \phi_i}\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\frac{\partial \boldsymbol{D}_{bs}(\boldsymbol{\phi})}{\partial \phi_i}\right)\\ &\quad \times\left(\boldsymbol{D}_{bs}^H(\boldsymbol{\phi})\boldsymbol{D}_{bs}(\boldsymbol{\phi}) + \rho\boldsymbol{I}\right)^{-1}\boldsymbol{D}_{bs}^H(\boldsymbol{\phi}),\end{aligned} \tag{64}$$

and

$$\begin{aligned}\frac{\partial v(\boldsymbol{\phi})}{\partial \boldsymbol{X}} &= \frac{\partial}{\partial \boldsymbol{X}}\operatorname{tr}\left(\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)^H\boldsymbol{X}\right)\\ &= \operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^H)\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^H),\\ \frac{\partial v(\boldsymbol{\phi})}{\partial \boldsymbol{X}^*} &= \frac{\partial}{\partial \boldsymbol{X}^*}\operatorname{tr}\left(\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)\operatorname{vec}(\boldsymbol{Y}_{\mathcal{N}}^T)^H\boldsymbol{X}\right) = 0,\end{aligned} \tag{65}$$

where

$$\frac{\partial \boldsymbol{D}_{bs}(\boldsymbol{\phi})}{\partial \phi_i} = \left[\boldsymbol{0}_{MT,L_\phi},\ldots,\frac{\partial \boldsymbol{D}(\phi_i)}{\partial \phi_i},\ldots,\boldsymbol{0}_{MT,L_\phi}\right], \tag{66}$$

$\boldsymbol{0}_{MT,L_\phi}$ is a zero matrix with dimension of $MT \times L_\phi$, and $\boldsymbol{D}(\phi_i)$ is defined in (26).

References

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[2] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.
[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3950–3600, Nov. 2010.
[4] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[5] N. Shlezinger and Y. C. Eldar, "On the spectral efficiency of noncooperative uplink massive MIMO system," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1956–1971, Mar. 2019.
[6] S. Jin, X. Liang, K. K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.
[7] Q. Zhang, S. Jin, K. K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, May 2014.

[8] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
[9] A. Younis, N. Abuzgaia, R. Mesleh, and H. Haas, "Quadrature spatial modulation for 5G outdoor millimeter-wave communications: Capacity analysis," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2882–2890, May 2017.
[10] 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.
[11] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Aug. 2002.
[12] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 886–901, Sep. 2019.
[13] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
[14] M. K. Samimi and T. S. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, 2014, pp. 3483–3489.
[15] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, May 2014.
[16] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Shanghai, China, 2013, pp. 2890–2895.
[17] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feed-back for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
[18] 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. 2016.
[19] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.
[20] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications.* Cambridge, U.K.: Cambridge Univ. Press, 2012.
[21] B. Wang, F. Gao, J. Shi, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, May 2018.
[22] R. J. Mailloux, *Phased Array Antenna Handbook.* Norwood, MA, USA: Artech House, 2007.
[23] S. K. Garakoui, E. A. M. Klumperink, B. Nauta, and F. E. van Vliet, "Phased-array antenna beam squinting related to frequency dependency of delay circuits," in *Proc. Eur. Radar Conf.*, Manchester, U.K., 2011, pp. 416–419.
[24] W. A. Ahmad, J. H. Lu, D. Kissinger, and H. J. Ng, "Beam squinting in wideband 60 GHz on-board series-fed differential patch arrays," in *Proc. IEEE Asia Pacific Microw. Conf.*, Kuala Lumpar, Malaysia, 2017, pp. 13–16.
[25] H. Mirzaei and G. V. Eleftheriades, "Eliminating beam-squinting in wideband linear series-fed antenna arrays using feed networks constructed by slow-wave transmission lines," *IEEE Antennas Wireless Propaga. Lett.*, vol. 15, pp. 798–801, Sep. 2015.
[26] B. Ijaz *et al.*, "A series-fed microstrip patch array with interconnecting CRLH transmission lines for WLAN applications," in *Proc. Eur. Conf. Antennas Propaga.*, Gothenburg, Sweden, 2013, pp. 2088–2091.
[27] M. Cai, J. N. Laneman, and B. Hochwald, "Beamforming codebook compensation for beam squint with channel capacity constraint," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, 2017, pp. 76–80.
[28] J. Dai, A. Liu, and H. C. So, "Non-uniform burst-sparsity learning for massive MIMO channel estimation," *IEEE Trans Signal Process.*, vol. 67, no. 4, pp. 1075–1087, Feb. 2019.
[29] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans Signal Process*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
[30] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no 8, pp. 3010–3022, Jul. 2015.
[31] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.
[32] J. Dai, A. Liu, and V. K. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2584–2599, May, 2018.
[33] E. Bjornson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
[34] 3GPP, "Study on channel model for frequency spectrum above 6 GHz," 3GPP, Sophia Antipolis, France, Tech. Rep. TR 38.900 V15.0.0, Jun. 2018.
[35] M. Morelli, C. C. J. Kuo, and M. O. Pun, "Synchronization techniques for orthogonal frequency division multiple access (OFDMA): A tutorial review," *Proc. IEEE*, vol. 95, no. 7, pp. 1394–1427, Jun. 2007.
[36] 3GPP, "Physical channels and modulation," 3GPP, Sophia Antipolis, France, TR 36.211 v14.2.0, Mar. 2017.
[37] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems.* Cambridge, U.K.: Cambridge Univ. Press, 2015.
[38] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Oct. 2012.
[39] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.
[40] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Process*, R. Chellappa and S. Theodoridis, Eds., vol. 7, San Francisco, CA, USA: Academic, 2018, pp. 509–581.
[41] L. Hu, J. Zhou, Z. Shi, and Q. Fu, "A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5744–5754, Nov. 2013.
[42] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.
[43] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Cambridge, MA, USA: Birkhaüser, 2013.
[44] 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.
[45] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.
[46] G. Barriac and U. Madhow, "Space–Time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.
[47] Y. D. J. Bultitude and T. Rautiainen, "IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models," NOKIA, Tech. Rep., 30 Sep. 2007.
[48] METIS, "Mobile and wireless communications enablers for the twenty-twenty information society," D1.4 METIS Channel Models, 2015, [Online]. Available: https://www.metis2020.com/documents/deliverables/
[49] D. Vasisht, S. Kumar, H. Rahul, and D. Katabi, "Eliminating channel feedback in next-generation cellular networks," in *Proc. ACM SIGCOMM*, Aug. 2016, pp. 398–411.
[50] J. M. Goldberg and J. R. Fonollosa, "Downlink beamforming for spatially distributed sources in cellular mobile communications," *Signal Process.*, vol. 65, pp. 181–197, Mar. 1998.
[51] Y. C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

Mingjin Wang (S'17) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2017. He is currently working toward the Ph.D. degree with Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing, China. His current research interests include signal processing for communications and broadband wireless communications, with a focus on Massive MIMO systems, array signal processing, compressed sensing, and deep learning in communications.

Feifei Gao (M'09–SM'14) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D. degree from the National University of Singapore, Singapore, in 2007. He was a Research Fellow with Institute for Infocomm Research (I2R), A*STAR, Singapore, in 2008 and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. He has authored/coauthored more than 120 refereed IEEE journal papers and more than 120 IEEE conference proceeding papers. His research interests include communication theory, signal processing for communications, array signal processing, and convex optimizations, with particular interests in MIMO techniques, multicarrier communications, cooperative communication, and cognitive radio networks.

Prof. Gao was an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, *International Journal on Antennas and Propagations*, and *China Communications*. He was the Symposium Co-Chair for the 2019 IEEE Conference on Communications, the 2018 IEEE Vehicular Technology Conference Spring (VTC), the 2015 ICC, the 2014 IEEE Global Communications Conference, the 2014 IEEE VTC Fall,

Nir Shlezinger (M'17) received the B.Sc., M.Sc., and Ph.D. degrees, in 2011, 2013, and 2017, respectively, from Ben-Gurion University, Beersheba, Israel, all in electrical and computer engineering. From 2017 to 2019, he was a Postdoctoral Researcher with the Technion, Israel Institute of Technology, and is currently a Postdoctoral Researcher with the Signal Acquisition Modeling, Processing, and Learning Lab, Weizmann Institute of Science. From 2009 to 2013, he worked as a Research and Development Engineer with Yitran Communications. His research interests include communications, information theory, signal processing, and machine learning.

Mark F. Flanagan (M'03–SM'10) received the B.E. and Ph.D. degrees in electronic engineering from University College Dublin, Dublin, Ireland, in 1998 and 2005, respectively.

From 1998 to 1999, he was a Project Engineer with Parthus Technologies Ltd. From 2006 to 2008, he was a Postdoctoral Research Fellow with the University of Zurich, Switzerland, the University of Bologna, Italy, and The University of Edinburgh, U.K. In 2008, he was appointed as SFI Stokes Lecturer in electronic engineering with University College Dublin, where he is currently an Associate Professor. In 2014, he was a Visiting Senior Scientist with the Institute of Communications and Navigation, German Aerospace Center, under a DLR-DAAD Fellowship. His research interests include information theory, wireless communications, and signal processing.

Dr. Flanagan is a member of the IEEE Communications Society and IEEE Information Theory Society. He regularly serves for the Technical Program Committees of several major IEEE International Conferences, and is currently serving as TPC Co-Chair for the Communication Theory Symposium of IEEE Conference on Communications 2020. He is an Executive Editor for IEEE COMMUNICATIONS LETTERS.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, 1995, 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002.

She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was previously a Professor with the Department of Electrical Engineering, the Technion, where she held the Edwards Chair in Engineering. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University, and was a Visiting Professor with Stanford. She is a member of the Israel Academy of Sciences and Humanities (elected 2017), an IEEE Fellow and a EURASIP Fellow. She is also an author of the book "Sampling Theory: Beyond Bandlimited Systems" and coauthor of four other books published by Cambridge University Press. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging and optics.

Dr. Eldar has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology program with the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair with the Technion, the Murieland David Jacknow Award for Excellence in Teaching, and the Technions Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues including the SIAM outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, was selected as one of the 50 most influential women in Israel and in Asia, and is a highly cited researcher. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor-in-Chief of Foundations and Trends in Signal Processing, a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the EURASIP Journal of Signal Processing, the SIAM Journal on Matrix Analysis and Applications, and the SIAM Journal on Imaging Sciences. She was Co-Chair and Technical Co-Chair of several international conferences and workshops.