

Estimating Doubly-Selective Channels for Hybrid mmWave Massive MIMO Systems: A Doubly-Sparse Approach

Shijian Gao^{ID}, *Student Member, IEEE*, Xiang Cheng^{ID}, *Senior Member, IEEE*, and Liuqing Yang^{ID}, *Fellow, IEEE*

Abstract—In mmWave massive multiple-input multiple-output (mMIMO) systems, hybrid (digital/analog) structure has been a prevalent option to balance system cost and performance. To facilitate transceiver design in hybrid mmWave mMIMO, acquiring an accurate channel state information is critical. To this end, a novel doubly-sparse approach is proposed to estimate doubly-selective mmWave channels under hybrid mMIMO. Via the judiciously designed training pattern, the well-utilized beamspace sparsity alongside the under-investigated delay-domain sparsity that mmWave channels exhibit can be jointly exploited to assist channel estimation. Thanks to our careful two-stage (random-probing and steering-probing) design, the proposed channel estimator possesses strong robustness against the double (frequency and time) selectivity whilst enjoying the benefits brought by the exploitation of double sparsity. Compared with existing alternatives, our proposed mmWave channel estimator not only works in doubly-selective channels, but also largely reduces the training overhead, storage demand as well as computational complexity.

Index Terms—mmWave, hybrid massive multiple-input multiple-output, channel estimation, double selectivity, double sparsity.

I. INTRODUCTION

THANKS to the abundant frequency resources at millimeter-wave (mmWave) band, mmWave communications have been recognized as one of key technologies for the 5G & beyond wireless systems [2]–[5]. However, a major concern impeding the wide deployment of mmWave systems comes from the severe propagation loss [6]–[9]. Fortunately, as a much shorter wavelength at mmWave band allows the

deployment of massive antennas at the transceiver, the propagating attenuation can therefore be compensated for by the large array gains [10]–[12].

To facilitate the transceiver design, the top priority is to acquire an accurate channel state information (CSI) [13]. However, compared to the centimeter-wave (cmWave) MIMO systems, channel estimation for mmWave mMIMO faces unprecedented challenges. First, mmWave mMIMO typically adopts a hybrid structure for power and cost concerns [7], [14], so the high-dimensional channel matrix has to be recovered via very few RF chains. Since the latter essentially determines the number of effective training symbols that can be transmitted simultaneously, it can take significant amount of time to transmit sufficient training symbols for mMIMO. When it comes to the mobile scenarios, the problem becomes even more challenging, because the channel turns out to be time-varying in the presence of Doppler.

As mmWave channels exhibit limited scattering, a unique sparsity holds in beamspace under mMIMO. Thanks to this sparsity, it may not be necessary to estimate the channel matrix element by element. Instead, one can resort to the compressed sensing (CS) theory to reduce the training overhead while ensuring a high accuracy. Following this idea, in [15], a hierarchical beam training coupled with orthogonal matching pursuit (OMP) [16] is devised to estimate static narrowband mmWave channels. In [17], block-OMP (BOMP) is applied to estimate narrowband & time-varying mmWave channels. The static wideband channel estimation in the line-of-sight (LoS) scenarios is considered in [18], and the relevant work has been further extended to the non-LoS (NLoS) scenarios like [19] and [20], where OMP is applied either in the time-domain or the frequency-domain to assist channel estimation.

Due to the wideband nature of mmWave, the narrowband channel model suffers from severe limitations, motivating us to focus on the wideband channel model. Generally speaking, existing wideband channel estimation works can be divided into two main categories: time-domain estimation vs. frequency-domain. The former is to estimate all channel taps jointly, while the latter is to estimate individual subcarriers independently. By exploiting the sparsity in beamspace, both schemes achieve similar performance in the sense of the normalized mean square error (NMSE), with a largely reduced training overhead compared to the least-squares (LS) estimator. However, to effectively exploit the sparsity so

Manuscript received October 2, 2019; revised March 11, 2020; accepted May 10, 2020. Date of publication May 27, 2020; date of current version September 10, 2020. This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Project 2019B010153003, and in part by the National Science Foundation under Grant CPS-1932413 and Grant ECCS-1935915. This article was presented at the IEEE International Conference on Communications, Shanghai, China, May 20–24, 2019. The associate editor coordinating the review of this article and approving it for publication was X. Yuan. (*Corresponding author: Xiang Cheng.*)

Shijian Gao and Liuqing Yang are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523-1373 USA (e-mail: sjgao@colostate.edu; lqyang@engr.colostate.edu).

Xiang Cheng is with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, School of Electronics Engineering and Computing Sciences, Peking University, Beijing 100080, China (e-mail: xiangcheng@pku.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2020.2995699

1536-1276 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

that OMP could be applied, either the demanding storage requirement or the heavy computational burden is inevitable. On top of that, these works have not taken into account of the Doppler effects, rendering their feasibility in mobility scenarios questionable. To address these issues, there is an urgent need for a more generalized and more efficient channel estimation approach.

To achieve significant reduction in training overhead and computational complexity, we resort to the under-exploited delay-domain sparsity in combination with the well-known beamspace sparsity. Aiming at a high-performance and easy-to-implement channel estimator, a novel channel estimator is proposed by exploiting the double sparsity. As a matter of fact, the idea of using either the delay-domain sparsity or the double sparsity can be also found in some works, such as [21], [22], and [23]. However, these works are not specifically designed for hybrid mMIMO, and their studied channels have not taken time selectivity into account. In fact, once the time selectivity is involved, how to exploit either the delay-domain sparsity or the beamspace sparsity becomes a thorny problem. Furthermore, the introduction of the hybrid structure makes channel estimation a totally different topic as before, because the design flexibility is severely restricted by the hardware constraints.

To address the aforementioned deficiencies, we propose a novel doubly-sparse approach to estimate mmWave mMIMO channels. Specifically, the so-called DSDS channel estimator comprises the following steps:

- To deal with the sparsity in delay domain, a special training pattern is judiciously designed to successfully separate each channel tap. Based on the energy detector, only a small proportion of channel taps will be identified effective regardless of Doppler effects and awaits a further processing.
- To deal with the beamspace sparsity, an enhanced OMP algorithm termed as A-BOMP is proposed to recover the beam direction. Given the maximum Doppler, A-BOMP can adjust basis matching & residue update with properly determined iterations, such that a high accuracy can be guaranteed even under strong Doppler effects.
- To jointly estimate the amplitudes and Doppler, repetitive steering-probing is applied based on the estimated beam direction. As a result, both the amplitudes and Doppler can be reliably estimated with low training overhead.

Compared with existing work, the doubly-sparse approach can remarkably improve the estimation accuracy, and largely reduce the training overhead, storage demand as well as computational complexity. As many implementing issues are also considered in the specific design, the proposed channel estimator has a great potential to be applied in practice.

The rest of this paper is organized as follows: Section II describes the system and channel models. Sections III and IV introduce how to exploit the beamspace and delay-domain sparsity, respectively. Section IV explains the estimation of amplitudes and Doppler. Extensive numerical results and discussions are presented in Section V, followed by conclusions in Section VI.

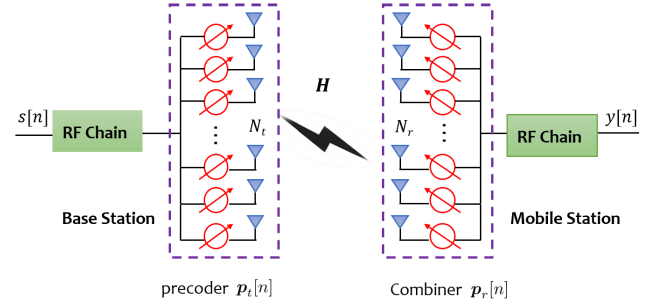


Fig. 1. The system model of hybrid mmWave mMIMO transceivers with one RF chain deployed at either end.

Notation: In the remainder of the paper, a , $\mathbf{a}$ and $\mathbf{A}$ represent a scalar, a vector and a matrix, respectively. $|\mathcal{A}|$ is the cardinality of the support $\mathcal{A}$. $\mathbf{A}[m, n]$, $\mathbf{A}[m, :]$ and $\mathbf{A}[:, m]$ are denoted as the (m, n) entry, the m -th row, and the m -th column of $\mathbf{A}$, respectively. $\mathbf{A}'$, $\mathbf{A}^*$ and $\mathbf{A}^\dagger$ denote the transpose, the hermitian transpose and the pseudo-inverse of $\mathbf{A}$, respectively. $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ represent the floor and ceiling operation, respectively. $\text{cal}(\cdot)$ represents the cardinality. $\text{diag}(\cdot)$ and $\text{vec}(\cdot)$ represent the operations of diagonalization and vectorization, respectively. $\mathbb{E}$ stands for expectation. $\mathcal{CN}(0, \sigma^2)$ represents the distribution of a circularly symmetric complex Gaussian random variable with variance σ^2 .

II. SYSTEM AND CHANNEL DESCRIPTION

A. System Model

A mmWave mMIMO system is considered, where N_t and N_r antennas are employed at the transmitter (Tx) and receiver (Rx), respectively. Since the proposed channel estimation approach does not rely on the channel reciprocity, we simply assume channel estimation is implemented at Rx. For the power consumption and hardware cost concerns, mmWave mMIMO typically adopts a hybrid structure, in which the number of RF chains at the transceivers is much smaller than that of the antennas. Similar to [7], [24], a fully-connected hybrid structure is studied here, where the RF chains and antennas are connected via a digitally controlled analog phase shifter (APS) network. Suppose each APS component has a resolution of b bits, then all adjustable angles are contained in

$$\mathcal{B} = \{0, 2\pi/2^b, \dots, 2\pi(2^b - 1)/2^b\} \quad (1)$$

with $|\mathcal{B}| = 2^b$. Accordingly, the angular quantization function is expressed as

$$\mathcal{Q}(x) = \mathcal{B}(i^*), i^* = \arg \min_i \text{mod}(x - \mathcal{B}(i), 2\pi). \quad (2)$$

As in [25], let the transceivers each employ a single RF chain as shown in Fig. 1. Note that, since we focus on channel estimation in this paper, this setup is without loss of generality and can be readily generalized to cope with arbitrary number of RF chains at the transceivers.

B. Geometric Channel

In this paper, we adopt the modified Sen-Matolak channel model [26], which is an extension of the narrowband geometric

model by taking the path delay and the Doppler effect into account. Denote the maximum number of delay taps as N_c . At time instant n , the sampled version of the tap- d channel ($0 \leq d < N_c$) is given by

$$\mathbf{H}_d(n) = \sum_{p=1}^P \sqrt{\frac{N_t N_r}{P}} \alpha_p h(dT_s - \tau_p) \mathbf{a}_r(\theta_p) \mathbf{a}_t^*(\phi_p) e^{j\omega_p n} \quad (3)$$

where $\alpha_p \sim \mathcal{CN}(0, 1)$ is the complex gain of the p -th path; $h(\cdot)$ is the pulse shaping filter response; τ_p is the propagation delay of the p -th path that obeys a uniform distribution on $[0, (N_c - 1)T_s]$; θ_p and ϕ_p represent the angle of arrival (AoA) and angle of departure (AoD), respectively, both of which being modeled as uniformly distributed variables on $[0, 2\pi)$. Define the system carrier frequency to be f_c , the velocity of light to be c_v , and the maximum relative velocity to be v_m . Then the normalized Doppler shift is $\omega_p = 2\pi f_c v_m T_s \sin(\theta_p)/c_v$. For notational simplicity, an array response generating function is defined as

$$\mathbf{f}_N(y) = \frac{1}{\sqrt{N}} [1, e^{j2\pi y}, \dots, e^{j2\pi(N-1)y}]'. \quad (4)$$

With half-wavelength uniform linear arrays (ULAs) employed at the transceivers, we have $\mathbf{a}_t(\phi) = \mathbf{f}_{N_t}(\sin(\phi)/2)$ and $\mathbf{a}_r(\theta) = \mathbf{f}_{N_r}(\sin(\theta)/2)$.

C. Beamspace Representation

To simplify expression, one can rewrite the geometric model into the following compact form [17]

$$\mathbf{H}_d(n) = \mathbf{A}_R \text{diag}(\mathbf{g}_d(n)) \mathbf{A}_T^* \quad (5)$$

where $\mathbf{A}_T = [\mathbf{a}_t(\phi_1), \mathbf{a}_t(\phi_2), \dots, \mathbf{a}_t(\phi_P)] \in \mathcal{C}^{N_t \times P}$ and $\mathbf{A}_R = [\mathbf{a}_r(\theta_1), \mathbf{a}_r(\theta_2), \dots, \mathbf{a}_r(\theta_P)] \in \mathcal{C}^{N_r \times P}$ are steering matrices that remain unchanged during the channel estimation stage. The time-varying effects are incorporated in $\mathbf{g}_d(n)$ given by

$$\sqrt{\frac{N_t N_r}{P}} [\alpha_1 h(dT_s - \tau_1) e^{j\omega_1 n}, \dots, \alpha_P h(dT_s - \tau_P) e^{j\omega_P n}]' \quad (6)$$

which contains the path gains at time instant n .

In Eq. (5), $\mathbf{A}_T$, $\mathbf{A}_R$ as well as $\mathbf{g}_d(n)$ are all associated with the physical channel taps, which are not always resolvable due to the finite resolution of the receiver in time and space, and thus cannot be directly estimated. To seek an equally general but more practical representation, we first construct the Tx-end and Rx-end angular dictionary matrices as in [15]

$$\mathbf{D}_t = [\mathbf{f}_{N_t}(0), \mathbf{f}_{N_t}(1/G_t), \dots, \mathbf{f}_{N_t}((G_t - 1)/G_t)] \quad (7a)$$

$$\mathbf{D}_r = [\mathbf{f}_{N_r}(0), \mathbf{f}_{N_r}(1/G_r), \dots, \mathbf{f}_{N_r}((G_r - 1)/G_r)] \quad (7b)$$

where G_t and G_r represent the size of corresponding dictionaries. Taking $\mathbf{D}_t$ as an example, it contains the steering vectors ranging from $[0, 2\pi]$ with resolution $2\pi/G_t$. As G_t approaches infinity, the resolution becomes zero, thus leading to a continuous dictionary. For practical implementation, most work show that setting G_t as 2 ~ 4 times the array size can

provide sufficient resolution for separating the AoAs/AoDs of the propagation paths. Based on the dictionary matrices, the channel representation in Eq. (5) can be re-expressed as

$$\mathbf{H}_d(n) = \mathbf{A}_R \text{diag}(\mathbf{g}_d(n)) \mathbf{A}_T^* = \mathbf{D}_r \overline{\mathbf{H}}_d(n) \mathbf{D}_t^*. \quad (8)$$

Under the mMIMO setup, P propagation paths result in P dominant non-zero entries in $\overline{\mathbf{H}}_d$. As $\mathbf{D}_r$ and $\mathbf{D}_t$ are irrelevant to $\mathbf{H}_d$, $\overline{\mathbf{H}}_d$ essentially gathers the entire channel information that was originally contained by $\mathbf{A}_T$, $\mathbf{A}_R$ and $\mathbf{g}_d(n)$. Specifically, by omitting the time instant and assuming on-grid AoA/AoD pairs, $\forall p \in [1, P]$, $n_p = \frac{\phi_p}{2\pi/G_t}$, $m_p = \frac{\theta_p}{2\pi/G_r}$, $\overline{\mathbf{H}}_d(m_p, n_p) = \mathbf{g}_d[p]$. From this sense, $\overline{\mathbf{H}}_d$ can be interpreted as the channel representation in beamspace. Because a limited scattering effect in mmWave propagation leads to $P \ll G_r G_t$ [24], $\overline{\mathbf{H}}_d$ exhibits an evident sparse nature.¹

Revisiting Eq. (3), the prior information available at both ends are N_t , N_r , N_c , the steering pattern of $\mathbf{a}_{t/r}$, while the remaining parameters are unknown to the transceivers, and thus have to be recovered via channel estimation. In the following, we will heavily rely on the beamspace representation to recover the beam direction (AoA & AoD), the beam amplitude, as well as the associated Doppler shift.

D. Input-Output Relationship

Let $s(n)$ be the training symbol at instant- n . At the Tx, $s(n)$ is first processed at the APS network, and the transmitted signal is $\mathbf{x}(n) = \mathbf{p}_t(n)s(n) \in \mathcal{C}^{N_t \times 1}$. Since each APS component can only adjust the phase, the probing vector $\mathbf{p}_t(n)$ bears the form as

$$\mathbf{p}_t(n) = \sqrt{1/N_t} [e^{j\alpha_1(n)}, e^{j\alpha_2(n)}, \dots, e^{j\alpha_{N_t}(n)}]' \quad (9)$$

with $\alpha_i(n) \in \mathcal{B}$, $\forall i \in [1, N_t]$.

After channel propagation, the received signal is

$$\mathbf{r}(n) = \sum_{d=0}^{N_c-1} \mathbf{H}_d(n) \mathbf{x}(n-d) + \boldsymbol{\eta}(n) \quad (10)$$

which is the convolution of multiple time-varying channel taps. $\boldsymbol{\eta}(n) \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$ is the receiver noise vector. $\mathbf{r}(n)$ then goes through the Rx-end APS network, whose function is described by an $N_r \times 1$ probing vector $\mathbf{p}_r(n)$, so the received sample after APS becomes

$$y(n) = \sum_{d=0}^{N_c-1} \mathbf{p}_r^*(n) \mathbf{H}_d(n) \mathbf{p}_t(n-d) s(n-d) + \xi(n). \quad (11)$$

where $\xi(n) = \mathbf{p}_r^*(n) \boldsymbol{\eta}(n) \sim \mathcal{CN}(0, \sigma^2)$ remains white. Let $\overline{\mathbf{p}}_t(n) = \mathbf{D}_t \mathbf{p}_t(n)$ and $\overline{\mathbf{p}}_r(n) = \mathbf{D}_r \mathbf{p}_r(n)$. Based on the beamspace representation in Eq. (8), we have

$$y(n) = \sum_{d=0}^{N_c-1} \overline{\mathbf{p}}_r^*(n) \overline{\mathbf{H}}_d(n) \overline{\mathbf{p}}_t(n-d) s(n-d) + \xi(n). \quad (12)$$

¹In practice, the off-grid leakage may lead to extra non-zero entries in $\overline{\mathbf{H}}_d$. Since the leakage is typically very weak under mMIMO, the ensemble of dominant entries in $\overline{\mathbf{H}}_d$ is still similar to $\mathbf{g}_d$. Regardless of whether the non-zero entries of $\overline{\mathbf{H}}_d$ corresponds to a single channel path p or to some leakage terms, these entries are the resolvable ones that can be estimated.

Without loss of generality, we consider the general I-O relationship for the first frame only unless otherwise specified. The length- N_f training frame is simply denoted as $[s(0), s(1), \dots, s(N_f - 1)]$, and its specific form will be explained later. By concatenating all received samples, we write the I-O relationship in matrix form shown in Eq. (13) at the bottom of this page, with

$$\bar{\mathbf{P}}_r^* = \begin{bmatrix} \bar{\mathbf{p}}_r^*(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{p}}_r^*(1) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{\mathbf{p}}_r^*(N_f - 1) \end{bmatrix}$$

and

$$\bar{\mathbf{P}}_t = \begin{bmatrix} \bar{\mathbf{p}}_t(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{p}}_t(1) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{\mathbf{p}}_t(N_f - 1) \end{bmatrix}.$$

III. EXPLOIT DELAY-DOMAIN SPARSITY

As described in Section II, mmWave channels exhibit sparsity in beamspace. Apart from this well-known sparsity, this section will further show that mmWave channels exhibit sparsity in the delay domain as well. We first analyze why existing approaches fail to exploit the delay-domain sparsity, and then explain how one can effectively benefit this largely overlooked sparsity.

A. Sparsity in Delay Domain

To eliminate inter-frame interferences (IFIs) in block transmission, a commonly adopted approach amounts to zero-padding (ZP) a guard interval with length at least $(N_c - 1)$ to each frame. For example, the data-frame length is 512 in IEEE 802.11ad, while the prefix length can be up to 128. However, a long delay spread with large N_c due to the high symbol rate does not mean a rich multi-path environments. In fact, mmWave channels have very few dominant paths². Hence, a majority of the channel taps are actually too weak to be considered, rendering sparsity in the delay domain. To gain some intuitive insight, we plot the colormap of a randomly generated channel in Fig. 2, where the double sparsity in both the beamspace and delay domain can be clearly observed.

²typically 8 ~ 12 even in “rich” scattering environments, and is much less in other environments [28].

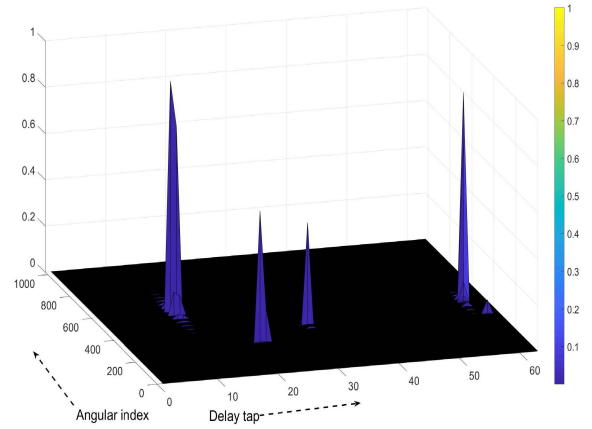


Fig. 2. The delay-beamspace colormap of a randomly generated mmWave channel with $N_t = N_r = 32$, $N_c = 64$ and $P = 5$.

B. Conventional Training Pattern

The currently adopted training pattern [20] is given by

$$[s(0), s(1), \dots, s(N_f - 1)] = [s_0, s_1, s_2, \dots, s_{N-1}, \underbrace{0, \dots, 0}_{N_c}]. \quad (14)$$

Specifically, each frame contains $N_f = N + N_c$ symbols, where N and N_c are the length of the data sequence and ZP, respectively. Clearly, the I-O relationship of this pattern still follows the general one in Eq. (13), but some specifics need to be clarified.

In wideband mmWave systems, symbols are pumped out at a very high rate, thus leaving insufficient buffer time for the APS network reconfiguration except for the ZP interval [19]. As a result, the probing vectors remain unchanged over the entire frame, that is

$$\mathbf{p}_{t/r}(n) = \mathbf{p}_{t/r}(0), \quad \forall n \in [0, N_f]. \quad (15)$$

Accordingly, $\bar{\mathbf{P}}_r^* = \mathbf{I}_{N_f} \otimes \bar{\mathbf{p}}_r^*(0)$ and $\bar{\mathbf{P}}_t = \mathbf{I}_{N_f} \otimes \bar{\mathbf{p}}_t(0)$.

Although the introduction of ZP ensures IFI-free, N_c channel taps remain unresolvable after convoluting with the training sequence. In consequence, channel estimation requires joint processing across all taps, leading to high storage demand and heavy computational burden. More importantly, exploiting the delay-domain sparsity becomes an intractable task.

$$\begin{aligned} \mathbf{y} &= [y(0), y(1), \dots, y(N_f - 1)]' \\ &= \bar{\mathbf{P}}_r^* \begin{bmatrix} \bar{\mathbf{H}}_0(0) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \bar{\mathbf{H}}_0(1) & \mathbf{0} & \cdots & \mathbf{0} \\ \bar{\mathbf{H}}_{N_c-1}(N_c - 1) & \cdots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \cdots & \ddots & \mathbf{0} \\ \mathbf{0} & \vdots & \bar{\mathbf{H}}_{N_c-1}(N_f - 1) & \cdots & \bar{\mathbf{H}}_0(N_f - 1) \end{bmatrix} \begin{bmatrix} s(0) \\ \vdots \\ s(N_c - 1) \\ \vdots \\ s(N_f - 1) \end{bmatrix} + \boldsymbol{\xi}. \end{aligned} \quad (13)$$

C. Proposed Training Pattern

To avoid these limitations, there is a great urgency to devise a new training pattern, by which the delay-domain sparsity could be exploited to facilitate channel estimation, and the pattern itself must be friendly to implementation. To this end, a new training pattern is designed as follows

$$\begin{aligned} & [s(0), s(1), \dots, s(N_f - 1)] \\ &= \left[\underbrace{s_0, 0, \dots, 0}_{(1)} \left| \underbrace{s_1, 0, \dots, 0}_{(2)} \right| \dots \left| \underbrace{s_{L-1}, 0, \dots, 0}_{(L)} \right] \end{aligned} \quad (16)$$

As can be seen, each frame is further divided into $L = N_f/N_c$ subframes.³ Owing to the ZP in each subframe, sufficient buffer time is left to reconfigure the APS network after each non-zero training symbol. In other words, the probing vectors can be updated L times per frame, i.e.,

$$\mathbf{p}_{t/r}(n) = \mathbf{p}_{t/r}(N_c \lfloor n/N_c \rfloor), \forall n \in [0, N_f). \quad (17)$$

An interesting fact is that, when it comes to the frequency-selective channel estimation in conventional MIMO setup, the training pattern in Eq. (16) has been proved optimal in the sense of both the mean squared error (MSE) and system mutual information [27]. Before taking a closer look at the I-O relationship with the new pattern, we first make the following definition.

Random-probing vector: At the random-probing stage, the probing vectors are generated by randomly adjusting the angle of each APS component from $\mathcal{B}$. The resultant vector is termed as the random-probing vector and denoted as

$$\mathbf{p}_{t/r}^R(l) = \mathbf{p}_{t/r}(lN_c + n_c), (l < L, n_c < N_c). \quad (18)$$

Applying random probing is simply because no prior CSI is available at this stage. Note that, the above definition implies that $\mathbf{p}_{t/r}^R$ possesses both the randomness and subframe-updatability property. Applying a similar notational change to $\bar{\mathbf{p}}_{t/r}$, $\bar{\mathbf{P}}_r^*$ becomes Eq. (19), shown at the bottom of this page, and $\bar{\mathbf{P}}_t$ is obtained likewise. Substituting the new $\bar{\mathbf{P}}_r^*$ and $\bar{\mathbf{P}}_t$, together with the training frame into Eq. (13), the received signal becomes

$$\begin{aligned} y(lN_c + n_c) &= (\bar{\mathbf{p}}_r^R(l))^* \bar{\mathbf{H}}_{n_c}(lN_c + n_c) \bar{\mathbf{p}}_t^R(l) s_l \\ &\quad + \xi(lN_c + n_c). \end{aligned} \quad (20)$$

Clearly, the received samples are now associated with a single channel tap. Hence, our proposed pattern facilitates separating

³Without loss of generality, L is assumed to be an integer here. If N_f/N_c is not an integer, one can simply use $L = \lfloor N_f/N_c \rfloor$.

channel taps and thus rendering it possible to exploit the delay-domain sparsity easily. Since the success of tap separation does not rely on the non-zero training symbol s_l in Eq. (16), s_l is set as 1 in the rest of paper without loss of generality.

D. Identification of Effective Taps

To determine the existence of tap- d channel, we gather all $\bar{\mathbf{H}}_d$ -related samples, i.e., $y(lN_c + d)$, $\forall l \in [0, L - 1]$. If at least one dominant path exists in the tap- d channel, $y(lN_c + d)$ includes both the signal and noise parts. Otherwise, $y(lN_c + d)$ contains noise only. Hence, detecting the existence of the tap- d channel is a binary hypothesis testing problem that can be dealt with via energy detector. We first average the power of all samples associated with $\bar{\mathbf{H}}_d$, and get the test statistics (TS) and its normalized version nTS as

$$Y_d = \frac{1}{L} \sum_{l=0}^{L-1} |y(lN_c + d)|^2 \quad (21a)$$

$$\bar{Y}_d = \frac{Y_d - \sigma^2}{\max_{0 \leq m < N_c} (Y_m - \sigma^2, 0)}. \quad (21b)$$

When applying CS, random probing is necessary in estimating both the time-invariant and time-varying channels. While for the latter, another important function of random probing is to remain robust against Doppler.

Proposition 1 [Validity of test statistics with Doppler]: With sufficient random probings, the test statistics Y_d is approximately irrelevant to the channel's time variation.

Proof: Let $n = lN_c + d$ and $g_{d,p}(n)$ be the p -th element of $\mathbf{g}_d(n)$, then

$$\begin{aligned} y(n) &= \sum_{d=0}^{N_c-1} \bar{\mathbf{p}}_r^*(n) \bar{\mathbf{H}}_d(n) \bar{\mathbf{p}}_t(n-d) + \xi(n) \\ &= \sum_{p=1}^P (\mathbf{p}_r^R(l))^* \mathbf{a}_r(\theta_p) \mathbf{a}_t^*(\phi_p) \mathbf{p}_t^R(l) g_{d,p}(n) + \xi(n). \end{aligned} \quad (22)$$

Denote $\rho_p(l) = (\mathbf{p}_r^R(l))^* \mathbf{a}_r(\theta_p) \mathbf{a}_t^H(\phi_p) \mathbf{p}_t^R(l)$. By using $g_{d,p}(n) = g_{d,p}(0) e^{j\omega_p n}$, we have

$$\begin{aligned} & |y(n)|^2 \\ &= \sum_{p=1}^P |\rho_p(n) g_{d,p}(0)|^2 + 2\mathcal{R} \left\{ \sum_{p=1}^P \rho_p(l) g_{d,p}(0) \xi(n) e^{j\omega_{p1} n} \right\} \\ &\quad + 2\mathcal{R} \left\{ \sum_{p_1} \sum_{p_2} \rho_{p_1}(l) g_{d,p_1}(0) \rho_{p_2}^*(l) g_{d,p_2}^*(0) e^{j(\omega_{p_1} - \omega_{p_2}) n} \right\}. \end{aligned} \quad (23)$$

Since $\mathbf{p}_t^R$ and $\mathbf{p}_r^R$ are random probing vectors with zero mean, it can be readily verified $\mathbb{E}\{\rho_p(l)\} = 0$, $\forall p \in [1, P]$.

$$\bar{\mathbf{P}}_r^* = \begin{bmatrix} \mathbf{I}_{N_c} \otimes (\bar{\mathbf{p}}_r^R(0))^* & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_c} \otimes (\bar{\mathbf{p}}_r^R(1))^* & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{N_c} \otimes (\bar{\mathbf{p}}_r^R(L-1))^* \end{bmatrix} \quad (19)$$

By averaging sufficient $|y(n)|^2$ terms, the last two terms in Eq. (23) approach zero, thus the TS becomes irrelevant to ω_p . ■

Proposition 1 guarantees the exploitation of the delay-domain sparsity regardless of Doppler effects. With the energy detector, the effective taps can be roughly selected as

$$\mathcal{P}_1 = \left\{ d \mid \bar{Y}_d \geq \mu \right\} \cap \left\{ d \mid Y_d > \sigma^2 \right\} \quad (24)$$

where μ is the threshold.⁴ To avoid extreme cases where $\text{cal}(\mathcal{P}_1)$ is either 0 or unreasonably large, a tuning procedure is added, and the ultimately determined taps are given by

$$\mathcal{P} = \begin{cases} \mathcal{P}_1, & 0 < \text{cal}(\mathcal{P}_1) \leq A \\ \{d \mid \bar{Y}_d \geq \bar{\lambda}_A\}, & \text{cal}(\mathcal{P}_1) > A \\ \{d \mid Y_d \geq \lambda_A\}, & \text{cal}(\mathcal{P}_1) = 0. \end{cases} \quad (25)$$

with λ_A and $\bar{\lambda}_A$ representing the A -th largest TS and nTS, respectively.

Up till now, we have accomplished the first part of the random-probing stage. Summarizing, the main steps can be described as follows:

- Transmit judiciously designed training frames with random APS probing.
- Calculate the TS/nTS for each channel tap based on the corresponding received samples.
- Determine the non-negligible channel taps based on the energy detector $\mathcal{P}$.

IV. EXPLOITING THE BEAMSPACE SPARSITY

As outlined in Section II, the beamspace channel exhibits sparsity under mMIMO settings. Therefore, instead of estimating the original geometric channel matrix $\mathbf{H}_d$ with dimension $N_t N_r$, we estimate the sparse beamspace channel $\bar{\mathbf{H}}_d$. Since the time variation imposes a great difficulty in recovering the exact values of non-zero entries from $\bar{\mathbf{H}}_d$, we focus on locating the non-zero entries (essentially the angle support) first in this section, and leaving the estimation of exact values to the next section.

A. Sparse Transformation

After tap detection, n_c out of N_c taps are recognized effective, with their indices collected by $\mathcal{D} = \{d_1, d_2, \dots, d_D\}$. Using the samples already obtained at the random-probing stage, we proceed to determine the angle support for those taps belonging to $\mathcal{D}$. It has to be stressed that this step does not require extra training frames.

Due to the similarity, we take tap- d_i ($d_i \in \mathcal{D}$) for example, and the subscript of d_i is omitted for brevity. To apply CS, let us first derive the sparse representation for received samples. Stacking all $\bar{\mathbf{H}}_d$ -related samples from $\mathbf{y}$ yields

$$\mathbf{y}_d = [y(d), y(N_c + d), \dots, y((L-1)N_c + d)]'. \quad (26)$$

⁴Evidently, the energy detector is somewhat heuristic. Recall that the energy detector actually plays the role of a binary classifier, a promising direction is to seek the power of deep neural networks. Specifically, given the channel model, a bunch of synthesized data can be generated to train the network for classification (tap detection) in a supervised manner. The offline trained network could then be used for online prediction.

Denoting $n_l = lN_c + d$ ($\forall l \in [0, L)$) and using matrix equality $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$, $y(n_l)$ can be rewritten as

$$y(n_l) = \underbrace{\left((\bar{\mathbf{p}}_t^R(l))' \otimes (\bar{\mathbf{p}}_r^R(l))^* \right)}_{\boldsymbol{\psi}(l)} \underbrace{\left(\text{vec}(\bar{\mathbf{H}}_d(n_l)) \right)}_{\bar{\mathbf{h}}_d(n_l)} + \xi(n_l). \quad (27)$$

Neglecting the noises temporarily for brevity, $\mathbf{y}_d$ can be further expressed as

$$\mathbf{y}_d = \underbrace{\begin{bmatrix} \psi(0) & 0 & \cdots & 0 \\ 0 & \psi(1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi(L-1) \end{bmatrix}}_{\boldsymbol{\Psi}} \underbrace{\begin{bmatrix} \bar{\mathbf{h}}_d(n_0) \\ \bar{\mathbf{h}}_d(n_1) \\ \vdots \\ \bar{\mathbf{h}}_d(n_{L-1}) \end{bmatrix}}_{\bar{\mathbf{h}}_d}. \quad (28)$$

In the special case of time-invariant channels, all $\bar{\mathbf{h}}_d(n_l)$'s are exactly the same [17], thus giving rise to

$$\mathbf{y}_d = \begin{bmatrix} \psi(0) \\ \psi(1) \\ \vdots \\ \psi(L-1) \end{bmatrix} \bar{\mathbf{h}}_d(n_0). \quad (29)$$

Determining the angle support of $\bar{\mathbf{H}}_d$ is equivalent to locating non-zero entries from the $G_t G_r$ -dimensional vector $\bar{\mathbf{h}}_d$. Since $P \ll G_t G_r$, the “localization” can be effectively solved via OMP, through which $\mathcal{O}(P \log G_t G_r)$ instead of $\mathcal{O}(G_t G_r)$ samples suffice to guarantee a high accuracy.

However, Eq. (29) is no longer valid in the presence of Doppler shifts, motivating us to restudy the more general Eq. (28). Because $\bar{\mathbf{h}}_d$ remains sparse for $LP \ll LG_t G_r$, a natural option would be OMP as well. Reminisce that the variations of AoAs/AoDs are negligible during the channel estimation, thus a common angle support is shared by all $\bar{\mathbf{h}}_d(n_l)$'s. However, OMP cannot exploit such a unique structure because it treats $\mathbf{h}_d(n_l)$ as a generic sparse vector. Fortunately, by utilizing the unique property of $\bar{\mathbf{h}}_d(n_l)$, a more general block-sparse representation can be derived. Specifically, constructing such a permutation matrix $\mathbf{P}$ satisfying $\mathbf{P}[:, (i-1)G_t G_r + j] = \mathbf{I}_{G_t G_r}[:, (j-1)G_t G_r + i]$ [29], $\mathbf{y}_d$ can be decomposed as

$$\mathbf{y}_d = (\boldsymbol{\Psi} \mathbf{P}) \cdot (\mathbf{P}' \bar{\mathbf{h}}_d). \quad (30)$$

The “new” sparse signal and sensing matrix then become

$$\tilde{\mathbf{h}}_d = \mathbf{P}' \bar{\mathbf{h}}_d = [\tilde{\mathbf{h}}_{d,1}', \tilde{\mathbf{h}}_{d,2}', \dots, \tilde{\mathbf{h}}_{d,G_t G_r}']' \quad (31a)$$

$$\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} \mathbf{P} = [\tilde{\boldsymbol{\Psi}}_1, \tilde{\boldsymbol{\Psi}}_2, \dots, \tilde{\boldsymbol{\Psi}}_{G_t G_r}] \quad (31b)$$

where $\tilde{\mathbf{h}}_{d,i} = [\bar{\mathbf{h}}_{d,i}(n_0), \bar{\mathbf{h}}_{d,i}(n_1), \dots, \bar{\mathbf{h}}_{d,i}(n_{L-1})]^*$ and $\tilde{\boldsymbol{\Psi}}_i = \text{diag}[\psi_i(0), \psi_i(1), \dots, \psi_i(L-1)]$, $\forall i \in [1, G_t G_r]$, with $\bar{\mathbf{h}}_{d,i}(n_j)$ and $\psi_i(l)$ being the i -th entry of $\bar{\mathbf{h}}_d(n_j)$ and $\boldsymbol{\psi}(l)$, respectively. Unlike the original $\bar{\mathbf{h}}_d$, the rearranged $\tilde{\mathbf{h}}_d$ exhibits block sparsity [30]. More importantly, the block sparsity of $\tilde{\mathbf{h}}_d$ equals to the sparsity of $\bar{\mathbf{h}}_d(n_l)$.

Towards the issue of block or structured compressed sensing, quite a few methods are available in the existing literature. To be more specific, these methods can be categorized into

the Bayesian type and non-Bayesian type. For the former, representative solutions include block-sparse Bayesian learning [31], [32], message passing based compresses sensing [33], [34], etc; while for the latter, representative solutions include structured Lasso [35], block OMP [17], etc. In this paper, we choose the OMP-based method similar to many related works [15]–[20]. It should be mentioned that this choice has no optimality guarantee, and other methods could be attempted in the future.

Although block OMP seems to be a powerful tool to estimate $\mathbf{h}_d$ in Eq. (28), accurate identification of the non-zero support still encounters two major difficulties:

- P.1 How to properly set the number of iterations when applying CS algorithms.
- P.2 How to avoid the potential degradation resulting from the strong Doppler effects.

To address these problems, we propose an algorithm termed as adaptive-block OMP (A-BOMP) that will be detailed next.

B. A-BOMP

When recovering the sparse signal via CS, a proper number of iterations is equal to (or a slightly higher than) the signal sparsity. Unfortunately, the actual sparsity of $\mathbf{h}_d$ is unknown. To reduce the risks of estimation loss, most works adopt large iterations. However, once the iterations severely mismatch the signal sparsity, it may result in increased computational complexity and potential over-fitting errors. Albeit not knowing the sparsity either, we will show that, it is possible to set iterations properly after tap identification.

Since D out of N_c taps are regarded effective, the number of beams should be no greater than D , thus the signal sparsity is upper bounded by D . Surprisingly, the upper bound could be set even smaller for implementation. To verify this, we first provide the following result

Lemma 1: For the wideband channel with N_c taps, the probability that k out of K ($k \leq K$) beams reside within one tap is approximated as

$$P(K, k) = C_K^k \left(\frac{1}{N_c} \right)^k \left(\frac{N_c - 1}{N_c} \right)^{K-k}. \quad (32)$$

A brief illustration is made under $N_c = 128$ and $K = 10$. In this case, $P(10, 4) < 10^{-5}$, implying that it is virtually impossible for one tap containing over 3 beams, so k is expected to be smaller than 4, regardless to say 10 for $P(10, 10) < 10^{-18}$. Combing above discussions, a proposition is made below to provide guidance on iterations setting:

Proposition 2 [Number of iterations]: Let P_T be a small threshold (e.g., 10^{-3}) and D be the number of effective taps after tap identification. A proper iterations can be set as $k-1$, where k is the smallest integer satisfying $P(D, k) < P_T$.

To address P.2, DPC-BEM model was used in [17] to capture the variations before implementing BOMP. This approach can dramatically lower the deterioration, but has two drawbacks. First, the estimation performance is heavily dependent on the basis order. Secondly, to construct orthogonal DPC basis, a large-scale eigenvalue decomposition (EVD) has to be involved [36] with complexity $\mathcal{O}(L^3)$. To lower complexity

Algorithm 1 Proposed A-BOMP Algorithm

Input: Received signal $\mathbf{y}_d$ and sensing matrix $\tilde{\Psi}_d$, maximum block-sparsity $\mathcal{K}$, group size S , group number $G = \frac{L}{S}$, and error threshold ϵ .

Output: The AoA support set $\tilde{\mathcal{A}}_d$ and the corresponding AoD support set $\tilde{\mathcal{D}}_d$.

- 1: Initialization: The residue $\mathbf{r}_d = \mathbf{y}_d$, iteration index $C = 0$, $\mathcal{A}_d/\tilde{\mathcal{A}}_d$ and $\mathcal{D}_d/\tilde{\mathcal{D}}_d$ are set to be empty, $\beta = \infty$, $\Phi = \emptyset$ and $x = x_0 = 0$.
- 2: **while** $C < \mathcal{K}$ and $\beta > \epsilon$ **do**
- 3: $C = C + 1$;
- 4: $g_i = \arg \max_g \sum_{i=1}^G \frac{\|[\tilde{\Psi}_{d,g}^* \mathbf{r}_d]((i-1)S+1:iS)\|_1}{\|\tilde{\Psi}_{d,g}\|_F}$
- 5: $n_R = \lceil g_i/G_t \rceil$ and $n_T = g_i - (n_R - 1)G_t$.
- 6: **if** $\exists i, \text{ mod } (\lfloor n_D/n_R - \mathcal{D}(i)/\mathcal{A}(i) \rfloor, G_t/G_r) \leq 1$ **then**
- 7: **goto** 2
- 8: **end if**
- 9: $\mathcal{A} = \{\mathcal{A}, n_R\}$, $\mathcal{D} = \{\mathcal{D}, n_T\}$
- 10: $\hat{\mathbf{A}}_T = [\mathbf{f}_{N_t}(\frac{n_T-1}{G_t} + \frac{2j_T}{G_t^2})]_{j_T \in [-\frac{G_t}{2}: \frac{G_t}{2}-1]}$
- 11: $\hat{\mathbf{A}}_R = [\mathbf{f}_N(\frac{n_R-1}{G_r} + \frac{2j_R}{G_r^2})]_{j_R \in [-\frac{G_r}{2}: \frac{G_r}{2}-1]}$
- 12: $\hat{\psi}[n_i] = (\mathbf{p}'_A[n_i] \otimes \mathbf{w}_A^*[n_i])(\hat{\mathbf{A}}_T^* \otimes \hat{\mathbf{A}}_R)_{i=1 \sim L}$
- 13: $\hat{\Psi}_{d,i} = \text{diag}[\hat{\psi}[n_1](i), \dots, \hat{\psi}[n_{G_t G_r}](i)]_{i=1 \sim G_t G_r}$
- 14: $\hat{g}_i = \arg \max_g \sum_{i=1}^G \frac{\|[\tilde{\Psi}_{d,g}^* \mathbf{r}_d]((i-1)S+1:mS)\|_1}{\|\tilde{\Psi}_{d,g}\|_F}$
- 15: $\hat{n}_R = \lceil \hat{g}_i/G_t \rceil$ and $\hat{n}_T = \hat{g}_i - (\hat{n}_R - 1)G_t$.
- 16: $\Phi = [\Phi, M(\mathbf{f}_{N_t}^*(\frac{G_t(n_T-1)+\hat{n}_T}{G_t^2})) \otimes \mathbf{f}_{N_r}(\frac{G_r(n_R-1)+2\hat{n}_R}{G_r^2})]$
- 17: **for** $j = 1 : G$ **do**
- 18: $\mathbf{j} = (j-1)S + 1 : mS$
- 19: $x = x + \|\Phi(\mathbf{j}, :)^\dagger \mathbf{y}(\mathbf{j})\|_2$
- 20: $\mathbf{r}_d(\mathbf{j}) = \mathbf{y}_d(\mathbf{j}) - \Phi(\mathbf{j}, :)\Phi(\mathbf{j}, :)^{\dagger} \mathbf{y}(\mathbf{j})$
- 21: **end for**
- 22: $\beta = |x - x_0|/x$
- 23: $x_0 = x$, $x = 0$
- 24: $\tilde{\mathcal{A}}_d = \{\tilde{\mathcal{A}}_d, \frac{2\pi G_r(n_R-1)+\hat{n}_R}{G_r^2}\}$
- 25: $\tilde{\mathcal{D}}_d = \{\tilde{\mathcal{D}}_d, \frac{2\pi G_t(n_T-1)+\hat{n}_T}{G_t^2}\}$
- 26: **end while**

while remaining robustness against Doppler, A-BOMP is proposed with its pseudo-code presented in **Algorithm 1**. In A-BOMP, each outer iteration consists of three parts:

- S₁** (Lines.4-9) *partial basis matching*: select the angle pair having the largest sum of grouping correlations, and make sure that there is no overlapping with the selected ones.
- S₂** (Lines.10-16) *resolution refinement*: re-construct sensing matrix associated with the selected angle pair, and implement estimation procedure like **S₁** to refine resolution.
- S₃** (Lines.17-21) *partial residue update*: estimate the coefficients by the least-squared (LS) estimator, then update the residue $\mathbf{r}_d$ by subtracting the projection of each group.

In A-BOMP, another key parameter is the group size S (equivalent to the group number G). In the presence of Doppler, the size of the non-zero support always exceeds the number of measurements. In this case, accurately localizing the non-zero support is already challenging, not to mention to the recovery of the entire vector. The only exception is in the absence of Doppler, where the uncertainty could vanish when sufficient training frames are available. The great shortage of measurements forces us to “shrink” the non-zero support. To this end, we divide $\tilde{\mathbf{h}}_{d,i}$ defined in Eq. (31a) into S groups, and those entries belonging to one group are highly correlated thus being treated equally. Therefore, the group division essentially performs the signal compression, and S is nothing but the coherent interval.

Proposition 3 [Determination of the group size]: Let τ denote a high-correlation coefficient (e.g., 0.707). A proper group size can be set as the largest S satisfying $\cos(\omega_{max}N_cS) \leq \tau$ and $\omega_{max}N_cS \leq \pi/2$.

Proposition 3 indicates that a smaller ω_{max} results in a larger S . For $\omega_{max} = 0$, i.e., a time-invariant channel, A-BOMP degenerates to BOMP as $G = L/S = 1$. Besides, one can readily verify that estimating $\tilde{\mathbf{h}}_d$ via BOMP and estimating $\tilde{\mathbf{h}}_d(n_i)$ via OMP are equivalent. Compared to BOMP, A-BOMP only introduces a few small-scale matrix inversions, and simulations show that such minimal computational cost will bring in a significantly improved accuracy.

Based on the output of A-BOMP, the steering matrices for tap- d channel are estimated as

$$\tilde{\mathbf{A}}_{r,d} = [\mathbf{f}_{N_r}(\tilde{\mathcal{A}}_d(1)/2\pi), \dots, \mathbf{f}_{N_r}(\tilde{\mathcal{A}}_d(c_d)/2\pi)] \quad (33a)$$

$$\tilde{\mathbf{A}}_{t,d} = [\mathbf{f}_{N_t}(\tilde{\mathcal{D}}_d(1)/2\pi), \dots, \mathbf{f}_{N_t}(\tilde{\mathcal{D}}_d(c_d)/2\pi)] \quad (33b)$$

with $c_d = \text{cal}(\tilde{\mathcal{A}}_d)$. The approximate beamspace representation for tap- d channel bear a form as

$$\tilde{\mathbf{H}}_d(n) = \tilde{\mathbf{A}}_{r,d} \text{diag}(\tilde{\mathbf{g}}_d(n)) \tilde{\mathbf{A}}_{t,d}^* \quad (34)$$

where $\tilde{\mathbf{g}}_d(n)$ consists of unknown path gains. Despite that both the amplitudes and angle support can be simultaneously obtained via OMP in time-invariant channels, for the more general time-varying channels, an additional stage is still necessary to estimate the amplitudes and Doppler.

So far, we have completed the second part of the random-probing stage. Summarizing, the main steps are listed as

- Stack the receive samples for each identified tap.
- Transform the samples into a generic block-sparse form.
- Determine the iterations and group size for A-BOMP.
- Apply A-BOMP to estimate the angle support.

V. JOINT ESTIMATION OF PATH GAIN & DOPPLER

At the random-probing stage, the effective taps are identified with their angle support obtained as well. In this section, we proceed to estimate the remaining unknown path gain/Doppler at the so-termed steering-probing stage.

A. Steering Probing Design

To accurately estimate path gains and Doppler shifts, steering-probing will be implemented based on the estimated

beam direction. Specifically, for tap- d channel, construct set $\mathcal{I}_d$ whose i -th element is $(\mathcal{A}_d(i), \mathcal{D}_d(i))$. Because of the off-grid issues in beamspace and delay domain, different $\mathcal{I}_d$'s may share the same element, thus we get their union as

$$\mathcal{I} = \mathcal{I}_{d_1} \cup \mathcal{I}_{d_2} \cup \dots \cup \mathcal{I}_{d_D}. \quad (35)$$

Further, all AoAs and AoDs are individually extracted from $\mathcal{I}$ and captured by $\mathcal{I}_A$ and $\mathcal{I}_D$, respectively. To facilitate beam-forming, only the discrete AoD indices need to be fed back. Without causing ambiguity, we reset the time instant at the steering-probing stage, and making the following definition.

Steering-probing vector: At the steering-probing stage, denote $\mathbf{p}_t^S(n)$ and $\mathbf{p}_r^S(n)$ to be the RF vectors at time instant n . To improve receive SNR, $p_{p,t}^S(n)$ (the p -th element of $\mathbf{p}_t^S(n)$) and $p_{q,r}^S(n)$ (the q -th element of $\mathbf{p}_r^S(n)$) are designed as [17]

$$p_{p,t}^S(n) = \frac{1}{\sqrt{N_t}} e^{jQ((p-1)\mathcal{I}_D(\hat{n}))}, p \in [1, N_t] \quad (36a)$$

$$p_{q,r}^S(n) = \frac{1}{\sqrt{N_r}} e^{jQ((q-1)\mathcal{I}_A(\hat{n}))}, q \in [1, N_r] \quad (36b)$$

with $\hat{n} = \text{mod}(\lfloor n/N_c \rfloor, \text{cal}(\mathcal{I}))$. As can be seen from Eq. (36), the probing vectors repeat every $\text{cal}(\mathcal{I})$ subframes, so that that each beam will be steered once during each polling.

B. Path Gain/Doppler Estimation

At the i -th polling, stacking all tap- d related samples yields

$$\mathbf{y}_{d,i} = [y(d + n_{i,0}), \dots, y(d + n_{i,|\mathcal{I}|-1})]' \quad (37)$$

where $n_{i,j} = N_c j + \text{cal}(\mathcal{I}) N_c i$, $\forall j \in [0, \text{cal}(\mathcal{I})]$. Using the compact beamspace representation obtained in Eq. (34), each sample in $\mathbf{y}_{d,i}$ is approximately equivalent to

$$\begin{aligned} y(n_{i,j} + d) &\approx (\mathbf{p}_r^S(n_{i,j}))^* \tilde{\mathbf{A}}_{r,d} \text{diag}(\tilde{\mathbf{g}}_d(n_{i,j} + d)) \\ &\quad \times \tilde{\mathbf{A}}_{t,d}^* \mathbf{p}_t^S(n_{i,j}) + \xi(n_{i,j} + d) \\ &= \text{vec}'(\text{diag}(\tilde{\mathbf{g}}_d(n_{i,j} + d))) \mathbf{m}_d(n_{i,j}) \\ &\quad + \xi(n_{i,j} + d) \end{aligned} \quad (38)$$

where $\mathbf{m}_d(n_{i,j}) = ((\mathbf{p}_t^S(n_{i,j}))^* \tilde{\mathbf{A}}_{t,d}^*) \otimes ((\mathbf{p}_r^S(n_{i,j}))^* \tilde{\mathbf{A}}_{r,d})$. By capturing the amplitudes with the one sampling in the middle of current polling, $\mathbf{y}_{d,i}$ can be approximately represented as

$$\mathbf{y}_{d,i} \approx \underbrace{\begin{bmatrix} \mathbf{m}_d(n_{i,0}) \\ \mathbf{m}_d(n_{i,1}) \\ \vdots \\ \mathbf{m}_d(n_{i,\text{cal}(\mathcal{I})-1}) \end{bmatrix}}_{\mathbf{M}_{d,i}} \text{vec}(\text{diag}(\mathbf{g}_d(\bar{n}_i))) + \boldsymbol{\xi}_{d,i} \quad (39)$$

with $\bar{n}_i = (n_{i,0} + n_{i,\text{cal}(\mathcal{I})-1})/2 + d$ and $\boldsymbol{\xi}_{d,i} = [\xi(n_{i,0} + d), \xi(n_{i,1} + d), \dots, \xi(n_{i,\text{cal}(\mathcal{I})-1} + d)]'$. Let $\tilde{\mathbf{M}}_{d,i} = [\mathbf{M}_{d,i}[:, 1^2], \mathbf{M}_{d,i}[:, 2^2], \dots, \mathbf{M}_{d,i}[:, C_d^2]]$, then Eq. (39) equals to

$$\mathbf{y}_{d,i} \approx \tilde{\mathbf{M}}_{d,i} \mathbf{g}_d(\bar{n}_i) + \boldsymbol{\xi}_{d,i} \quad (40)$$

Since $\text{cal}(\mathcal{I}) \geq c_d$, $\mathbf{g}_d(\bar{n}_i)$ can be recovered by LS estimator:

$$\hat{\mathbf{g}}_d(\bar{n}_i) = \tilde{\mathbf{M}}_{d,i}^\dagger \mathbf{y}_{d,i} = \mathbf{g}_d(\bar{n}_i) + \tilde{\mathbf{M}}_{d,i}^\dagger \boldsymbol{\xi}_{d,i}. \quad (41)$$

Once getting a new $\hat{g}_d$, we pick its j -th element, which is the estimated amplitudes of the j -th beam in current polling. The polling lasts for $R = \lfloor L/\text{cal}(\mathcal{I}) \rfloor$ times,⁵ so a pseudo time series is finally obtained as

$$\hat{g}_{d,j} = [\hat{g}_{d,j}(\bar{n}_0), \hat{g}_{d,j}(\bar{n}_1), \dots, \hat{g}_{d,j}(\bar{n}_{R-1})]'. \quad (42)$$

Lemma 2: Through repetitive polling, the pseudo time series $\hat{g}_{d,j}$ has an equal sampling interval thus can be modeled as finite noisy samples of a single-tone sinusoid.

Many techniques have been proposed over the years for the frequency estimation of a complex sinusoid in complex additive white Gaussian noise. Here we adopt the WNALP estimator known for its computational efficiency and near-optimal performance [37]. The detailed procedures for path gain/Doppler estimation are described as below

- Set $M_0 = \lfloor R/2 \rfloor$.
- Calculate the autocorrelation of $\hat{g}_{d,j}$ as

$$R(m) = \frac{1}{R-m} \sum_{i=m+1}^{2M_0} \hat{g}_{d,j}(\bar{n}_i) \hat{g}_{d,j}^*(\bar{n}_{i-m}) \quad (43)$$

- Calculate the smoothing coefficient w_m as

$$w_m = \frac{3((M_0 - m)(2M_0 - m + 1) - M_0^2)}{M_0(4M_0^2 - 1)} \quad (44)$$

- Estimate the Doppler shift as

$$\hat{\omega}_{d,j} = \frac{1}{N_c \text{cal}(\mathcal{I})} \sum_{m=1}^{M_0} w_m \text{angle}(R(m) R^*(m-1)) \quad (45)$$

- Estimate the amplitudes as

$$\begin{aligned} \hat{g}_{d,j}(d) &= \frac{e^{-j\hat{\omega}_{d,j} \frac{N_c \text{cal}(\mathcal{I})}{2}}}{R} \sum_{i=1}^R \hat{g}_{d,j}(\bar{n}_i) e^{-j\hat{\omega}_{d,j} N_c \text{cal}(\mathcal{I})(i-1)} \\ &= \frac{1}{R} \sum_{i=1}^R \hat{g}_{d,j}(\bar{n}_i) e^{-j\hat{\omega}_{d,j} N_c \text{cal}(\mathcal{I})(i-1/2)}. \end{aligned} \quad (46)$$

The rest beams can be estimated similarly thus being omitted.

Summarizing, the steering-probing stage is carried out as:

- 1) Perform beamforming polling based on the union of the angle supports.
- 2) Estimate the path gains/Doppler for each tap via WNALP estimator.

VI. DISCUSSIONS AND SIMULATIONS

A. Implementing Discussions

Implementing procedures: The implementation of the proposed DSDS channel estimator entails four key components:

- 1: Send multiple delta-like training pilots to separate channel taps with random probing applied.

⁵Similar to the random-probing stage, we introduce the steering-probing state based on one frame consisting of L subframes. In practice or numerical comparisons, one can simply replace L with the actual number of subframes, i.e., $\text{cal}(\mathcal{I})R$.

- 2: Identify significant channel taps via energy detector to exploit the delay-domain sparsity regardless of Doppler.
- 3: Identify the direction of significant beams via A-BOMP with effective mechanism applied to combat time-variation.
- 4: Apply steering-probing to estimate the amplitudes and Doppler using high-quality received samples.

The proposed channel estimator is tailored for a general doubly-selective channels. In practice, the investigated channel may not exhibit double selectivity, so one can use part of the above steps to accommodate these special cases.

Storage demand: The major storage demand in channel estimation comes from the sensing matrix. Suppose p_1 frames ($p_1 L$ subframes actually) are allocated at the random-probing stage, then the size of sensing matrix in [19] is $C_1 = p_1 N \times U N_c G_t G_r$, with U being the up-sampling ratio. Although our estimation is conducted at each tap independently, the sensing matrix is shared by all taps with size $C_2 = p_1 L \times G_t G_r$. For $N_c = 128$ and $N = 512$, C_1 is more than 13000 times larger than C_2 .

Computational complexity: The major computational complexity comes from the OMP-based algorithm, which comprises three parts: basis matching, orthogonal projection, and residue update. For a Q -dimensional sparse vector recovered via V measurements, the involved flops for these parts at iteration- k are $(2V-1)Q$, $4kV$, and $2kV$ flops, respectively [38]. The total flops of [19] and ours are $p_1 N V (2U N_c G_t G_r + 3p_1 N)$ and $p_1 L V G (2G_t G_r + 3p_1 L)$, both in the order of $\mathcal{O}(p_1^2)$. Thanks to our extremely small-scale sensing matrix, even for $p_1 < 300$, the former is still more than ten times larger than the upper-bound of the latter.

Sensing matrix construction: To ensure a reliable recovery via OMP, the sensing matrix should best satisfy the restricted isometry property (RIP). According to [39], the optimal sensing matrix in terms of the RIP is the independent and identically distributed (IID) Gaussian matrix. Unfortunately, due to the constant-modulus limitation of APS, the optimal sensing matrix remains an open topic. In this work, we follow [17], [18] and randomly adjust APS obeying a uniform distribution.

From a single RF chain to multiple RF chains: Although the DS-DS channel estimator is introduced based on a single RF chain, it can be readily generalized to multiple RF chains, because the proposed estimator is relevant to RF precoder only without a dedicated digital precoding design similar to [25]. Besides a slight change in the number of effective measurements each time, the algorithm can be carried out for arbitrary number of RF chains without any modification.

B. Simulation Verifications

In this subsection, extensive numerical results are presented to verify the advantages of the proposed approach over existing works. In simulations, the system carrier frequency f_c is 60 GHz. The number of antennas is $N_t = N_r = 32$. The dictionary sizes are $G_t = G_r = 64$. $h(\cdot)$ is the raised-cosine filter with the roll-off factor $\beta = 1$. Each channel realization is generated according to Eq. (3) with P ranging from 1 to 4. One-stage refinement is applied for all cases when applying OMP-based methods. If not specified, the resolution of APS is 2-bit. Other simulation parameters include $N_c = 16$, $N = 64$,

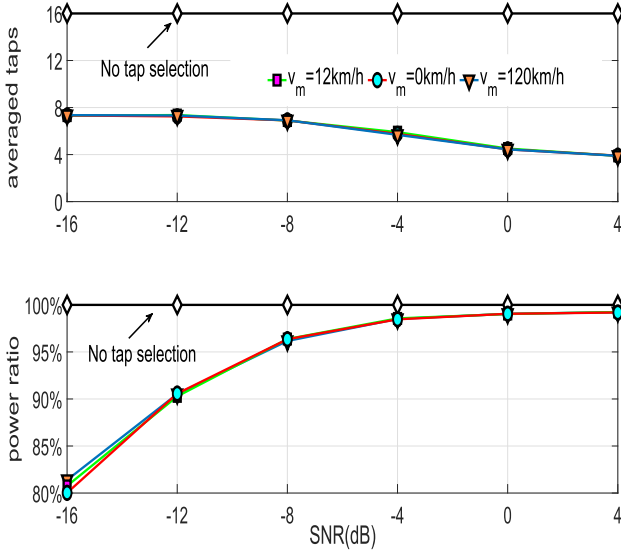


Fig. 3. The averaged selected taps after tap identification.

$T_s = 50\text{ns}$, $A = 8$, $P_T = 10^{-3}$, $\epsilon = 0.01$ and $\mu = 0.03$. The SNR (averaged TSNR) is defined as $\frac{L}{N\sigma^2}$. The estimation performance is weighted by the normalized MSE (NMSE) given by

$$\varepsilon = \frac{\sum_{d=0}^{N_c-1} \|\mathbf{H}_d - \widehat{\mathbf{H}}_d\|_F^2}{\sum_{d=0}^{N_c-1} \|\mathbf{H}_d\|_F^2}. \quad (47)$$

Each curve is on the average of 1000 channel realizations.

1) *Verification of the Functionality of Tap Identification:* To verify the effectiveness of tap identification, we plot the averaged selected taps together with their power ratio in Fig. 3. $P = 3$ and 40 frames are allocated at the random-probing stage. Three different v_m 's: 0, 12km/h, and 120km/h are considered. We see that the tap identification is regardless of Doppler effects. Fewer taps are selected with the increase of SNR, and reduction is 75% at 0dB. As this work uses the raised-cosine pulse-shaper, the delay-domain also suffers from off-grid issues due to side-lobe leakage, so the number of identified taps is slightly larger than that of actual paths. Note that, the large reduction in taps to be processed is not at the cost of power loss. As can be seen from Fig. 3, the averaged power ratio soon exceeds 97% at medium SNR. The effectiveness of tap identification is attributed to the delay-domain sparsity of mmWave channels.

2) *NMSE Comparisons in Static & Wideband Channels:* We then compare the double-sparse approach (DSA) with state-of-the-art beamspace-sparse approach (BSA) [19] at the same averaged SNR in Fig. 4. The channel is generated with 3 paths and $\omega_m = 0$. For DSA, 40 training frames are allocated at the random-probing stage with repeating beamforming polling for $R = 4$ times at the steering-probing stage. 60 training frames are allocated for BSA and the regularized LS-estimator. For BSA, its sensing matrix size is 16384×131072 , requiring a memory space over 18GB, in contrast to ours with size 200×4096 occupying 9Mb memory space. Due to the great shortage of training frames, the LS estimator without utilizing any sparsity performs the worst.

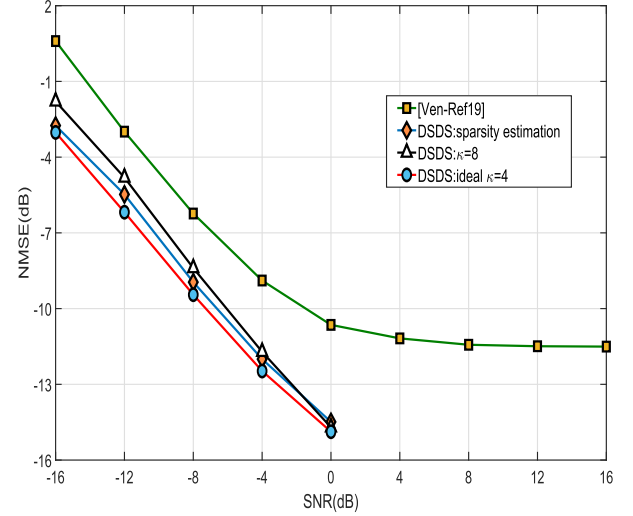


Fig. 4. The NMSE comparisons among different schemes in static wideband channels.

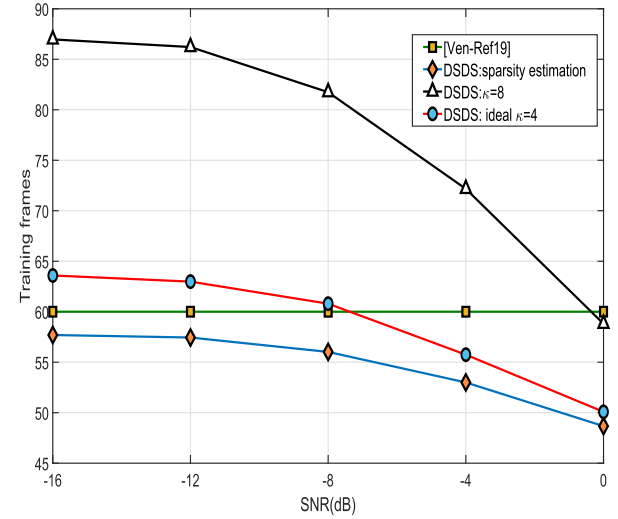


Fig. 5. The total training frames consumed by different schemes in static wideband channels.

BSA performs much better than LS but still much worse than DSA. In addition, the shortage of probings makes the NMSE curve of BSA soon becomes flat. Even under the same peak SNR, we see that DSA still outperforms BSA at medium-to-high SNR region, implying that the benefits brought by DSA outweigh the power inefficiency of the proposed training pattern.

In Fig. 5, we further plot the averaged consumed training frames of different approaches. From two figures, it is clear that improper iterations ($\mathcal{K} = 8$) will result in additional training overhead without making any substantial performance improvement. Following proposition 2, iterations can be properly set for A-BOMP (A-BOMP is equivalent to OMP here), and the resultant NMSE performance is very close to the ideal benchmark ($\mathcal{K} = 4$). With pre-determined iterations, DSA requires the least training overhead, with a reduction of 20% compared to BSBA at high SNR.

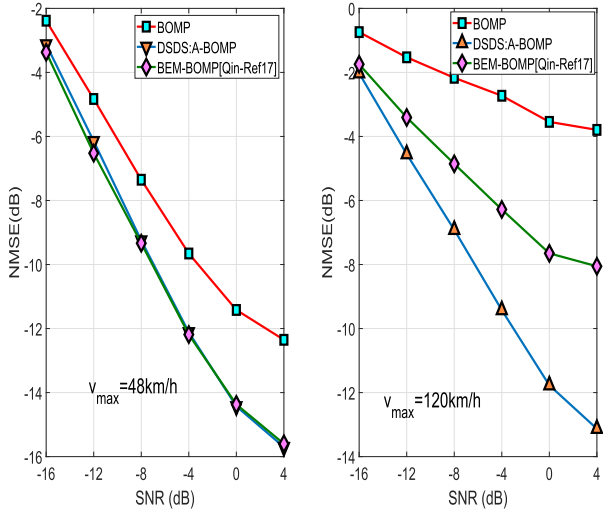


Fig. 6. The NMSE comparisons in “frequency-flat” & time-varying channels in modest mobility.

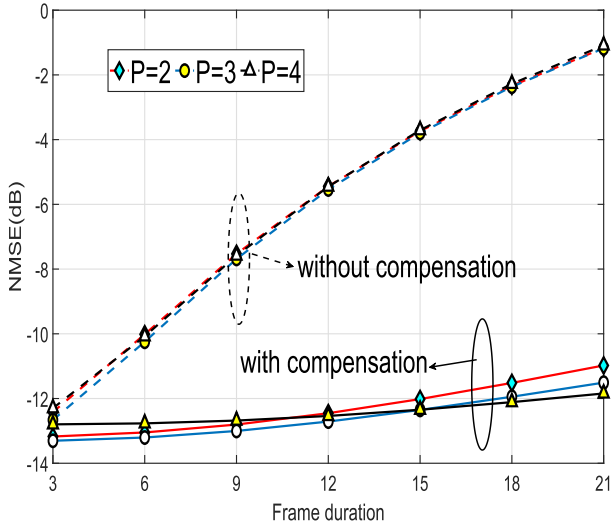


Fig. 7. The NMSE versus the number of paths.

3) *NMSE Comparisons in “Frequency-Flat” & Time-Varying Channels:* In Fig. 6, the channel is generated with $P = 3$ with $v_m = 48\text{km/h}$ and $v_m = 120\text{km/h}$, respectively. We compare the NMSEs with the angle support recovered via A-BOMP and DPC-BOMP [17], respectively. Since each tap channel is “frequency-flat” & time-varying (FTV), the results are essentially the comparison with state-of-the-art FTV channel estimator [17]. $p_1 = 60$ frames are allocated at the random-probing stage with repeating beamforming polling $R = 4$ times at the steering-probing stage. The DPC-basic order is 2 as in [17]. In modest mobility ($v_m = 48\text{km/h}$), A-BOMP and DPC-BOMP achieve similar performances, both outperforming BOMP remarkably. In high mobility ($v_m = 120\text{km/h}$), the advantage of A-BOMP over DPC-BOMP becomes notable. As described before, A-BOMP avoids the large-scale EVD required in DPC-BOMP, demonstrating that it is more efficient and superior.

4) *NMSE Performance in Doubly-Selective Channels:* To thoroughly evaluate the functionality of DSA, we fix

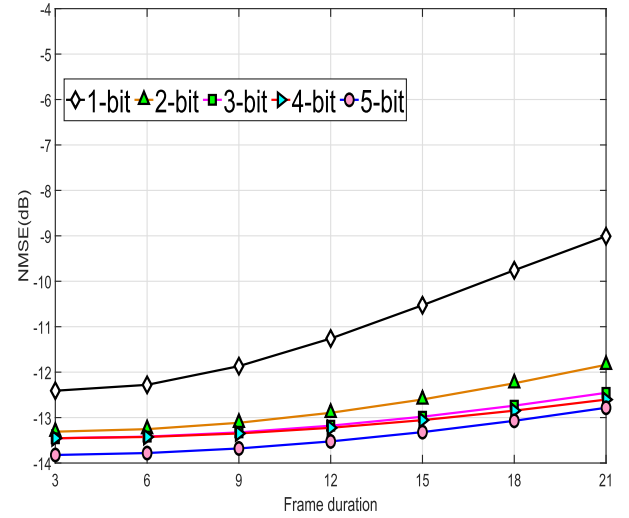


Fig. 8. The NMSE versus the resolution of APS.

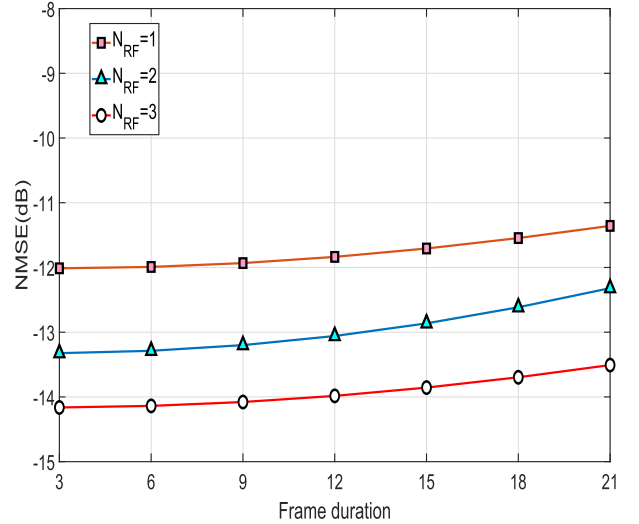


Fig. 9. The NMSE versus the number of RF chains.

$\text{SNR} = -1\text{dB}$ and $v_m = 55\text{km/h}$, and simulate NMSE versus frame duration under various conditions.

In Fig. 7, we compare the NMSEs by varying the number of paths. Other parameters are set as $p_1 = 60$, $R = 4$, and $b = 2$. The results show that, without Doppler compensation, the NMSE is soon to exceed -10dB , resulting in a great discrepancy with the actual channels. By compensating for the Doppler using the estimate, superb tracking ability can be guaranteed over up to 20 frames. Furthermore, there is a minimal performance degradation when increasing P from 2 to 4. In contrast, [19] suffers from a nearly 2dB degradation in a similar setup, implying that the proposed estimator is more robust against frequency selectivity.

In Fig. 8, we compare the NMSEs by varying the resolution (referring to b) of APS. Other parameters are set as $p_1 = 60$, $b = 2$, $R = 4$, and $P = 3$. A remarkable performance gap is noticed with the ultra-coarse 1-bit APS. However, increasing b by 1 bit will lead to a huge improvement. The performance gap compared to a finer APS (3~5-bit) in terms of the NMSE

is very small (only 0.5dB), implying that the proposed channel estimator is insensitive to the resolution of APS.

In Fig. 9, we compare the NMSEs by varying the number of RF chains. Other parameters are set as $P = 3$, $b = 2$, $R = 6$, and $p_1 = 30$. As can be seen, multiple RF chains can lower the estimation error compared to the single RF chain. This is because multiple RF chains can generate more random beam probing patterns, which in turn benefits the recovery of the angle support using CS. We need to mention that throughout the estimation, all non-zero symbols are set as one. Actually, if the peak to average power ratio (PAPR) is not a significant concern, one can potentially set these symbols as the Gaussian distributed variables like [18] to strengthen the randomness.

VII. CONCLUSION

In this paper, we investigated the doubly-selective channel estimation for hybrid mmWave mMIMO systems. With the help of judiciously designed training pattern and analog probing, the beamspace sparsity and the delay-domain sparsity can be jointly exploited to facilitate estimation. The proposed channel estimator demonstrates strong robustness against double selectivity, without imposing any additional constraints on the hybrid structure itself. Compared with existing works, our proposed doubly-sparse approach is demonstrated to be a more general and superior solution to channel estimation under hybrid mmWave mMIMO.

REFERENCES

- [1] S. Gao, X. Cheng, and L. Yang, "Making wideband channel estimation feasible for mmwave massive MIMO: A doubly sparse approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, May 2019, pp. 1–6.
- [2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [3] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [4] S. Wu, C.-X. Wang, E.-H.-M. Aggoune, M. M. Alwakeel, and X. You, "A general 3-D non-stationary 5G wireless channel model," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3065–3078, Jul. 2018.
- [5] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-frequency mmwave massive MIMO channel measurements and characterization for 5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1591–1605, Jul. 2017.
- [6] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.
- [7] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [8] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, Aug. 2018.
- [9] Y. Liu, C.-X. Wang, J. Huang, J. Sun, and W. Zhang, "Novel 3-D nonstationary mmwave massive MIMO channel models for 5G high-speed train wireless communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2077–2086, Mar. 2019.
- [10] C. Chen, Y. Dong, X. Cheng, and L. Yang, "Low-resolution PSs based hybrid precoding for multiuser communication systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6037–6047, Jul. 2018.
- [11] S. Gao, Y. Dong, C. Chen, and Y. Jin, "Hierarchical beam selection in mmwave multiuser MIMO systems with one-bit analog phase shifters," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2016.
- [12] S. Gao, X. Cheng, and L. Yang, "Spatial multiplexing with limited RF chains: Generalized beamspace modulation (GBM) for mmwave massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2029–2039, Sep. 2019.
- [13] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [14] S. Gao, X. Cheng, L. Yang, and R. Zhang, "Zero-forcing based limited feedback hybrid precoding in mmwave communications," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2019, pp. 362–366.
- [15] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [16] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [17] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9435–9448, Oct. 2018.
- [18] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [19] K. Venugopal, A. Alkhateeb, N. Gonzalez Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [20] J. Fernandez, N. Prelcic, K. Venugopal, and R. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [21] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions Wireless Commun.*, vol. 6, no. 5, pp. 1743–1763, May 2007.
- [22] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
- [23] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
- [24] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [25] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 6493–6497.
- [26] S. Buzzi and C. D'Andrea, "On clustered statistical MIMO millimeter wave channel simulation," 2016, *arXiv:1604.00648*. [Online]. Available: <http://arxiv.org/abs/1604.00648>
- [27] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.
- [28] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1445–1455, Mar. 2014.
- [29] X. He, R. Song, and W. P. Zhu, "Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2990–3004, May 2016.
- [30] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.
- [31] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.
- [32] Z.-Q. He, X. Yuan, and L. Chen, "Super-resolution channel estimation for massive MIMO via clustered sparse Bayesian learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6156–6160, Jun. 2019.
- [33] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.

- [34] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2018.
- [35] A. Liu, V. K. N. Lau, and W. Dai, "Exploiting burst-sparsity in massive MIMO with partial channel support information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7820–7830, Nov. 2016.
- [36] T. Zemen and C. F. Mecklenbrauker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.
- [37] A. B. Awoseyila, C. Kasparis, and B. G. Evans, "Improved single frequency estimation with wide acquisition range," *Electron. Lett.*, vol. 44, no. 3, pp. 245–247, Jan. 2008.
- [38] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
- [39] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.



Shijian Gao (Student Member, IEEE) received the B.Sc. degree in electrical engineering from Nankai University in 2014 and the M.Sc. degree in electrical engineering from Peking University in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins. His research interests are in the areas of wireless communications and related fields.



Xiang Cheng (Senior Member, IEEE) received the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., and the University of Edinburgh, Edinburgh, U.K., in 2009. He is currently a Professor with Peking University. His general research interests are in the areas of channel modeling, wireless communications, and data analytics, subject on which he has published more than 200 journal articles and conference papers, six books, and holds ten patents. He was a recipient of the IEEE Asia-Pacific Outstanding Young Researcher Award in 2015 and a co-recipient of the 2016 IEEE JSAC Best Paper Award: Leonard G. Abraham Prize, the NSFC Outstanding Young Investigator Award, and the First-Rank and Second-Rank Award in Natural Science, Ministry of Education, in China. He has received the Best Paper Awards at IEEE ITST'12, ICC'13, ITSC'14, ICC'16, ICNC'17, GLOBECOM'18, ICCS'18, and ICC'19. He has served as the symposium leading chair, the co-chair, and a member of the technical program committee for several international conferences. He is also an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE WIRELESS COMMUNICATIONS LETTERS, and the *Journal of Communications and Information Networks*. He is an IEEE Distinguished Lecturer.



Liuqing Yang (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2004. She is currently a Professor with Colorado State University. Her general interests are in communications and networking—subjects on which she has published more than 330 journal articles and conference papers, four book chapters, and five books. She was a recipient of the ONR Young Investigator Program (YIP) Award in 2007, and the NSF Faculty Early Career Development (CAREER) Award in 2009, the Best Paper Award at the IEEE ICUWB'06, ICC'13, ITSC'14, GLOBECOM'14, ICC'16, WCSP'16, GLOBECOM'18, ICCS'18, and ICC'19. She is the Editor-in-Chief for *IET Communications*, an Executive Editorial Committee (EEC) Member of the IEEE TRANSACTIONS ON COMMUNICATIONS, and a Senior Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. She has also served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE INTELLIGENT SYSTEMS, and *PHYCOM: Physical Communication*, and as program chair, track/symposium or TPC chair for many conferences.