

Received March 29, 2018, accepted May 28, 2018, date of publication June 4, 2018, date of current version July 6, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2843783

Sparse Channel Estimation for Massive MIMO-OFDM Systems Over Time-Varying Channels

QIBO QIN¹, LIN GUI¹, (Member, IEEE), BO GONG¹, AND SHENG LUO²

¹Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

²College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518000, China

Corresponding author: Lin Gui (guulin@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grants 61471236, 61420106008, and 61671295, in part by the 111 Project under Grant B07022, in part by the Shanghai Key Laboratory of Digital Media Processing, and in part by the Shanghai Pujiang Program under Grant 16PJD029.

ABSTRACT For downlink orthogonal frequency division multiplexing-based massive multiple-input-multiple-output transmission over the time-varying channel, channel estimation is very challenging due to numerous channel coefficients to be estimated. To track this problem, this paper proposes a novel sparse channel estimation scheme, exploiting the sparsity in the delay domain and the high correlation in the spatial domain. By utilizing the basis expansion model (BEM) to model the time variation and the generalized-spatial BEM to model the spatial correlation, we are able to significantly reduce the number of the coefficients to be estimated. Then, the channel estimation is formulated into a compressive sensing problem with a quasi-block-sparse matrix to be recovered. Motivated by the quasi-block sparsity, a novel quasi-block simultaneous orthogonal matching pursuit (QBSO) algorithm is developed to recover the channel, which seeks nonzero channel tap positions at the first stage and calculates the channel coefficients at the second stage. Moreover, an adaptive-QBSO algorithm is further designed to improve the recovery accuracy, where the measurement matrix is adaptively designed based on the estimated virtual angle of departure, and thus, the sparsity representation can be strengthened. We also discuss the design of pilot pattern including pilot values and positions. Simulation results are provided to validate the effectiveness of our proposed channel estimation scheme.

INDEX TERMS Time-varying channel estimation, massive MIMO, compressive sensing, generalized-spatial basis expansion model, pilot design.

I. INTRODUCTION

As a promising technology for the next-generation wireless communication systems, massive multiple-input-multiple-output (MIMO) has drawn considerable interest in both academia and industry due to its high spectral and energy efficiency, reliable linkage, and high spatial resolution [1]–[5]. To reach these potential gains, accurate downlink channel state information (CSI) is required at the user equipment (UE) for signal detection, as well as at the base station (BS) for precoding. However, large number of antennas at the BS pose a great challenge in obtaining the downlink CSI due to numerous channel parameters to be estimated, especially over time-varying channels which introduce the inter-carrier interference (ICI) to the orthogonal frequency division multiplexing (OFDM) system. Several

estimators for MIMO systems were proposed in [6] and [7], which obtained satisfactory estimation accuracy over time-varying channels. However, these estimators are based on the assumption of rich multipath channels, and incur high pilot overhead when the number of transmit antennas becomes large.

In wideband wireless communications, growing experimental studies have verified that many practical channels exhibit sparsity in the delay domain, where the delay spread could be very large but the number of distinguishable multipath delays is usually small [8]–[10]. Exploiting this feature, the compressive sensing (CS) theory [11] can be leveraged for the sparse channel estimation to reduce the pilot overhead [12]. Recently, some CS-based estimators for time-varying channels have been designed [13]–[16].

A position-based interference elimination method was proposed in [13] to eliminate the ICI by exploiting the train position information. Cheng *et al.* [14] utilized the basis expansion model (BEM) to capture the time variation and decoupled the channel estimation into an ICI-free structure owing to a sparse pilot pattern. A joint multi-symbol estimator was designed in [15], where a block-based simultaneous orthogonal matching pursuit (BSOMP) algorithm was designed to improve both estimation accuracy and spectrum efficiency. The strategy was extended to the massive MIMO scenario in [16], which exploited the common channel support of the closely deployed antennas and developed a block discrete stochastic optimization (BDSO) algorithm to optimize the pilot positions. However, the aforementioned works [13]–[16] only looked into the sparsity in the delay domain, without taking into account the spatial correlation of the massive MIMO channel.

For a typical cellular configuration, since the BS is usually elevated highly with limited number of surrounding scatterers, the angle spread of the outgoing rays at the BS corresponding to a specific UE is narrow [17]. Hence, different paths between the BS and the UE may be highly correlated in the spatial domain. Taking into account this feature, a joint spatial division and multiplexing (JSDM) scheme was proposed in [18], which exploited the low-rank structure of the channel covariance matrix (CCM) and designed a downlink precoder to reduce the dimension of the effective channel matrix. Meanwhile, Yin *et al.* [19] proposed a CCM-aided uplink channel estimation scheme, which utilized the dominant eigenvectors to span the massive MIMO channel in the spatial domain. However, the acquisition of CCM is difficult in practical application, especially for time-varying scenarios where the angle spread would change rapidly. Besides, these CCM-aided methods involve in high complexity due to the need for singular value decomposition (SVD).

To effectively exploit the spatial correlation with low complexity, the virtual channel representation (VCR) [20] has been leveraged to acquire sparsity in the virtual angle domain [21], [22]. Rao and Lau [21] exploited the joint angle sparsity of multi-user channel matrices, where the measurements were observed at the UEs whereas the channel recovery was performed at the BS. Meanwhile, a CS-based adaptive CSI acquisition scheme was proposed in [22], which looked into the common sparsity support shared by channels of multiple subcarriers. However, these estimators exploiting the sparsity in the virtual angle domain would face with the power leakage and bring a performance loss when estimating the continuously distributed angles. Recently, the spatial BEM (SBEM) was built in [23]–[25] to model the massive MIMO channels with fewer parameter dimensions, which can be viewed as a modified VCR but represented the channel in a sparser form [5].

In this paper, we propose a novel sparse channel estimation scheme for massive MIMO-OFDM downlink transmission over time-varying channels, exploiting the sparsity in the delay domain and high correlation in the spatial domain.

The main contributions of the paper are summarized as follows:

- We propose a generalized-SBEM (G-SBEM) to capture the spatial correlation of the massive MIMO channels. Compared with the SBEM in [5], basis functions of the G-SBEM are not limited to be orthogonal, and as a consequence, the G-SBEM is more flexible in designing the dictionary matrix and able to improve the resolution of angles of departure (AoDs). Based on the G-SBEM and the sparsity in the delay domain, we formulate the time-varying channel estimation into a CS problem with a quasi-block-sparse matrix to be recovered.
- Motivated by the quasi-block-sparse structure in our formulated problem, a novel quasi-block SOMP (QBSO) algorithm is developed to recover the channel, which seeks nonzero channel tap (NCT) positions at the first stage, and calculates channel coefficients at the second stage. Furthermore, an adaptive-QBSO algorithm is designed to refine the estimation accuracy, which first estimates a virtual AoD corresponding to each cluster, and then strengthens the sparsity representation by adaptively designing the measurement matrix.
- We formulate the pilot design into two separated problems with some theoretical analyses. Specifically, the pilot values are designed to minimize the sub-coherence of the measurement matrix, and the pilot positions are designed to minimize the block-coherence of the measurement matrix.

Simulation results demonstrate that our developed algorithms can effectively recover NCT positions, and verify that the proposed estimator can achieve higher channel estimation accuracy than conventional schemes for massive MIMO-OFDM systems over time-varying channels.

The remainder of this paper is organized as follows. Section II introduces the background, including the system model and the channel model for the time-varying massive MIMO system. The formulation of the channel estimation problem is presented in Section III. Section IV describes the proposed pilot design scheme and channel estimation algorithms, where the complexity analysis is also discussed. In Section V, simulation results are provided to demonstrate the superior performance of our proposed scheme. Finally, some concluding remarks are given in Section VI.

Notations: For a given matrix \mathbf{A} , \mathbf{A}^T , \mathbf{A}^* , \mathbf{A}^H , \mathbf{A}^\dagger and $\|\mathbf{A}\|_F$ denote its transpose, conjugate, conjugate transpose, pseudo inverse, and Frobenius norm, respectively. $[\mathbf{A}]_{k,n}$ is the (k, n) -th entry of \mathbf{A} . $[\mathbf{A}]_{\mathcal{P},:}$ denotes the sub-matrix of \mathbf{A} with rows indexed by \mathcal{P} , and $[\mathbf{A}]_{:, \mathcal{P}}$ denotes the sub-matrix of \mathbf{A} with columns indexed by \mathcal{P} . $[\mathbf{a}]_{\mathcal{P}}$ denotes the sub-vector of \mathbf{a} with elements indexed by \mathcal{P} . $\|\mathbf{a}\|_2$ is the ℓ_2 norm. $\text{diag}(\mathbf{a})$ denotes a diagonal matrix with \mathbf{a} on its main diagonal. \mathbf{I}_N is an $N \times N$ identity matrix, and $\mathbf{0}_{M \times N}$ denotes the $M \times N$ all-zero matrix. $\mathbb{C}^{M \times N}$ represents the set of $M \times N$ matrices in the complex field. $\mathcal{CN}(\boldsymbol{\alpha}, \mathbf{R})$ is a Gaussian random vector with mean $\boldsymbol{\alpha}$ and covariance \mathbf{R} . $|\mathcal{P}|$ denotes the cardinality

of the set \mathcal{P} . $E(z)$ represents the average of z . \otimes denotes the Kronecker product.

II. BACKGROUND

A. MIMO-OFDM SYSTEM MODEL

Let us consider a downlink massive MIMO system with cyclic prefix (CP)-OFDM transmission, where the BS is equipped with N_t antennas and the UE is equipped with a single antenna. The CP is set to be long enough to avoid the inter-symbol interference (ISI). Considering N subcarriers for each OFDM symbol, we denote $\mathbf{x}^{n_t} = (x^{n_t}[1], \dots, x^{n_t}[N])^T$ as the transmitted signal of the n_t -th ($1 \leq n_t \leq N_t$) transmit antenna. After being sent over the channel, the signal received at the UE can be expressed as

$$\mathbf{y} = \sum_{n_t=1}^{N_t} \underbrace{\mathbf{F} \mathbf{H}_T^{n_t} \mathbf{F}^H}_{\mathbf{H}_F^{n_t}} \mathbf{x}^{n_t} + \mathbf{n}, \quad (1)$$

where $\mathbf{y} = (y[1], \dots, y[N])^T \in \mathbb{C}^{N \times 1}$, \mathbf{F} denotes an N -point normalized discrete Fourier transform (DFT) matrix with $[\mathbf{F}]_{n,m} = \frac{1}{\sqrt{N}} \exp(-j2\pi nm/N)$, $\mathbf{H}_T^{n_t} \in \mathbb{C}^{N \times N}$ and $\mathbf{H}_F^{n_t} \in \mathbb{C}^{N \times N}$, respectively, denote the time domain and frequency domain channel matrix from the n_t -th antenna at the BS to the UE, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{N \times 1}, \sigma_n^2 \mathbf{I}_N)$ is the additive noise vector.

The (n, m) -th entry of $\mathbf{H}_T^{n_t}$ can be denoted as

$$[\mathbf{H}_T^{n_t}]_{n,m} = h_{n,\text{mod}(n-m,N)}^{n_t}, \quad 1 \leq n \leq N, \quad 1 \leq m \leq N, \quad (2)$$

where $h_{n,l}^{n_t}$ denotes the downlink channel coefficient at the n -th ($1 \leq n \leq N$) instant of the l -th ($1 \leq l \leq L$) channel tap, $L = \lceil \tau_{\max}/T_s \rceil$ is the length of the channel, with τ_{\max} denoting the maximum delay and T_s denoting the sampling period. Note that $h_{n,l}^{n_t} = 0$ for $L < l \leq N$.

In (1), the number of channel coefficients to be estimated is LNN_t , which is typically very large for wideband massive MIMO systems. Thus a large number of pilots are required to obtain the CSI. In the literature, the complex exponential BEM (CE-BEM) has been widely used to model the time-varying channel, which is able to represent the channels with fewer coefficients. By utilizing the CE-BEM, the l -th channel tap of the n_t -th transmit antenna can be expressed as

$$\mathbf{h}_l^{n_t} = \underbrace{(\mathbf{b}_1, \dots, \mathbf{b}_Q)}_{\mathbf{B}_{\text{CE}}} \begin{pmatrix} c_{1,l}^{n_t} \\ \vdots \\ c_{Q,l}^{n_t} \end{pmatrix} + \boldsymbol{\xi}_l^{n_t}, \quad (3)$$

where $\mathbf{h}_l^{n_t} \triangleq (h_{1,l}^{n_t}, \dots, h_{N,l}^{n_t})^T \in \mathbb{C}^{N \times 1}$, Q ($Q \ll N$) denotes the CE-BEM order, $\mathbf{B}_{\text{CE}} = (\mathbf{b}_1, \dots, \mathbf{b}_Q) \in \mathbb{C}^{N \times Q}$ contains Q orthonormal bases, $c_{q,l}^{n_t}$ represents the corresponding CE-BEM coefficient, and $\boldsymbol{\xi}_l^{n_t} \in \mathbb{C}^{N \times 1}$ denotes the CE-BEM modeling error. To be specific, the q -th basis \mathbf{b}_q can be expressed as

$$\mathbf{b}_q = \left(1, \dots, e^{j\frac{2\pi}{N}n\left(q-\frac{Q+1}{2}\right)}, \dots, e^{j\frac{2\pi}{N}(N-1)\left(q-\frac{Q+1}{2}\right)} \right)^T. \quad (4)$$

Substituting (3) into (1), we can rewrite the received signal as [16]

$$\mathbf{y} = \sum_{n_t=1}^{N_t} \left(\sum_{q=1}^Q \mathbf{I}_N^{(q-\frac{Q+1}{2})} \text{diag}(\mathcal{V} \mathbf{c}_q^{n_t}) \right) \mathbf{x}^{n_t} + \mathbf{z}, \quad (5)$$

where $\mathbf{I}_N^{(q)} \in \mathbb{C}^{N \times N}$ denotes a permutation matrix obtained from \mathbf{I}_N by shifting its column circularly $|q|$ -times to the left if $q > 0$ and to the right otherwise, $\mathcal{V} \in \mathbb{C}^{N \times L}$ is a sub-matrix consisting of the first L columns of $\sqrt{N} \mathbf{F}$, $\mathbf{c}_q^{n_t} = (c_{q,1}^{n_t}, \dots, c_{q,L}^{n_t})^T \in \mathbb{C}^{L \times 1}$, and \mathbf{z} includes the additive noise and the CE-BEM modeling error. The derivation of (5) can be referred to the Appendix.

In this paper, we focus on pilot-aided channel estimation, where M effective pilots are distributed among N subcarriers. In order to reduce the pilot overhead for the massive MIMO system, all antennas are set to share the same pilot positions. We denote the effective pilot index set as $\mathcal{P}_{\text{eff}} = \{p_1, \dots, p_M\} \subset \{1, \dots, N\}$, with $|\mathcal{P}_{\text{eff}}| = M$. By the assumption that Q is an odd number and the number of guard pilots before and after each effective pilot is $Q - 1$, the received signals can be expressed as [16]

$$[\mathbf{y}]_{\mathcal{P}_q} = \sum_{n_t=1}^{N_t} \text{diag}([\mathbf{x}^{n_t}]_{\mathcal{P}_{\text{eff}}}) \mathbf{U} \mathbf{c}_q^{n_t} + [\mathbf{z}]_{\mathcal{P}_q}, \quad 1 \leq q \leq Q. \quad (6)$$

where $\mathbf{U} = [\mathcal{V}]_{\mathcal{P}_{\text{eff}}} \in \mathbb{C}^{M \times L}$, and the set \mathcal{P}_q is defined as

$$\mathcal{P}_q = \mathcal{P}_{\text{eff}} - \left(\frac{Q+1}{2} - q \right), \quad 1 \leq q \leq Q. \quad (7)$$

To be compact, (6) is further rewritten in a matrix form as (8), as shown at the bottom of the next page, where $\mathbf{R} = ([\mathbf{y}]_{\mathcal{P}_1}, \dots, [\mathbf{y}]_{\mathcal{P}_Q}) \in \mathbb{C}^{M \times Q}$, and $\mathbf{Z} = ([\mathbf{z}]_{\mathcal{P}_1}, \dots, [\mathbf{z}]_{\mathcal{P}_Q}) \in \mathbb{C}^{M \times Q}$.

Remark 1: The total number of pilot subcarriers of each OFDM symbol is $P = (2Q - 1)M$, which is independent of the number of transmit antennas.

B. CHANNEL MODEL

The parametric channel model is adopted in this paper, where the channel matrix is expressed as a function of spatial parameters [24]–[26]. As shown in Fig. 1, the propagation from the BS to the UE is assumed to be composed of K active clusters, and each cluster consists of some rays. Since antennas at the BS are closely-spaced in typical massive MIMO systems, the channels between all transmit antennas and the UE share the common scatterers [27]. Generally, we assume each active cluster associates with one specific channel tap, and delay variations of different rays corresponding to the same cluster are indistinguishable [26].¹ Under this assumption, there are K NCTs and the corresponding index set can be denoted as $\mathcal{L} = \{l_1, \dots, l_k, \dots, l_K\}$ with $1 \leq l_k \leq L$.

¹Note that for the case where some rays of a cluster corresponds to different channel taps, our proposed channel estimation scheme can be also applied by dividing this cluster into several sub-clusters.

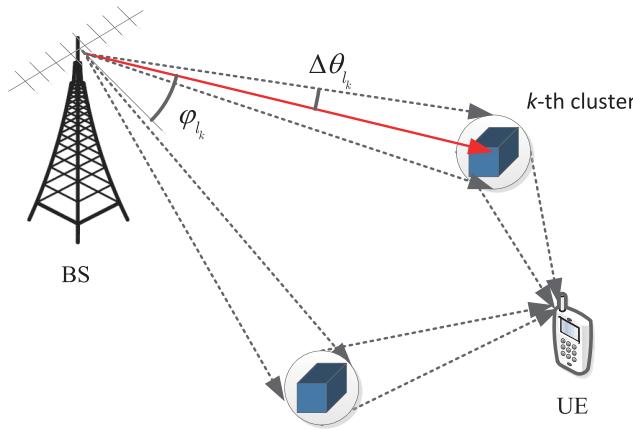


FIGURE 1. Downlink massive MIMO transmission channel model with large number of antennas at the BS and a single antenna at each UE.

Assuming that clusters are uncorrelated, we focus on modeling the l_k -th ($l_k \in \mathcal{L}$) channel tap $\mathbf{h}_{n,l_k} = (h_{n,l_k}^1, \dots, h_{n,l_k}^{N_t}) \in \mathbb{C}^{1 \times N_t}$, which can be expressed as

$$\mathbf{h}_{n,l_k} = \sqrt{\frac{N_t}{KR_{l_k}}} \sum_{r=1}^{R_{l_k}} \alpha_{l_k}^r e^{j2\pi v_{l_k}^r (n-1)T_s} \mathbf{a}^H(\theta_{l_k}^r), \quad (9)$$

where R_{l_k} is the number of rays of the k -th active cluster, $\alpha_{l_k}^r$ denotes the complex gain of the r -th ray, $v_{l_k}^r$ is the Doppler shift, $\theta_{l_k}^r$ is the AoD, and $\mathbf{a}(\theta_{l_k}^r) \in \mathbb{C}^{N_t \times 1}$ represents the array response vector. For the uniform linear array (ULA) at the BS, $\mathbf{a}(\theta)$ is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N_t}} \left(1, e^{-j\frac{2\pi}{\lambda} d \sin(\theta)}, \dots, e^{-j(N_t-1)\frac{2\pi}{\lambda} d \sin(\theta)} \right)^T, \quad (10)$$

where λ is the signal wavelength, and d is the distance between neighboring antenna elements, which usually satisfies $d = \frac{\lambda}{2}$.

From Fig. 1, each AoD of the k -th cluster can be expressed as

$$\theta_{l_k}^r = \varphi_{l_k} + \tilde{\theta}_{l_k}^r, \quad 1 \leq r \leq R_{l_k}, \quad (11)$$

where φ_{l_k} is the central AoD of the k -th cluster, $\tilde{\theta}_{l_k}^r$ denotes the relative AoD shift from the central AoD, which satisfies $\tilde{\theta}_{l_k}^r \in [-\Delta\theta_{l_k}/2, \Delta\theta_{l_k}/2]$, with $\Delta\theta_{l_k}$ denoting the double-side angle spread of the k -th cluster. Note that $\Delta\theta_{l_k}$ is usually limited in a narrow region due to the limited scattering around the BS [5], [24].

III. PROBLEM FORMULATION

As the angle spread of each cluster is small, vectors $\{\mathbf{a}(\theta_{l_k}^r)\}_{r=1}^{R_{l_k}}$ would be highly correlated. The SBEM was built

in [5] to capture the spatial correlation and reduce the effective dimensions of the massive MIMO channel with limited orthogonal bases. And it has been verified that by utilizing N_t -point DFT matrix to model the channel, most channel power concentrates on a few DFT entries. However, the resolution of AoDs is limited for SBEM when finite number of transmit antennas are deployed in practical systems, which may cause large error in channel estimation.

Derived from SBEM, we propose a generalized-SBEM (G-SBEM) to represent the channels with several non-orthogonal bases. Specifically, the l -th ($l \in \mathcal{L}$) channel tap in (9) can be expressed by $\gamma_l < N_t$ bases as

$$\mathbf{h}_{n,l} = \sum_{i=1}^{\gamma_l} \tilde{h}_{n,l}^i \mathbf{f}^H(i), \quad (12)$$

where γ_l is the G-SBEM order, $\mathbf{f}(i) \in \mathbb{C}^{N_t \times 1}$ is the i -th ($1 \leq i \leq \gamma_l$) basis, and $\tilde{h}_{n,l}^i$ denotes the corresponding G-SBEM coefficient. Here, $\{\mathbf{f}(i)\}_{i=1}^{\gamma_l}$ are no longer limited to be orthogonal to each other, which makes it more flexible to design bases.

To generate $\mathbf{f}(i)$, we first design the set of quantized AoDs as

$$\Gamma = \{\bar{\theta}_g : \sin \bar{\theta}_g = \frac{2}{G_t}(g-1) - 1, g = 1, \dots, G_t\}, \quad (13)$$

where $\bar{\theta}_g$ denotes the quantized AoDs, and G_t ($G_t \geq N_t$) is the number of grid points. Here, we assume no knowledge of AoDs region and $\{\sin \bar{\theta}_g\}$ is designed to be uniformly distributed in $[-1, 1]$. Collecting all possible array response vectors corresponding to $\bar{\theta}_g$, we define the dictionary matrix $\mathcal{A} \triangleq (\mathbf{a}(\bar{\theta}_1), \dots, \mathbf{a}(\bar{\theta}_{G_t})) \in \mathbb{C}^{N_t \times G_t}$. Specifically, the g -th column of \mathcal{A} can be expressed as

$$[\mathcal{A}]_{:,g} = \frac{1}{\sqrt{N_t}} (1, e^{-j2\pi(\frac{g-1}{G_t}-\frac{1}{2})}, \dots, e^{-j2\pi(N_t-1)(\frac{g-1}{G_t}-\frac{1}{2})})^T. \quad (14)$$

By comparing (9) with (12), it tends to be effectively to set the G-SBEM bases as forms of array response vectors. Thus, γ_l columns of \mathcal{A} are chosen as the G-SBEM bases, denoted as

$$(\mathbf{f}(1), \dots, \mathbf{f}(\gamma_l)) = [\mathcal{A}]_{:, \mathcal{B}_l}, \quad (15)$$

where \mathcal{B}_l is an index set with $|\mathcal{B}_l| = \gamma_l$. By substituting (15) into (12), we obtain

$$\mathbf{h}_{n,l} = [\tilde{\mathbf{h}}_{n,l}]_{:, \mathcal{B}_l} [\mathcal{A}^H]_{\mathcal{B}_l,:} = \tilde{\mathbf{h}}_{n,l} \mathcal{A}^H, \quad (16)$$

where $\tilde{\mathbf{h}}_{n,l} \in \mathbb{C}^{1 \times G_t}$ is a γ_l -sparse vector, containing γ_l nonzero entries $\{\tilde{h}_{n,l}^i\}_{i=1}^{\gamma_l}$. Within one OFDM symbol, it is

$$\mathbf{R} = \left(\text{diag}([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}) \mathbf{U}, \dots, \text{diag}([\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}}) \mathbf{U} \right) \begin{pmatrix} \mathbf{c}_1^1 & \cdots & \mathbf{c}_Q^1 \\ \vdots & \ddots & \vdots \\ \mathbf{c}_1^{N_t} & \cdots & \mathbf{c}_Q^{N_t} \end{pmatrix} + \mathbf{Z} \quad (8)$$

safe to assume that \mathcal{B}_l is unaltered, due to the fact that AoDs depend on the large scale properties of the scattering environment and vary slowly [27]. Hence, the channel matrix of the l -th channel tap can be given by

$$\mathbf{H}_l = [\tilde{\mathbf{H}}_l]_{:, \mathcal{B}_l} [\mathcal{A}^H]_{\mathcal{B}_l, :} = \tilde{\mathbf{H}}_l \mathcal{A}^H, \quad (17)$$

where $\mathbf{H}_l = (\mathbf{h}_{1,l}^T, \dots, \mathbf{h}_{N_t,l}^T)^T \in \mathbb{C}^{N_t \times N_t}$, and $\tilde{\mathbf{H}}_l = (\tilde{\mathbf{h}}_{1,l}^T, \dots, \tilde{\mathbf{h}}_{N_t,l}^T)^T \in \mathbb{C}^{N_t \times G_t}$, including γ_l nonzero columns.

To finish the problem formulation, we define the CE-BEM coefficient matrix $\mathbf{C} \in \mathbb{C}^{LN_t \times Q}$ as

$$\mathbf{C} \triangleq \begin{pmatrix} c_{1,1}^1 & \cdots & c_{1,1}^{N_t} & \cdots & c_{1,L}^1 & \cdots & c_{1,L}^{N_t} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{Q,1}^1 & \cdots & c_{Q,1}^{N_t} & \cdots & c_{Q,L}^1 & \cdots & c_{Q,L}^{N_t} \end{pmatrix}^T. \quad (18)$$

Then, recalling (3), the matrix \mathbf{C} can be expressed in terms of the G-SBEM as

$$\begin{aligned} \mathbf{C} &= (\mathbf{B}_{CE}^H \mathbf{H}_1, \dots, \mathbf{B}_{CE}^H \mathbf{H}_L)^T \\ &= (\mathbf{B}_{CE}^H \tilde{\mathbf{H}}_1 \mathcal{A}^H, \dots, \mathbf{B}_{CE}^H \tilde{\mathbf{H}}_L \mathcal{A}^H)^T \\ &= (\mathbf{I}_L \otimes \mathcal{A}^*) \begin{pmatrix} \tilde{\mathbf{H}}_1^T \mathbf{B}_{CE}^* \\ \vdots \\ \tilde{\mathbf{H}}_L^T \mathbf{B}_{CE}^* \end{pmatrix}. \end{aligned} \quad (19)$$

Substituting (19) into (8), we finally formulate the channel estimation problem as (20), as shown at the bottom of the this page, where $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_L] \in \mathbb{C}^{LN_t \times LN_t}$ denotes a permutation matrix with $\mathbf{P}_i = [\mathbf{e}_1, \mathbf{e}_{L+1}, \dots, \mathbf{e}_{(N_t-1)L+1}]$, \mathbf{e}_k is the k -th column of the identity matrix \mathbf{I}_{LN_t} , $\Psi \in \mathbb{C}^{M \times LN_t}$, $\mathbf{S} = [\mathbf{S}_1^T, \dots, \mathbf{S}_L^T]^T \in \mathbb{C}^{LG_t \times Q}$ with

$$\mathbf{S}_l \triangleq \tilde{\mathbf{H}}_l^T \mathbf{B}_{CE}^* \in \mathbb{C}^{G_t \times Q}. \quad (21)$$

In (20), we denote $\Phi \triangleq \Psi(\mathbf{I}_L \otimes \mathcal{A}^*) \in \mathbb{C}^{M \times LG_t}$ as the measurement matrix.

Given (20), instead of estimating numerous channel taps $\{\mathbf{H}_l\}_{l=1}^L$, the task of channel estimation is transformed into identifying the coefficient matrix \mathbf{S} . A natural approach is to use the least square (LS) method, which requires the number of measurements $M \geq LG_t$. However, for the wideband massive MIMO systems, LG_t is usually a large integer such that the LS method requires tremendous pilot overhead. Alternatively, the sparsity of \mathbf{S} motivates us to apply the CS technique to solve the underdetermined problem when $M < LG_t$. To illustrate the sparsity of \mathbf{S} , we give the following theorem

Theorem 1: For the formulated problem in (20), \mathbf{S} is a quasi-block-sparse matrix with no more than $\sum_{l \in \mathcal{L}} \gamma_l$ nonzero rows.

Proof: Based on our previous analysis, we have

$$[\tilde{\mathbf{H}}_l]_{:,g} = 0, \text{ for } l \notin \mathcal{L} \text{ or } g \notin \mathcal{B}_l. \quad (22)$$

Then, according to (21) we have

$$[\mathbf{S}_l]_{g,:} = 0, \text{ for } l \notin \mathcal{L} \text{ or } g \notin \mathcal{B}_l. \quad (23)$$

Thus the number of nonzero rows in \mathbf{S} is less than $\sum_{l \in \mathcal{L}} |\mathcal{B}_l| = \sum_{l \in \mathcal{L}} \gamma_l$. Note that nonzero rows in \mathbf{S} appears in the sub-matrix \mathbf{S}_l ($l \in \mathcal{L}$), but may not be continuously distributed. This feature is referred as quasi-block sparsity in this paper. For better clarification, we show the quasi-block-sparse structure of \mathbf{S} in Fig. 2, where the gray parts denote nonzero blocks and the white parts denote zero blocks, and the black dots in gray parts denote nonzero entries and white dots represent zero entries. ■

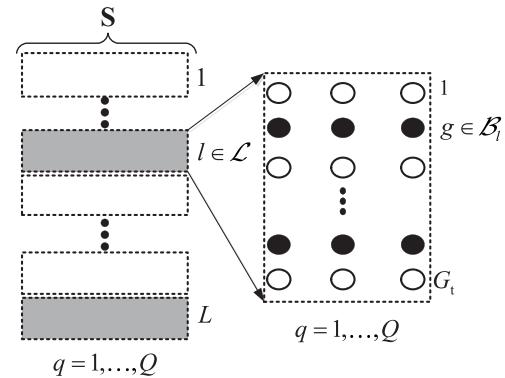


FIGURE 2. Quasi-block-sparse structure of the coefficient matrix \mathbf{S} to be estimated.

Conventional CS methods, such as the OMP algorithm [28] and SOMP algorithm [14], can be applied to recover the sparse matrix \mathbf{S} . However, it is of great interest to develop novel recovery algorithms consistent with the quasi-block sparsity to improve the channel estimation accuracy.

IV. PROPOSED ALGORITHMS

In this section, two novel algorithms based on CS theory are developed to recover the quasi-block-sparse coefficient matrix. Before introducing recovery algorithms, we discuss the pilot design and formulate it into two separated problems, including designing the pilot positions and pilot values.

$$\begin{aligned} \mathbf{R} &= \underbrace{\left(\text{diag}([\mathbf{x}^1]_{\mathcal{P}_{eff}}) \mathbf{U}, \dots, \text{diag}([\mathbf{x}^{N_t}]_{\mathcal{P}_{eff}}) \mathbf{U} \right)}_{\Psi} \mathbf{P} \mathbf{P}^T \underbrace{\begin{pmatrix} \mathbf{c}_1^1 & \cdots & \mathbf{c}_Q^1 \\ \vdots & \ddots & \vdots \\ \mathbf{c}_1^{N_t} & \cdots & \mathbf{c}_Q^{N_t} \end{pmatrix}}_{\mathbf{C}} + \mathbf{Z} \\ &= \Psi(\mathbf{I}_L \otimes \mathcal{A}^*) \mathbf{S} + \mathbf{Z} \end{aligned} \quad (20)$$

A. PILOT DESIGN SCHEME

To design the pilot pattern, there are two issues, listed as below

- 1) Design the effective pilot positions, i.e., $\mathcal{P}_{\text{eff}} = \{p_1, \dots, p_M\} \subset \{1, \dots, N\}$.
- 2) Design the effective pilot values, i.e., $\{[\mathbf{x}^{n_t}]_{\mathcal{P}_{\text{eff}}}\}_{n_t=1}^{N_t}$.

These two issues will be solved by investigating the property of the measurement matrix, as discussed in the following part.

For the formulated problem in (20), conventional CS theory has proved that the sparse matrix \mathbf{S} can be exactly recovered so long as the measurement matrix Φ satisfies the restricted isometry property (R.I.P.). However, no existing method can check this property during polynomial time. Alternatively, the coherence of the measurement matrix is commonly used to analyze the recovery accuracy [14]. The coherence of Φ is defined as

$$\mu(\Phi) = \max_{1 \leq i \neq j \leq LG_t} \frac{|\phi_i^H \phi_j|}{\|\phi_i\|_2 \|\phi_j\|_2}, \quad (24)$$

where ϕ_i is the i -th column of Φ . To achieve high recovery probability, the measurement matrix with low coherence is desired [14].

Inspired by the quasi-block-sparse structure of \mathbf{S} , we represent the measurement matrix Φ as a concatenation of L measurement sub-matrices, expressed as

$$\Phi = (\underbrace{\phi_1, \dots, \phi_{G_t}}_{\Phi_1}, \dots, \underbrace{\phi_{(L-1)G_t+1}, \dots, \phi_{LG_t}}_{\Phi_L}), \quad (25)$$

where $\Phi_l \in \mathbb{C}^{M \times G_t}$ denotes the l -th measurement sub-matrix. And we define the block-coherence of Φ according to [29] as

$$\mu_B(\Phi) = \max_{1 \leq i \neq j \leq L} \frac{\rho(\Phi_i^H \Phi_j)}{\|\Phi_i\|_2 \|\Phi_j\|_2}, \quad (26)$$

where $\rho(\mathbf{A})$ denotes the spectrum norm of \mathbf{A} , expressed as $\rho(\mathbf{A}) = \lambda_{\max}^{1/2}(\mathbf{A}^H \mathbf{A})$, with $\lambda_{\max}(\cdot)$ denoting the largest eigenvalue. While $\mu_B(\Phi)$ quantifies the global property of the sensing matrix Φ , the local property is characterized by the sub-coherence of Φ [30], defined as

$$\vartheta(\Phi) = \max_l \max_{i \neq j} \frac{|\phi_i^H \phi_j|}{\|\phi_i\|_2 \|\phi_j\|_2}, \quad \phi_i, \phi_j \in \Phi_l \\ = \max_l \mu(\Phi_l), \quad (27)$$

where $\mu(\Phi_l)$ is the coherence of the sub-matrix Φ_l .

Remark 2: The coherence $\mu(\Phi)$ is usually used to analyze the recovery probability of conventional CS problems. However, for the formulated CS problem with a quasi-block-sparse matrix to be recovered, we utilize the block-coherence $\mu_B(\Phi)$ to identify the cross-correlation between two measurement sub-matrices, and the sub-coherence $\vartheta(\Phi)$ to identify the cross-correlation between two columns of each measurement sub-matrix.

Now, we are going to design \mathcal{P}_{eff} and $\{[\mathbf{x}^{n_t}]_{\mathcal{P}_{\text{eff}}}\}_{n_t=1}^{N_t}$ by simultaneously minimizing $\mu_B(\Phi)$ and $\vartheta(\Phi)$. And the pilot

design can be stated as a multiobjective optimization problem, expressed as

$$\begin{aligned} & \arg \min_{\{\mathcal{P}_{\text{eff}}, [\mathbf{x}^{n_t}]_{\mathcal{P}_{\text{eff}}}\}} \{\min \mu_B(\Phi), \min \vartheta(\Phi)\} \\ & \text{s.t. } |k_i - k_j| \geq 2Q - 1, \quad \forall i, j, i \neq j \\ & |\mathbf{x}^{n_t}(p_m)| = 1, \quad 1 \leq m \leq M, \end{aligned} \quad (28)$$

where the first constraint is to guarantee $(2Q - 2)$ guard pilots around each effective pilot, and the second constraint is for the pilot power limitation. However, this problem is difficult to solve due to two optimization objectives. To simplify the pilot design problem, we first introduce the following theorem

Theorem 2: For the proposed model, the sub-coherence of the measurement matrix Φ is equal to the coherence of the measurement sub-matrix Φ_l ($1 \leq l \leq L$), i.e.,

$$\vartheta(\Phi) = \mu(\Phi_1) = \dots = \mu(\Phi_L). \quad (29)$$

Proof: Recalling (20), the l -th measurement sub-matrix can be expressed as

$$\begin{aligned} \Phi_l &= \left([\text{diag}([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}) \mathbf{U}]_{:,l}, \dots, [\text{diag}([\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}}) \mathbf{U}]_{:,l} \right) \mathcal{A}^* \\ &= \text{diag}([\mathbf{U}]_{:,l}) \left([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}, \dots, [\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}} \right) \mathcal{A}^*. \end{aligned} \quad (30)$$

For better clarification, we give \mathbf{U} as

$$\mathbf{U} = \begin{pmatrix} 1 & w^{p_1-1} & \dots & w^{(p_1-1)(L-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{p_M-1} & \dots & w^{(p_M-1)(L-1)} \end{pmatrix}, \quad (31)$$

with $w \triangleq \exp(-i\frac{2\pi}{N})$. According to (30) and (31), we obtain

$$\begin{cases} \Phi_l = ([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}, \dots, [\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}}) \mathcal{A}^*, & l = 1 \\ \Phi_l = \text{diag}([\mathbf{U}]_{:,l}) \Phi_1, & 2 \leq l \leq L. \end{cases} \quad (32)$$

Then, the coherence of the l -th ($2 \leq l \leq L$) measurement sub-matrix can be derived as

$$\begin{aligned} \mu(\Phi_l) &= \max_{1 \leq i \neq j \leq G_t} \left| (\text{diag}([\mathbf{U}]_{:,l}) \phi_i)^H \text{diag}([\mathbf{U}]_{:,l}) \phi_j \right| \\ &= \max_{1 \leq i \neq j \leq G_t} \left| \phi_i^H \phi_j \right| \\ &= \mu(\Phi_1), \end{aligned} \quad (33)$$

where ϕ_i and ϕ_j ($1 \leq i, j \leq G_t$) are two arbitrary columns of Φ_1 . By combining (27) and (33), we can finally obtain (29). ■

According to **Theorem 2**, we have

$$\vartheta(\Phi) = \mu \left(([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}, \dots, [\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}}) \mathcal{A}^* \right), \quad (34)$$

which indicates that $\vartheta(\Phi)$ is independent of the pilot positions and merely depends on the pilot values. Therefore, to simplify the pilot design, we consider the following two separated

problems to give a near-optimal solution to the problem (28), expressed as

$$\begin{aligned} \arg \min_{\{\mathbf{x}^{n_t}\}_{\mathcal{P}_{\text{eff}}}^{N_t}} & \vartheta(\Phi) \\ \text{s.t. } & |\mathbf{x}^{n_t}(p_m)| = 1, \quad 1 \leq m \leq M, \end{aligned} \quad (35)$$

and

$$\begin{aligned} \arg \min_{\mathcal{P}_{\text{eff}}} & \mu_B(\Phi) \\ \text{s.t. } & |k_i - k_j| \geq 2Q - 1, \quad \forall i, j, i \neq j. \end{aligned} \quad (36)$$

Namely, we first design the pilot values according to (35), and then based on the designed pilot values, the pilot positions are designed according to (36). The procedures are as follows:

- To solve the problem in (35), we set $[\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}} = [1, 1, \dots, 1]$ and the pilot values of the n_t -th ($2 \leq n_t \leq N_t$) antenna as [31]

$$\mathbf{x}^{n_t}(p_m) = \mathbf{x}^1(p_m) e^{-j \frac{2\pi}{N} (p_m - 1) \sigma_{n_t}}, \quad m = 1, \dots, M, \quad (37)$$

where σ_{n_t} is an integer to be optimized. Then we can minimize $\vartheta(\Phi)$ only by tuning σ_{n_t} , which can be solved by the methods in [31]. For convenience, here we consider the transmit pilot signal $([\mathbf{x}^1]_{\mathcal{P}_{\text{eff}}}, \dots, [\mathbf{x}^{N_t}]_{\mathcal{P}_{\text{eff}}})$ as a sub-matrix consists of N_t columns of M -point DFT matrix.

- Based on the designed pilot values, we are able to utilize the classical discrete stochastic optimization (DSO) technique [14] to solve the problem in (36), where the only change should be made is that the optimization function in our problem is the block-coherence $\mu_B(\Phi)$, rather than the coherence $\mu(\Phi)$ in [14].

B. CHANNEL ESTIMATION ALGORITHM

Two novel recovery algorithms are introduced in this subsection, consistent with the quasi-block sparsity of the formulated problem. The proposed QBSO algorithm is listed in **Algorithm 1** including two stages, where we seek all NCT positions at the first stage, and calculate the quasi-block-sparse matrix \mathbf{S} at the second stage. And the proposed adaptive-QBSO algorithm is listed in **Algorithm 2**, which can adaptively design the measurement based on the estimated virtual AoD. Here, we assume the number of clusters K is a priori knowledge [5].²

We first introduce the QBSO algorithm. In the i -th iteration of stage 1, we search the index $g_i \in \{1, \dots, LG_t\}$ with the corresponding normalized $[\Phi]_{:,g_i}$ being best matched to the residual \mathbf{r} . This greedy procedure guarantees that $\lceil g_i/G_t \rceil$ indicates the NCT position (i.e., nonzero-block position) with high probability. Once finding an NCT position, we update Ω_B by adding $\lceil g_i/G_t \rceil$, and then set $g_k^{\text{opt}} = g_i$ as the optimal

²In our proposed algorithm, we assume the number of clusters K is a priori knowledge. As for unknown K , we can utilize the statistical bound on the number of clusters, which is a result of the large-scale scattering environment and easy to obtain [21].

Algorithm 1 Proposed QBSO Algorithm

Input: $\mathbf{R}, \Phi = [\Phi_1, \Phi_2, \dots, \Phi_L], K, \gamma_l, \eta_0$.

Output: The estimated coefficient matrix \mathbf{S} .

Stage 1: Find the nonzero channel tap positions.

- 1: Initialize $\mathbf{r} = \mathbf{R}$, $\Omega = \emptyset$, $\Omega_B = \emptyset$, $\Theta = \emptyset$.
- 2: **for** $i = 1$ to $\sum_{l \in \mathcal{L}} \gamma_l$ **do**
- 3: $g_i \leftarrow \arg \max_{g \in \{1, \dots, LG_t\}} \|\mathbf{r}^H [\Phi]_{:,g}\|_2 / \|[\Phi]_{:,g}\|_2$.
- 4: $\Omega \leftarrow \Omega \cup \{g_i\}$, $\Theta \leftarrow \Theta \cup [\Phi]_{:,g_i}$.
- 5: **if** $|\Omega_B| < K$ and $\lceil g_i/G_t \rceil \notin \Omega_B$ **then**
- 6: $g_k^{\text{opt}} \leftarrow g_i$, $\hat{l}_k \leftarrow \lceil g_i/G_t \rceil$, $\Omega_B \leftarrow \Omega_B \cup \{\hat{l}_k\}$.
- 7: **else if** $|\Omega_B| = K$ **then**
- 8: Break.
- 9: **end if**
- 10: Update the residual as $\mathbf{r} \leftarrow \mathbf{R} - \Theta \Theta^\dagger \mathbf{R}$.
- 11: **end for**
- 12: Update the measurement matrix $\Phi \leftarrow (\Phi_{\hat{l}_1}, \dots, \Phi_{\hat{l}_K})$.

Stage 2: Calculate the coefficient matrix \mathbf{S} .

- 13: Initialize $\mathbf{S} = \mathbf{0}_{LG_t \times Q}$, $\mathbf{r} = \mathbf{R}$, $\Omega = \emptyset$, $\Theta = \emptyset$, $\Delta\eta = \infty$, $\eta^1 = 0$.
 - 14: **while** $j \leq \sum_{\hat{l}_k} \gamma_{\hat{l}_k}$ and $\Delta\eta > \eta_0$ **do**
 - 15: $g_j \leftarrow \arg \max_{g \in \{1, \dots, KG_t\}} \|\mathbf{r}^H [\Phi]_{:,g}\|_2 / \|[\Phi]_{:,g}\|_2$.
 - 16: Update $\Omega \leftarrow \Omega \cup \{g_j\}$, and $\Theta \leftarrow \Theta \cup [\Phi]_{:,g_j}$.
 - 17: Update the residual as $\mathbf{r} \leftarrow \mathbf{R} - \Theta \Theta^\dagger \mathbf{R}$.
 - 18: $\eta^{j+1} \leftarrow \|\mathbf{r}\|_F / \|\Theta \Theta^\dagger \mathbf{R}\|_F$, $\Delta\eta \leftarrow |\eta^{j+1} - \eta^j|$.
 - 19: $j \leftarrow j + 1$.
 - 20: **end while**
 - 21: $\Omega_0 = \{\hat{l}_k - 1\} * G_t + g : g = 1, 2, \dots, G_t, \hat{l}_k \in \Omega_B\}$
 - 22: Compute nonzero coefficients as $[[\mathbf{S}]_{\Omega_0,:}]_{\Omega,:,:} = \Theta^\dagger \mathbf{R}$.
-

index corresponding to the k -th cluster. Note that parameters g_k^{opt} and \hat{l}_k calculated in step 6 will be used in the adaptive-QBSO algorithm. The loop of stage 1 breaks out when K NCT positions are found out, i.e., $|\Omega_B| = K$. Then, in step 12 the measurement matrix is updated by deleting the measurement sub-matrices corresponding to zero blocks of \mathbf{S} . Since zero blocks of \mathbf{S} have no effect on \mathbf{R} , updating the measurement matrix will not change the results.

In stage 2 of the QBSO algorithm, the quasi-block-sparse matrix \mathbf{S} is calculated based on the updated measurement matrix. The nonzero-row positions of \mathbf{S} with the index Ω are searched from KG_t rows. This procedure will be stopped until it satisfies $j > \sum_{\hat{l}_k} \gamma_{\hat{l}_k}$ or $\Delta\eta \leq \eta_0$ with $\Delta\eta$ calculated in step 18. Finally, we obtain the nonzero coefficients of \mathbf{S} in step 22, and the coefficients out of the support vector Ω are set as 0.

Moreover, inspired by the narrow angle spread of each cluster, we propose a novel adaptive-QBSO algorithm, which can adaptively design the measurement matrix and thus improve the resolution of AoDs. In step 1, $\{g_k^{\text{opt}}\}_{k=1}^K$ and $\{\hat{l}_k\}_{k=1}^K$ are obtained by stage 1 of the QBSO algorithm. Then, the virtual AoD θ_k^{vir} is calculated in step 4, which is expected to locate in the practical AoDs region corresponding to the k -th cluster. Next, the estimated AoDs region is denoted as

Algorithm 2 Proposed Adaptive-QBSO Algorithm

Input: $\mathbf{R}, \Phi = [\Phi_1, \Phi_2, \dots, \Phi_L], \Psi, K, \gamma_l, \eta_0, \Delta\theta$.
Output: The estimated coefficient matrix \mathbf{S} .

- 1: Perform Stage 1 of **Algorithm 1** to obtain $\{g_k^{\text{opt}}\}_{k=1}^K$ and $\{\hat{l}_k\}_{k=1}^K$.
- 2: **for** $k = 1$ to K **do**
- 3: $w_k \leftarrow \text{mod}(g_k^{\text{opt}} - 1, G_t) + 1$.
- 4: $\theta_k^{\text{vir}} \leftarrow \arcsin(\frac{2}{G_t}(w_k - 1) - 1)$.
- 5: $\hat{\theta}_k^{\max} \leftarrow \theta_k^{\text{vir}} + \Delta\theta, \hat{\theta}_k^{\min} \leftarrow \theta_k^{\text{vir}} - \Delta\theta$.
- 6: Update Γ_k as (38).
- 7: Set $\mathcal{A}_k \leftarrow (\mathbf{a}(\bar{\theta}_1), \dots, \mathbf{a}(\bar{\theta}_{G_t})), \bar{\theta}_{g_t} \in \Gamma_k$.
- 8: $\mathcal{W}_k \leftarrow [\Psi]_{:, (\hat{l}_k-1)N_t+1:\hat{l}_k N_t} \mathcal{A}_k^*$.
- 9: **end for**
- 10: Update the measurement matrix $\Phi \leftarrow (\mathcal{W}_1, \dots, \mathcal{W}_K)$.
- 11: Perform Stage 2 of **Algorithm 1** to calculate the coefficient matrix \mathbf{S} .

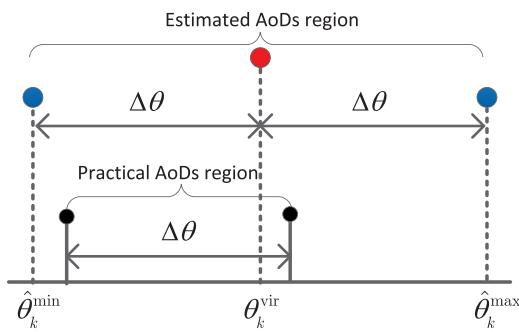


FIGURE 3. The estimated AoDs region covers the practical AoDs region when θ_k^{vir} locates in the practical AoDs region.

$[\hat{\theta}_k^{\min}, \hat{\theta}_k^{\max}] = [\theta_k^{\text{vir}} - \Delta\theta, \theta_k^{\text{vir}} + \Delta\theta]$. As shown in Fig. 3, the estimated AoDs region certainly covers the practical AoDs region when θ_k^{vir} locates in the practical AoDs region. Based on $[\hat{\theta}_k^{\min}, \hat{\theta}_k^{\max}]$, we update the set of quantized AoDs of the k -th cluster as

$$\Gamma_k = \{\bar{\theta}_g : \sin \bar{\theta}_g = (\sin \hat{\theta}_k^{\max} - \sin \hat{\theta}_k^{\min}) \frac{g-1}{G_t} + \sin \hat{\theta}_k^{\min}, g = 1, \dots, G_t\}, \quad (38)$$

where $\{\sin \bar{\theta}_g\}$ is designed to be uniformly distributed in $[\sin \hat{\theta}_k^{\min}, \sin \hat{\theta}_k^{\max}]$. Next, the k -th dictionary matrix \mathcal{A}_k is designed in step 7, and the new measurement matrix Φ is designed in step 10. Note that \mathbf{R} is the received pilots, which is independent of which measurement matrix we have designed. Once real AoDs locate in the assumed AoDs region, \mathbf{R} can be utilized to estimate \mathbf{S} . Based on the designed Φ , stage 2 of **Algorithm 1** can be further applied to calculate the coefficient matrix \mathbf{S} .

For conventional OMP and SOMP algorithms, the quasi-block sparsity of \mathbf{S} cannot be exploited and nonzero positions need to be searched from LG_t rows in one stage. Nevertheless, in the proposed QBSO algorithm, after we obtain NCT positions in stage 1, the number of unknown rows for stage 2 is

reduced to KG_t . We can expect the estimation accuracy to be improved when $K \ll L$. For the adaptive-QBSO algorithm, the AoDs of each cluster is restricted to a smaller region according to the virtual AoD, and then new measurement matrix is updated to improve the resolution of AoDs. Hence, the adaptive-QBSO algorithm is expected to further improve the channel estimation accuracy.

After estimating the sparse matrix \mathbf{S} based on the proposed algorithms, we can obtain the CE-BEM coefficients \mathbf{C} according to (19) and (21). Next, we apply the smoothing treatment via discrete prolate spheroidal sequences (DPSSs) [14] to the estimated CE-BEM coefficients. Finally, based on (3) channel vector $\mathbf{h}_l^{n_t}$ ($1 \leq n_t \leq N_t, 1 \leq l \leq L$) is calculated.

C. COMPLEXITY ANALYSIS

In this subsection, we briefly discuss the computational complexity of **Algorithm 1** and **Algorithm 2** in terms of complex multiplications. For simplicity, we denote γ as the mean value of $\{\gamma_l : l \in \mathcal{L}\}$.

For **Algorithm 1**, the main computational burden comes from step 3 and step 10 in stage 1, and step 15 and step 17 in stage 2. Note that the number of iterations in stage 1 is $\mathcal{O}(K)$ as it jumps out of the loop when $|\Omega_B| = K$. In step 3, the inner product between the residual \mathbf{r}^H and all vectors $[\Phi]_{:,g}, g \in \{1, \dots, LG_t\}$, has the complexity $\mathcal{O}(QMLG_t)$; in step 10, we perform the residual update with the complexity $\mathcal{O}(K^2M + KQM)$ when using the Gram-Schmidt algorithm. Similarly, the complexity is $\mathcal{O}(KQMG_t)$ in step 15 and $\mathcal{O}(\gamma^2 K^2M + \gamma KQM)$ in step 17. Thus, the total complexity of the **Algorithm 1** with K iterations in stage 1 and $\mathcal{O}(\gamma K)$ iterations in stage 2 is $\mathcal{O}(KQMLG_t + \gamma K^2QMG_t + \gamma^3 K^3M)$ due to $\gamma < G_t$. As a comparison, the SOMP algorithm [14] which only uses stage 2 has the complexity $\mathcal{O}(\gamma KLQMG_t + \gamma^3 K^3M)$.

Algorithm 2 contains both stage 1 and stage 2 of **Algorithm 1**, as shown in step 1 and step 11. Besides, the complexity is $\mathcal{O}(KMN_t G_t)$ in step 8 with K iterations. Thus, the total complexity of the **Algorithm 2** is $\mathcal{O}(KQMLG_t + \gamma K^2QMG_t + \gamma^3 K^3M + KMN_t G_t)$.

V. SIMULATION RESULTS

In this section, MATLAB simulations are carried out to compare the performance of our proposed algorithms with conventional estimators. We list the simulation parameters in **Table 1**. The K NCT positions are randomly distributed in $[1, L]$. Without loss of generality, we assume the number of rays of each active cluster is $R_l = 30$, and the G-SBEM order is $\gamma_l = 5$ for $l \in \mathcal{L}$. Note that we usually have $M = 4K\gamma_l$ to guarantee the recovery accuracy in practice [14], where M is the number of effective pilots that the system can afford. The central AoDs of the active clusters are assumed to take continuous values and uniformly distributed in $[-\pi/3, \pi/3]$. Besides, the threshold parameter η_0 for stop criterion in our algorithms is set as 10^{-3} . The variation of the channel is characterized by the normalized maximum

TABLE 1. Parameters of the simulation.

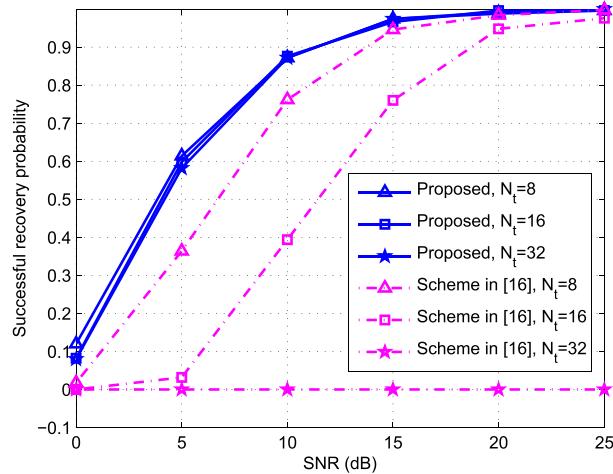
Parameters	Values
Number of subcarriers	$N = 2048$
Length of CIR	$L = 90$
Number of active clusters	$K = 4$
Number of effective pilots	$M = 80$
CE-BEM order	$Q = 3$
Subcarrier spacing	$\Delta f = 15 \text{ KHz}$
Carrier frequency	$f_c = 3 \text{ GHz}$
Modulation	QPSK

Doppler shift (NMDS), calculated as $v_M = f_c v / (c \Delta f)$, with v denoting the velocity of the UE, and c denoting the speed of the light. Unless otherwise mentioned, we set $N_t = 32$, $G_t = 2N_t$, $v = 350 \text{ km/h}$ (the corresponding NMDS $v_M = 0.065$), and the double-side angle spread is supposed to be $\Delta\theta = \Delta\theta_l = 4^\circ$ for $l \in \mathcal{L}$.

A. ESTIMATION PERFORMANCE OF NCT POSITIONS AND VIRTUAL AoDs

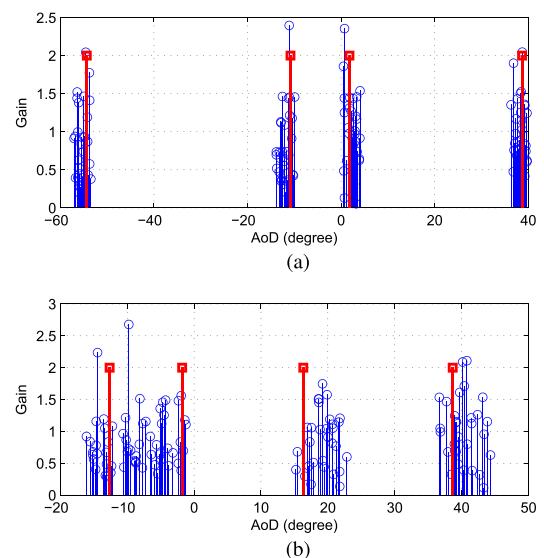
In this subsection, we evaluate the proposed scheme by recovery accuracy of NCT positions, and then carry out simulations to show the relation between virtual AoDs and practical AoDs.

In Fig. 4, we plot the successful recovery probability of NCT positions, as a function of the signal to noise ratio (SNR). The number of transmit antennas N_t is set to 8, 16 and 32. For comparison purposes, we also present the simulation results of the scheme in [16], which merely exploits the sparsity in the delay domain without taking into account the spatial correlation. As shown in Fig. 4, the recovery probability of scheme in [16] decreases with increasing number of antennas. What's worse, the scheme in [16] cannot work in the case $N_t = 32$ under current parameters, due to the fact that the nonzero rows to be recovered in the model of [16] is KN_t , larger than the number of measurements M .

**FIGURE 4.** Successful recovery probability of NCT positions against SNR with $v = 350 \text{ km/h}$.

However, it is shown that our proposed algorithm is robust to the number of transmit antennas in the estimation accuracy of NCT positions. This is because in our scheme, in order to obtain NCT positions we typically require to seek K nonzero rows of \mathbf{S} (only 1 nonzero row corresponding to each submatrix \mathbf{S}_l , $l \in \mathcal{L}$), which is much less than M and independent of N_t .

Simulation results are depicted in Fig. 5 to show the distribution of practical AoDs and virtual AoDs, where the blue lines denote the practical AoDs and the red lines denote the virtual AoDs. Two cases with different double-side angle spreads $\Delta\theta = 4^\circ$ and $\Delta\theta = 8^\circ$ are simulated. It is shown that all virtual AoDs locate in the practical AoDs region. This observation indicates the effectiveness of the developed algorithm for estimating virtual AoDs. Hence, we can expect that the updated measurement matrix based on virtual AoDs can improve the resolution of AoDs and thus refine the estimation accuracy.

**FIGURE 5.** Practical AoDs and virtual AoDs with $v = 350 \text{ km/h}$, $N_t = 32$, $\text{SNR} = 25 \text{ dB}$. (a) $\Delta\theta = 4^\circ$. (b) $\Delta\theta = 8^\circ$.

B. NMSE OF THE PROPOSED ESTIMATION SCHEME

To evaluate the channel estimation performance, we calculate the normalized mean square error (NMSE) defined as

$$\text{NMSE (dB)} = 10 \log_{10} \left(E \left[\frac{\sum_{l=1}^L \|\mathbf{H}_l - \hat{\mathbf{H}}_l\|_F^2}{\sum_{l=1}^L \|\mathbf{H}_l\|_F^2} \right] \right), \quad (39)$$

where \mathbf{H}_l is the true channel parameter and $\hat{\mathbf{H}}_l$ is its estimate.

In Fig. 6, we compare the NMSE performance of our proposed scheme with the scheme in [16], where cases $N_t = 8$ and $N_t = 64$ are simulated. As expected, the proposed adaptive-QBSO algorithm achieves superior performance as it adaptively updates the measurement matrix to improve the resolution of AoDs. Both of proposed algorithms achieve

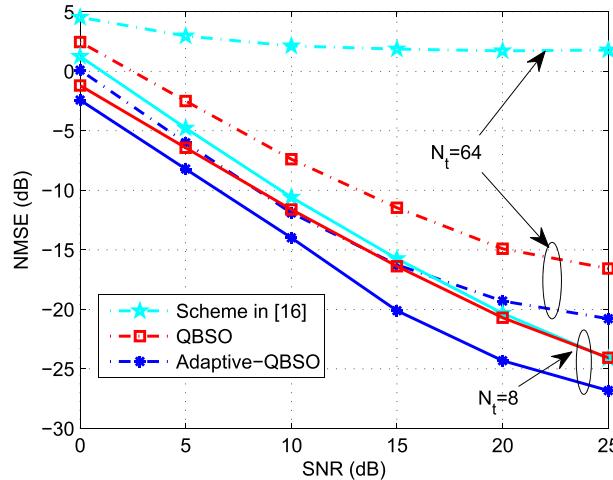


FIGURE 6. NMSE performance against SNR with $v = 350 \text{ km/h}$, $N_t = 8$ and $N_t = 64$.

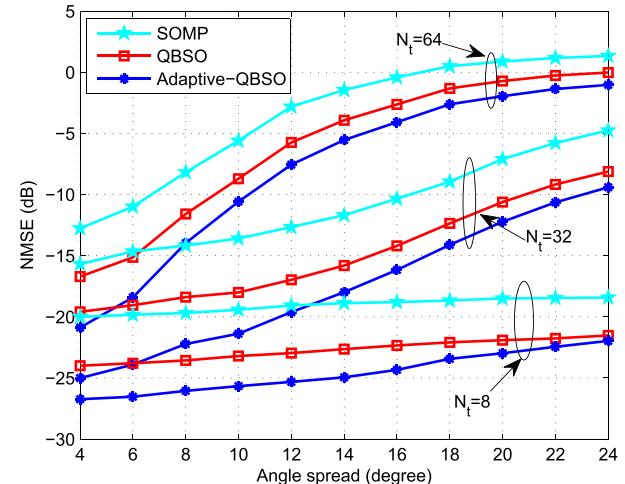


FIGURE 8. NMSE performance against the angle spread with $v = 350 \text{ km/h}$, $\text{SNR} = 25 \text{ dB}$.

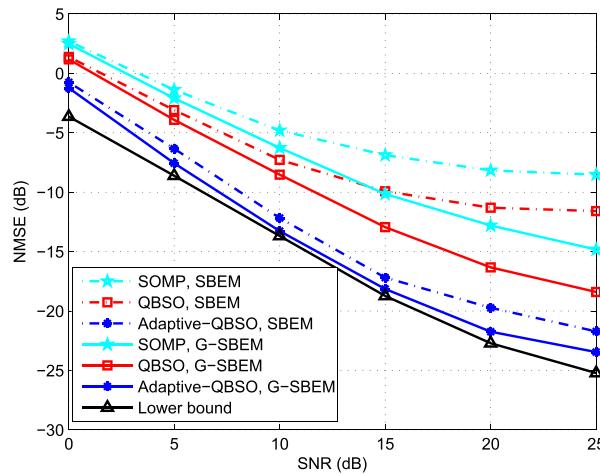


FIGURE 7. NMSE performance against SNR with $v = 500 \text{ km/h}$, $N_t = 32$.

higher channel estimation accuracy than the scheme in [16]. For example, at $\text{NMSE} = -10 \text{ dB}$ in the case $N_t = 8$, the QBSO algorithm and the adaptive-QBSO algorithm, respectively, achieve an SNR gain of about 1 dB and 3 dB compared with the scheme in [16]. Moreover, the scheme in [16] performs poorly for all the SNR range in the case $N_t = 64$ due to large error in the estimation accuracy of NCT positions, which has already been explained in Fig. 4.

In Fig. 7, we show the NMSE comparison of different algorithms with $v = 500 \text{ km/h}$ (the corresponding NMDS $v_M = 0.093$). Both the SBEM with $G_t = N_t$ and the G-SBEM with $G_t = 2N_t$ are simulated. The lower bound is also presented as the benchmark, where we assume the prior knowledge of AoDs. For curves of the SBEM, the SOMP and QBSO algorithms approach certain error floors due to the power leakage in the spatial domain, while the adaptive-QBSO algorithm performs better as it can strengthen the sparsity representation by adaptively updating the measurement matrix. In contrast, the performances of all curves in

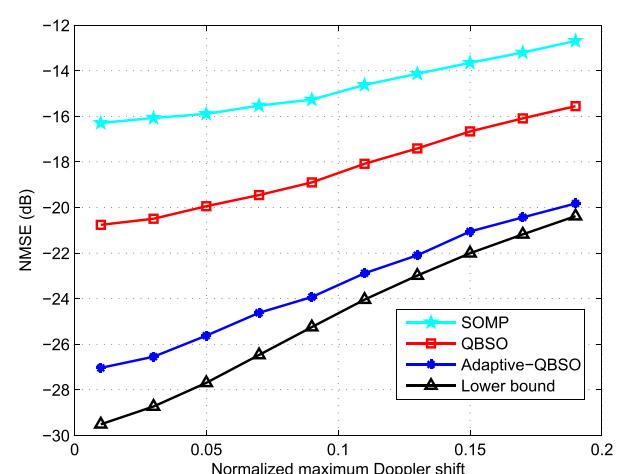


FIGURE 9. NMSE performance against the normalized maximum Doppler shift with $N_t = 32$, $\text{SNR} = 25 \text{ dB}$.

G-SBEM are improved compared with the SBEM scheme. This is due to the fact that the G-SBEM is able to improve the resolution of AoDs by utilizing more than N_t non-orthogonal bases. It is clearly seen that our proposed adaptive-QBSO algorithm is superior to others, and able to approach the lower bound.

Fig. 8 plots the NMSE performance versus the angle spread $\Delta\theta$ for various numbers of antennas $N_t \in \{8, 32, 64\}$. We can observe that the larger angle spread introduces more NMSE performance degradation. This is due to the fact that the spatial correlation degrades with increasing angle spread, and as a consequence, the G-SBEM would result in larger modeling error with limited bases. Nevertheless, our proposed algorithms achieve a notably higher channel estimation accuracy than the conventional SOMP algorithm for all concerned $\Delta\theta$ range and different N_t , which verifies that our proposed scheme is highly promising in practical application for massive MIMO systems.

Further, we examine the robustness of our proposed algorithms against the NMDS ν_M in Fig. 9, where we set $N_t = 32$ and $SNR = 25$ dB, and ν_M varies from 0.01 to 0.19. It is shown that the NMSE curves monotonically increase with ν_M , due to the CE-BEM modeling error getting larger when Doppler shift increases. A superiority of our proposed adaptive-QBSO algorithm can be observed among all the NMDS range, which verifies that our proposed scheme exhibits robustness against the Doppler shift.

VI. CONCLUSION

In this paper, we have proposed a new downlink channel estimation scheme for the massive MIMO-OFDM systems over time-varying channels. Both the sparsity in the delay domain and the high correlation in the spatial domain have been exploited. We designed the G-SBEM to model the spatial correlation and formulated the channel estimation as a CS problem with a quasi-block-sparse matrix to be recovered. Two novel algorithms referred as QBSO and adaptive-QBSO algorithms were developed to recover the sparse channels. The adaptive-QBSO algorithm is able to adaptively design the measurement matrix based on the estimated virtual AoDs, and thus the resolution of AoDs can be improved. Simulation results have shown that our proposed channel estimation algorithms can achieve better NMSE performance than conventional schemes.

APPENDIX

PROOF OF (5)

Based on (2) and (3), $\mathbf{H}_T^{n_t}$ can be expressed as

$$\mathbf{H}_T^{n_t} = \sum_{q=1}^Q \text{diag}(\mathbf{b}_q) \mathcal{C}_q^{n_t} + \xi^{n_t}, \quad (40)$$

where $\mathcal{C}_q^{n_t}$ is a circulant matrix with the first column being $((\mathbf{c}_q^{n_t})^T, \mathbf{0}_{1 \times (N-L)})^T$. Due to its circularity, $\mathcal{C}_q^{n_t}$ can be diagonalized, expressed as

$$\mathcal{C}_q^{n_t} = \mathbf{F}^H \text{diag}(\mathcal{V} \mathbf{c}_q^{n_t}) \mathbf{F}. \quad (41)$$

Accordingly, (1) can be denoted in terms of the CE-BEM as

$$\mathbf{y} = \sum_{n_t=1}^{N_t} \left(\sum_{q=1}^Q \mathbf{F} \text{diag}(\mathbf{b}_q) \mathbf{F}^H \text{diag}(\mathcal{V} \mathbf{c}_q^{n_t}) \right) \mathbf{x}^{n_t} + \mathbf{z}. \quad (42)$$

Since \mathbf{b}_q is one of the rows of \mathbf{F} , we have

$$\begin{aligned} \mathbf{F} \text{diag}(\mathbf{b}_q) \mathbf{F}^H &= \mathbf{I}_N^{(q-\frac{Q+1}{2})} \mathbf{F} \mathbf{F}^H \\ &= \mathbf{I}_N^{(q-\frac{Q+1}{2})}. \end{aligned} \quad (43)$$

Substitute (43) into (42), and we finally obtain (5).

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

- [2] 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.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [4] B. Gong, L. Gui, Q. Qin, and X. Ren, "Compressive sensing-based detector design for SM-OFDM massive MIMO high speed train systems," *IEEE Trans. Broadcast.*, vol. 63, no. 4, pp. 714–726, Dec. 2017.
- [5] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.
- [6] N. Aboutorab, W. Hardjawana, and B. Vučetić, "A new iterative Doppler-assisted channel estimation joint with parallel ICI cancellation for high-mobility MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1577–1589, May 2012.
- [7] K. Muralidhar and D. Sreedhar, "Pilot design for vector state-scalar observation Kalman channel estimators in doubly-selective MIMO-OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 147–150, Apr. 2013.
- [8] 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.
- [9] C. Qi, G. Yue, L. Wu, Y. Huang, and A. Nallanathan, "Pilot design schemes for sparse channel estimation in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1493–1505, Apr. 2015.
- [10] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2015.
- [11] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [12] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Novel compressed sensing-based channel estimation algorithm and near-optimal pilot placement scheme," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2590–2603, Apr. 2016.
- [13] X. Ren, W. Chen, B. Gong, Q. Qin, and L. Gui, "Position-based interference elimination for high-mobility OFDM channel estimation in multicell systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7986–8000, Sep. 2017.
- [14] P. Cheng et al., "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4173–4185, Oct. 2013.
- [15] Q. Qin, L. Gui, B. Gong, X. Ren, and W. Chen, "Structured distributed compressive channel estimation over doubly selective channels," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 521–531, Sep. 2016.
- [16] B. Gong, L. Gui, Q. Qin, X. Ren, and W. Chen, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9149–9161, Oct. 2017.
- [17] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.
- [18] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [19] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Mar. 2013.
- [20] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [21] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
- [22] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.
- [23] H. Xie, F. Gao, S. Zhang, and S. Jin, "Spatial-temporal BEM and channel estimation strategy for massive MIMO time-varying systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [24] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.

- [25] J. Zhao, F. Gao, W. Jia, J. Zhao, and W. Zhang, "Channel tracking for massive MIMO systems with spatial-temporal basis expansion model," in *Proc. IEEE Int. Commun. Conf. (ICC)*, May 2017, pp. 1–5.
- [26] A. Forenza, D. J. Love, and R. W. Heath, Jr., "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.
- [27] Z. Chen and C. Yang, "Pilot decontamination in wideband massive MIMO systems by exploiting channel sparsity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5087–5100, Jul. 2016.
- [28] 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.
- [29] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K, Jun. 2015, pp. 4558–4563.
- [30] Y. C. Eldar, P. Kuppinger, and H. Boelskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [31] R. Mohammadian, A. Amini, and B. H. Khalaj, "Deterministic pilot design for sparse channel estimation in MISO/multi-user OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 129–140, Jan. 2017.



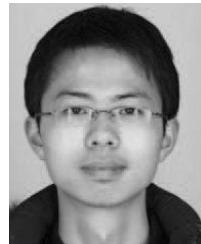
LIN GUI (M'08) received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2002. Since 2002, she has been with the Institute of Wireless Communication Technology, Shanghai Jiao Tong University, Shanghai, China, where she is currently a Professor. Her current research interests include HDTV and wireless communications.



BO GONG received the B.S. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2013. She is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. Her current research interests include doubly selective channel estimation, massive multiple-input-multiple-output systems, and compressive sensing.



SHENG LUO received the B.E. and M.Sc. degrees in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2009 and 2012, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2017. He is currently an Assistant Professor with the School of Computer and Software Engineering, Shenzhen University. His current research interests are in the areas of cooperative communication, buffer-aided relaying, wireless information and power transfer, and millimeter-wave communication.



QIBO QIN received the B.S. degree in electronic and information engineering from Xian Jiao Tong University, Xian, China, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include doubly selective channel estimation, massive multiple-input multiple-output systems, and millimeter-wave communications.