

Research Article

Distributed Structured Compressive Sensing-Based Time-Frequency Joint Channel Estimation for Massive MIMO-OFDM Systems

Wenjie Zhang , Hui Li, Rong Jin, Shanlin Wei, Wei Cheng, Weisi Kong , and Penglu Liu

School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China

Correspondence should be addressed to Wenjie Zhang; z714549544@126.com

Received 19 January 2019; Revised 31 March 2019; Accepted 2 April 2019; Published 2 May 2019

Academic Editor: Adrian Kliks

Copyright © 2019 Wenjie Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In massive multi-input multi-output orthogonal frequency division multiplexing (MIMO-OFDM) systems, accurate channel state information (CSI) is essential to realize system performance gains such as high spectrum and energy efficiency. However, high-dimensional CSI acquisition requires prohibitively high pilot overhead, which leads to a significant reduction in spectrum efficiency and energy efficiency. In this paper, we propose a more efficient time-frequency joint channel estimation scheme for massive MIMO-OFDM systems to resolve those problems. First, partial channel common support (PCCS) is obtained by using time-domain training. Second, utilizing the spatiotemporal common sparse property of the MIMO channels and the obtained PCCS information, we propose the priori-information aided distributed structured sparsity adaptive matching pursuit (PA-DS-SAMP) algorithm to achieve accurate channel estimation in frequency domain. Third, through performance analysis of the proposed algorithm, two signal power reference thresholds are given, which can ensure that the signal can be recovered accurately under power-limited noise and accurately recovered according to probability under Gaussian noise. Finally, pilot design, computational complexity, spectrum efficiency, and energy efficiency are discussed as well. Simulation results show that the proposed method achieves higher channel estimation accuracy while requiring lower pilot sequence overhead compared with other methods.

1. Introduction

Due to the scarcity of radio spectrum in the microwave band, how to improve spectrum efficiency (SE) has become one of the significant problems for future wireless communication systems. By leveraging the outstanding advantages of multiplexing and diversity, massive multi-input multi-output (MIMO) or large-scale antenna technology can achieve high SE and energy efficiency (EE) and has become one of the key technologies for fifth-generation (5G) wireless communications [1, 2]. In addition, orthogonal frequency division multiplexing (OFDM) has been widely adopted in modern wireless communication systems due to its excellent antimultipath capability [3]. Consequently, massive MIMO-OFDM will be a new standard for 5G.

In massive MIMO-OFDM systems, channel state information (CSI) is indispensable for precoding/combine, channel equalization and coherent detection, etc., which makes the accurate channel estimation crucial to improving system performance [3, 4]. In conventional MIMO-OFDM systems, orthogonal pilot training scheme is adopted for accurate downlink CSI acquisition. In massive MIMO-OFDM systems, however, since hundreds of transmit antennas are installed at BS, the acquisition of high-dimensional CSI results in the prohibitively high pilot overhead, much less the problem of computational complexity caused by high-dimensional matrix operation [5]. Meanwhile, quantization accuracy of the CSI fed back from users to BS, which is affected by the number of BS antennas, is a question worth considering as well [6, 7].

To solve those problems, many low-rank channel estimation approaches were proposed such as low-rank channel covariance matrices- (CCMs-) based method [8, 9], array signal processing-based method [10, 11], and compressive sensing- (CS-) based method [12–23]. Although the low-rank property of CCMs can make the dimensions of MIMO channels greatly reduced, high-dimensional CCMs are difficult to be obtained. Moreover, high-dimensional matrix operations involved in the singular value decomposition (SVD) or eigenvalue decomposition (EVD) lead to high computational complexity. Array signal processing-based methods have many advantages, but it is particularly applicable to millimeter-wave massive MIMO systems which have high angular resolution [4]. Fortunately, the CS-based channel estimation method can be adopted for massive MIMO-OFDM systems due to the sparse characteristics of MIMO wireless channels [24].

Basically, there are three categories of CS-based channel estimation methods for MIMO-OFDM systems: time-domain training channel estimation (TTCE) method [12, 13], frequency-domain pilot training channel estimation (FTCE) method [17–19], and time-frequency joint training channel estimation (TFTCE) method [21–23]. TTCE exploits the interblock interference-free (IBI-free) region in the redundant portion of the received time-domain training sequences (TSs) for channel estimation. Although this scheme has high spectral efficiency, the design of the time-domain TSs is rather difficult when the number of antennas is large enough, because it not only needs to ensure the orthogonality between different TSs but also makes sure that the sensing matrices satisfy the restricted isometry property (RIP) condition [14, 15]. Moreover, accuracy of the estimation is seriously affected by channel length [16]. Unlike the TTCE method, the FTCE method uses orthogonal or nonorthogonal frequency-domain pilots for channel estimation. Although the difficulty of designing training sequences is greatly reduced, the spectral efficiency is decreased because the useful information provided by the preamble is not effectively utilized. Taking the advantages of the above two schemes, Dai et al. [5] proposed TFTCE scheme, while the CS method is not adopted for higher accuracy estimation. On this basis, Ding et al. [21] proposed a TFTCE method based on compressed sensing. However, due to the interference of nonorthogonal time-domain TSs, it cannot be applied to massive MIMO systems. Ding et al. and Fan et al. [22, 23] exploited identical TSs in time domain and frequency domain over a transmission frame, while changing TSs can bring better performance [25]. More importantly, these methods did not make effective use of the sparse characteristics of MIMO channels.

In this paper, we propose a distributed structured compressive sensing-based time-frequency joint channel estimation method for massive MIMO-OFDM systems. Our contributions in this paper are summarized as follows:

- (i) By using spatio-temporal common sparse characteristics of wireless MIMO channel, we propose a priori information aided distributed structured sparsity adaptive matching pursuit algorithm (PA-

DS-SAMP). The proposed algorithm regards the partial channel common support as a priori aided information and combines SAMP algorithm with structured multimeasurement vector (MMV) ideas [26]. The proposed algorithm improves the spectral efficiency and the accuracy of the channel estimation, while without knowing the sparsity of MIMO channels.

- (ii) The performance of the proposed algorithm is analyzed. First, the signal power thresholds which can ensure accurate recovery of signals under power-limited noise and according to probability accurate recovery under additive white Gaussian noise (AWGN) are obtained though strict derivation. Next, we give the pilot design scheme that is well suited for massive MIMO systems. Finally, computational complexity, spectrum efficiency, and energy efficiency are given, which show that not only the computational complexity is less than other algorithms but also the spectrum efficiency and energy efficiency are greatly improved.

The remainder of this paper is organized as follows. The distributed massive MIMO-OFDM system model with block-sparse channel is described in Section 2. In Section 3, a distributed structured compressive sensing-based channel estimation algorithm is proposed. Next, Section 4 presents the performance analysis of the proposed algorithm. Simulation results are presented in Section 5 to demonstrate the performance of the proposed algorithm. Finally, the paper is concluded in Section 6.

Notation. Throughout our discussion, lowercase bold-face letters indicate vector and uppercase boldface letters indicate matrix; $\mathbf{0}$ and \mathbf{I} denote the zeros matrix of arbitrary size and identity matrix; $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^\dagger$, and $|\cdot|_c$ stand for the transport, conjugate transport, matrix inversion, Moore–Penrose matrix inversion, and cardinality, respectively. $\|\cdot\|_p$ and $\|\cdot\|_F$ represent ℓ_p norm and Frobenius norm; operator \otimes represents circular convolution; $\mathbf{A}[i]$ expresses the i th block of matrix or vector \mathbf{A} and $\|\cdot\|_{p,q}$ represents ℓ_p/ℓ_q mixed norm of matrix or vector \mathbf{A} .

2. Distributed System Model with Block-Sparse Channel

Consider a massive MIMO-OFDM system with N_t transmit antennas configured at base station (BS) to serve K signal-antenna mobile terminals, where $N_t \gg K$. In this paper, we focus on the case of downlink transmission in which each frame consisting of several time-frequency training OFDM (TFT-OFDM) symbols is transmitted from BS to mobile terminals. As described in Figure 1, the i th ($1 \leq i \leq R$) TFT-OFDM symbol $\mathbf{s}_i^{(p)} = [(\mathbf{c}_i^{(p)})^T, (\mathbf{x}_i^{(p)})^T]^T$ transmitted from the p th ($1 \leq p \leq N_t$) antenna consists of a length- M time-domain TS $\mathbf{c}_i^{(p)} = [\mathbf{c}_{i,1}^{(p)}, \mathbf{c}_{i,2}^{(p)}, \dots, \mathbf{c}_{i,M}^{(p)}]^T$ and a following length- N OFDM data block $\mathbf{x}_i^{(p)} = [x_{i,1}^{(p)}, x_{i,2}^{(p)}, \dots, x_{i,N}^{(p)}]^T$. Unlike cyclic prefix OFDM (CP-OFDM) and zero-padding OFDM (ZP-OFDM), known pseudonoise (PN) instead of unknown CP or ZP is adopted as guard interval $\mathbf{c}_i^{(p)}$ in

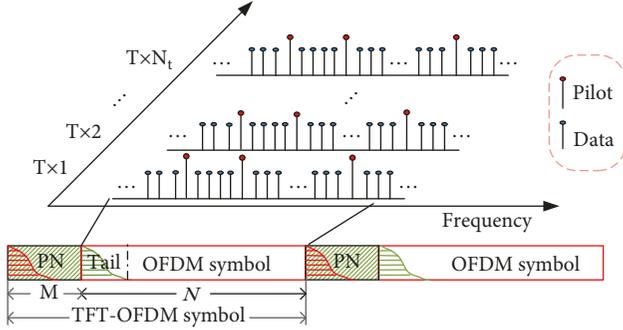


FIGURE 1: Structure of a TFT-OFDM symbol.

TFT-OFDM to alleviate interblock interference (IBI) caused by multipath.

At receiver, the TFT-OFDM symbols transmitted from each BS antenna are mixed together at each mobile terminal so that the i th received TFT-OFDM symbol vector in time domain can be written as [5]

$$\mathbf{r}_i = \sum_{p=1}^{N_t} (\mathbf{A}_{i,ISI}^{(p)} \mathbf{s}_i^{(p)} + \mathbf{A}_{i,IBI}^{(p)} \mathbf{x}_{i-1}^{(p)}) + \mathbf{n}_i, \quad (1)$$

where $\mathbf{A}_{i,ISI}^{(p)}$ and $\mathbf{A}_{i,IBI}^{(p)}$ are $N \times N$ Toeplitz lower and upper triangular matrices with the first column vector $[h_{i,1}^{(p)}, h_{i,2}^{(p)}, \dots, h_{i,L}^{(p)}, 0, \dots, 0]^T$ and the first row vector $[0, \dots, 0, h_{i,L}^{(p)}, \dots, h_{i,2}^{(p)}]$, respectively. At the same time, $\mathbf{h}_i^{(p)} = [h_{i,1}^{(p)}, h_{i,2}^{(p)}, \dots, h_{i,L}^{(p)}]^T \in \mathbb{C}^{L \times 1}$ denotes the channel impulse response (CIR) vector from the p th antenna to receiver with the length L and \mathbf{n}_i is the complex additive white Gaussian noise (AWGN) vector with zero mean and the variance of $\sigma^2 \mathbf{I}$.

After cyclicity reconstruction operated by the overlapping and adding (OLA) algorithm [5], the received pilot sequence of the i th OFDM symbol can be expressed through discrete Fourier transform (DFT) process as

$$\begin{aligned} \mathbf{y}_i &= \sum_{p=1}^{N_t} \text{diag}\{(\mathbf{F}\mathbf{x}_i^{(p)})|_{\xi_i}\} \mathbf{F}_L |_{\xi_i} \mathbf{h}_i^{(p)} + \mathbf{w}_i \\ &= \sum_{p=1}^{N_t} \text{diag}\{\mathbf{p}_i^{(p)}\} \mathbf{F}_L |_{\xi_i} \mathbf{h}_i^{(p)} + \mathbf{w}_i \\ &= \sum_{p=1}^{N_t} \mathbf{\Phi}_i^{(p)} \mathbf{h}_i^{(p)} + \mathbf{w}_i \\ &= \mathbf{\Phi}_i \mathbf{h}_i + \mathbf{w}_i, \end{aligned} \quad (2)$$

where \mathbf{F} is the $N \times N$ DFT matrix and \mathbf{F}_L is the $N \times L$ matrix consisted of the first L column of \mathbf{F} , ξ_i denotes the index set of subcarriers allocated to pilots whose elements are uniquely selected repeatedly from the set of $\{1, 2, \dots, N\}$, $(\mathbf{F}\mathbf{x}_i^{(p)})|_{\xi_i}$ and $\mathbf{F}_L|_{\xi_i}$ are the submatrix or subvector by selecting the rows of $(\mathbf{F}\mathbf{x}_i^{(p)})$ or \mathbf{F}_L according to ξ_i , $\mathbf{h}_i = [(\mathbf{h}_i^{(1)})^T, \dots, (\mathbf{h}_i^{(p)})^T, \dots, (\mathbf{h}_i^{(N_t)})^T]^T \in \mathbb{C}^{LN_t \times 1}$, $\mathbf{\Phi}_i^{(p)} = \text{diag}\{\mathbf{p}_i^{(p)}\} \mathbf{F}_L|_{\xi_i} = [\boldsymbol{\varphi}_{i,1}^{(p)}, \dots, \boldsymbol{\varphi}_{i,L}^{(p)}]$, $\mathbf{\Phi}_i = [\mathbf{\Phi}_i^{(1)}, \dots, \mathbf{\Phi}_i^{(N_t)}] \in$

$\mathbb{C}^{N_p \times LN_t}$, $N_p = |\xi_i|_c$ is the number of pilot subcarriers, and \mathbf{w}_i is the complex AWGN.

Wireless channels have sparse characteristics [24, 27]. The presence of scatterers and reflectors in wireless communication environment results in a multipath channel with several significantly resolvable propagation paths. Because the significant scatterers between transmitter and receiver are usually scarce, the number of the paths containing the majority of channel energy is so small that they are sparsely distributed in the resolvable paths. In other words, the gains of most paths approach to zeros except for a small number of significant paths so that the channel exhibits sparsity in delay domain [12]. The sparsity support of the sparse channel $\mathbf{h}_i^{(p)}$ is defined

$$D_i^{(p)} = \text{supp}\{\mathbf{h}_i^{(p)}\} = \left\{ l : |h_{i,l}^{(p)}| \geq p_{\text{th}}^{(p)}, \quad 1 \leq l \leq L \right\}, \quad (3)$$

where $p_{\text{th}}^{(p)}$ is the noise floor which can be determined according to [28]. The sparsity level of the channel is denoted as $S_i^{(p)} = |D_i^{(p)}|_c$.

Because antennas usually share the same scattering environment, all of the channels between one user and different antennas deployed on base station share common sparsity [17]. Therefore, the nonzero entries of the CIR vector can occur in clusters so that the CIR vector \mathbf{h}_i can be transformed to a block-sparse CIR vector

$$\begin{aligned} \tilde{\mathbf{h}}_i &= \begin{bmatrix} h_{i,1}^{(1)}, \dots, h_{i,1}^{(N_t)}, \dots, h_{i,l}^{(1)}, \dots, h_{i,l}^{(N_t)}, \dots, h_{i,L}^{(1)}, \dots, h_{i,L}^{(N_t)} \\ \underbrace{(\tilde{\mathbf{h}}_i[1])^T}_{(\tilde{\mathbf{h}}_i[1])^T} \quad \underbrace{(\tilde{\mathbf{h}}_i[l])^T}_{(\tilde{\mathbf{h}}_i[l])^T} \quad \underbrace{(\tilde{\mathbf{h}}_i[L])^T}_{(\tilde{\mathbf{h}}_i[L])^T} \end{bmatrix}^T \\ &= [(\tilde{\mathbf{h}}_i[1])^T, \dots, (\tilde{\mathbf{h}}_i[l])^T, \dots, (\tilde{\mathbf{h}}_i[L])^T]^T \in \mathbb{C}^{LN_t \times 1}, \end{aligned} \quad (4)$$

where $\tilde{\mathbf{h}}_i[l] = [h_{i,l}^{(1)}, h_{i,l}^{(2)}, \dots, h_{i,l}^{(N_t)}]^T \in \mathbb{C}^{N_t \times 1}$, $1 \leq l \leq L$. The blocking process of CIR vector is shown in Figure 2. At the same time, the observation matrix should have corresponding changes as follows:

$$\begin{aligned} \tilde{\Psi}_i &= \begin{bmatrix} \boldsymbol{\varphi}_{i,1}^{(1)}, \dots, \boldsymbol{\varphi}_{i,1}^{(N_t)}, \dots, \boldsymbol{\varphi}_{i,l}^{(1)}, \dots, \boldsymbol{\varphi}_{i,l}^{(N_t)}, \dots, \boldsymbol{\varphi}_{i,L}^{(1)}, \dots, \boldsymbol{\varphi}_{i,L}^{(N_t)} \\ \underbrace{\boldsymbol{\varphi}_{i,1}^{(1)}, \dots, \boldsymbol{\varphi}_{i,1}^{(N_t)}}_{\tilde{\boldsymbol{\varphi}}_i[1]} \quad \underbrace{\boldsymbol{\varphi}_{i,l}^{(1)}, \dots, \boldsymbol{\varphi}_{i,l}^{(N_t)}}_{\tilde{\boldsymbol{\varphi}}_i[l]} \quad \underbrace{\boldsymbol{\varphi}_{i,L}^{(1)}, \dots, \boldsymbol{\varphi}_{i,L}^{(N_t)}}_{\tilde{\boldsymbol{\varphi}}_i[L]} \end{bmatrix} \\ &= [\tilde{\Psi}_i[1], \dots, \tilde{\Psi}_i[l], \dots, \tilde{\Psi}_i[L]] \in \mathbb{C}^{N_p \times LN_t}, \end{aligned} \quad (5)$$

where $\tilde{\Psi}_i[l] = [\boldsymbol{\varphi}_{i,l}^{(1)}, \boldsymbol{\varphi}_{i,l}^{(2)}, \dots, \boldsymbol{\varphi}_{i,l}^{(N_t)}] \in \mathbb{C}^{N_p \times N_t}$, $1 \leq l \leq L$. Thus, equation (2) can be re-expressed as

$$\mathbf{y}_i = \tilde{\Psi}_i \tilde{\mathbf{h}}_i + \mathbf{w}_i. \quad (6)$$

Besides, the path delays of practical wireless channels vary much slower than the path gains as shown in many studies [12]. This slow variation results that the CIRs over consecutive TFT-OFDM symbols share the same sparsity pattern [17]. Consequently, this spatio-temporal common sparse characteristic of MIMO channel over R TFT-OFDM symbols can be expressed as

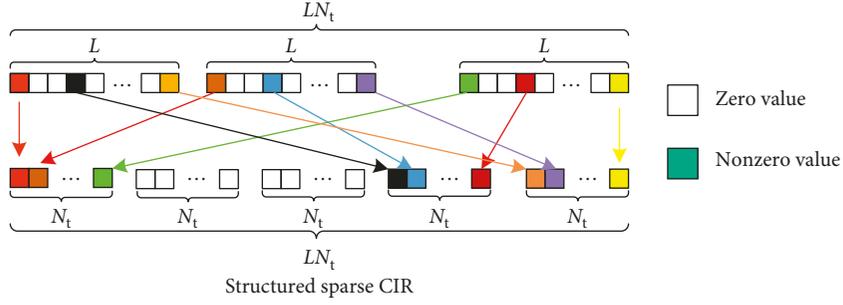


FIGURE 2: Structured block-sparse signal.

$$\begin{cases} S_i^{(1)} = S_i^{(2)} = \dots = S_i^{(N_t)} = S, \\ S_1^{(p)} = S_2^{(p)} = \dots = S_R^{(p)} = S, \\ \mathbb{D}_i^{(1)} = \mathbb{D}_i^{(2)} = \dots = \mathbb{D}_i^{(N_t)} = \mathbb{D}, \\ \mathbb{D}_1^{(p)} = \mathbb{D}_2^{(p)} = \dots = \mathbb{D}_R^{(p)} = \mathbb{D}. \end{cases} \quad (7)$$

Defining the spatio-temporal *Common Sparse Support* \mathbb{D} of $\tilde{\mathbf{H}} \triangleq \{\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_R\}$ as

$$\mathbb{D} \triangleq \left\{ l: \frac{1}{RN_t} \sum_{i=1}^R \|\tilde{h}_i[l]\|_F > p_{\text{th}}, \quad 1 \leq l \leq L \right\}, \quad (8)$$

where $p_{\text{th}} = (1/N_t) \sum_{p=1}^{N_t} p_{\text{th}}^{(p)}$ and $p_{\text{th}}^{(p)}$ is determined as (3). For simplicity, we assume the same threshold p_{th} is used over each frame. Thus, the spatio-temporal *Common Sparse Level* S of the channel vector set $\tilde{\mathbf{H}}$ is $S = |\mathbb{D}|_c$.

Because the observation matrices are different, but CIRs have the same sparse characteristics, this system model can be called a *distributed system model*. The structure of the distributed system model with common block-sparse channel is described in Figure 3.

3. Distributed Structured Compressive Sensing-Based Channel Estimation

In this section, we will introduce a time-frequency joint channel estimation method to obtain more accurate estimation accuracy. Firstly, time-domain training is used to obtain partial common path delay and then accurate channel estimation is achieved using the proposed PA-DS-SAMP algorithm.

3.1. Acquisition of Partial Channel Common Support. Although time-domain training cannot realize accurate channel estimation, it can be used to acquire some prior information of channels such as partial channel delay information. Using the overlap-add method, the superimposed signal with the sum of the main part and tail of received time-domain TS can be expressed as

$$\mathbf{d}_i = \mathbf{r}_{i,\text{main}} + \mathbf{r}_{i,\text{tail}}, \quad 1 \leq i \leq R, \quad (9)$$



FIGURE 3: Distributed system model with block-sparse channel.

where $\mathbf{r}_{i,\text{main}}$ and $\mathbf{r}_{i,\text{tail}}$ are the received signals associated with PN part and the tail part expressed as (10) and R is the number of symbols in which the channel delay remains unchanged.

$$\begin{cases} \mathbf{r}_{i,\text{main}} = \mathbf{G}_{i,\text{main}} \mathbf{h}_i + \mathbf{n}_{i,\text{main}}, \\ \mathbf{r}_{i,\text{tail}} = \mathbf{G}_{i,\text{tail}} \mathbf{h}_i + \mathbf{n}_{i,\text{tail}}, \end{cases} \quad (10)$$

where the measurement matrices $\mathbf{G}_{i,\text{main}} = [\mathbf{G}_{i,\text{main}}^{(1)}, \dots, \mathbf{G}_{i,\text{main}}^{(N_t)}]$ and $\mathbf{G}_{i,\text{tail}} = [\mathbf{G}_{i,\text{tail}}^{(1)}, \dots, \mathbf{G}_{i,\text{tail}}^{(N_t)}]$ are determined by $\mathbf{G}_{i,\text{main}}^{(p)}$ and $\mathbf{G}_{i,\text{tail}}^{(p)}$ which are given in [12]. Thus, we can get

$$\mathbf{d}_i = \sum_{p=1}^{N_t} \mathbf{C}_i^{(p)} \mathbf{h}_i^{(p)} + \sum_{p=1}^{N_t} \mathbf{X}_{i,\text{DBI}}^{(p)} \mathbf{h}_i^{(p)} + \mathbf{n}_{i,\text{total}}, \quad 1 \leq i \leq R. \quad (11)$$

The matrix $\mathbf{C}_i^{(p)}$ is generated by cyclically shifting the basic TS $\mathbf{c}_i^{(p)}$ and the data block interference (DBI) matrix $\mathbf{X}_{i,\text{DBI}}^{(p)}$ can be denoted as

$$\mathbf{X}_{i,\text{DBI}}^{(p)} = \begin{bmatrix} \mathbf{x}_{i,1}^{(p)} & \mathbf{x}_{i-1,N}^{(p)} & \mathbf{x}_{i-1,N-1}^{(p)} & \dots & \mathbf{x}_{i-1,N-L+2}^{(p)} \\ \mathbf{x}_{i,2}^{(p)} & \mathbf{x}_{i,1}^{(p)} & \mathbf{x}_{i-1,N}^{(p)} & \dots & \mathbf{x}_{i-1,N-L+3}^{(p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{i,L}^{(p)} & \mathbf{x}_{i,L-1}^{(p)} & \mathbf{x}_{i,L-2}^{(p)} & \dots & \mathbf{x}_{i,1}^{(p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{i,N}^{(p)} & \mathbf{x}_{i,N-1}^{(p)} & \mathbf{x}_{i,N-2}^{(p)} & \dots & \mathbf{x}_{i,N-L+1}^{(p)} \end{bmatrix}_{N \times L}. \quad (12)$$

Taking advantage of the circular convolution, the coarse CIR associated with the u th antenna can be estimated as

$$\begin{aligned}
 \bar{\mathbf{h}}_i^{(u)} &= \frac{1}{M} \mathbf{c}_i^{(u)} \otimes \mathbf{d}_i \\
 &= \mathbf{h}_i^{(u)} + \frac{1}{M} \mathbf{c}_i^{(u)} \otimes \sum_{\substack{p=1 \\ p \neq u}}^{N_i} \mathbf{C}_i^{(p)} \mathbf{h}_i^{(p)} \\
 &\quad + \frac{1}{M} \mathbf{c}_i^{(u)} \otimes \sum_{p=1}^{N_i} \mathbf{X}_{i,\text{DBI}}^{(p)} \mathbf{h}_i^{(p)} + \mathbf{v}_i^{(u)}.
 \end{aligned} \tag{13}$$

In equation (13), the second part and the third part are the interference caused by nonorthogonal PN sequences and data, respectively. The data interference is mainly due to the superposition of the previous TFT-OFDM data block and the current TFT-OFDM symbol data block. Fortunately, there is no correlation between the random OFDM data block and the fixed TS sequence so that the data interference from others antennas has little effect on the delay estimation of the channel, especially in large-scale antenna systems [29]. However, the PN sequences interference from different antennas will increase linearly with the number of antennas when the cyclic matrices of the PN sequences associated with different antennas are not orthogonal with each other. To the best of our knowledge, however, there is no a set of sequences which have good autocorrelation and the cyclic matrices of which are orthogonal to each other. It is worth noting that this problem can be resolved once the same PN sequence is used for each antenna. Consequently, we have average CIR

$$\begin{aligned}
 \bar{\mathbf{h}}_i &= \frac{1}{M} \mathbf{c}_i \otimes \mathbf{d}_i \\
 &= \sum_{p=1}^{N_i} \mathbf{h}_i^{(p)} + \underbrace{\frac{1}{M} \mathbf{c}_i \otimes \sum_{p=1}^{N_i} \mathbf{X}_{i,\text{DBI}}^{(p)} \mathbf{h}_i^{(p)}}_{\text{Data interference+Noise}} + \mathbf{v}_i^{(u)}.
 \end{aligned} \tag{14}$$

As a perfect sequence, Zadoff-Chu sequence which has good autocorrelation and cross-correlation is widely used in the OFDM system for synchronization and channel estimation [30]. A length- M Zadoff-Chu sequence $\{\gamma(m)\}_{m=0}^{M-1}$ is defined as [30]

$$\gamma(m) = (-1)^{mP} \exp\{j\pi mP(m+2g)\}, \quad 0 \leq m \leq M-1, \tag{15}$$

where M is the sequence length, g is an integer, and P is an integer prime to M . By taking the inverse discrete Fourier transform (IDFT) of a diagonal matrix which the diagonal elements comprise the Zadoff-Chu sequence γ , matrix \mathbf{C} is constructed with a lot of excellent correlation properties in each row vector as well. For each row sequence c_i of the constructed matrix \mathbf{C} , the following correlation properties are held [30]:

$$R_i(\tau) = \sum_{m=0}^{M-1} c_i(m) (c_i(m+\tau)_M)^* = \begin{cases} M, & \tau = 0, \\ 0, & \tau \neq 0, \end{cases} \tag{16}$$

$$\begin{aligned}
 R_{k,l}(\tau) &= \sum_{m=0}^{M-1} c_k(m) (c_l(m+\tau))^* \\
 &= \begin{cases} M \exp\left\{\frac{j\pi[P\tau^2 - 2(l-Pg)]\tau}{M}\right\}, \\ \quad \text{mod}(l-k-P\tau, M) = 0, \\ 0, & \text{mod}(l-k-P\tau, M) \neq 0, \end{cases}
 \end{aligned} \tag{17}$$

where $c(\cdot)_M$ represents a periodic cycle of length M .

It has been proved that exploiting constant TS within the transmission frame is not an optimal scheme for channel tracking and changing TS can bring better channel estimation performance [5, 25]. Thus, we adopt changing time-domain training scheme which different PN sequences generated by each row sequence of \mathbf{C} for the acquisition of accurate CSI over one frame of the transmission signals. The partial common support of MIMO channels over R TFT-OFDM symbols can be calculated as follows:

$$\mathbb{D}^0 = \left\{ l : \frac{1}{RN_t} \sum_{i=1}^R |\bar{h}_{i,l}| \geq p_{\text{th}}, \quad 1 \leq l \leq L \right\}, \tag{18}$$

where p_{th} is the same as that in (8) an can ensure the reliability of the channel common support.

3.2. Acquisition of Accurate Path Gain Estimation. In this section, we use the frequency-domain method in combination with the partial channel common support as a priori information for accurate channel estimation.

Since the support set of CIR remains unchanged over several consecutive TFT-OFDM symbols, it can be estimated simultaneously by using multiple measurement vector (MMV) CS methods [31]. In general, MMV methods are used to solve sparse signal recovery problems in which measurement vectors are associated with the same sensing matrix, while generalized MMV (GMMV) CS methods are required for the distributed systems with different sensing matrices. The problem of sparse channel state acquisition can be formulated as the following optimization problem:

$$\begin{aligned}
 \min & \left(\sum_{i=1}^R \|\tilde{\mathbf{h}}_i\|_{2,0}^2 \right)^{(1/2)}, \\
 \text{s.t.} & \quad \mathbf{y}_i = \tilde{\Psi}_i \tilde{\mathbf{h}}_i + \mathbf{w}_i, \quad 1 \leq i \leq R.
 \end{aligned} \tag{19}$$

To solve this problem, the PA-DS-SAMP algorithm listed in Algorithm 1, which is developed from the DSAMP algorithm [18], is proposed.

The initialization of the PA-DS-SAMP algorithm is implemented in Steps 1~3. In Step 5, the correlations of the residuals with the column vector of sensing matrices are

Input:

- (1) Initial support \mathbb{D}^0 , initial channel sparsity $S_0 = |\mathbb{D}^0|_c$,
- (2) Noisy measurements \mathbf{y}_i and sensing matrices $\tilde{\Psi}_i$, $1 \leq i \leq R$,
- (3) Step size selection threshold Γ , initial step size s ,
- (4) Termination threshold p_{th} , noise power σ^2 , iteration index $k = 1$.

Initialization:

- (1) $\tilde{\mathbf{h}}_i = \mathbf{0} \in \mathbb{C}^{LN_i \times 1}$,
- (2) Residual $\mathbf{b}_i^0 = \mathbf{y}_i$,
- (3) Increased sparsity $S_I = s$,
- (4) $\mathbb{D}_B = \mathbb{D}^0$, $S_B = S_0$.

Iterations:

- (5) $\mathbf{a}_i = (\tilde{\Psi}_i)^H \mathbf{b}_i^{k-1}$, $1 \leq i \leq R$ /*correlation*/
- (6) $\mathbb{T}^k = \arg \max_l \left\{ \sum_{i=1}^R \|\mathbf{a}_i[l]\|_2^2 \right\}_{l=1}^L, S_B + S_I$ /*rough support obtain*/
- (7) $\mathbb{C}^k = \mathbb{T}^k \cup \mathbb{D}^{k-1}$ /*support union*/
- (8) $\mathbf{z}_i^k = (\tilde{\Psi}_{i, \mathbb{C}^k})^\dagger \mathbf{b}_i^{k-1}$ /*coarse estimate*/
- (9) $\mathbb{S}^k = \arg \max_l \left\{ \sum_{i=1}^R \|\mathbf{z}_i^k[l]\|_2^2 \right\}_{l=1}^L, S_B + S_I$ /*full support*/
- (10) $\mathbb{D}^k = \mathbb{D}_B \cup \arg \max_{l \in \{\mathbb{S}^k - \mathbb{S}^k \cap \mathbb{D}_B\}} \left\{ \sum_{i=1}^R \|\mathbf{z}_i^k[l]\|_2^2 \right\}_{l=1}^L, S_I$ /*support pruning*/
- (11) $\tilde{\mathbf{h}}_i = (\tilde{\Psi}_{i, \mathbb{D}^k})^\dagger \mathbf{y}_i$ /*block estimate*/
- (12) $\mathbf{b}_i^k = \mathbf{y}_i - \tilde{\Psi}_{i, \mathbb{D}^k} \tilde{\mathbf{h}}_i$ /*residue update*/
- (13) $l_{min} = \arg \min_l \left\{ \sum_{i=1}^R \|\tilde{\mathbf{h}}_i[l]\|_2^2 \right\}, l \in \mathbb{D}^k$ /*minimum power index*/
- (14) If $\sum_{i=1}^R \|\tilde{\mathbf{h}}_i\|_{l_{min}}^2 / L < p_{th}$ then quit
- (15) Else if $\sum_{i=1}^R \|\mathbf{b}_i^k\|_2^2 > \sum_{i=1}^R \|\mathbf{b}_i^{k-1}\|_2^2$ then
 - (16) if $\|\tilde{\mathbf{h}}_i - \tilde{\mathbf{h}}_i^{k-1}\|_2^2 < \Gamma$ then
 - (17) $s = \lceil s/2 \rceil$, $S_I = S_I + s$, $r_i^k = r_i^{k-1}$ /*step adaptation*/
 - (18) Else $S_I = S_I + s$, $r_i^k = r_i^{k-1}$ /*sparsity level update*/
 - (19) End if
- (20) Else $k = k + 1$ /*iteration index update*/
- (21) end
- (22) **Until** $\sum_{i=1}^R \|\mathbf{b}_i^k\|_2^2 / R \leq \sigma^2$
- (23) $\hat{\mathbf{h}}_i = \tilde{\mathbf{h}}_i^k$

Output: Estimation of structured channel impulse response $\hat{\mathbf{h}}_i$, $1 \leq i \leq R$.

ALGORITHM 1: PA-DS-SAMP algorithm.

calculated over each TFT-OFDM symbol and the indexes of the $S_B + S_I$ largest nonzero entries are selected in Step 6, where $\arg \max\{A, b\}$ represents the indexes of b maximum values in the set A and S_B and S_I presents basic sparsity and increased sparsity, respectively. Steps 7~10 are the support pruning for more precise sparsity support. Note that Step 9 does not guarantee that the basic support is fully contained so that Step 10 is needed to refine the effective support and ensure $S_B + S_I$ entries in the final support. Step 11 estimates $\tilde{\mathbf{h}}_i$ associated with the effective support \mathbb{D}^k using LS, and Step 12 calculate the residuals over each symbol. The index of the minimum power in the block of $\tilde{\mathbf{h}}_i$ is found in Step 13, and the iteration is terminated if the condition that the average energy of signal is less than a certain threshold holds, which is expressed in Step 14. Steps 15~21 are the step adaptive process, and the advantages of this approach will be explained below, where \cdot denotes upper bound integer operation. The whole iteration will be terminated if the residual average

energy is less than a certain threshold given in Step 22. Finally, the estimation of structured channel impulse response is obtained in Step 23.

Compared with DSAMP algorithm [18] and the classic SAMP algorithm [32], the proposed PA-DS-SAMP algorithm has four distinctive features:

- (i) Since the signal to be estimated has block structure, the structured sparsity is exploited in the proposed PA-DS-SAMP algorithm for block signal recovery. However, this useful signal feature is not used in either DSAMP algorithm or SAMP algorithm.
- (ii) In the PA-DS-SAMP algorithm, the sparsity support consists of basic support and additional support. Only additional support is updated while the basic support does not change in each iteration operation. The reason is that the basic support achieved by time-domain training has high reliability.

- (iii) The third difference is that the step size can be adaptively adjusted in PA-DS-SAMP algorithm, which has no discussion in DSAMP algorithm. Due to the tradeoff between recovery speed and recovery efficiency [33], an adaptive step size selection scheme is utilized which is based on the experiences that large step size is preferable for signal with flat magnitude and small step size is suitable for signal with fast magnitude attenuation.
- (iv) Finally, we give the termination threshold of the algorithm through strict mathematical derivation, which plays an important role in the performance of the algorithm.

Remark 1. After removing the large gain taps of CIR by using a priori information of channel delay, the remaining tap amplitudes have not much difference so that large step size is suitable. However, with large magnitude taps being reconstructed after a few stages of the algorithm, the number of taps to be reconstructed is less and less. In other words, the energy of the reconstructed CIR approaches to remain stable and the estimated sparsity level is close to the true sparsity level. Therefore, in order to better determine the more accurate sparseness, the step should be gradually reduced. According to this behavior, a large step size is set at the several stages of beginning in PA-DS-SAMP algorithm to expedite the convergence. Then, the step size decreases adaptively to provide fine tuning in following stages when the change of the reconstructed signal is less than a certain threshold.

4. Performance Analysis

In this part, we discuss the performance of the proposed algorithm. Firstly, we discuss the convergence of the algorithm, getting the convergence condition and termination threshold of the algorithm. Secondly, pilot design for sensing matrices is discussed. Then, the computational complexity of proposed algorithm is compared with OMP, SOMP, DSAMP algorithm, DS-SAMP algorithm, and PA-SOMP algorithm. Finally, pilot overhead for CSI acquisition is obtained. Before performance analysis, we first introduce two important definitions and two useful lemmas, which will play a key role in the subsequent analysis and discussion.

Definition 1 [34]. Block-Coherence of Ψ is defined as

$$\mu_B = \max_{k,l \neq k} \frac{1}{N_t} \rho(\mathbf{M}[k,l]), \quad (20)$$

where $\mathbf{M}[k,l] = \Psi_k^H \Psi_l$ and N_t is the size of the block.

Definition 2 [34]. Subcoherence of Ψ is defined as

$$v = \max_l \max_{m,n \neq m} \left| \left(\Phi_l^{(m)} \right)^H \Phi_l^{(n)} \right|, \quad (21)$$

$$\Phi_l^{(m)}, \Phi_l^{(n)} \in \Psi[l],$$

If $N_t = 1$, then $v = 0$.

Lemma 1. Let Ψ be a $KN_t \times LN_t$ matrix and $\Psi[k,l]$ be the (k,l) th $N_t \times N_t$ block of Ψ , then

$$\|\Psi\|_{2,\infty}^2 \leq \max_k \left(\sum_l \rho(\Psi[k,l]) \right)^2 \triangleq \rho_r^2(\Psi), \quad (22)$$

$$\|\Psi\|_{2,1}^2 \leq \max_k \left(\sum_l \rho(\Psi[k,l]) \right)^2 \triangleq \rho_c^2(\Psi), \quad (23)$$

$$\rho_r(\Psi) = \rho_c(\Psi^H). \quad (24)$$

Lemma 1 is obtained directly from Lemma 1 [34].

Lemma 2. [35]. Suppose that the noise vector \mathbf{w} follows Gaussian distribution $\mathbf{w} \sim N(0, \sigma^2 \mathbf{I}_n)$; if bounded set $B_2 =$

$$\left\{ \mathbf{w} : \|\mathbf{w}\|_2 \leq \sigma \sqrt{n + 2\sqrt{n \log n}} \right\} \text{ is satisfied, then we have}$$

$$P(\mathbf{w} \in B_2) \geq 1 - \frac{1}{n}. \quad (25)$$

4.1. Convergence Analysis of Proposed PA-DS-SAMP Algorithm. The PA-DS-SAMP algorithm begins with the residual initialization as $\mathbf{b}_i^0 = \mathbf{y}_i$, and the significant part is to choose at least one correct column block of the sensing matrix in each iteration. At the k th stage ($k \geq 1$), the block best matched to \mathbf{b}_i^{k-1} is chosen according to $l_k = \arg \max_l \sum_{i=1}^R \|(\tilde{\Phi}_i[l])^H \mathbf{b}_i^{k-1}\|$. Suppose \mathbb{D} denotes the support set of the signal $\tilde{\mathbf{h}}_i$ and \mathbb{D}^c denotes the complementary set of \mathbb{D} . Theorem 1 reveals the convergence conditions of the proposed algorithm in the absence of noise. We first discuss the noise-free case, then discuss power-limited noise $\|\mathbf{w}_i\|_2 \leq \epsilon$, and finally extend it to Gaussian white noise case $\mathbf{w}_i \sim N(0, \sigma^2 \mathbf{I}_{N_p})$.

Theorem 1. For block-sparse signals $\tilde{\mathbf{h}}_i \in \mathbb{C}^{LN_t}$ with the blocks length of N_t which satisfy $\mathbf{y}_i = \tilde{\Psi}_i \tilde{\mathbf{h}}_i$ ($1 \leq i \leq R$), a sufficient condition for the PA-DS-SAMP algorithm to recover $\tilde{\mathbf{h}}_i$ is that

$$\max_{1 \leq i \leq R} \left\{ \rho_c \left(\left(\tilde{\Psi}_{i,\mathbb{D}^c} \right)^\dagger \tilde{\Psi}_{i,\mathbb{D}} \right) \right\} < 1. \quad (26)$$

Proof. See Appendix A. □

The sufficient condition (26) depends on $\tilde{\Psi}_{i,\mathbb{D}}$ which is determined by the nonzero position of the signal $\tilde{\mathbf{h}}_i$, but the nonzero position of the $\tilde{\mathbf{h}}_i$ is not known in advance. Therefore, Theorem 1 is not practical. Following theorem under certain conditions on *Block Coherence* and *Subcoherence* associated with $\tilde{\Psi}_i$ can make sure the (A.1) holds.

Theorem 2. *Sufficient condition (26) is satisfied if for every μ_{Bi} , there is*

$$SN_t < \frac{1}{2} \left(\mu_{Bi} + N_t - (N_t - 1) \frac{\nu_i}{\mu_{Bi}} \right), \quad 1 \leq i \leq R. \quad (27)$$

The proof of the theorem can refer to the literature [34]. Based on Theorem 1 and Theorem 2, we will discuss the condition in which the proposed algorithm can select the correct atom for each step in the case of power-limited noise.

Theorem 3. *Suppose that observed noise follows additive white Gaussian noise $\mathbf{w}_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{N_p})$; if the remaining energy of the signal $p_{res} = \max_{l \in \mathbb{D}/\mathbb{D}^{k-1}} \{\sum_{i=1}^R \|\tilde{\mathbf{h}}_i[l]\|_2^2\}$ satisfy $p_{res} \geq (S - k + 1) \min_{1 \leq i \leq R} \{((2\sqrt{R}\epsilon)/(1 - (N_t - 1)\nu_i - (2S - 1)N_t\mu_{Bi}))^2\}$, the proposed algorithm with the stopping rule $\sum_{i=1}^R \|\mathbf{b}_i^k\|_2^2 > \epsilon^2$ selects the correct atom with probability at least $1 - (1/RN_p)$, where $\epsilon = \sigma \sqrt{RN_p + 2\sqrt{RN_p \log RN_p}}$.*

Proof. See Appendix B. \square

Remark 2. Theorem 3 gives the double threshold condition for the signal to be fully recovered according to a certain probability. Although noise power and empirical threshold

involved in signal power are usually used as terminated conditions, these thresholds are somewhat conservative. When the signal-to-noise ratio is low, it leads to the incorrect supports. Theorem 3 can provide a degree of quantification of correct recovery so as to guide the choice of thresholds.

4.2. Pilot Design Discussion for Sensing Matrices. According to (2), the design of sensing matrices $\tilde{\Psi}_i$ is related to pilot sequences $\{\mathbf{p}_i^{(p)}\}_{p=1}^{N_t}$ and pilot placement ξ_i . In CS theory, RIP is treated as a sufficient condition that the sparse signal can be reliably and stably recovered. However, RIP of a sensing matrix is so difficult to calculate that mutual incoherence property (MIP) which is stronger than RIP is widely used as a framework for sparse signal recovery. Besides, from Theorem 2, we can see that the MIP of sensing matrix is directly related to the termination threshold of the proposed algorithm. The smaller the ν_i and the μ_{Bi} , the lower the threshold of the recoverable signal and the more accurate the signal recovery. Therefore, our purpose is to design sensing matrices with good cross-correlation.

Using the Central Limit Theorem, Gao [17] proposed a constant envelope complex exponential random phase pilot $[\mathbf{p}_i^{(p)}]_{t,p} = e^{j\theta_{i,t,p}}$, $1 \leq i \leq R$, $1 \leq t \leq N_p$, $1 \leq p \leq N_t$, where $\theta_{i,t,p}$ has the independent and identically distributed (i.i.d.) uniform distribution $U[0, 2\pi]$, and proved that the pilot vectors between different antennas and different OFDM symbols have asymptotic orthogonality when pilots are placed at equal intervals.

$$\begin{aligned} \lim_{N_p \rightarrow \infty} \frac{(\boldsymbol{\Phi}_{i,l_1}^{(p_1)})^H \boldsymbol{\Phi}_{i,l_2}^{(p_2)}}{N_p} &\stackrel{(a)}{=} \lim_{N_p \rightarrow \infty} \frac{(\mathbf{p}_i^{(p_1)} * (\mathbf{F}_L|_{\xi_i})^{(l_1)})^H (\mathbf{p}_i^{(p_2)} * (\mathbf{F}_L|_{\xi_i})^{(l_2)})}{N_p} \\ &\stackrel{(b)}{=} \lim_{N_p \rightarrow \infty} \frac{\sum_{t=1}^{N_p} \exp\{j2\pi/N \tilde{l} I_i(t) + j\tilde{\theta}_{i,t,p}\}}{N_p} \\ &\stackrel{(c)}{=} \lim_{N_p \rightarrow \infty} \frac{\sum_{t=1}^{N_p} \exp\{j(2\pi/N) \tilde{l} I_i(t)\} \mathbb{E}[\exp\{j\tilde{\theta}_{i,t,p}\}]}{N_p} \\ &\stackrel{(d)}{=} 0, \end{aligned} \quad (28)$$

where $\tilde{l} = l_2 - l_1$, $\tilde{\theta}_{i,t,p} = \theta_{i,t,p_2} - \theta_{i,t,p_1}$, and $I_i(t)$ is the t th index of the set ξ_i . The equation (c) takes advantage of independent condition between $(2\pi/N) \tilde{l} I_i(t)$ and $j\tilde{\theta}_{i,t,p}$.

When $t_1 \neq t_2$, we have $E[\exp\{j\tilde{\theta}_{i,t,p}\}] = 0$; otherwise, $\lim_{N_p \rightarrow \infty} ((\boldsymbol{\Phi}_{i,l_1}^{(p_1)})^H \boldsymbol{\Phi}_{i,l_2}^{(p_2)})/N_p = \lim_{N_p \rightarrow \infty} (\sum_{t=1}^{N_p} \exp\{j(2\pi/N) \tilde{l} I_i(t)\} / N_p) = 0$.

Therefore, when the pilot number is large enough, sensing matrices $\tilde{\Psi}_i$ will have good cross-correlation. As discussed below, the number of pilots is proportional to the number of antennas. So, for massive MIMO systems, the column vectors of the sensing matrices satisfy approximately orthogonality. Besides, according to Theorem 2, we have inequality $N_t < ((1 + \nu_i)/(2S - 1)\mu_{Bi} + \nu_i)$, which means that smaller ν_i and μ_{Bi} bring a larger number of antenna.

4.3. Computational Complexity of the DS-SAMP Algorithm. The computational complexity of the algorithm is analyzed in this section. Firstly, the computational complexity of each iteration in each step is given in Table 1.

Obviously, among the steps of the PA-DS-SAMP algorithm, matrix inversion operated for least squares (LS) estimation contributes the dominant computational complexity. Gao [18] compared the number of complex multiplications of OMP, SAMP, and DSAMP algorithms in each iteration. Compared with these four algorithms, it is clear that the DS-SAMP algorithm almost has the same computational complexity due to almost the same calculation steps they have. However, with a priori common support, the number of iterations of the proposed algorithm will be greatly reduced so that the total amount of calculation will be significantly smaller.

4.4. Spectrum Efficiency and Energy Efficiency. Block-structured processing of signals can reduce pilot overhead, and the minimum pilot overhead will be derived below.

Theorem 4. [31]. For $\tilde{\Psi}_i$, $1 \leq i \leq R$, whose elements obey an i.i.d. continuous distribution, the minimization problem $\min_{\tilde{\mathbf{h}}_i} (\sum_{i=1}^R \|\tilde{\mathbf{h}}_i\|_{2,0})^{1/2}$, s.t. $\mathbf{y}_i = \tilde{\Psi}_i \tilde{\mathbf{h}}_i$ will have the unique solution if

$$S < \frac{\text{spark}(\tilde{\Psi}_1) - 1 + \text{rank}\{\tilde{\mathbf{Y}}\}}{2}, \quad (29)$$

where $\tilde{\mathbf{Y}} = [\mathbf{y}_1, \Gamma_2^{-1} \mathbf{y}_1, \dots, \Gamma_R^{-1} \mathbf{y}_R]$, Γ_i^{-1} , $2 \leq i \leq R$ is a full rank matrix satisfying $(\tilde{\Psi}_i)_{\Xi} = \Gamma_i (\tilde{\Psi}_1)_{\Xi}$ and Ξ is the common support of $\tilde{\mathbf{h}}_i$. After coarse CIR estimation, system equation becomes

$$\begin{aligned} \mathbf{y}_i &= \tilde{\Psi}_i (\tilde{\mathbf{h}}_{i,\mathbb{D}} - \tilde{\mathbf{h}}_{i,\mathbb{D}^0}) + \mathbf{w}_i \\ &= \tilde{\Psi}_i \tilde{\mathbf{h}}_{\mathbb{D}/\mathbb{D}^0} + \mathbf{w}_i, \end{aligned} \quad (30)$$

where \mathbb{D}/\mathbb{D}^0 denotes the difference set between \mathbb{D} and \mathbb{D}^0 . Due to the temporal common sparsity, the sparsity of the signal that needs to be recovered becomes $S_1 = N_t (|\mathbb{D}|_c - S_0)$ and the equation (29) can be transformed into $\text{spark}(\tilde{\Psi}_1) > 2N_t (|\mathbb{D}|_c - S_0) + 1 - \text{rank}(\tilde{\mathbf{Y}}_{\mathbb{D}/\mathbb{D}^0})$. Besides, we know that $2 \leq \text{spark}(\tilde{\Psi}_1) \leq N_p + 1$ and $\text{rank}(\tilde{\mathbf{Y}}_{\mathbb{D}/\mathbb{D}^0}) \leq \min\{S_1, R\}$; thus, we have

$$\begin{aligned} N_p &> 2N_t (|\mathbb{D}|_c - S_0) - \min\{S_1, R\} \\ &= 2N_t (S - S_0) - \min\{S_1, R\} + 1. \end{aligned} \quad (31)$$

In addition, in order to recovery the signal with known support, $N_t S_0$ pilots are needed. Therefore, the total number of pilots needs to meet the following condition:

$$\begin{aligned} N_p &> 2N_t (|\mathbb{D}|_c - S_0) - \min\{S_1, R\} + N_t S_0 \\ &\geq N_t (2S - S_0) - \min\{S_1, R\}. \end{aligned} \quad (32)$$

Subsequently, we can get the smallest required pilot overhead $N_p = N_t (2S - S_0) - \min\{S_1, R\} + 1$ which is less than the minimum pilot overhead required for unstructured CS methods.

Due to the overhead caused by the time-domain guard interval and the frequency-domain pilots, the spectral

efficient of the proposed method normalized by the ideal case without any overhead [5, 13] can be expressed in the percentage notation as

$$\eta_{SE} = \frac{N - N_p}{N + M} = \frac{N - N_t (2S - S_0) + \min\{S_1, R\} - 1}{N + M}. \quad (33)$$

Table 2 compares the spectral efficiency of several commonly used algorithms [5], where the typical wireless digital television system with the total subcarriers number $N = 4096$ is adopted. Besides, the channel model with six resolvable paths is defined by ITU, which means that the sparsity $S = 6$. Without loss of generality, the guard interval length is $M = 256$ and the number TFFT-OFDM symbols in a frame is $R = 10$. Suppose the initial sparsity is $S_0 = 2$; for large-scale 16×16 MIMO system, the spectral efficient can be calculated as $\eta_{SE} = 91.89\%$. As can be seen from Table 2, the proposed method has the highest spectral efficiency compared with other methods.

Besides, the PN sequence power and pilot power are boosted to achieve more reliable channel estimation in the TFFT-OFDM scheme. Therefore, the energy efficiency can be expressed in the percentage notation as [12]

$$\eta_{EE} = \frac{N}{N + \alpha^2 M} \times \frac{N - N_p}{N - N_p + \beta^2 N_p} \times 100\%, \quad (34)$$

where α and β denote the amplitude factor imposed on the time-domain NP and frequency-domain pilots, respectively. Generally, $\beta = 4/3$ and $\alpha = \sqrt{2}$ have been specified such as DVB-T2 standard and DTMB standard. Therefore, energy efficiency can be calculated as $\eta_{EE} = 85.21\%$. Table 3 summarizes the energy efficiency comparison for different OFDM schemes. It is clear that the proposed method has the highest energy efficiency. The reason is that less pilot overhead leads to higher energy efficiency.

5. Simulation Results

In this section, in order to verify the effectiveness of the proposed algorithm, we perform performance analysis by comparing the proposed algorithm with other seven schemes: OMP, simultaneous OMP (SOMP), distributed SAMP (DSAMP) [18], and priori information aided SOMP (PA-SOMP) [22] through computer simulation. Simulation system is configured according to the most commonly used wireless broadcasting systems with 40 antennas and the parameters are set as centric carrier frequency $f_c = 835$ MHz, signal sampling frequency bandwidth $f_s = 1/T_s = 7.56$ MHz, Doppler shift $f_d = 80$ Hz, DFT size $N = 4096$, and guard interval length $M = 256$. The percentage of pilot overhead can be calculated as $\eta = N_p / (N + M)$. The typical multiple channels with 6-taps named ITU B is used to evaluate system performance and the specific parameters can be referred as literature [12].

Figure 4 shows the mean square error (MSE) performance comparison of five channel estimation methods for the massive MIMO-OFDM system where per frame of the transmitted signal contains 10 TFFT-OFDM symbols. The Cramer-Rao Lower Bound (CRLB) is used as a benchmark

TABLE 1: Computational complexity in one iteration.

Operations	Steps	Computational complexity
Signal proxy	5	$O(RNLN_t)$
Normal operation	6, 8, 12, 13, 14, 15, 23	$O(R)$
Identification or pruning	6, 8, 9, 12	$O(L)$
LS operation	8, 10	$O(R(2LN_t(S_0 + S_1)^2 + (S_0 + S_1)^3))$
Residual computation	11	$O(RNLN_t)$

TABLE 2: Spectral efficiency comparison.

TFT-OFDM MIMO [5]	TFT-OFDM MIMO [22]	FT-MIMO [5]	Proposed TFT-OFDM MIMO
88.67%	86.76%	75%	91.89%

TABLE 3: Energy efficiency comparison.

TFT-OFDM MIMO [5]	TFT-OFDM MIMO [22]	FT-MIMO [5]	Proposed TFT-OFDM MIMO
82.90%	77.25%	55.91%	85.21%

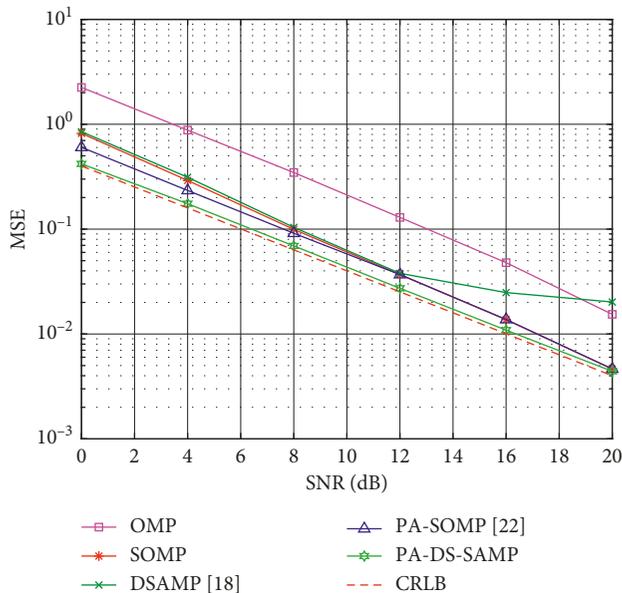


FIGURE 4: MSE performance comparison of different channel estimation methods.

for comparison. The number of pilots is 600, that is, the percentage of pilot overhead accounts for 14.6 percent. It is clear that the proposed algorithm has 1.5~2 dB signal-to-noise ratio advantage for other methods when the MSE is 10^{-1} and OMP algorithm has the worst performance. However, when the MSE is less than 10^{-2} , the performance of several algorithms tends to be the same. The reason is that those algorithms can choose the correct support when the signal-to-noise ratio is high. Therefore, we can conclude that the proposed algorithm has great advantages in the case of low and medium signal-to-noise ratio (SNR).

Figure 5 shows the bit error rate (BER) of each algorithm for the massive MIMO-OFDM system. Zero-forcing (ZF) equalizer is adopted at mobile terminals for signal detection by using the estimated channel state information. Modulation scheme adopts quadrature amplitude modulation

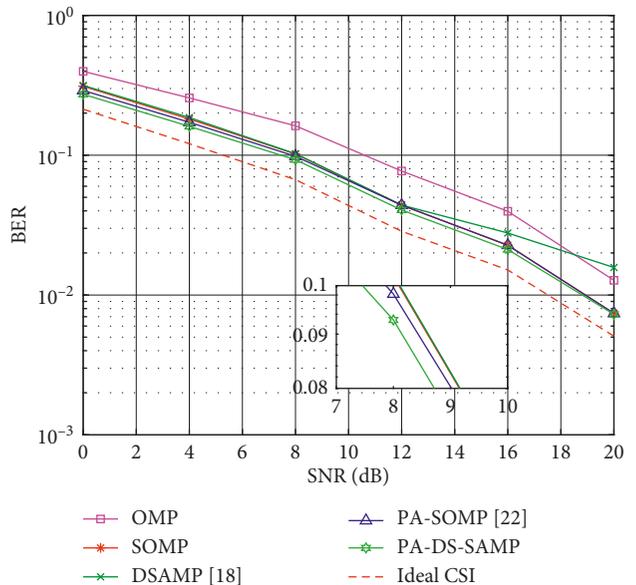


FIGURE 5: BER performance comparison of different channel estimation methods.

(QAM), and Gray code is used for source code. The BER with ideal CSI is plotted as the benchmark. Although the bit error rate curve of several algorithms is relatively close to each other, it can be seen from the performance curves in the subgraph that the proposed algorithm has the best bit error rate compared to other algorithms. This result is consistent with the analysis results in Figure 4.

To further evaluate the performance of the proposed method, Figures 6 and 7 show the MSE performance and BER performance with the variety of the number of measurements under the fixed SNR of 20 dB, respectively. As the number of pilots increases, it can be observed that the MSE curve of the proposed algorithm is closer to the MSE curve of CRLB. This means that the proposed algorithm can get the right support due to signal structured processing. The proposed algorithm can achieve an MSE accuracy of 10^{-2} with the 9.77% pilot

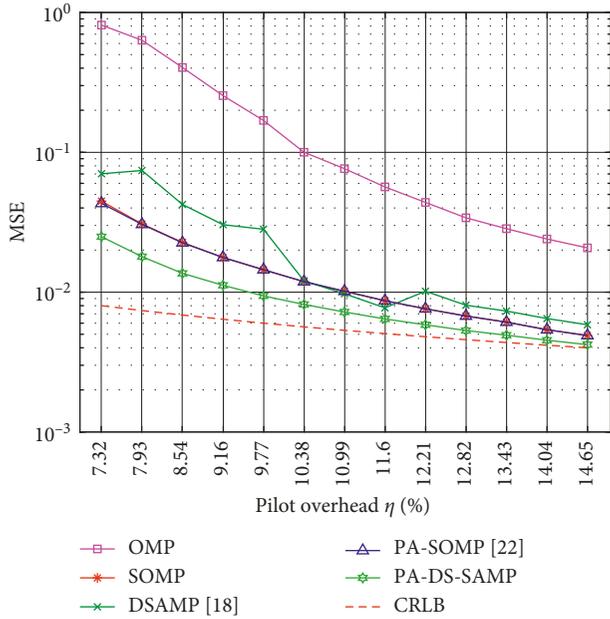


FIGURE 6: MSE performance comparison of different channel estimation methods under different numbers of measurement.

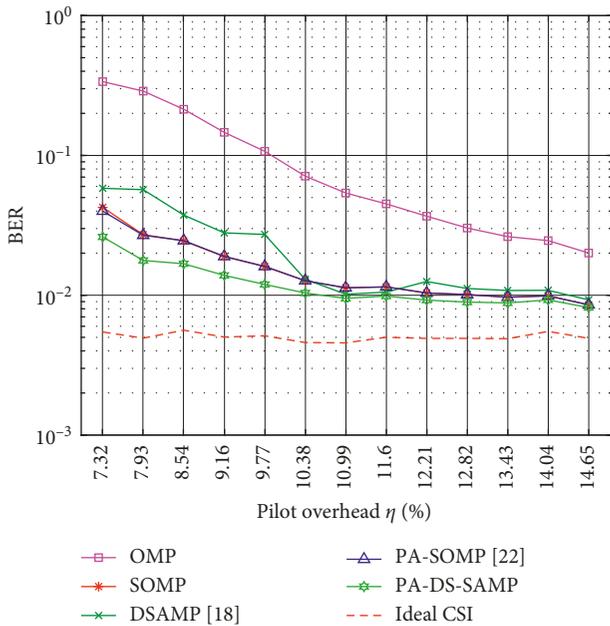


FIGURE 7: BER performance comparison of different channel estimation methods under different numbers of measurement.

overhead, while SOMP and PA-SOMP algorithms require pilot overhead of 10.99% at least and DSAMP needs more. As expected, OMP algorithm has the worst estimation performance consistent with the previous analysis. From the performance curves of the figures and the analysis above, we conclude that the proposed algorithm requires fewer pilots overhead at the same estimation accuracy.

In Figure 8, the correct signal recovery probabilities which are functions of the pilot overhead are drawn to evaluate the performance of the five methods. Due to the presence of additive white noise in the measurement signal,

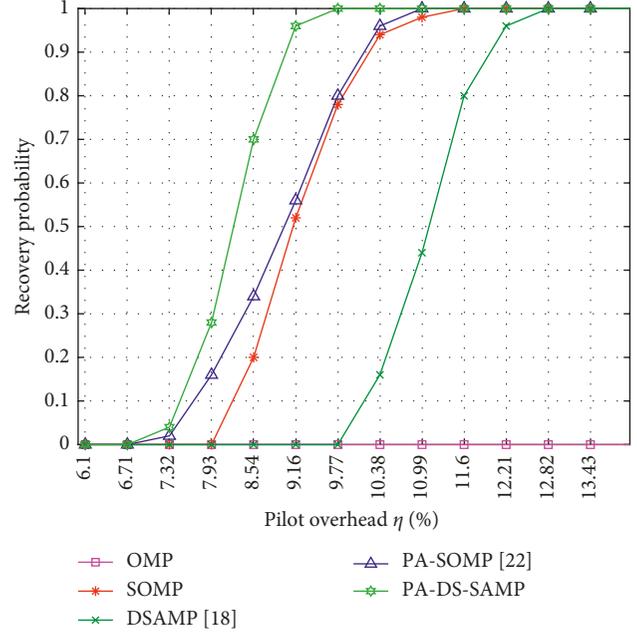


FIGURE 8: Correct signal recovery probability under different numbers of measurement.

the correct recovery of the signal is probabilistic. For more convenient discussion, the recovery probability of signal is regarded as 1 when the MSE of the signal estimation is lower than 10^{-2} . It is clear that the proposed method can accurately recover the signal when the pilot overhead is 9.77%, while the DSAMP, SOMP, and PA-SOMP algorithms need up to 12.82%, 11.6%, and 10.99% pilot overhead, respectively. This result is consistent with the performance analysis in Section 4. Therefore, compared to other algorithms, the proposed method can reduce about 2% to 3% pilot overhead so that spectral efficiency is improved.

6. Conclusions

In this paper, we proposed a time-frequency joint channel estimation scheme for massive MIMO-OFDM systems. First, partial channel common support was achieved by using changing time-domain training sequence. In order to obtain accurate CSI, we proposed a priori information aided distributed structured sparsity adaptive matching pursuit (PA-DS-SAMP) channel estimation algorithm by using obtained common path delay information. Then, we analyzed the performance of the proposed algorithm, including the convergence analysis, pilot design, computational complexity, and spectrum and energy efficiencies. Through convergence analysis, we gave two signal power thresholds which can ensure signal is fully recovered and recovered fully based on probability under the power-limited noise and the Gaussian white noise cases, respectively. In addition, pilot design results showed that the pilot sequences of different antennas tend to be orthogonal with the increase in number of antennas. Furthermore, the computational cost of the proposed algorithm is significantly reduced, and the spectrum and energy efficiencies are improved. Experimental

simulation showed that compared with other algorithms, the proposed algorithm not only increases the estimation accuracy but also greatly reduces the pilot overhead.

Appendix

A. Proof of Theorem 1

Assume that the next chosen indexes will at least contain a block in $\tilde{\Psi}_{i,\mathbb{D}}$ if \mathbf{b}_i^{k-1} is in $\mathbb{R}(\tilde{\Psi}_{i,\mathbb{D}})$. This is equivalent to the following inequality:

$$\max_{d \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_i[d])^H \mathbf{b}_i^{k-1} \right\|_2^2 > \max_{l \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_i[l])^H \mathbf{b}_i^{k-1} \right\|_2^2. \quad (\text{A.1})$$

Due to $\sum_{i=1}^R \left\| (\tilde{\Psi}_i[d])^H \mathbf{b}_i^{k-1} \right\|_2^2 = \left\| \tilde{\Psi}[d] \mathbf{b}^{k-1} \right\|_2^2$, where $\tilde{\Psi}[d] = \text{diag}\{\tilde{\Psi}_1[d], \dots, \tilde{\Psi}_R[d]\}$, $\tilde{\Psi}_{\mathbb{D}} = \text{diag}\{\tilde{\Psi}_{1,\mathbb{D}}, \dots, \tilde{\Psi}_{R,\mathbb{D}}\}$, and $\mathbf{b}^{k-1} = [(\mathbf{b}_1^{k-1})^H, \dots, (\mathbf{b}_R^{k-1})^H]^H$, inequality (A.1) can be converted to the following form:

$$\max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \mathbf{b}^{k-1} \right\|_2^2 \right\} > \max_{l \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[l])^H \mathbf{b}^{k-1} \right\|_2^2 \right\}, \quad (\text{A.2})$$

and it also can be expressed as the following form:

$$Z_{k-1} = \frac{\max_{l \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_i[l])^H \mathbf{b}_i^{k-1} \right\|_2^2}{\max_{d \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_i[d])^H \mathbf{b}_i^{k-1} \right\|_2^2} = \frac{\left\| (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2}{\left\| (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2} < 1. \quad (\text{A.3})$$

As a generalization of the matrix norm, the mixed matrix norm is defined as $\|\mathbf{A}\|_{2,p} = \max_{x \neq 0} (\|\mathbf{A}\mathbf{x}\|_{2,p} / \|\mathbf{x}\|_{2,p})$. According to the properties of the pseudoinverse, $\tilde{\Psi}_{\mathbb{D}} (\tilde{\Psi}_{\mathbb{D}})^\dagger$ is the orthogonal projector onto $\mathcal{R}(\tilde{\Psi}_{\mathbb{D}})$. Thus, it holds that $\tilde{\Psi}_{\mathbb{D}} (\tilde{\Psi}_{\mathbb{D}})^\dagger \mathbf{b}^{k-1} = \mathbf{b}^{k-1}$. Since $\tilde{\Psi}_{\mathbb{D}} (\tilde{\Psi}_{\mathbb{D}})^\dagger$ is the Hermitian matrix, the following equation holds $((\tilde{\Psi}_{\mathbb{D}})^\dagger)^H (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} = \mathbf{b}^{k-1}$. So, we have

$$\begin{aligned} Z_{k-1} &= \frac{\left\| (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2}{\left\| (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2} \\ &= \frac{\left\| (\tilde{\Psi}_{\mathbb{D}})^H (\tilde{\Psi}_{\mathbb{D}})^\dagger (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2}{\left\| \tilde{\Psi}_{\mathbb{D}} \mathbf{b}^{k-1} \right\|_{2,\infty}^2} \\ &\leq \frac{\left\| (\tilde{\Psi}_{\mathbb{D}})^H (\tilde{\Psi}_{\mathbb{D}})^\dagger \right\|_{2,\infty}^2 \left\| (\tilde{\Psi}_{\mathbb{D}})^H \mathbf{b}^{k-1} \right\|_{2,\infty}^2}{\left\| \tilde{\Psi}_{\mathbb{D}} \mathbf{b}^{k-1} \right\|_{2,\infty}^2} \\ &= \left\| (\tilde{\Psi}_{\mathbb{D}})^H (\tilde{\Psi}_{\mathbb{D}})^\dagger \right\|_{2,\infty}^2 \\ &\leq \rho_r^2 \left((\tilde{\Psi}_{\mathbb{D}})^H (\tilde{\Psi}_{\mathbb{D}})^\dagger \right)^H. \end{aligned} \quad (\text{A.4})$$

The first inequality is obtained by using the definition of mixed matrix norm $\|\mathbf{A}\mathbf{x}\|_{2,p} \leq \|\mathbf{A}\|_{2,p} \|\mathbf{x}\|_{2,p}$, and the second inequality is obtained by Lemma 1. Since $\tilde{\Psi}_{\mathbb{D}} = \text{diag}\{\tilde{\Psi}_1[\mathbb{D}], \dots, \tilde{\Psi}_R[\mathbb{D}]\}$ and $\tilde{\Psi}_{\mathbb{D}} = \text{diag}\{\tilde{\Psi}_1[\mathbb{D}], \dots, \tilde{\Psi}_R[\mathbb{D}]\}$, it results in a block diagonal matrix $(\tilde{\Psi}_{\mathbb{D}})^H ((\tilde{\Psi}_{\mathbb{D}})^\dagger)^H = \text{diag}\left\{ \tilde{\Psi}_i^H[\mathbb{D}] (\tilde{\Psi}_i^H[\mathbb{D}] \tilde{\Psi}_i[\mathbb{D}])^{-1} \tilde{\Psi}_i^H[\mathbb{D}] \right\}$, $1 \leq i \leq R$ so that formula (A.4) can be transformed into

$$\begin{aligned} Z_{k-1} &\leq \rho_r^2 \left((\tilde{\Psi}_{\mathbb{D}})^H (\tilde{\Psi}_{\mathbb{D}})^\dagger \right)^H \\ &= \max_{1 \leq i \leq R} \left\{ \rho_r^2 \left((\tilde{\Psi}_{i,\mathbb{D}})^H (\tilde{\Psi}_{i,\mathbb{D}})^\dagger \right)^H \right\} \\ &= \max_{1 \leq i \leq R} \left\{ \rho_c^2 \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\mathbb{D}} \right) \right\} \\ &\leq \max_{1 \leq i \leq R} \left\{ \rho_c \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\mathbb{D}} \right) \right\} \\ &\leq 1. \end{aligned} \quad (\text{A.5})$$

Thus, if $\max_{1 \leq i \leq R} \left\{ \rho_c \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\mathbb{D}} \right)^H \right\} < 1$ holds, PA-DS-SAMP algorithm can pick up a correct new block in each step. If the correct block can be selected for each iteration, the recursive refining process of the estimate of support set, which includes adding the correct block and removing the wrong block, will lead to estimated subspaces with strictly decreasing distance from the measurement vector so that the residuals will be decreasing. After a limited iteration, the algorithm reaches convergence.

B. Proof of Theorem 2

By using orthogonal projection matrix $\mathbf{P}_i^k = \tilde{\Psi}_{i,\mathbb{D}^k} (\tilde{\Psi}_{i,\mathbb{D}^k})^\dagger$, we have

$$\begin{aligned} \mathbf{b}_i^{k-1} &= (\mathbf{I} - \mathbf{P}_i^{k-1}) \mathbf{y}_i \\ &= (\mathbf{I} - \mathbf{P}_i^{k-1}) \tilde{\Psi}_i \tilde{\mathbf{h}}_i + (\mathbf{I} - \mathbf{P}_i^{k-1}) \mathbf{w}_i \\ &= \Delta \tilde{\mathbf{h}}_i^{k-1} + \Delta \mathbf{w}_i^{k-1}. \end{aligned} \quad (\text{B.1})$$

Let $M_{k-1,1} = \max_{d \in \mathbb{D}} \left\{ \sum_{i=1}^R \left\| (\tilde{\Psi}_i[d])^H \Delta \tilde{\mathbf{h}}_i^{k-1} \right\|_2^2 \right\} = \max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_2^2 \right\}$, $M_{k-1,2} = \max_{l \in \mathbb{D}} \left\{ \sum_{i=1}^R \left\| (\tilde{\Psi}_i[l])^H \Delta \tilde{\mathbf{h}}_i^{k-1} \right\|_2^2 \right\} = \max_{l \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[l])^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_2^2 \right\}$, and $N_{k-1,1} = \max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \Delta \mathbf{w}^{k-1} \right\|_2^2 \right\}$, then $N_{k-1,2} = \max_{l \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[l])^H \Delta \mathbf{w}^{k-1} \right\|_2^2 \right\}$, where $\Delta \tilde{\mathbf{h}}^{k-1} = [(\Delta \tilde{\mathbf{h}}_1^{k-1})^T, \dots, (\Delta \tilde{\mathbf{h}}_R^{k-1})^T]^T$ and $\Delta \mathbf{w}^{k-1} = [(\Delta \mathbf{w}_1^{k-1})^T, \dots, (\Delta \mathbf{w}_R^{k-1})^T]^T$.

The left side and right side of equation (A.2) are changed into

$$\begin{aligned}
\max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \mathbf{b}^{k-1} \right\|_2^2 \right\} &= \max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \left(\Delta \tilde{\mathbf{h}}^{k-1} + \Delta \mathbf{w}^{k-1} \right) \right\|_2^2 \right\} \\
&\geq \left(\sqrt{M_{k-1,1}} - \sqrt{N_{k-1,1}} \right)^2, \\
\max_{l \in \bar{\mathbb{D}}} \left\{ \left\| (\tilde{\Psi}[l])^H \mathbf{b}^{k-1} \right\|_2^2 \right\} &= \max_{l \in \bar{\mathbb{D}}} \left\{ \left\| (\tilde{\Psi}[l])^H \left(\Delta \tilde{\mathbf{h}}^{k-1} + \Delta \mathbf{w}^{k-1} \right) \right\|_2^2 \right\} \\
&\leq \left(\sqrt{M_{k-1,2}} + \sqrt{N_{k-1,2}} \right)^2.
\end{aligned} \tag{B.2}$$

Thus, in Step 6, the sufficient condition that at least one correct atom should be selected is $\sqrt{M_{k-1,1}} - \sqrt{M_{k-1,2}} \geq \sqrt{N_{k-1,1}} + \sqrt{N_{k-1,2}}$. Besides,

$$\begin{aligned}
Z_{k-1} M_{k-1,1} &= \left\| (\tilde{\Psi}_{\bar{\mathbb{D}}})^H \left((\tilde{\Psi}_{\mathbb{D}})^\dagger \right)^H \right\|_{2,\infty}^2 \max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}_d)^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_2^2 \right\} \\
&= \left\| (\tilde{\Psi}_{\bar{\mathbb{D}}})^H \left((\tilde{\Psi}_{\mathbb{D}})^\dagger \right)^H \right\|_{2,\infty}^2 \left\| (\tilde{\Psi}_{\mathbb{D}})^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_{2,\infty}^2 \\
&\geq \left\| (\tilde{\Psi}_{\bar{\mathbb{D}}})^H \left((\tilde{\Psi}_{\mathbb{D}})^\dagger \right)^H (\tilde{\Psi}_{\mathbb{D}})^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_{2,\infty}^2 \\
&= \left\| (\tilde{\Psi}_{\bar{\mathbb{D}}})^H \Delta \tilde{\mathbf{h}}^{k-1} \right\|_{2,\infty}^2 \\
&= M_{k-1,2}.
\end{aligned} \tag{B.3}$$

The last equation is obtained by the projection theorem. Thus, the sufficient condition for $\sqrt{M_{k-1,1}} - \sqrt{M_{k-1,2}} \geq \sqrt{N_{k-1,1}} + \sqrt{N_{k-1,2}}$ to be hold is $\sqrt{M_{k-1,1}} - \sqrt{Z_{k-1} M_{k-1,1}} \geq \sqrt{N_{k-1,1}} + \sqrt{N_{k-1,2}}$ and we have

$$\sqrt{M_{k-1,1}} \geq \frac{\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,1}}}{1 - \sqrt{Z_{k-1}}}. \tag{B.4}$$

According to [34], there is inequality (B.5) established:

$$\rho_c \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\bar{\mathbb{D}}} \right) \leq \frac{SN_{t\mu_{Bi}}}{1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}}. \tag{B.5}$$

Thus, we have

$$\begin{aligned}
Z_{k-1} &\leq \max_{1 \leq i \leq R} \left\{ \rho_c^2 \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\bar{\mathbb{D}}} \right) \right\} \\
Z_{k-1} &\leq \max_{1 \leq i \leq R} \left\{ \rho_c^2 \left((\tilde{\Psi}_{i,\mathbb{D}})^\dagger \tilde{\Psi}_{i,\bar{\mathbb{D}}} \right) \right\} \\
&\leq \max_{1 \leq i \leq R} \left\{ \frac{SN_{t\mu_{Bi}}}{1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}} \right\}^2.
\end{aligned} \tag{B.6}$$

Substituting inequality (B.7) into inequality (B.5), we have a sufficient condition that (A.2) is hold, which means that at least one correct atom should be selected in the current cycle

$$\begin{aligned}
\sqrt{M_{k-1,1}} &\geq \frac{\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,1}}}{1 - \max_{1 \leq i \leq R} \left\{ (SN_{t\mu_{Bi}}) / (1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}) \right\}} \\
&= \frac{\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,1}}}{\max_{1 \leq i \leq R} \left\{ 1 - (SN_{t\mu_{Bi}}) / (1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}) \right\}} \\
&= \min_{1 \leq i \leq R} \left\{ \frac{1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}}{1 - (N_t - 1)v_i - (2S - 1)N_{t\mu_{Bi}}} \right\} \\
&\quad \cdot \left(\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,1}} \right).
\end{aligned} \tag{B.7}$$

Meanwhile, we have

$$\begin{aligned}
M_{k-1,1} &= \max_{d \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_{i,d})^H \Delta \tilde{\mathbf{h}}_i^{k-1} \right\|_2^2 \\
&= \max_{d \in \mathbb{D}} \sum_{i=1}^R \left\| (\tilde{\Psi}_{i,d})^H (\mathbf{I} - \mathbf{P}_i^{k-1}) \tilde{\Psi}_i \tilde{\mathbf{h}}_i \right\|_2^2 \\
&= \max_{d \in \mathbb{D}/\mathbb{D}^{k-1}} \sum_{i=1}^R \left\| (\tilde{\Psi}_{i,d})^H (\mathbf{I} - \mathbf{P}_i^{k-1}) \tilde{\Psi}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_2^2 \\
&= \sum_{i=1}^R \left\| (\tilde{\Psi}_{i,\mathbb{D}/\mathbb{D}^{k-1}})^H (\mathbf{I} - \mathbf{P}_i^{k-1}) \tilde{\Psi}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_{2,\infty}^2 \\
&\geq (S - k + 1)^{-1} \sum_{i=1}^R \left\| (\tilde{\Psi}_{i,\mathbb{D}/\mathbb{D}^{k-1}})^H (\mathbf{I} - \mathbf{P}_i^{k-1}) \tilde{\Psi}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_2^2 \\
&\geq (S - k + 1)^{-1} \sum_{i=1}^R \lambda_{\min}^2 \left((\tilde{\Psi}_{i,\mathbb{D}})^H \tilde{\Psi}_{i,\mathbb{D}} \right) \left\| \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_2^2.
\end{aligned} \tag{B.8}$$

The second inequality is obtained by using Lemma 5 in [35]. According to Gershgorin circle theorem, we have $\lambda_{\min} \left((\tilde{\Psi}_{i,\mathbb{D}})^H \tilde{\Psi}_{i,\mathbb{D}} \right) \geq 1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}}$. Besides, $\mu_i \geq \mu_{Bi}$; thus, we have

$$\begin{aligned}
M_{k-1,1} &\geq (S - k + 1)^{-1} \sum_{i=1}^R \left(1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}} \right)^2 \\
&\quad \cdot \left\| \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_2^2 \\
&\geq (S - k + 1)^{-1} \max_{1 \leq i \leq R} \left\{ \left(1 - (N_t - 1)v_i - (S - 1)N_{t\mu_{Bi}} \right)^2 \right\} \\
&\quad \cdot \sum_{i=1}^R \left\| \tilde{\mathbf{h}}_{i,\mathbb{D}/\mathbb{D}^{k-1}} \right\|_2^2.
\end{aligned} \tag{B.9}$$

Combining formulas (B.8) and (B.9), the sufficient condition for selecting at least one correct atom in the current step is

$$\begin{aligned}
& \min_{1 \leq i \leq R} \left\{ \left(\frac{1 - (N_t - 1)v_i - (S - 1)N_t \mu_{Bi}}{1 - (N_t - 1)v_i - (2S - 1)N_t \mu_{Bi}} \right)^2 \right\} \\
& \cdot \left(\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,2}} \right)^2 \\
& < (S - k + 1)^{-1} \max_{1 \leq i \leq R} \left\{ \left(1 - (N_t - 1)v_i - (S - 1)N_t \mu_{Bi} \right)^2 \right\} \\
& \cdot \sum_{i=1}^R \|\tilde{\mathbf{h}}_{i, \mathbb{D}/\mathbb{D}^{k-1}}\|_2^2.
\end{aligned} \tag{B.10}$$

Thus, we have

$$\begin{aligned}
& \sum_{i=1}^R \|\tilde{\mathbf{h}}_{i, \mathbb{D}/\mathbb{D}^{k-1}}\|_2^2 \geq (S - k + 1) \\
& \cdot \frac{\min_{1 \leq i \leq R} \left\{ \left(\frac{1 - (N_t - 1)v_i - (S - 1)N_t \mu_{Bi}}{1 - (N_t - 1)v_i - (N_t - 1)N_t \mu_{Bi}} \right)^2 \right\}}{\max_{1 \leq i \leq R} \left\{ \left(1 - (N_t - 1)v_i - (S - 1)N_t \mu_{Bi} \right)^2 \right\}} \\
& \cdot \left(\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,2}} \right)^2 \\
& = (S - k + 1) \min_{1 \leq i \leq R} \left\{ \left(\frac{\sqrt{N_{k-1,2}} + \sqrt{N_{k-1,2}}}{1 - (N_t - 1)v_i - (2S - 1)N_t \mu_{Bi}} \right)^2 \right\}.
\end{aligned} \tag{B.11}$$

On the other hand, there are

$$\begin{aligned}
N_{k-1,1} &= \max_{d \in \mathbb{D}} \left\{ \left\| (\tilde{\Psi}[d])^H \Delta \mathbf{w}^{k-1} \right\|_2^2 \right\} \\
&= \max_{d \in \mathbb{D}} \left\{ \sum_{i=1}^R \left\| (\tilde{\Psi}_i[d])^H (\mathbf{I} - \mathbf{P}_i^{k-1}) \mathbf{w}_i \right\|_2^2 \right\} \\
&\leq R\epsilon^2,
\end{aligned} \tag{B.12}$$

and $N_{k-1,1} \leq R\epsilon^2$. Therefore, we have

$$\begin{aligned}
& \sum_{i=1}^R \|\tilde{\mathbf{h}}_{i, \mathbb{D}/\mathbb{D}^{k-1}}\|_2^2 = (S - k + 1) \\
& \cdot \min_{1 \leq i \leq R} \left\{ \left(\frac{2\sqrt{R}\epsilon}{1 - (N_t - 1)v_i - (2S - 1)N_t \mu_{Bi}} \right)^2 \right\}.
\end{aligned} \tag{B.13}$$

This means that if the remaining signal energy is large enough, the algorithm will choose a correct atom at this step. For conservation, we get the next result that if

$$\begin{aligned}
& \max_{l \in \mathbb{D}/\mathbb{D}^{k-1}} \left\{ \sum_{i=1}^R \|\tilde{\mathbf{h}}_i[l]\|_2^2 \right\} \\
& \geq (S - k + 1) \min_{1 \leq i \leq R} \left\{ \left(\frac{2\sqrt{R}\epsilon}{1 - (N_t - 1)v_i - (2S - 1)N_t \mu_{Bi}} \right)^2 \right\}
\end{aligned} \tag{B.14}$$

holds, then the correct atom will be selected in the next step. When all atoms are properly selected, the lowest energy of the selected atoms is $\min_{1 \leq i \leq R} \{ ((2\sqrt{R}\epsilon)/(1 - (N_t - 1)v_i - (2S - 1)N_t \mu_{Bi}))^2 \}$.

According to lemma 2, we have $P\{\|\mathbf{w}\|_2 \leq \sigma \sqrt{RN_p + 2\sqrt{RN_p \log RN_p}}\} \geq 1 - (1/(RN_p))$. This means that, if (B.14) is hold, the proposed algorithm with the stopping rule $\sum_{i=1}^R \|\mathbf{b}_i^k\|_2^2 > \sigma^2 (RN_p + 2\sqrt{RN_p \log RN_p})$ selects the correct atom with probability at least $1 - (1/RN_p)$.

Abbreviations

MIMO:	Multiple-input multiple-output
OFDM:	Orthogonal frequency division multiplexing
CSI:	Channel state information
PCCS:	Partial channel common support
SE:	Spectrum efficiency
EE:	Energy efficiency
5G:	Fifth generation
CCMs:	Channel covariance matrices
CS:	Compressive sensing
BS:	Base station
SVD:	Singular value decomposition
EVD:	Eigenvalue decomposition
TTCE:	Time-domain training channel estimation
FTCE:	Frequency-domain channel estimation
TFTCE:	Time-frequency joint training channel estimation
TS:	Training sequence
RIP:	Restricted isometry property
MIP:	Mutual incoherence property
MMV:	Multimeasurement vector
AWGN:	Additive white Gaussian noise
IBI:	Interblock interference
DBI:	Data block interference
PN:	Pseudo noise
CP-OFDM:	Cyclic prefix OFDM
ZP-OFDM:	Zero padding OFDM
TFT-OFDM:	Time-frequency training ODFM
CIR:	Channel impulse response
LS:	Least squares
OLA:	Overlapping and adding
DFT:	Discrete Fourier transform
IDFT:	Inverse discrete Fourier transform
MMV:	Multiple measurement vectors
GMMV:	Generalized MMV
OMP:	Orthogonal matching pursuit
PA-SOMP:	Priori information aided SOMP
SOMP:	Simultaneous OMP
SAMP:	Sparsity adaptive matching pursuit
DSAMP:	Distributed SAMP
DS-SAMP:	Distributed structured SAMP
PA-DS-SAMP:	Priori information aided distributed structured SAMP
SNR:	Signal-to-noise ratio

ZF:	Zero-forcing
QAM:	Quadrature amplitude modulation
BER:	Bit error rate
MSE:	Mean square error
CRLB:	Cramer–Rao lower bound.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

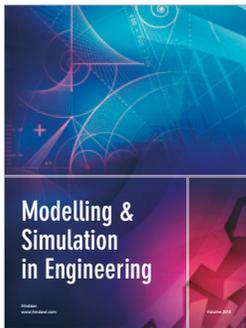
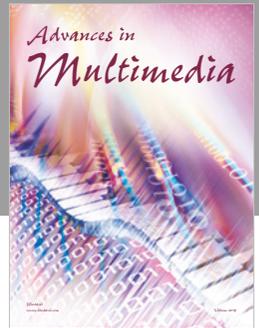
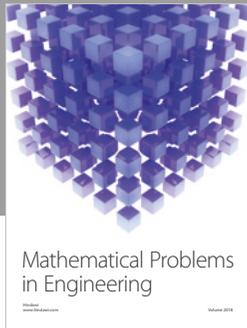
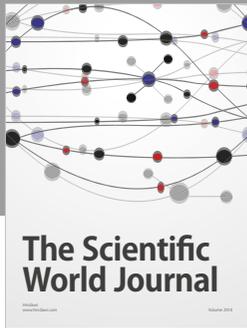
Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 61571364), Innovation Foundation for Doctoral Dissertation of Northwestern Polytechnical University (Grant no. CX201833), and Shaanxi Provincial Natural Science Foundation (Grant no. 2017JM6073).

References

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [2] M. Thomas L, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–12, 2015.
- [3] Y. Liu, Z. Tan, H. Hu, L. J. Cimini Jr., and G. Y. Li, "Channel estimation for OFDM," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, pp. 1891–1908, 2014.
- [4] 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, 2016.
- [5] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 251–263, 2013.
- [6] D. Love, R. Heath Jr., V. N. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.
- [7] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, 2015.
- [8] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation for multi cell multiuser MIMO systems with very large antenna arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3249–3252, Kyoto, Japan, March 2012.
- [9] W. Xu, W. Xiang, Y. Jia, Y. Li, and Y. Yang, "Downlink performance of Massive-MIMO systems using EVD-based channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3045–3058, 2017.
- [10] A. Liao, Z. Gao, Y. Wu, H. Wang, and M.-S. Wang, "2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding," *IEEE Access*, vol. 5, pp. 24747–24757, 2017.
- [11] L. Zhao, G. Geraci, T. Yang, D. W. K. Ng, and J. Yuan, "A tone-based AoA estimation and multiuser precoding for millimeter wave massive MIMO," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5209–5225, 2017.
- [12] L. Dai, J. Wang, Z. Wang, P. Tsiaflakis, and M. Moonen, "Energy-efficient OFDM based on simultaneous multi-channel reconstruction," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 6047–6059, 2013.
- [13] W. Wang, S. Liu, J. Song, and F. Yang, "Structured compressive sensing-based non-orthogonal time-domain training channel state information acquisition for multiple input multiple output systems," *IET Communications*, vol. 10, no. 6, pp. 685–690, 2016.
- [14] X. Ma, F. Yang, W. Ding, and J. Song, "Novel approach to design time-domain training sequence for accurate sparse channel estimation," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 512–520, 2016.
- [15] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mm wave communications: a new approach for sparse channel estimation in massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1486–1497, 2017.
- [16] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 313–321, 2014.
- [17] Z. Gao, L. Dai, W. Dai, B. Shim, Z. Wang, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 601–617, 2016.
- [18] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, 2015.
- [19] X. Rao and V. K. N. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4914–4924, 2015.
- [20] Q. Qin, L. Gui, B. Gong, X. Ren, and W. Chen, "Block distributed compressive sensing-based doubly selective channel estimation and pilot design for large-scale MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9149–9161, 2017.
- [21] W. Ding, F. Yang, W. Dai, and J. Song, "Time-frequency joint sparse channel estimation for MIMO-OFDM systems," *IEEE Communications Letters*, vol. 19, no. 1, pp. 58–61, 2015.
- [22] W. Ding, F. Yang, S. Liu, X. Wang, and J. Song, "Non-orthogonal time-frequency training-sequence-based CSI acquisition for MIMO systems," *IEEE Transactions on Vehicular Technology*, July, vol. 65, no. 7, pp. 5714–5719, 2016.
- [23] Y. Fan, H. Li, S. Song, W. Kong, and W. Zhang, "Structured compressed sensing-based time-frequency joint channel estimation for MIMO-OFDM systems," in *Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA*, pp. 2006–2010, Wuhan, China, June 2018.
- [24] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: useful tips and tricks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1527–1550, 2017.

- [25] O. Rousseeaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum-likelihood channel estimation with short training sequences," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2945–2955, 2005.
- [26] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.
- [27] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: a new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
- [28] F. Wan, W.-P. Zhu, and M. N. S. Swamy, "Semi-blind most significant tap detection for sparse channel estimation of OFDM systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 3, pp. 703–713, 2010.
- [29] L. Dai, Z. Wang, and Z. Yang, "Time-frequency training OFDM with high spectral efficiency and reliable performance in high speed environments," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 695–707, 2012.
- [30] C.-P. Li and W.-C. Huang, "A constructive representation for the fourier dual of the Zadoff-Chu sequences," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4221–4224, 2007.
- [31] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [32] T. Do, G. Lu, N. Nam, and T. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 581–587, Pacific Grove, CA, USA, October 2008.
- [33] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Novel compressed sensing-based channel estimation algorithm and near-optimal pilot placement scheme," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2590–2603, 2016.
- [34] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [35] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.



Hindawi

Submit your manuscripts at
www.hindawi.com

