IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 2, 2018, accepted March 13, 2018, date of publication March 22, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2818281

Distributed Compressed Sensing Aided Sparse Channel Estimation in FDD Massive MIMO System

RUOYU ZHANG, HONGLIN ZHAO, AND JIAYAN ZHANG
Communication Research Center, Harbin Institute of Technology, Harbin 150080, China

Corresponding author: Ruoyu Zhang (15B905028@hit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671176 and in part by the Fundamental Research Funds for the Center Universities under Grant HIT.MKSTISP.2016 13

ABSTRACT Massive multi-input multi-output (MIMO), which employs large number of antennas at the base station, can significantly boost the spectral efficiency and multiplexing gain. To fully exploit the huge array gain, the accurate channel state information is required at the transmitter side. However, the associated training overhead for downlink channel estimation consumes large amount of communication resource, especially for frequency division duplexing massive MIMO system. To address this issue, a distributed compressed sensing (DCS)-aided channel estimation approach is proposed, which fully exploits slow variation of the channel statistics in consecutive frames and spatially common sparsity within multiple subchannels in the frequency domain. Specifically, by exploiting the slow variation of the channel statistics, a hybrid training structure is proposed to probe the channel in the current frame based on the support information in previous frame. Then, a DCS-aided channel estimation algorithm, which combines least square method and DCS method, is proposed to estimate the two parts of channel vector in angular domain among different subcarriers. In addition, to effectively acquire the support information at the beginning of communication, a prior information estimation method is proposed by exploiting the uplink-downlink angular reciprocity. Simulation results demonstrate that the proposed approach outperforms the counterparts and is capable to significantly reduce the training overhead for channel estimation.

INDEX TERMS Massive MIMO, frequency division duplexing, sparse channel estimation, distributed compressed sensing.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems, which deploy a large number of antennas at the base station (BS) [1], have been extensively investigated by both academia and industry in recent years. It has been shown that with the increasing number of antennas, the energy consumption by each antenna inversely increases, while the high spectral efficiency can be enhanced by orders of magnitude [2]. Such advantages promote the massive MIMO to be a key technology for the next generation cellular communication [3].

To attain the benefits of massive MIMO, BS requires the accurate channel state information (CSI) to perform precoding and resource allocation operations [4]. However, it is challenging to acquire the downlink CSI, especially for frequency division duplex (FDD) massive MIMO system, since the high dimensional channel matrix incurs large amount of pilot overhead. This issue is not intractable in time division duplex (TDD) based massive MIMO system, since the downlink CSI can be obtained via the uplink channel training by exploiting the channel reciprocity property. Consequently, the length of pilot for uplink channel training is only proportional to the number of users, instead of number of antennas at the BS [5]. However, the complicated system calibration is required for TDD reciprocity and leads to inaccurate downlink CSI [6]. On the other hand, most of the contemporary cellular networks have adopted the FDD protocol, which is more effective for the delay-sensitive and symmetric traffic applications [7], thus it is vital to develop an effective downlink CSI estimation approach to achieve backward compatibility with current cellular networks [8].

In conventional small-scale MIMO systems, CSI is obtained by orthogonal pilots with least square (LS)-based channel estimators [9], [10]. The pilot can be orthogonal in either time domain or frequency domain [11], while the associated pilot length is proportional to the number of antennas at the BS, which is inefficient for massive MIMO system with large number of antennas. To address this issue, the pilot was designed based on the channel statistics [12] and correlated nature of massive MIMO channels [6]. These approaches require the long-term channel statistics, i.e., channel covariance matrix, which is difficult to acquire in practice and consumes training time and memory storage. Although the method in [6] can be also applied when the channel statistics is unavailable at the BS, it is appropriate for the channels with slow fading property and high temporal correlation.

An alternative way to effectively acquire the CSI in FDD massive MIMO system is based on compressed sensing (CS) [13], which enables to estimate a high dimensional vector from compressive measurements with overwhelming probability, as long as the high dimensional vector is sparse or approximately sparse [14]. Accordingly, CS theory is naturally applied to reduce the required pilot overhead of channel estimation [15]–[18]. The essential reason behind this is that the massive MIMO channels between the BS and users exhibit limited scatterers and small angle spread when compared to the large number of transmit antennas, which enables wireless channels to be represented by a sparse form in the virtual angular domain [19]–[21]. In [22], a multi-user scenario was considered, in which the channel matrices share some common scatterers and the common sparsity property among geographically neighbouring users. A joint orthogonal matching pursuit recovery algorithm was proposed to exploit the hidden joint sparsity in the channel matrices of multiple users. The user grouping strategy was considered in [23], where the channel matrices of a user group are jointly recovered by exploiting the beam block sparsity.

However, the above approaches only consider the static massive MIMO channels, where the channel sparsity information and support set remain unchanged in consecutive time slots. In [24], Rao and Lau further considered the sparse channel estimation with temporal correlation in massive MIMO systems, where the channels in several frames share some common channel paths, i.e., common support set. In this case, the prior support and the prior support quality information of sparse channel were taken into consideration to further reduce the required pilot training overhead. The prior support information was also theoretically considered in [25], where the impact on the required training overhead was examined within a weighted l_1 minimization framework, and a sharp estimate of the reduced overhead size was analytically obtained. In [26], by exploiting the slowly changing property of channel statistics [27], a CS-Aided approach was proposed to reduce the pilot overhead, which simultaneously utilizes the LS and CS approaches.

However, the above approaches only consider the narrow band channel. In [28], a spatially common sparsity based adaptive channel estimation and feedback scheme was proposed for orthogonal frequency division multiplexing (OFDM) based FDD system, which can adaptively adjust the training overhead and pilot design for reliable downlink channel estimation. This approach fully exploits the common sparsity among the subchannels of different subcarriers, which is referred as spatially common sparsity. This is caused by the fact that the spatial propagation characteristics of channels are nearly unchanged with different sub-channels [21]. In addition, by exploiting the unchanged support set in adjacent time blocks, a distributed sparsity adaptive matching pursuit algorithm was proposed to jointly estimate the channels of multiple subcarriers and time blocks [29]. On the other hand, although the channel reciprocity is unavailable for FDD system, the angles of arrivals or departures (AoAs/AoDs) of the channel paths enjoy reciprocity due to the fact that the paths experience the same scatterers in the downlink (DL) and uplink (UL) transmission [30], [31]. Based on this idea, a dictionary learning based sparse representation of massive MIMO channel was considered in [32], whereby a joint dictionary learning algorithm was proposed to reduce pilot overhead by exploiting the joint sparse pattern of the UL and DL channels. In [33], by exploiting the reciprocity of the scattering function between UL and DL channels, the required pilot overhead for obtaining the support of DL channel can be reduced.

In this paper, a distributed compressed sensing (DCS)-aided channel estimation scheme is proposed for the DL channel estimation in FDD massive MIMO system. This is based on the the observation that the channel statistics changes with a low rate during the consecutive frames and this channel statistics remains frequency-invariant due to the similar propagation characteristics among subcarriers. The channel vectors share the spatially common sparsity and partially common sparsity among several consecutive frames. To fully exploit this channel structure, a hybrid training structure is proposed to probe the channel in the current frame based on the support information in previous frame. To estimate the channel in different subcarriers, a DCS-aided channel estimation algorithm is proposed, which combines LS and DCS methods to estimate the two parts of channel vector in angular domain of current frame. In addition, to effectively acquire the support information at the beginning of communication, a prior information estimation method is proposed by exploiting the UL-DL angular reciprocity. The contribution of this paper can be summarized as follows:

Hybrid training structure: To fully exploit the slow variation of channel statistics, a hybrid training structure is proposed based on the prior support information in previous frame. The proposed hybrid training structure is a generalization of the pilot design in [28].

DCS-aided channel estimation method: By leveraging the spatially common sparsity and prior support information in previous frame, the channel to be estimated in each subcarrier at the current frame can be divided into two parts: one is the dense part containing the elements indexed by the previous support, the other one is sparse part containing the elements indexed by the complementary set of support in the previous frame. Then, the DCS-aided channel estimation algorithm combines the LS method and DCS method, which are used to estimate the dense part and sparse part of channel coefficients in the angular domain, respectively.

Prior information estimation via UL training: To effectively acquire the prior support information, UL-DL angular reciprocity is exploited to obtain the initial support estimation via UL training, which can reduce the pilot overhead when compared to the DL training.

Notations: Vectors and matrices are written in lower-case and upper-case boldface, respectively; $|\cdot|_c$ denotes the cardinality of a set, while $\|\cdot\|_0$ and $\|\cdot\|_2$ denote the l_0 norm and l_2 norm, respectively. The matrix transpose, conjugate, conjugate transpose and inversion are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^{-1}$ respectively; The (i, j)-th entry of matrix $\mathbf{A}$ and i-th entry of vector $\mathbf{a}$ are denoted by $\{\mathbf{A}\}_{i,j}$ and $\{\mathbf{a}\}_i$, respectively; $\mathbf{I}$ is an identity matrix; $\mathbb{E}[\cdot]$ defines the expectation operation; $\{\mathbf{B}\}_{\Omega,:}$ is the sub-matrix of $\mathbf{B}$ by collecting the rows indexed by set Ω and $\{\mathbf{B}\}_{:,\Omega}$ denotes the sub-matrix of $\mathbf{B}$ by extracting the columns indexed by set Ω; $\setminus$ denotes the set subtraction operation.

II. SYSTEM MODEL

A. MASSIVE MIMO SYSTEM

Consider a typical massive MIMO system where the BS is equipped with M antennas and simultaneously serves U single-antenna users with $M \gg U$. The OFDM transmission is employed and the number of subcarriers in an OFDM symbol is K. The BS performs downlink channel training by broadcasting the pilot sequences. For the j-th time slot, $j = 1, 2, \cdots, T$, the received signal of the u-th user at the k subcarrier is written as

$$y_{u,k,j} = \sqrt{\rho}\mathbf{h}_{u,k}^T\mathbf{x}_{k,j} + z_{u,k,j}, \tag{1}$$

where $\mathbf{h}_{u,k} \in \mathbb{C}^{M\times 1}$ is the downlink channel vector between the BS and the u-th user, $z_{u,k,j} \in \mathbb{C}^{1\times 1}$ is the additive Gaussian white noise with the distribution $\mathcal{CN}(0, 1)$, ρ is the signal to noise ratio (SNR), $\mathbf{x}_{k,j} \in \mathbb{C}^{M\times 1}$ is the transmitted pilot vector which satisfies power constraint $\mathbb{E}\left[\|\mathbf{x}_{k,j}\|_2^2\right] = 1$. The concatenated received signal in all T time slots $\mathbf{y}_{u,k} = \left[y_{u,k,1}, y_{u,k,2}, \cdots, y_{u,k,T}\right] \in \mathbb{C}^{1\times T}$ can be given by

$$\mathbf{y}_{u,k} = \sqrt{\rho}\mathbf{h}_{u,k}^T\mathbf{X}_k + \mathbf{z}_{u,k}, \tag{2}$$

where $\mathbf{X}_k = \left[\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \cdots, \mathbf{x}_{k,T}\right] \in \mathbb{C}^{M\times T}$ and $\mathbf{z}_{u,k} = \left[z_{u,k,1}, z_{u,k,2}, \cdots, z_{u,k,T}\right] \in \mathbb{C}^{1\times T}$ are the aggregated pilot signal and noise, respectively. To effectively obtain the CSI, the LS-based approach performs channel estimation by

$$\mathbf{h}_{u,k} = \frac{1}{\sqrt{\rho}}\left(\mathbf{y}_{u,k}\mathbf{X}_k^\dagger\right)^T, \tag{3}$$

where $\mathbf{X}_k^\dagger = \mathbf{X}_k^H\left(\mathbf{X}_k\mathbf{X}_k^H\right)^{-1}$. This pseudo-inversion demands $T \geq M$, which leads to overwhelming pilot overhead for massive MIMO system.

B. CHANNEL MODEL

A physical channel model, which captures the propagation structure between BS and user, is considered. Specifically, the downlink channel associated with the uth user at the kth subcarrier is expressed as [25], [30], [34], and [35]

$$\mathbf{h}_{u,k} = \sum_{l=1}^{N_c}\alpha_{u,l}e^{-j2\pi d_{u,l}/\lambda_c}\mathbf{e}(\phi_{u,l})e^{-j2\pi\tau_{u,l}f_sk/K}, \tag{4}$$

where N_c is the number of channel paths, $\alpha_{u,l} \sim \mathcal{CN}(0, 1)$, $d_{u,l}$, $\phi_{u,l}$ and $\tau_{u,l}$ are the channel attenuation, physical distance between BS and user, AoD and propagation delay associated with the lth path for uth user, respectively, λ_c denotes the signal wavelength, f_s is the sampling rate of system, $\mathbf{e}(\phi_{u,l}) \in \mathbb{C}^{M\times 1}$ is the antenna array vector at the BS side along with the direction of $\phi_{u,l}$, which satisfies $\|\mathbf{e}(\phi_{u,l})\|_2^2 = 1$ and can be given by

$$\mathbf{e}(\phi_{u,l}) = \frac{1}{\sqrt{M}}\left[1, e^{j\frac{2\pi d}{\lambda_c}\sin\phi_{u,l}}, \cdots, e^{j\frac{2\pi d}{\lambda_c}(M-1)\sin\phi_{u,l}}\right]^T, \tag{5}$$

where d is the antenna spacing, when uniform linear array (ULA) is employed at the BS. To elaborate the channel sparsity in massive MIMO system, the channel vector in (4) can be further represented in the angular domain [21]:

$$\mathbf{h}_{u,k} = \mathbf{F}\bar{\mathbf{h}}_{u,k}, \tag{6}$$

where $\bar{\mathbf{h}}_{u,k} \in \mathbb{C}^{M\times 1}$ is the channel vector in the angular domain, $\mathbf{F} \in \mathbb{C}^{M\times M}$ is the unitary discrete Fourier transform (DFT) matrix, the qth column of which is given by

$$\{\mathbf{F}\}_{:,q} = \frac{1}{\sqrt{M}}\left[1, e^{-j\frac{2\pi q}{M}}, \cdots, e^{-j\frac{2\pi(M-1)q}{M}}\right]^T. \tag{7}$$

Recalling (4), we have

$$\bar{\mathbf{h}}_{u,k} = \mathbf{F}^H\mathbf{E}_u\boldsymbol{\beta}_{u,k}, \tag{8}$$

where $\mathbf{E}_u = \left[\mathbf{e}(\phi_{u,1}), \mathbf{e}(\phi_{u,2}), \cdots, \mathbf{e}(\phi_{u,N_c})\right] \in \mathbb{C}^{M\times N_c}$ and $\boldsymbol{\beta}_{u,k} = \left[\beta_{u,k,1}, \beta_{u,k,2}, \cdots, \beta_{u,k,N_c}\right]^T \in \mathbb{C}^{N_c\times 1}$ with $\beta_{u,k,l} = \alpha_{u,l}e^{-j2\pi d_{u,l}/\lambda_c}e^{-j2\pi\tau_{u,l}f_sk/K}$. For a given AoD $\phi_{u,l}$, the qth element of antenna steering vector in the angular domain is given by

$$\begin{aligned}\{\mathbf{F}\}_{:,q}^H\mathbf{e}(\phi_{u,l}) &= \frac{1}{M}\sum_{p=0}^{M-1}e^{j\frac{2\pi pq}{M}}e^{j\frac{2\pi d}{\lambda_c}p\sin\phi_{u,l}} \\ &= \frac{1}{M}\frac{\sin\left(\pi M\psi_{q,\phi}\right)}{\sin\left(\pi\psi_{q,\phi}\right)}e^{j\pi(M-1)\psi_{q,\phi}},\end{aligned} \tag{9}$$

where $\psi_{q,\phi} = \frac{q}{M} + \frac{d}{\lambda_c}\sin\phi_{u,l}$. $\frac{\sin(\pi M\psi_{q,\phi})}{\sin(\pi\psi_{q,\phi})}$ achieves the maximum value 1 when $\psi_{q,\phi} = 0$ and shows non-negligible amplitude when $\left|\psi_{q,\phi}\right| \leq \frac{1}{M}$, which indicates that the path direction can be resolved by the qth angular basis vector when the AoDs ϕ satisfies $\left|\frac{q}{M} + \frac{d}{\lambda_c}\sin\phi\right| \leq \frac{1}{M}$. Due to the randomness of $\boldsymbol{\beta}_{u,k}$, the qth element of $\bar{\mathbf{h}}_{u,k}$ is a random variable. Specifically, the mean of $\{\bar{\mathbf{h}}_{u,k}\}_q$ satisfies

$\mathbb{E}\left[\{\bar{\mathbf{h}}_{u,k}\}_q\right] = \mathbb{E}\left[\sum_{l=1}^{N_c}\{\mathbf{F}\}_{:,q}^H \mathbf{e}(\phi_{u,l})\beta_{u,k,l}\right] = 0$. Then the variance of $\{\bar{\mathbf{h}}_{u,k}\}_q$ is given by

$$\begin{aligned}\mathrm{var}\left[\{\bar{\mathbf{h}}_{u,k}\}_q\right] &= \mathbb{E}\left[\left|\{\bar{\mathbf{h}}_{u,k}\}_q\right|^2\right] - \mathbb{E}\left[\{\bar{\mathbf{h}}_{u,k}\}_q\right]^2 \\ &= \mathbb{E}\left[\left|\sum_{l=1}^{N_c}\{\mathbf{F}\}_{:,q}^H \mathbf{e}(\phi_{u,l})\beta_{u,k,l}\right|^2\right] \\ &= \sum_{l=1}^{N_c}\left|\{\mathbf{F}\}_{:,q}^H \mathbf{e}(\phi_{u,l})\right|^2 \mathbb{E}\left[\left|\beta_{u,k,l}\right|^2\right] \\ &= \frac{1}{M^2}\sum_{l=1}^{N_c}\mathbb{E}\left[\left|\beta_{u,k,l}\right|^2\right]\left|\frac{\sin(\pi M\psi_{q,\phi})}{\sin(\pi\psi_{q,\phi})}\right|^2, \end{aligned} \tag{10}$$

where the significant magnitude of $\bar{\mathbf{h}}_{u,k}$ is achieved when $\phi_{u,l} \in \Omega_{u,k,q}$, i.e.,

$$\Omega_{u,k,q} = \{\phi_{u,l} | \left|\psi_{q,\phi}\right| \leq \frac{1}{M}\}. \tag{11}$$

The above equation has an important implication: the magnitude of qth element in angular domain channel $\bar{\mathbf{h}}_{u,k}$ is usually small and achieves a relative value only when the AoDs satisfies (11). The support set of channel vector of u-th at the k-th subcarrier is defined as

$$\Omega_{u,k} = \{q \in [1:M] | \left|\{\bar{\mathbf{h}}_{u,k}\}_q\right| \geq \varepsilon\}, \tag{12}$$

where ε is a constant to extract the significant elements in $\bar{\mathbf{h}}_{u,k}$ and discard the negligible elements caused by power leakage. In particular, the threshold ϵ is determined according to the desired sparsity level of channel coefficients based on the uplink observation. It can be known that $S = \left|\Omega_{u,k}\right|_c \ll M$, which illustrates channel vector $\bar{\mathbf{h}}_{u,k}$ is sparse.

Also, according to (12), for a given user u, the support set of channel vector in angular domain is irrelevant with the index of subcarriers in the frequency domain, since the location of significant entries are only determined by the scatterers environment, that is AoDs $\phi_{u,l}$, instead of the signal frequency within the system bandwidth [28]. The support information remains stationary among different subcarriers since it is uncorrelated with the subcarrier frequency. In other words, although the non-zeros indices of virtual channel vector change with the location of users, the different subcarriers share the common support information which varies simultaneously, which is referred as spatially common sparsity [28] [32]:

$$\Omega_{u,1} = \Omega_{u,2} = \cdots = \Omega_{u,K}. \tag{13}$$

This sparse property can be interpreted in Fig. 1, where there are only limited number of scatterers between the BS and uth user. For the channel vector in the angular domain, a large number of entries in $\bar{\mathbf{h}}_{u,k}$ are either zeros or close to zero. In the frequency domain, the indices of non-zeros remains unchanged within the system bandwidth due to the similar propagation characteristics among different subcarriers.

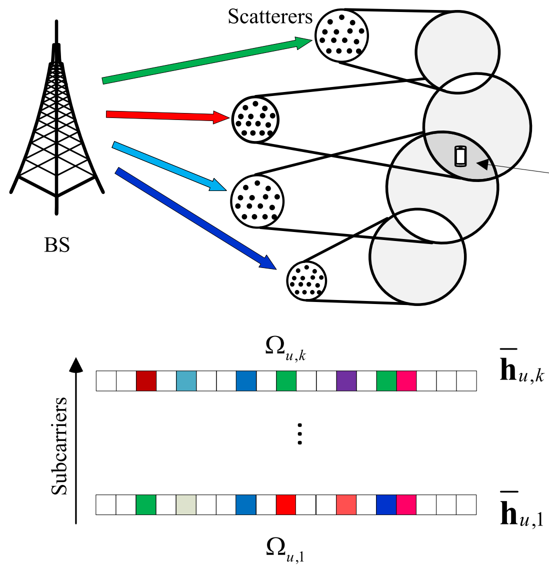

FIGURE 1. **The scatter environment of signal transmission between BS and uth user, and the associated the channel supports in different subcarriers are also illustrated below.**

C. TIME-VARYING SUPPORT

In practice, there are several communication frames between the BS to the users, where a sequence of channels are required to estimate [36]. During the consecutive frames, the channel statistics may change with a slow rate, which is manifested by the appearance and disappearance of scatterers, or move from a visible region to another visible region. This channel property has been modelled as a birth-death evolution in terms of scatterers in [27]. The birth-death of scatterers incurs the variation of support set in angular domain channel, that is the indices of non-zeros change with time in a slow rate. Specifically, in the frame t, the channel vector of a given user can be represented as

$$\mathbf{h}_k^{(t)} = \{\mathbf{F}\}_{:,\Omega_k^{(t)}}\{\bar{\mathbf{h}}_k^{(t)}\}_{\Omega_k^{(t)}}, \tag{14}$$

where the user index u in $\mathbf{h}_k^{(t)}$, $\Omega_k^{(t)}$, $\bar{\mathbf{h}}_k$ is omitted for notations simplification, $\Omega_k^{(t)}$ is the support set at the frame t, which slightly changes from the previous frame $t-1$. Denote the maximum number of non-zeros which are newly added in frame t to the support in previous frame $t-1$ as S_{n}, then we have

$$\left|\Omega_k^{(t)} \setminus \Omega_k^{(t-1)}\right|_c \leq S_{\mathrm{n}}, \tag{15}$$

where $\Omega_k^{(t-1)}$ are the support set at the previous frame $t-1$. (15) reveals that there are at most S_{n} indices of non-zero elements that belong to support set $\Omega_k^{(t)}$ but not belong to $\Omega_k^{(t-1)}$. On the other hand, the the slight change of support set between two consecutive frames means

$$S_{\mathrm{n}} \ll \left|\overline{\Omega}_k^{(t-1)}\right|_c, \tag{16}$$

where $\overline{\Omega}_k^{(t-1)} = \{1, 2, \cdots, M\} \setminus \Omega_k^{(t-1)}$ is the complement set of $\Omega_k^{(t-1)}$ at the frame $t-1$. (15) and (16) reflect the slow variation of propagation environment between the BS and a specific user. It should be noticed that this slow variation is shared by different subcarriers, since the support set change is caused by the appearance and disappearance of scatterers, which is uncorrelated with the frequency band, as illustrated in Fig. 2.

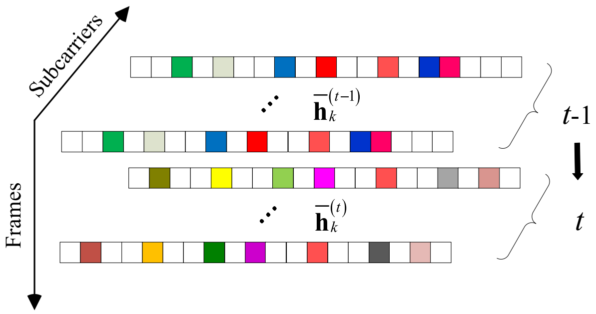

FIGURE 2. The support variation of channel vector in angular domain from frame $t-1$ to frame t. The support variation property is shared by different subcarriers.

III. DISTRIBUTED COMPRESSED SENSING AIDED CHANNEL ESTIMATION

A. DOWNLINK CHANNEL PROBING

During the information transmission, the whole process includes channel training and data transmission. The fewer training period used means that more channel uses can be available for the downlink signal transmission. The pilot signal with length T is used to probe the channel at the tth frame. According to (2), the received signal at the kth subcarrier during the tth training frame can be given by

$$\mathbf{y}_k^{(t)} = \sqrt{\rho}\left(\mathbf{h}_k^{(t)}\right)^T \mathbf{X}_k^{(t)} + \mathbf{z}_k^{(t)} = \sqrt{\rho}\left(\overline{\mathbf{h}}_k^{(t)}\right)^T \mathbf{F}^T \mathbf{X}_k^{(t)} + \mathbf{z}_k^{(t)}. \tag{17}$$

Since the angular channel vector $\overline{\mathbf{h}}_k$ is sparse, the required pilot period can be reduced. Furthermore, when the support information $\Omega_k^{(t)}$ of $\overline{\mathbf{h}}_k$ is known at the receiver, then length of pilot training can decrease to $T \geq S$, instead of $T \geq M$. In this case, the pilot matrix can be constructed as $\mathbf{X}_k^{(t)} = \{\mathbf{F}\}_{:,\Omega_k^{(t)}} \overline{\mathbf{X}}_k^{(t)}$, where $\{\mathbf{F}\}_{:,\Omega_k^{(t)}} \in \mathbb{C}^{M\times S}$ is the sub-matrix extracted from DFT matrix based on $\Omega_k^{(t)}$, $\overline{\mathbf{X}}_k^{(t)} \in \mathbb{C}^{S\times T}$ is the pilot sequences. Then the angular channel vector $\overline{\mathbf{h}}_k$ can be estimated in a LS manner with only $T \geq S$ training length.

On the other hand, due to the unknown support set $\Omega_k^{(t)}$ in practice, the channel estimation problem becomes an underdetermined problem when $T < M$, which is difficult to effectively obtain the channel station information $\mathbf{h}_k^{(t)}$. Fortunately, thanks to the sparse property of angular channel, the CS has been extensively applied to reduce the pilot overhead [16]. Specifically, the problem can be formulated as

$$\begin{aligned} &\underset{\overline{\mathbf{h}}_k^{(t)}}{\operatorname{argmin}} \quad \left\|\overline{\mathbf{h}}_k^{(t)}\right\|_0 \\ &\text{s.t. } \mathbf{y}_k^{(t)} = \sqrt{\rho}\left(\overline{\mathbf{h}}_k^{(t)}\right)^T \mathbf{F}^T \mathbf{X}_k^{(t)} + \mathbf{z}_k^{(t)}. \end{aligned} \tag{18}$$

The sparse vector $\overline{\mathbf{h}}_k^{(t)}$ can be recovery with high probability with $T = \mathcal{O}(S\log(M/S))$via sparse recovery solvers when the measurement matrix $\mathbf{X}_k^{(t)}$ satisfies restricted isometry property (RIP) [37]. The above formulation is a single measurement vector (SMV) problem. To effectively reduce the pilot overhead, the joint sparse property can be exploited according to (13). Since the different pilot signals are used to probe the channel at every subcarrier, the channel estimation problem can be further formulated as a generalized multiple measurement vectors (GMMV) [28] problem, which is a generalization of multiple measurement vector (MMV) problem in [38] by diversifying the measurement matrix. The formulation can be given as follows:

$$\begin{aligned} &\underset{\overline{\mathbf{h}}_k^{(t)},\forall k}{\operatorname{argmin}} \quad \left(\sum_{k=1}^{K}\left\|\overline{\mathbf{h}}_k^{(t)}\right\|_0^2\right)^{1/2} \\ &\text{s.t. } \mathbf{y}_k^{(t)} = \sqrt{\rho}\left(\overline{\mathbf{h}}_k^{(t)}\right)^T \mathbf{F}^T \mathbf{X}_k^{(t)} + \mathbf{z}_k^{(t)} \quad \text{and} \quad (13). \end{aligned} \tag{19}$$

When the sparse channel vectors in different subcarriers share the common support, the channel support recovery probability can be further promoted.

B. TRAINING CONSTRUCTION

However, the above CS based approach exploits the common sparsity either in frequency domain or several consecutive frames, the time varying support has not been exploited. The support in previous frame $t-1$ can be regarded as the prior support information. The support set in previous frame $\Omega_k^{(t-1)}$ can be used to effectively perform channel probing in current frame based on the channel model with time varying support elaborated in II-C. Specifically, the angular domain channel to be estimated at current frame t can be decomposed into two sub-vectors based on the prior support $\Omega_k^{(t-1)}$, i.e.,

$$\overline{\mathbf{h}}_{k,d}^{(t)} = \mathbf{I}_{\Omega_k^{(t-1)},:} \overline{\mathbf{h}}_k^{(t)}, \tag{20}$$

and

$$\overline{\mathbf{h}}_{k,s}^{(t)} = \mathbf{I}_{\overline{\Omega}_k^{(t-1)},:} \overline{\mathbf{h}}_k^{(t)}, \tag{21}$$

where $\mathbf{I} \in \mathbb{C}^{M\times M}$ is the identity matrix, $\overline{\mathbf{h}}_{k,d}^{(t)} \in \mathbb{C}^{S_n\times 1}$ is expected to be a dense vector with most of elements being non-zeros, which is non-sparse due to $\left|\text{supp}\left(\overline{\mathbf{h}}_{k,d}^{(t)}\right)\right|_c \geq S - S_n$, $\overline{\mathbf{h}}_{k,s}^{(t)} \in \mathbb{C}^{(M-S_n)\times 1}$ is expected to be the sparse part of angular domain channel due to the slow variation of channel statistics, i.e., $\left|\text{supp}\left(\overline{\mathbf{h}}_{k,s}^{(t)}\right)\right|_c \leq S_n$. To explain this, we consider the channel vector $\overline{\mathbf{h}}_k^{(t-1)}$ in the angular domain at the previous frame $t-1$. The elements of $\overline{\mathbf{h}}_k^{(t-1)}$ indexed by $\Omega_k^{(t-1)}$ are all non-zeros, while the elements of $\overline{\mathbf{h}}_k^{(t-1)}$ indexed by $\overline{\Omega}_k^{(t-1)}$ are all zeros or approximately zeros. Then at the current frame t, where the channel statistics changes and incurs to the slow variation of support, the majority elements in $\overline{\mathbf{h}}_{k,d}^{(t)}$ are still non-zeros and the entries become zero or approximately zeros only at most $S_{\rm n}$ indices. Correspondingly, the vector $\overline{\mathbf{h}}_{k,d}^{(t)}$ is a dense vector, instead of sparse vector, since $\overline{\mathbf{h}}_{k,d}^{(t)}$ is a vector extracted from the support $\Omega_k^{(t-1)}$ in previous frame. $\overline{\mathbf{h}}_{k,s}^{(t)}$ is a sparse vector containing at most $S_{\rm n}$ significant elements in dimension $T-S$.

The pilot should be designed to exploit the slow variation of support set in consecutive frames, which is illustrated in (20) and (21). The channel vector to be estimated in current frame t can be divided into two parts. Consequently, the pilot can be decomposed into two parts to adaptively probe the unknown CSI for a given subcarrier k. Specifically, the pilot is constructed as

$$\mathbf{X}_k^{(t)} = \left[\mathbf{X}_{k,d}^{(t)} \quad \mathbf{X}_{k,s}^{(t)}\right], \tag{22}$$

where $\mathbf{X}_{k,d}^{(t)} \in \mathbb{C}^{M\times S}$ is the first part of pilot and $\mathbf{X}_{k,s}^{(t)} \in \mathbb{C}^{M\times(T-S)}$ denotes the second part, which are utilized to probe the angular domain coefficients $\overline{\mathbf{h}}_{k,d}^{(t)}$ and $\overline{\mathbf{h}}_{k,s}^{(t)}$, respectively. Since the former part $\overline{\mathbf{h}}_{k,d}^{(t)}$ is a dense vector, the associated pilot $\mathbf{X}_{k,d}^{(t)}$ is constructed as

$$\mathbf{X}_{k,d}^{(t)} = \{\mathbf{F}\}_{:,\Omega_k^{(t-1)}}\overline{\mathbf{X}}_{k,d}^{(t)}, \tag{23}$$

where $\overline{\mathbf{X}}_{k,d}^{(t)} \in \mathbb{C}^{S\times S}$ is the orthonormal matrix which satisfies $\overline{\mathbf{X}}_{k,d}^{(t)}\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^H = \rho\mathbf{I}$. For the pilot of latter part which is utilized to estimate sparse vector $\overline{\mathbf{h}}_{k,s}^{(t)}$, it can be designed as

$$\mathbf{X}_{k,s}^{(t)} = \{\mathbf{F}\}_{:,\overline{\Omega}_k^{(t-1)}}\left(\mathbf{\Phi}_{k,s}^{(t)}\right)^T, \tag{24}$$

where $\mathbf{\Phi}_{k,s}^{(t)} \in \mathbb{C}^{(T-S)\times(M-S)}$ is the measurement matrix satisfying the RIP to guarantee the successful recovery of sparse vector. Denote the index set for pilot subcarrier as $\Pi = \{\xi_1, \xi_2, \cdots, \xi_{K_P}\}$. As shown in the left part of Fig. 3, when the channel sparsity fails to be exploit, the pilot overhead is $P_{total1} = K_P T_1$ with $T_1 \geq M$ [10]. For the pilot placement in [28], the non-orthogonal pilot are investigated, where the BS antennas occupy the identical placement of frequency-domain subcarriers. The associated time-frequency resource consumption of pilot is $P_{total2} = K_P T_2$. The time slots T_2 mainly dependent on the sparsity level S with a logarithm coefficient $\log(M/S)$. The proposed pilot structure is the hybrid version between the conventional time-orthogonal pilot placement and non-orthogonal pilot, the associated time slots assigned are S and $T-S$, respectively. The $T-S$ is mainly determined by the varying sparsity $S_{\rm n}$ with a logarithm coefficient $\log((M-S)/S_{\rm n})$, which is shown in the right part of Fig. 3. It should be noted when number of subcarriers is $K=1$, the proposed training construction scheme reduce to the case in [26], where a narrowband channel was considered. By contrast, the proposed training construction can effectively probe the frequency selective channel in practical wideband massive MIMO systems, in which the diversity property of measurement matrix among multiple subcarriers can be further exploited. On the other hand, when the prior support sparsity level $S=0$, the required time slots for channel training is T, which reduces to the case in [28]. Therefore, the proposed training construction is a generalization of cases in [26] and [28].

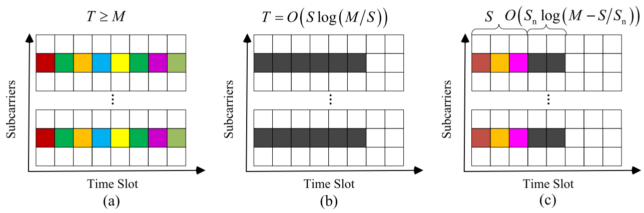

FIGURE 3. (a) Orthogonal pilot in time-domain, (b) Non orthogonal pilot [28], (c) Proposed hybrid pilot structure.

C. PRIOR INFORMATION ESTIMATION VIA UPLINK PILOT

The pilot design requires the prior information related to support set in the previous channel block, which is vital to both the downlink channel estimation and pilot design. Instead of assuming that it can be known in advance via off-line channel measurement in [24] or obatained according to the downlink estimation in the first block [26], we propose to exploit the angular reciprocity to obtain this information. Specifically, we consider the UL channel between the user and BS, which is given as

$$\mathbf{h}_k^{\rm UL} = \sum_{l=1}^{N_c}\alpha_l^{\rm UL}e^{-j2\pi d_l/\lambda_c^{\rm UL}}\mathbf{e}^{\rm UL}(\phi_l)e^{-j2\pi\tau_l f_s k/K}, \tag{25}$$

where UL denotes the notation of uplink transmission. $\alpha_l^{\rm UL} \sim \mathcal{CN}(0,1)$ is the channel attenuation in the UL transmission and independent with the attenuation in the DL channel. $\phi_{u,l}$ is the AoAs seen by the BS, $\lambda_c^{\rm UL} = c/f_c^{\rm UL}$ denotes the signal wavelength. $\mathbf{e}^{\rm UL}(\phi_l) \in \mathbb{C}^{M\times 1}$ is the antenna array vector at the BS side along with the direction of $\phi_{u,l}$ and is given by

$$\mathbf{e}^{\rm UL}(\phi_l) = \frac{1}{\sqrt{M}}\left[1, e^{j\frac{2\pi d}{\lambda_c^{\rm UL}}\sin\phi_l}, \cdots, e^{j\frac{2\pi d}{\lambda_c^{\rm UL}}(M-1)\sin\phi_l}\right]^T. \tag{26}$$

Express the UL channel into angular domain by utilizing the DFT matrix $\mathbf{F}$, then we have

$$\overline{\mathbf{h}}_k^{\rm UL} = \mathbf{F}^H\mathbf{E}^{\rm UL}\boldsymbol{\beta}_k^{\rm UL}, \tag{27}$$

where $\mathbf{E}^{\rm UL} = \left[\mathbf{e}^{\rm UL}(\phi_1), \mathbf{e}^{\rm UL}(\phi_2), \cdots, \mathbf{e}^{\rm UL}(\phi_{N_c})\right] \in \mathbb{C}^{M\times N_c}$ and $\boldsymbol{\beta}_k^{\rm UL} = \left[\beta_{k,1}^{\rm UL}, \beta_{k,2}^{\rm UL}, \cdots, \beta_{k,N_c}^{\rm UL}\right]^T \in \mathbb{C}^{N_c\times 1}$ with $\beta_{k,l}^{\rm UL} = \alpha_l^{\rm UL}e^{-j2\pi d_l/\lambda_c^{\rm UL}}e^{-j2\pi\tau_l f_s k/K}$. Due to the zero mean of

$\mathbb{E}\left[\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q\right]$, the variance of $\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q$ is given by

$$\begin{aligned}\mathrm{var}\left[\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q\right] &= \mathbb{E}\left[\left|\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q\right|^2\right] - \mathbb{E}\left[\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q\right]^2 \\ &= \mathbb{E}\left[\left|\sum_{l=1}^{N_c}\{\mathbf{F}\}_{:,q}^H \mathbf{e}^{\mathrm{UL}}(\phi_l)\beta_{k,l}^{\mathrm{UL}}\right|^2\right] \\ &= \frac{1}{M^2}\sum_{l=1}^{N_c}\mathbb{E}\left[\left|\beta_{k,l}^{\mathrm{UL}}\right|^2\right]\left|\frac{\sin\left(\pi M \psi_{q,\phi}^{\mathrm{UL}}\right)}{\sin\left(\pi \psi_{q,\phi}^{\mathrm{UL}}\right)}\right|^2,\end{aligned} \tag{28}$$

where $\psi_{q,\phi}^{\mathrm{UL}} = \frac{q}{M} + \frac{d}{\lambda_c^{\mathrm{UL}}}\sin\phi_l$. Similar to the DL transmission, $\frac{\sin\left(\pi M \psi_{q,\phi}^{\mathrm{UL}}\right)}{\sin\left(\pi \psi_{q,\phi}^{\mathrm{UL}}\right)}$ achieves significant amplitude when $\left|\psi_{q,\phi}^{\mathrm{UL}}\right| \leq \frac{1}{M}$. In other words, the path direction can be resolved by the qth angular basis vector when the AoAs ϕ satisfies $\left|\frac{q}{M} + \frac{d}{\lambda_c^{\mathrm{UL}}\sin\phi}\right| \leq \frac{1}{M}$. Under this condition, the elements of $\bar{\mathbf{h}}_k^{\mathrm{UL}}$ can achieve significant magnitude. Accordingly, the support set of channel vector in the UL at the k-th subcarrier is defined as

$$\Omega_k^{\mathrm{UL}} = \{q \in [1:M] \,|\, \left|\{\bar{\mathbf{h}}_k^{\mathrm{UL}}\}_q\right| \geq \varepsilon\}. \tag{29}$$

It can be observed that the UL support is also determined by the AoAs ϕ, which is identical with that in the DL channel. Further, with the increase of number of BS antennas, the intervals $\left|\psi_{q,\phi}^{\mathrm{UL}}\right| \leq \frac{1}{M}$ become narrower and the number of intervals increases. The resolution of the angular domain is able to resolve the angles AoDs/AoAs finely by the large antenna array at the BS, which leads to the common spatial structure and is treated as the angle reciprocity in the angular domain [28], [30], [33]. The identical AoAs/AoDs manifest the same indices of nonzeros elements in angular domain channel between UL and DL, i.e., $\mathrm{supp}(\bar{\mathbf{h}}_k^{\mathrm{UL}}) = \mathrm{supp}(\bar{\mathbf{h}}_k)$. For the UL training, the received signal at the BS always has high dimension and contains large number of measurements from the perspective of CS due to $M \gg T^{UL}$, where T^{UL} is the number of time slots for UL pilot signal. In other words, by exploiting the high dimension received signal, the UL training can be used to reduce large amount of pilot overhead if we use the UL training to obtain the support information that is identical with the DL channel. Then the prior support set can be obtained with much less UL pilot overhead due to the constraint of $\mathrm{supp}(\bar{\mathbf{h}}_k^{\mathrm{UL}}) = \mathrm{supp}(\bar{\mathbf{h}}_k)$. This is different from the prior support acquisition in [24], which assumes the support set is known in advance by off-line channel measurement. And in [26], the initial support set is obtained by performing channel probing and performing the downlink estimation, which consumes large amount of pilot overhead and has inferior performance.

To obtain the prior support information via UL channel training, the user transmits the UL pilot $\mathbf{x}_k^{\mathrm{UL}} \in \mathbb{C}^{1\times T^{\mathrm{UL}}}$ to the BS through the kth subcarrier. The orthogonal pilot satisfies $\mathbf{x}_k^{\mathrm{UL}}\left(\mathbf{x}_k^{\mathrm{UL}}\right)^H = T^{\mathrm{UL}}$. Accordingly, the received signal at the BS is expressed as

$$\begin{aligned}\mathbf{Y}_k^{\mathrm{UL}} &= \sqrt{\rho^{\mathrm{UL}}}\mathbf{h}_k^{\mathrm{UL}}\mathbf{x}_k^{\mathrm{UL}} + \mathbf{Z}_k^{\mathrm{UL}} \\ &= \sqrt{\rho^{\mathrm{UL}}}\mathbf{F}\bar{\mathbf{h}}_k^{\mathrm{UL}}\mathbf{x}_k^{\mathrm{UL}} + \mathbf{Z}_k^{\mathrm{UL}},\end{aligned} \tag{30}$$

where $\mathbf{Y}_k^{\mathrm{UL}} \in \mathbb{C}^{M\times T^{\mathrm{UL}}}$, ρ^{UL} is the UL transmit power, and $\mathbf{Z}_k^{\mathrm{UL}} \in \mathbb{C}^{M\times T^{\mathrm{UL}}}$ is the associated noise matrix with elements distributed by i.i.d $\mathcal{CN}(0, 1)$. At the BS side, the channel is estimated via

$$\begin{aligned}\hat{\mathbf{h}}_k^{\mathrm{UL}} &= \frac{1}{\sqrt{\rho^{\mathrm{UL}}}T^{\mathrm{UL}}}\mathbf{Y}_k^{\mathrm{UL}}\left(\mathbf{x}_k^{\mathrm{UL}}\right)^H \\ &= \mathbf{h}_k^{\mathrm{UL}} + \frac{1}{\sqrt{\rho^{\mathrm{UL}}T^{\mathrm{UL}}}}\mathbf{z}_k^{\mathrm{UL}},\end{aligned} \tag{31}$$

where $\mathbf{z}_k^{\mathrm{UL}} \in \mathbb{C}^{M\times 1}$ is the normalized Gaussian white noise vector. The channel vector in the angular domain can be acquired by

$$\hat{\bar{\mathbf{h}}}_k^{\mathrm{UL}} = \mathbf{F}^H\hat{\mathbf{h}}_k^{\mathrm{UL}}. \tag{32}$$

Once (32) is obtained by the UL training, the support can be calculated by the evaluating the magnitude of elements in $\hat{\bar{\mathbf{h}}}_k^{\mathrm{UL}}$, i.e.,

$$\hat{\Omega}_k^{\mathrm{UL}} = \{q \in [1:M] \,|\, \left|\{\hat{\bar{\mathbf{h}}}_k^{\mathrm{UL}}\}_q\right| \geq \varepsilon\}. \tag{33}$$

Only the entries with magnitude exceeding a certain threshold are labelled as the non-zeros elements, the indices of which construct the support set. The channel coefficients within support set dominate the total energy of channel vector in angular domain. In essence, the reason why UL training can reduce pilot overhead is that at a certain time slot, the received UL signal at the BS is an $M \times 1$ vector, while the received DL signal at the user side is a 1×1 scalar (only 1 antenna at the user side). To elaborate the benefit of UL training, we can consider an extreme case, when only 1 training slot is available, BS has M measurements to estimate the channel. By contrast, the user only has 1 measurement to estimate an M dimensional channel, which is impossible to obtain a satisfactory channel estimation. Therefore, the UL training can alleviate the required number of time slots for channel estimation when compared to the DL training.

D. PROPOSED DISTRIBUTED COMPRESSED SENSING AIDED CHANNEL ESTIMATION SCHEME

Based on the proposed training construction, which exploits the prior support information in the previous channel block, at the user side, the received training signal in the kth subcarrier of frequency domain also includes two parts, i.e.,

$$\mathbf{y}_k^{(t)} = \left[\mathbf{y}_{k,d}^{(t)}, \quad \mathbf{y}_{k,s}^{(t)}\right], \tag{34}$$

where $\mathbf{y}_{k,d}^{(t)} \in \mathbb{C}^{1\times S}$, $\mathbf{y}_{k,s}^{(t)} \in \mathbb{C}^{1\times(T-S)}$, and can be respectively expressed as

$$\begin{aligned}\mathbf{y}_{k,d}^{(t)} &= \sqrt{\rho}\left(\mathbf{h}_k^{(t)}\right)^T \mathbf{X}_{k,d}^{(t)} + \mathbf{z}_{k,d}^{(t)} \\ &= \sqrt{\rho}\left(\overline{\mathbf{h}}_k^{(t)}\right)^T \mathbf{F}^T\{\mathbf{F}\}_{:,\Omega_k^{(t-1)}}\overline{\mathbf{X}}_{k,d}^{(t)} + \mathbf{z}_{k,d}^{(t)} \\ &= \sqrt{\rho}\left(\overline{\mathbf{h}}_{k,d}^{(t)}\right)^T \overline{\mathbf{X}}_{k,d}^{(t)} + \mathbf{z}_{k,d}^{(t)},\end{aligned} \tag{35}$$

and

$$\begin{aligned}\mathbf{y}_{k,s}^{(t)} &= \sqrt{\rho}\left(\mathbf{h}_k^{(t)}\right)^T \mathbf{X}_{k,s}^{(t)} + \mathbf{z}_{k,s}^{(t)} \\ &= \sqrt{\rho}\left(\overline{\mathbf{h}}_k^{(t)}\right)^T \mathbf{F}^T\{\mathbf{F}\}_{:,\overline{\Omega}_k^{(t-1)}}\left(\mathbf{\Phi}_{k,s}^{(t)}\right)^T + \mathbf{z}_{k,s}^{(t)} \\ &= \sqrt{\rho}\left(\overline{\mathbf{h}}_{k,s}^{(t)}\right)^T \left(\mathbf{\Phi}_{k,s}^{(t)}\right)^T + \mathbf{z}_{k,s}^{(t)}.\end{aligned} \tag{36}$$

For the first dense part, the $\overline{\mathbf{h}}_{k,d}^{(t)}$ can be estimated from $\mathbf{y}_{k,d}^{(t)}$ via

$$\begin{aligned}\widehat{\overline{\mathbf{h}}}_{k,d}^{(t)} &= \frac{1}{\sqrt{\rho}}\left(\mathbf{y}_{k,s}^{(t)}\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^H\right)^T \\ &= \overline{\mathbf{h}}_{k,d}^{(t)} + \frac{1}{\sqrt{\rho}}\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^* \left(\mathbf{z}_{k,s}^{(t)}\right)^T,\end{aligned} \tag{37}$$

where $\widehat{\overline{\mathbf{h}}}_{k,d}^{(t)} \in \mathbb{C}^{S\times 1}$ is the angular channel estimation of $\overline{\mathbf{h}}_{k,d}^{(t)}$ at the subcarrier k. The same processing method is repeated by the received signal at other subcariers.

The estimation of second part $\overline{\mathbf{h}}_{k,s}^{(t)}$ can be formulated as a sparse recovery problem by utilizing CS since $\overline{\mathbf{h}}_{k,s}^{(t)}$ is a sparse vector. Instead of recovering the sparse channel vector $\overline{\mathbf{h}}_{k,s}^{(t)}$ separately, the common sparsity shared by the subchannels associated with different subcarriers can be further exploited, so that the pilot overhead to estimate the channels with at most S_n non-zeros can be reduced. Consequently, the CSI acquisition for the sparse part at different subcarriers $k \in K_P$ can be formulated as the following optimization:

$$\begin{aligned}&\underset{\overline{\mathbf{h}}_{k,s}^{(t)},\forall k}{\operatorname{argmin}} \left(\sum_{k=1}^{K}\left\|\overline{\mathbf{h}}_{k,s}^{(t)}\right\|_0^2\right)^{1/2} \\ &\text{s.t. } \mathbf{y}_{k,s}^{(t)} = \sqrt{\rho}\left(\overline{\mathbf{h}}_{k,s}^{(t)}\right)^T\left(\mathbf{\Phi}_{k,s}^{(t)}\right)^T + \mathbf{z}_{k,s}^{(t)} \quad \text{and} \quad (13).\end{aligned} \tag{38}$$

The above formulation can be solved by a DCS [29] method, which is summarized in the Algorithm 1. According to the received training signal during $T-S$ time slots among K_P subcarriers, the common support set can be jointly determined with higher probability. The first step is to initialize the parameters used in the iterations, including residual vector $\mathbf{r}_k^0$, the support set Γ^0, channel coefficients in the angular domain $\boldsymbol{\alpha}_k = \mathbf{0}$. The superscript denotes the iteration index. Then the algorithm enters the iteration stage. In each iteration, the distributed correlation is calculated for different subcarrier k. Based on the correlation between residual vector and training signal, the index of coefficients in angular domain is obtained by jointly choosing the maximum correlation value among all K_P subcarriers. Then the estimated index in step 4 is stored into the support set at step 5. The channel coefficients with corresponding support set at subcarrier k is calculated at step 6 using LS and update the residual vector in a distributed way. The residual at each subcarrier is distributed updated in step 7 and is input to the next iteration. Different from the traditional CS-based channel estimation method, the proposed method divides the unknown sparse channel into two parts, which are dense part and sparse part. Instead of estimating the whole channel vector in angular domain, the Algorithm 1 only estimates the sparse part of channel vector, which characters more sparse property according to the slow variation of channel statistics in (16).

Algorithm 1 DCS Channel Estimation Method at Frame t

Input: Received training signal $\mathbf{y}_{k,s}^{(t)}$, training signal $\mathbf{\Phi}_{k,s}^{(t)}$, $k = 1, 2, \cdots, K_P$

1: **Initialization:** residual $\mathbf{r}_k^0 = \left(\mathbf{y}_{k,s}^{(t)}\right)^T \in \mathbb{C}^{(M-S)\times 1}$, support set $\Gamma^0 = \phi$, estimation coefficients in the angular domain $\boldsymbol{\alpha}_k = \mathbf{0} \in \mathbb{C}^{(M-S)\times 1}$, iteration index $i = 1$;

2: **repeat**

3: **Distributed correlation:** $\mathbf{c}_k^i = \left(\mathbf{\Phi}_{k,s}^{(t)}\right)^H \mathbf{r}_k^{i-1}$;

4: **Index of maximum correlation among all subcarriers:** $\gamma = \underset{1\le q\le M}{\arg\max} \sum_{k=1}^{K}\left|\{\mathbf{c}_k^i\}_q\right|^2$;

5: **Update support set:** $\Gamma^i = \Gamma^{i-1} \cup \gamma$;

6: **Distributed least square:** $\{\hat{\boldsymbol{\alpha}}_k^i\}_{\Gamma^i} = \left(\{\mathbf{\Phi}_{k,s}^{(t)}\}_{:,\Gamma^i}^H\{\mathbf{\Phi}_{k,s}^{(t)}\}_{:,\Gamma^i}\right)^{-1}\{\mathbf{\Phi}_{k,s}^{(t)}\}_{:,\Gamma^i}^H\left(\mathbf{y}_{k,s}^{(t)}\right)^T$;

7: **Distributed residual update:** $\mathbf{r}_k^i = \left(\mathbf{y}_{k,s}^{(t)}\right)^T - \mathbf{\Phi}_{k,s}^{(t)}\hat{\boldsymbol{\alpha}}_k^i$, $i = i+1$;

8: **until** $i > S_n$

Output: Channel coefficients $\widehat{\overline{\mathbf{h}}}_{k,s}^{(t)} = \frac{1}{\sqrt{\rho}}\hat{\boldsymbol{\alpha}}_k^i$ for all subcarrier $k = 1, 2, \cdots K_P$.

After obtaining the estimation of dense part $\widehat{\overline{\mathbf{h}}}_{k,d}^{(t)}$ and sparse part $\widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}$ at the current frame t, the estimation of the whole channel in the angular domain is acquired by

$$\widehat{\overline{\mathbf{h}}}_k^{(t)} = \mathbf{I}_{:,\Omega_k^{(t-1)}}\widehat{\overline{\mathbf{h}}}_{k,d}^{(t)} + \mathbf{I}_{:,\overline{\Omega}_k^{(t-1)}}\widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}, \tag{39}$$

and the estimated channel vector can be further obtained by

$$\hat{\mathbf{h}}_k^{(t)} = \mathbf{F}\widehat{\overline{\mathbf{h}}}_k^{(t)}. \tag{40}$$

To summarize, the proposed DCS-aided downlink channel estimate algorithm is listed in Algorithm 2, which contains 4 steps. Specifically, according to the $\Omega_k^{(t-1)}$, the first step divides the received signal into two parts, which contain the dense part and sparse part of angular domain channel, respectively. Then, the two parts of channel coefficients are estimated by the LS method in (37) and DCS method in Algorithm 1, respectively. The finial channel estimation is obtained via (39) in step 4. In summary, the Algorithm 2 is

Algorithm 2 DCS-Aided Downlink Channel Estimation Algorithm

Input: The received training signal from different subcarriers at current frame t, training matrix, prior support information $\Omega_k^{(t-1)}$;

Divide the received signal into two parts based on (34);

Estimate the dense part of angular domain channel coefficients according to (37);

Estimate the sparse part of angular domain channel coefficients according to Algorithm 1;

Combine the channel vector according to (39) for all subcarriers;

Output: The estimate channel in angular domain $\widehat{\overline{\mathbf{h}}}_k^{(t)}$.

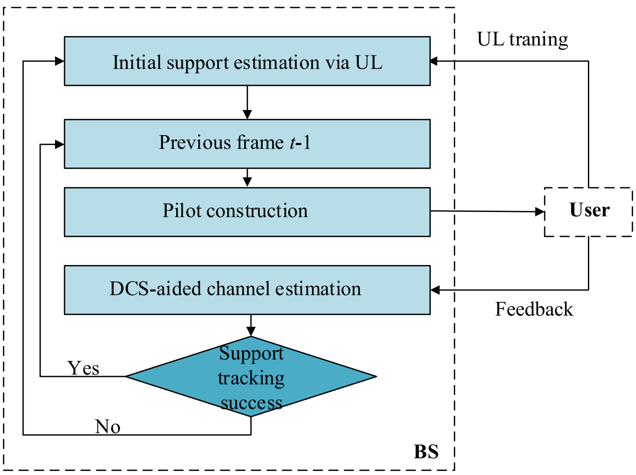

FIGURE 4. The diagram for exploiting the spatially common sparsity and slow variation of channel support in angular domain.

actually a hybrid algorithm, which combines the LS and DCS method together.

For the time-varying channel estimation, the support set in previous frame can be used to reduce the training overhead, which is illustrated in Fig. 4. When the support tracking fails or at the beginning of communication, the prior information is performed via the UL observation to acquire the accurate CSI with reduced pilot overhead. Then, this framework enters the channel estimation phase, where the support set information in previous frame is exploit to construct the DL training. The training construction can effectively probe the channel in current frame by exploiting the support information in previous frame. The received signal is directly feedback to the BS [22], and the DCS-aided channel estimation algorithm can be used to acquire the channel estimate and support information with less training overhead. In fact, the proposed channel estimation formulation is a combination of LS and GMMV based methods, which provides a general framework for FDD massive MIMO channel estimation. Specifically, when the number of subcarriers $K = 1$, the channel estimation formulation reduces to the SMV case, which is considered in [26]. On the other hand, when the sparsity variation $S_{\mathrm{n}} = S$, the proposed channel estimation scheme reduces to the case in [28], where only the common sparsity is exploited. This means the cases considered in [26] and [28] are the special cases of our proposed channel estimation scheme.

IV. PERFORMANCE ANALYSIS

In this section, the performance analysis of the proposed scheme is provided, which includes the hybrid pilot design, the error bound and the complexity analysis.

A. HYBRID PILOT DESIGN

According to the analysis in section III-B, the training signal in current frame t consists of two parts to probe the dense part and sparse part, respectively. For the first part of the training signal, the orthonormal pilot can be employed in each subcarrier, which satisfies $\overline{\mathbf{X}}_{k,d}^{(t)}\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^H = \left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^H\overline{\mathbf{X}}_{k,d}^{(t)} = \mathbf{I}$. For the second part, the sparse channel contains at most S_{n} non-zero elements and enables to reduce to pilot overhead for channel estimation. Due to $\mathbf{\Phi}_{k,s}^{(t)} \in \mathbb{C}^{(T-S)\times(M-S)}$ with $(T-S) \ll (M-S)$, to guarantee the stable recovery of the sparse part, the pilot signal $\mathbf{\Phi}_{k,s}^{(t)}$ should satisfy the following RIP condition.

Definition 1 (RIP [38]): A measurement matrix $\mathbf{\Upsilon}$ satisfies $\bar{s}$-order RIP with constant $\delta_{\bar{s}}$ $(0 < \delta_{\bar{s}} < 1)$ if

$$(1-\delta_{\bar{s}})\,\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Upsilon}\mathbf{x}\|_2^2 \leq (1+\delta_{\bar{s}})\,\|\mathbf{x}\|_2^2 \tag{41}$$

holds for all $\bar{s}$ sparse vector $\mathbf{x} \in \Sigma_{M,\bar{s}} = \{\mathbf{z} : \mathbf{z} \in \mathbb{C}^M, \|\mathbf{z}\|_0 \leq \bar{s}\}$.

It has been shown that the matrix with elements generated by independent identically distributed (i.i.d.) Gaussian distribution satisfies the RIP with overwhelming probability [38]. For $\mathbf{\Phi}_{k,s}^{(t)}$, the (a, b)th element $(1 \leq a \leq T-S, 1 \leq b \leq M-S)$ is generated by

$$\{\mathbf{\Phi}_{k,s}^{(t)}\}_{a,b} = e^{j2\pi a\theta_{b,k}}, \tag{42}$$

where $\theta_{b,k}$ is generated from i.i.d uniform distribution [0, 1]. In this case, $\{\mathbf{\Phi}_{k,s}^{(t)}\}$ is a non-equispaced Fourier matrix [39], which also satisfies the RIP. Denote the common support set of $\overline{\mathbf{h}}_{k,s}^{(t)}$ as $\Lambda_s^t = \mathrm{supp}(\overline{\mathbf{h}}_{k,s}^{(t)})$ with $\left|\Lambda_s^t\right|_c = S_{\mathrm{n}}$. For the measurement in each subcarrier k, the noiseless form of (36) can be given as

$$\begin{aligned}\left(\mathbf{y}_{k,s}^{(t)}\right)^T &= \sqrt{\rho}\mathbf{\Phi}_{k,s}^{(t)}\overline{\mathbf{h}}_{k,s}^{(t)} \\ &= \sqrt{\rho}\{\mathbf{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}\{\overline{\mathbf{h}}_{k,s}^{(t)}\}_{\Lambda_s^t},\end{aligned} \tag{43}$$

where $k \in \Pi$. If the pilot matrix $\mathbf{\Phi}_{k^*,s}^{(t)}$ at the k^*th subcarrier is chosen as the bridge, there must be full rank matrices $\mathbf{\Psi}_k$ with $k \in \Pi \setminus k^*$, satisfying $\mathbf{\Phi}_{k,s}^{(t)} = \mathbf{\Psi}_k\mathbf{\Phi}_{k^*,s}^{(t)}$, and

$$(\mathbf{\Psi}_k)^{-1}\left(\mathbf{y}_{k,s}^{(t)}\right)^T = \sqrt{\rho}\{\mathbf{\Phi}_{k^*,s}^{(t)}\}_{:,\Lambda_s^t}\{\overline{\mathbf{h}}_{k,s}^{(t)}\}_{\Lambda_s^t} = \sqrt{\rho}\mathbf{\Phi}_{k^*,s}^{(t)}\overline{\mathbf{h}}_{k,s}^{(t)}. \tag{44}$$

TABLE 1. Complexity of Algorithm 1 in iteration *i*.

Operation	Computational complexity
Step 3	$\mathcal{O}\left(2K_P(T-S)(M-S)\right)$
Step 4	$\mathcal{O}\left(K_P(M-S)\right)$
Step 5	$\mathcal{O}\left(1\right)$
Step 6	$\mathcal{O}\left(i^3+2i^2(T-S)+iK_P(T-S)\right)$
Step 7	$\mathcal{O}\left(2iK_P(T-S)\right)$
Overall	$\mathcal{O}\left((T-S)\left(K_P(M-S)+2i^2+iK_P\right)\right)$

Then, for the jointly channel estimation in (36) at different subcarrires, it can be re-formulated as

$$\mathbf{Y}_s^{(t)} = \sqrt{\rho}\boldsymbol{\Phi}_{k^*,s}^{(t)}\overline{\mathbf{H}}_s^{(t)}, \tag{45}$$

where $\mathbf{Y}_s^{(t)} = \left[(\boldsymbol{\Psi}_1)^{-1}\left(\mathbf{y}_{1,s}^{(t)}\right)^T, \cdots, \left(\mathbf{y}_{k^*,s}^{(t)}\right)^T \cdots, (\boldsymbol{\Psi}_{K_P})^{-1}\left(\mathbf{y}_{K_P,s}^{(t)}\right)^T\right]$, $\overline{\mathbf{H}}_s^{(t)} = \left[\overline{\mathbf{h}}_{1,s}^{(t)}, \cdots, \overline{\mathbf{h}}_{k^*,s}^{(t)}, \cdots, \overline{\mathbf{h}}_{K_P,s}^{(t)}\right]$. According to the Theorem 1 in [28], we have

$$2\, S_{\mathrm{n}} < \mathrm{spark}(\boldsymbol{\Phi}_{k^*,s}^{(t)}) - 1 + \mathrm{rank}(\mathbf{Y}_s^{(t)}), \tag{46}$$

where the $\mathrm{spark}(\boldsymbol{\Phi}_{k^*,s}^{(t)})$ denotes the smallest number of columns of matrix $\boldsymbol{\Phi}_{k^*,s}^{(t)}$ that are linearly dependent. According to (46), for the fixed S_{n}, the number of number of measurements can lead to $\mathrm{rank}(\mathbf{Y}_s^{(t)})$ increasing, which can reduce the number of time slots for estimating $\overline{\mathbf{H}}_s^{(t)}$.

B. ERROR BOUND

In this subsection, the mean square error (MSE) of channel estimation error by using the proposed scheme is analyzed. Although there are at most S_{n} support newly added to the channel vector of frame t, the support variation is assumed to be exactly S_{n} for the convenience. Denote the support set for dense part as $\Lambda_d^t = \mathrm{supp}(\mathbf{h}_{k,d}^{(t)})$ with $\left|\Lambda_d^t\right|_c = S - S_{\mathrm{n}}$. Specifically, the MSE at kth subcarrier is defined as

$$\mathrm{MSE}_k = \mathbb{E}\left[\left\|\mathbf{h}_k^{(t)} - \widehat{\mathbf{h}}_k^{(t)}\right\|_2^2\right], \tag{47}$$

then we have

$$\begin{aligned}
&\mathrm{MSE}_k \\
&= \mathbb{E}\left[\left\|\mathbf{F}\overline{\mathbf{h}}_k^{(t)} - \mathbf{F}\widehat{\overline{\mathbf{h}}}_k^{(t)}\right\|_2^2\right] = \mathbb{E}\left[\left\|\overline{\mathbf{h}}_k^{(t)} - \widehat{\overline{\mathbf{h}}}_k^{(t)}\right\|_2^2\right] \\
&= \mathbb{E}\left[\left\|\mathbf{I}_{:,\Omega_k^{(t-1)}}\left(\overline{\mathbf{h}}_{k,d}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,d}^{(t)}\right) + \mathbf{I}_{:,\overline{\Omega}_k^{(t-1)}}\left(\overline{\mathbf{h}}_{k,s}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}\right)\right\|_2^2\right] \\
&= \mathbb{E}\left[\left\|\left(\overline{\mathbf{h}}_{k,d}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,d}^{(t)}\right)\right\|_2^2\right] + \mathbb{E}\left[\left\|\left(\overline{\mathbf{h}}_{k,s}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}\right)\right\|_2^2\right].
\end{aligned} \tag{48}$$

For the first term, according to (37), it can be further expressed as

$$\begin{aligned}
&\mathbb{E}\left[\left\|\left(\overline{\mathbf{h}}_{k,d}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,d}^{(t)}\right)\right\|_2^2\right] \\
&\quad = \mathbb{E}\left[\left\|\{\overline{\mathbf{h}}_{k,d}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,d}^{(t)}\}_{:,\Lambda_d^t}\right\|_2^2\right] \\
&\quad = \mathbb{E}\left[\left\|\frac{1}{\sqrt{\rho}}\{\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^*\left(\mathbf{z}_{k,d}^{(t)}\right)^T\}_{:,\Lambda_d^t}\right\|_2^2\right] \\
&\quad = \frac{1}{\rho}\mathbb{E}\left[\mathrm{trace}\left(\left(\mathbf{z}_{k,d}^{(t)}\right)^*\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^T\left(\overline{\mathbf{X}}_{k,d}^{(t)}\right)^*\left(\mathbf{z}_{k,d}^{(t)}\right)^T\right)\right] \\
&\quad = \frac{S-S_{\mathrm{n}}}{\rho}.
\end{aligned} \tag{49}$$

Then, the second term can be calculated according to the output of Algorithm 1, where the support is assumed to be exactly estimated. In this case, the estimate of the channel coefficients at kth subcarrier can be obtained via the step 6 in Algorithm 1, i.e., $\widehat{\overline{\mathbf{h}}}_{k,s}^{(t)} = \frac{1}{\sqrt{\rho}}(\mathbf{I}_{M-S})_{:,\Lambda_s^t}(\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t})^{-1}\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\left(\mathbf{y}_{k,s}^{(t)}\right)^T$, then the second term can be expressed as

$$\begin{aligned}
&\mathbb{E}\left[\left\|\left(\overline{\mathbf{h}}_{k,s}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}\right)\right\|_2^2\right] \\
&\quad = \frac{1}{\rho}\mathbb{E}\left[\left\|(\mathbf{I}_{M-S})_{:,\Lambda_s^t}\left(\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}\right)^{-1}\right.\right. \\
&\qquad \left.\left.\cdot\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\left(\mathbf{z}_{k,s}^{(t)}\right)^T\right\|_2^2\right] \\
&\quad = \frac{1}{\rho}\mathrm{trace}\left((\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t})^{-1}\right).
\end{aligned} \tag{50}$$

Further, we have

$$\mathrm{trace}\left((\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t})^{-1}\right) = \sum_{s=1}^{S_{\mathrm{n}}}\frac{1}{\lambda_s}, \tag{51}$$

where λ_s is sth eigenvalue of $\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}^H\{\boldsymbol{\Phi}_{k,s}^{(t)}\}_{:,\Lambda_s^t}$. According to the Definition 1, the pilot matrix $\boldsymbol{\Phi}_{k,s}^{(t)}$ satisfies the RIP with constant $\delta_{S_{\mathrm{n}}}$, then the eigenvalues of $\boldsymbol{\Phi}_{k,s}^{(t)}$ have the following bound:

$$(1-\delta_{S_{\mathrm{n}}}) \le \lambda_s \le (1+\delta_{S_{\mathrm{n}}}). \tag{52}$$

Then the second term in (48) can be written by

$$\frac{S_{\mathrm{n}}}{\rho(1+\delta_{S_{\mathrm{n}}})} \le \mathbb{E}\left[\left\|\left(\overline{\mathbf{h}}_{k,s}^{(t)} - \widehat{\overline{\mathbf{h}}}_{k,s}^{(t)}\right)\right\|_2^2\right] \le \frac{S_{\mathrm{n}}}{\rho(1-\delta_{S_{\mathrm{n}}})}, \tag{53}$$

and consequently, the MSE can be bounded as

$$\frac{1}{\rho}\left(S - \frac{\delta_{S_{\mathrm{n}}}S_{\mathrm{n}}}{1-\delta_{S_{\mathrm{n}}}}\right) \le \mathrm{MSE}_k \le \frac{1}{\rho}\left(S + \frac{\delta_{S_{\mathrm{n}}}S_{\mathrm{n}}}{1-\delta_{S_{\mathrm{n}}}}\right). \tag{54}$$

It can be seen that the MSE performance is related to the SNR ρ, total sparsity level S, the support variation S_{n}. In addition, the decrease of S_{n} leads to the decrease of

$\frac{1}{\rho}\left(S+\frac{\delta_{S_n}S_n}{1-\delta_{S_n}}\right)$, since the constant δ_{S_n} also decreases with the order S_n. This reflects that fact the upper bound of estimation error increases when a large number of supports is shared by the consecutive frames. Also, support estimation accuracy Λ_s^t estimated by Algorithm 1 elevates with the number of subcarrires, incurring to better channel estimation performance.

C. COMPLEXITY ANALYSIS

In this subsection, the computational complexity of proposed DCS-aided channel estimation algorithm is analyzed. The computational complexity of Algorithm 1 in ith iteration is firstly investigated. Specifically, the distributed correlation in step 3 consumes $\mathcal{O}\left(2\, K_P(T-S)(M-S)\right)$ complex additions and multiplications. In step 4, the index is calculated according to the correlation of all the subcarriers and the complexity is $\mathcal{O}\left(K_P(M-S)\right)$. The complexity of support update is $\mathcal{O}(1)$. The distributed LS in step 6 consumes $\mathcal{O}\left(i^3+2i^2(T-S)+iK_P(T-S)\right)$. The complexity of residual update in step 7 is $\mathcal{O}\left(2\, iK_P(T-S)\right)$. Then, the overall complexity of ith iteration is $\mathcal{O}\left((T-S)\left(K_P(M-S)+2i^2+iK_P\right)\right)$. After S_n iterations, the total computational complexity for Algorithm 1 is $\mathcal{O}\left((T-S)(S_nK_P(M-S+\frac{S_n+1}{2})+\frac{S_n}{3}(2S_n^2+3S_n+1))\right)$. For the sake of convenience, the computational complexity for Algorithm 1 is listed in Table 1. While for the proposed Algorithm 2, the complexity for step 1 and step 2 can be omitted. The step 4 consumes $\mathcal{O}(M)$ complexity since it is a combination of sparse and dense part channel coefficients. Therefore, the overall complexity of proposed DCS-aided algorithm mainly depends on the complexity of Algorithm 1. It can be known the complexity linearly increases with training frames $T-S$, number of subcarriers K, and the number of antennas M.

V. SIMULATION RESULTS

The simulation are conducted to investigate the performance of the proposed algorithm. The BS employs ULA with $M=128$ antennas. $S=40$ and the support variation between consecutive frame is $S_n = 3$. $K = 2048$ and the number of subcarriers for pilot is $K_P = 4$. The number of users is $U = 1$ since the channel support variation is investigated between a specific user and the BS. The channel coefficients in the angular domain are generated from complex Gaussian distribution with mean 0 and unit variance accroding to [22]. In addition, the benchmark algorithms are adpoted as follows: Joint OMP (JOMP) in [22] is considered as a benchmark algorithm to individually estimate the channel from each subcarrier; Distributed sparsity adaptive matching pursuit (DSAMP) algorithm [28] jointly estimate the channel of different subcarriers. Static LS only exploits the static prior support information, which remains unchanged during the following frames. The support information can be inaccurate in the following frame due to the variation of channel statistics. Random LS is also employed as the benchmark algorithm, where the random means that the orthogonal pilots are randomly generated to probe the channel.

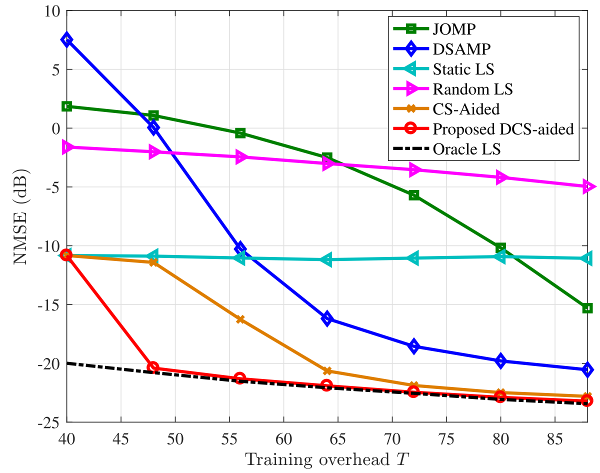

FIGURE 5. The NMSE performance versus length of training signals T, SNR = 20 dB, $S = 40$, $S_n = 3$, $M = 128$.

For JOMP and DSAMP algorithms, the pilot matrix is designed to only satisfy the RIP, while the pilot matrix satisfies the orthogonality for Static LS and Random LS. This is different from the proposed scheme, which is a combination of RIP-based and orthogonality-based pilot structure. The CS-Aided approach in [26] also adopts the hybrid pilot construction and is considered as a baseline scheme. Oracle LS indicates that the true support information is exactly known and the channel estimation is performed based on the support via LS method. This method is served as the lower bound of channel estimation. To evaluate the channel estimation performance, the normalized mean square error (NMSE) is considered as the performance metric and given by

$$\text{NMSE} = \frac{\sum_{k=1}^{K}\left\|\widehat{\mathbf{h}}_k^{(t)}-\mathbf{h}_k^{(t)}\right\|_2^2}{\sum_{k=1}^{K}\left\|\mathbf{h}_k^{(t)}\right\|_2^2}. \tag{55}$$

The normalized beamforming gain $\mathbb{E}\left[\frac{\left|(\mathbf{h}_k^{(t)})^H\widehat{\mathbf{h}}_k^{(t)}\right|^2}{\left\|\mathbf{h}_k^{(t)}\right\|_2^2\left\|\widehat{\mathbf{h}}_k^{(t)}\right\|_2^2}\right]$ is also calculated for different algorithms. 1000 independent experiments are computed by Monte Carlo simulation.

Fig. 5 compares the NMSE performance under various length of training signals when SNR = 20 dB. It can be observed that with the increase of training length, the performance of all the algorithms improves. However, the JOMP algorithm performs worst, since the sparsity level is not low enough and the JOMP algorithm fails to effectively recover the channel coefficient. DSAMP performs much better than that of JOMP, which indicates that the channel performance can be significantly improved by exploiting the spatially common sparsity in the frequency domain. An equivalent result is that the length of pilot can be reduced to achieve a desired NMSE level. Since the prior support information is exploited by the Static LS method, the NMSE of Static LS is even lower than that of DSAMP. By contrast, Random LS is agnostic to the prior information, so a satisfactory

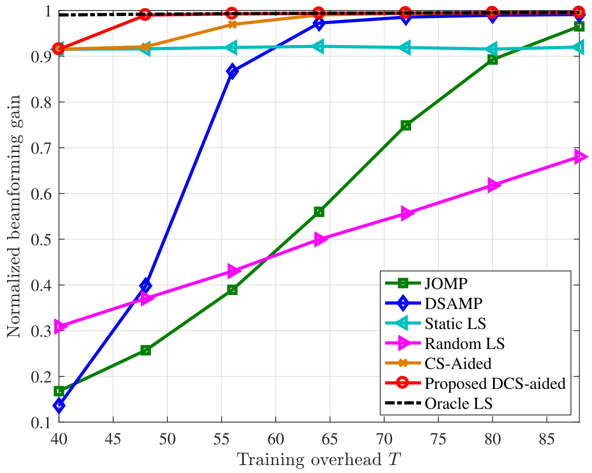

FIGURE 6. The normalized beamforming gain performance versus length of training signals T, SNR = 20 dB, $S = 40$, $S_n = 3$, $M = 128$.

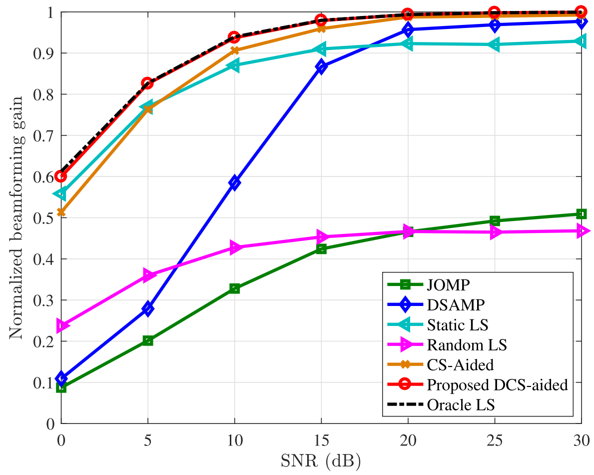

FIGURE 8. The normalized beamforming gain performance versus SNR, $T = 60$, $S = 40$, $S_n = 3$, $M = 128$.

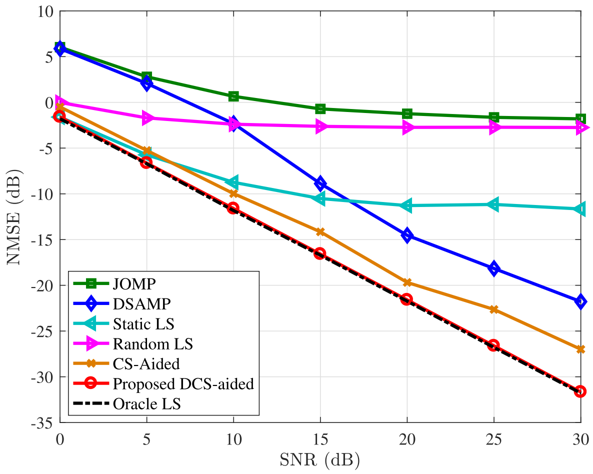

FIGURE 7. The NMSE performance versus SNR, $T = 60$, $S = 40$, $S_n = 3$, $M = 128$.

NMSE can be achieved when training length is large enough. The CS-Aided approach much better performance than the aforementioned algorithms, since it utilizes the prior information and the slow variation of support information based on CS approach. Despite of this, it still performs inferiorly to the proposed scheme. The proposed DCS-aided algorithm not only exploits the slow variation of channel statistics, but also exploits the spatially common sparsity, which can significantly improve the successful rate of support variation and achieve more accurate channel estimation. Compared to the CS-Aided method, the proposed algorithm further reduces the required training length and performs closely to the lower bound of NMSE. Accordingly, Fig. 6 presents the normalized beamforming gain of different algorithms. It can be observed that the better performance of NMSE leads to the higher beamforming gain. Among this, the proposed algorithm achieves higher gain when given a fixed training length and asymptotically approaches the optimal Oracle LS approach.

Fig. 7 illustrates the relationship between NMSE and SNR, where the length of training $T = 60$. The CS-based algorithms including JOMP and DSAMP algorithm perform worse when SNR is low. Even in high SNR regime, the NMSE performance is poor. This is due to the fact that the length of training signals is limited. By contrast, the LS-based algorithms, e.g., Static LS and Random LS algorithms perform better with the increase of SNR. However, for Static LS, there is a constant NMSE gap in high SNR regime, the reason of which is that the Static LS is agnostic to the support variation, and fails to adapt the support information in current frame. The proposed DCS-aided algorithm is a hybrid of DCS and LS method, which can adapt the channel variation by dividing the channel training into two parts. Moreover, for the recovery of sparse part, the received signals from different subcarriers can be combined to improve the NMSE performance by exploiting the spatially common sparsity. It can be observed that the proposed algorithm outperforms both the CS-based algorithm and LS-based algorithm in high SNR regime. There are three main reasons that the proposed DCS-aided algorithm achieves almost the same performance as Oracle-LS with the increase of SNR. Firstly, the algorithm not only exploits the slow variation of channel support, but also exploits the spatially common sparsity. Secondly, the training length is set $T = 60$, which is enough for the proposed DCS-aided algorithm to achieve accurate support estimation. Note in Fig. 5, when training length T is not enough, i.e., $T \in [40, 48]$, there is a big gap between the proposed algorithm and Oracle-LS algorithm. Thirdly, the exact size of support variation is provided for the proposed algorithm and other algorithms. The case of support variation mismatch will be investigated in Fig. 10. The normalized beamforming gain with various SNR is illustrated in Fig. 8. It can be seen that the proposed DCS-aided approach can achieve near optimal beamforming gain compared to the Oracle LS method. The performance gain is due to the fact that the algorithm not only exploits the slow variation of channel support, but also exploits the spatially common sparsity. The performance gain of spatially common sparsity also can be observed when comparing DSAMP and JOMP algorithm.

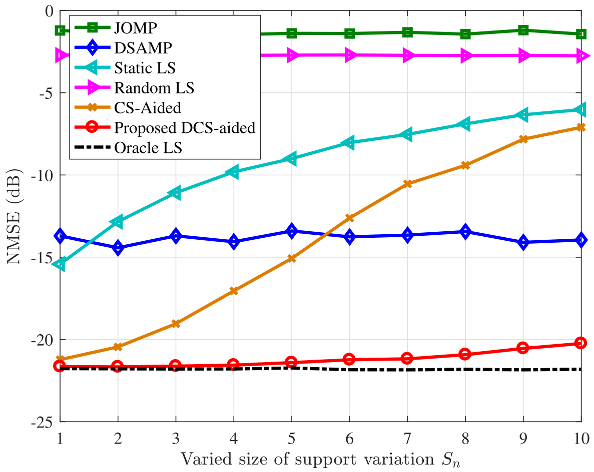

FIGURE 9. The NMSE performance versus the size of support variation S_n, $T = 60$, $S = 40$, SNR = 20 dB, $M = 128$.

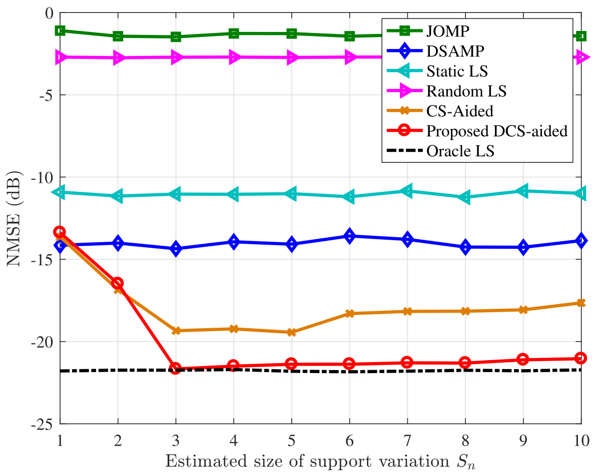

FIGURE 10. The NMSE performance versus unknown size of support variation, $T = 60$, $S = 40$, exact support variation $S_n = 3$, SNR = 20 dB, $M = 128$.

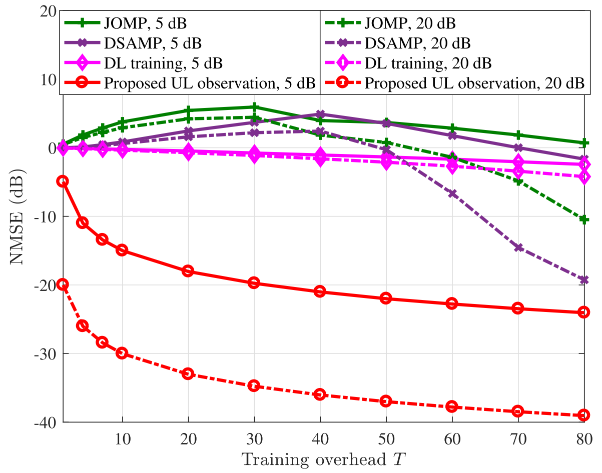

FIGURE 11. The NMSE performance versus length of training signals for estimating the initialized channel via UL and DL training, $S = 40$, $M = 128$, SNR = 5 dB, 20 dB, respectively.

Fig. 9 shows the NMSE performance when the size of support variation changes. The NMSE of JOMP and DSAMP algorithms remains almost unchanged regardless of the change of support variation level, since these CS-based algorithms are designed without considering the variation of channel statistics. The Random LS also experiences this trend it fails to adapt the channel statistics. It should be noted that the CS-Aided approach is sensitive to the increase of support variation. By contrast, the proposed algorithm is much more robust although the channel statistics has a more severe fluctuation. This can be attributed to the exploitation of spatially common sparsity.

The previous simulations assume that the size of support variation is exactly known, while Fig. 10 investigates the robust of algorithms when the estimated size of support variation $S_{\rm n}^{\rm e}$ is different with the exact $S_{\rm n}$. In this experiment, the exact support variation is set as $S_{\rm n} = 3$, while the estimated variation ranges from 1 to 10. It can be observed that proposed algorithm significantly outperforms both JOMP, Static LS and Random LS in terms of NMSE, although the mismatch of support variation occurs. When the estimated size of support variation exactly matches the true variation, i.e. $S_{\rm n}^{\rm e} = S_{\rm n}$, the proposed algorithm achieves the lowest NMSE. In the case of $S_{\rm n}^{\rm e} < S_{\rm n}$, the performance of proposed algorithm degrades significantly, and even worse than that of DSAMP when $S_{\rm n}^{\rm e} = 1$. By contrast, when $S_{\rm n}^{\rm e} > S_{\rm n}$, the performance gradually deteriorates with the increase of mismatch. But the degradation of proposed algorithm is marginal when support variation is overestimated. The robust of the proposed algorithm owes to both the hybrid training construction and the exploitation of spatially common sparsity. The important implication is that the overestimated support variation leads to marginal performance degradation but the underestimated support variation significantly deteriorates the NMSE performance.

We now examine the performance of prior information acquisition via UL training, where the prior support information is to be estimated. The benchmark schemes include the JOMP, DSAMP and DL channel estimation via Random LS method in [26]. Since the prior support is unavailable, the Static LS and CS-Aided algorithm can not be applied. By exploiting the angular reciprocity, the prior information can be obtained via the UL training. In Fig. 11, the same training overhead is used for channel estimation for various algorithms. The difference is that the proposed training method employs UL training while the other methods adopt the DL training to performance channel estimation. It can be seen the proposed channel estimation via UL observation can significantly outperform the counterparts. When SNR = 5 dB, to achieve NMSE = −10 dB, the proposed scheme only utilizes $T = 3$ training slots, while the counterpart algorithms can not achieve the goal even when $T = 80$. When SNR = 20 dB, the proposed scheme can achieve −20 dB NMSE by using only $T = 1$ training slot, while the DSAMP consumes $T = 80$ training slots to attain the target NMSE. This demonstrates the effectiveness of prior information acquisition via UL observation.

VI. CONCLUSION

A DCS-aided channel estimation approach has been proposed for FDD massive MIMO system, which fully exploits the slow variation of channel statistics in consecutive frames and spatially common sparsity within multiple subchannels, resulting to the significant pilot reduction for channel estimation. Specifically, a hybrid training structure is firstly proposed to probe the channel in the current frame based on the support information in previous frame, which exploits the slow variation of the channel statistics. Then, a DCS-aided channel estimation algorithm can be applied to estimate the channel vector in the current frame, which consists of dense part and sparse part from the angular domain. The two parts of channel among multiple subcarriers can be jointly estimated by the LS method and DCS method, respectively, to reduce the training overhead. For the prior support information, a prior information estimation method via UL training is proposed by exploiting the UL-DL angular reciprocity. Simulation results demonstrate that the proposed approach can estimate the channel with significant training overhead reduction.

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] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] 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] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[6] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[7] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[8] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[9] Y.-H. Nam, Y. Akimoto, Y. Kim, M.-I. Lee, K. Bhattad, and A. Ekpenyong, "Evolution of reference signals for LTE-advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 132–138, Feb. 2012.

[10] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[11] M. Simko, P. S. R. Diniz, Q. Wang, and M. Rupp, "Adaptive pilot-symbol patterns for MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4705–4715, Sep. 2013.

[12] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[13] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[14] 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.

[15] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[16] 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.

[17] R. Zhang, H. Zhao, S. Jia, and C. Shan, "Joint channel estimation algorithm based on structured compressed sensing for FDD multi-user massive MIMO," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 1202–1207.

[18] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 3rd Quart., 2017.

[19] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., Feb. 2007.

[20] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[22] 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.

[23] W. Huang, Y. Huang, W. Xu, and L. Yang, "Beam-blocked channel estimation for FDD massive MIMO with compressed feedback," *IEEE Access*, vol. 5, pp. 11791–11804, 2017.

[24] 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 Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

[25] J.-C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4145–4156, Jun. 2016.

[26] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.

[27] S. Wu, C. X. Wang, H. Haas, E. H. M. Aggoune, M. M. Alwakeel, and B. Ai, "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434–1446, Mar. 2015.

[28] 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.

[29] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk. (2009). "Distributed compressive sensing." [Online]. Available: https://arxiv.org/abs/0901.3403

[30] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[31] X. Luo, P. Cai, X. Zhang, D. Hu, and C. Shen, "A scalable framework for CSI feedback in FDD massive MIMO via DL path aligning," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4702–4716, Sep. 2017.

[32] Y. Ding and B. D. Rao, "Channel estimation using joint dictionary learning in FDD massive MIMO systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 185–189.

[33] M. B. Khalilsarai, S. Haghighatshoar, and G. Caire. (2017). "Efficient downlink channel probing and uplink feedback in FDD massive MIMO systems." [Online]. Available: https://arxiv.org/abs/1708.04444

[34] C. Sun, X. Q. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[35] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[36] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, document 25.996, 3GPP ETSI Release 9, Universal Mobile Telecommunications System, 2010. [Online]. Available: http://www.3gpp.org/DynaReport/25996.htm
[37] D. L. Donoho, ''Compressed sensing,'' *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[38] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, ''Block-sparse signals: Uncertainty relations and efficient recovery,'' *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[39] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9. Berlin, Germany: Walter de Gruyter, 2010.

RUOYU ZHANG received the B.E. degree in communication engineering from the Harbin Institute of Technology, Harbin, China, in 2014, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include compressed sensing, channel estimation, and massive MIMO system.

HONGLIN ZHAO received the B.E. and Ph.D. degrees in communication engineering from the Harbin Institute of Technology, Harbin, China, in 1991 and 1998, respectively. He is currently a Professor with the Communication Research Center, School of Electronics and Information Engineering, Harbin Institute of Technology. His research interests include spread spectrum communication, cognitive radio networks, and wireless broadband networks.

JIAYAN ZHANG received the Ph.D. degree in communication engineering from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently an Associate Professor with the Communication Research Center, School of Electronics and Information Engineering, Harbin Institute of Technology. His research interests include spread spectrum communication, high speed signal processing, and channel coding.