Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO

Zhen Gao, *Student Member, IEEE*, Linglong Dai, *Senior Member, IEEE*, Zhaocheng Wang, *Senior Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

Abstract—This paper proposes a spatially common sparsity based adaptive channel estimation and feedback scheme for frequency division duplex based massive multi-input multi-output (MIMO) systems, which adapts training overhead and pilot design to reliably estimate and feed back the downlink channel state information (CSI) with significantly reduced overhead. Specifically, a nonorthogonal downlink pilot design is first proposed, which is very different from standard orthogonal pilots. By exploiting the spatially common sparsity of massive MIMO channels, a compressive sensing (CS) based adaptive CSI acquisition scheme is proposed, where the consumed time slot overhead only adaptively depends on the sparsity level of the channels. In addition, a distributed sparsity adaptive matching pursuit algorithm is proposed to jointly estimate the channels of multiple subcarriers. Furthermore, by exploiting the temporal channel correlation, a closed-loop channel tracking scheme is provided, which adaptively designs the nonorthogonal pilot according to the previous channel estimation to achieve an enhanced CSI acquisition. Finally, we generalize the results of the multiple-measurement-vectors case in CS and derive the Cramér–Rao lower bound of the proposed scheme, which enlightens us to design the nonorthogonal pilot signals for the improved performance. Simulation results demonstrate that the proposed scheme outperforms its counterparts, and it is capable of approaching the performance bound.

Index Terms—Channel estimation, compressive sensing, feedback, frequency division duplex, massive multi-input multi-output, spatially common sparsity, temporal correlation.

I. Introduction

BY exploiting the increased degree of freedom in the spatial domain, massive multi-input multi-output (MIMO) can enhance the spectrum efficiency and energy efficiency by orders of magnitude [1], [2]. To harvest the benefits of massive MIMO, the base station (BS) needs the accurate channel state

information (CSI) in the downlink for beamforming, resource allocation, and other operations. However, it is challenging for the BS to acquire the accurate downlink CSI in frequency division duplex (FDD) based massive MIMO, since the overhead used for the downlink channel estimation and feedback can be prohibitively high. Most of the researches sidestep this challenge by assuming the time division duplex (TDD) protocol. In TDD based massive MIMO, the CSI in the uplink can be more easily acquired at the BS due to the limited number of users, and the channel reciprocity property can be exploited to realize the downlink channel estimation using the uplink channel estimation [1]–[4]. However, in TDD massive MIMO, the CSI acquired in the uplink may not be always accurate for the downlink due to the calibration error of radio frequency chains [5]. In addition, frequency division duplex (FDD) protocol still dominates current wireless networks, where the downlink channel estimation is necessary, since the channel reciprocity does not hold. Thus, it is of great importance to explore an efficient approach to enable massive MIMO to be backward compatible with current wireless networks [6]. In this paper, we focus on the reliable and efficient channel estimation and feedback for FDD massive MIMO.

Channel estimations in small-scale MIMO are usually based on orthogonal pilots [7]–[10]. In Long Term Evolution-Advanced (LTE-A), for example, pilots associated with different BS antennas occupy the different frequency-domain subcarriers [9]. Pilot signals can be also orthogonal in the time or code domain. However, the overhead of orthogonal pilots increases with the number of BS antennas, which becomes unaffordable for massive MIMO. For FDD massive MIMO, the work [11] proposed a pilot design for the downlink channel estimation by exploiting the channel statistics. However, the acquisition of the downlink channel covariance matrix is challenging in practice. An open-loop and closed-loop training based channel estimation scheme was proposed in [5]. Nevertheless, the long-term channel statistics required by the user may increase the training time and memory cost. In [6], a sparse channel estimation scheme was proposed to acquire CSI with significantly reduced pilot overhead by exploiting the sparsity of time-domain channel impulse response (CIR). But this time-domain sparsity of the channels may not exist when the number of scatterers at the user side becomes large. Furthermore, [5], [6], [11] do not consider the channel feedback to the BS. In order to obtain the fine-grain spatial channel structures, the conventional codebook based CSI feedback schemes may become impossible, since the dimension of codebook can be huge in massive MIMO. Hence the design, storage, and encoding of the high-dimensional

codebook can be difficult [12]. The compressive sensing (CS) based channel feedback schemes for massive MIMO were proposed to reduce the feedback overhead by exploiting the spatial correlation of CSI [12], [13]. However, these schemes do not consider downlink channel estimation. By exploiting the spatially joint sparsity of multiple users' channel matrices, the works [14], [15] proposed a joint orthogonal matching pursuit (OMP) based CSI acquisition scheme. However, this scheme cannot adaptively adjust the required overhead according to the sparsity level of the channels. Moreover, the spatially joint sparsity may disappear when the users are not spatially close. Even for the best case that multiple users' channel matrices share the spatially common sparsity, the sparse CSI acquisition problem is a multiple-measurement-vectors (MMV) problem, where the reduction in required overhead is limited.

Recent study and experiments have shown that the wireless channels between the BS and users exhibit a small angle spread seen from the BS [4], [16], [17]. Due to the small angle spread and large dimension of the channels, massive MIMO channels exhibit the sparsity in the virtual angular domain [18]. Moreover, since the spatial propagation characteristics of the wireless channels within the system bandwidth are nearly unchanged, such sparsity is shared by subchannels of different subcarriers when the widely used orthogonal frequency-division multiplexing (OFDM) is considered. This phenomenon is referred to as the *spatially common sparsity within the system bandwidth* [19]. Besides, due to the temporal correlation of the channels [19], massive MIMO channels are quasi-static in several adjacent time slots or one time block consisting of multiple time slots. Moreover, the support set of the sparse channels in the virtual angular domain is almost unchanged in multiple time blocks, which is referred to as the *spatially common sparsity during multiple time blocks*.

By exploiting the spatially common sparsity and the temporal correlation of massive MIMO channels, this paper proposes an adaptive channel estimation and feedback scheme with low overhead. The proposed scheme consists of two stages: a CS based adaptive CSI acquisition with the adaptive training overhead and a follow-up closed-loop channel tracking with the adaptive pilot design. Specifically, the BS transmits the proposed non-orthogonal pilot. The users simply feed back the received non-orthogonal pilot signals to the BS. According to the feedback signals, the CS based adaptive CSI acquisition scheme acquires the downlink CSI at the BS with the adaptive training time slot overhead. For this stage, a distributed sparsity adaptive matching pursuit (DSAMP) algorithm is proposed to acquire the CSI, whereby the spatially common sparsity of massive MIMO channels within the system bandwidth is exploited. By exploiting the spatially common sparsity of massive MIMO channels during multiple time blocks, the closed-loop channel tracking scheme is proposed to track the channels in the second stage. For this stage, the BS can adaptively adjust the pilot signals according to the previous acquired CSI, and a simple least squares (LS) algorithm is used to estimate the channels with improved performance. Additionally, we generalize the results for the conventional MMV to the generalized MMV (GMMV) and provide the Cramer-Rao lower bound (CRLB) of the proposed scheme, which enlightens us to design the non-or-

thogonal pilot signals. Simulation results verify that the proposed scheme is superior to its counterparts, and it is capable of approaching the performance bound. We now summarize our novel contributions.

- **CS based adaptive CSI acquisition scheme**: By exploiting the spatially common sparsity of massive MIMO channels within the system bandwidth, this scheme substantially reduces the required time slot overhead for channel estimation and feedback, where the required time slot overhead is only dependent on the channel sparsity level, rather than the number of BS antennas as in conventional CSI acquisition schemes.

- **Closed-loop channel tracking scheme**: By leveraging the spatially common sparsity of massive MIMO channels during multiple time blocks, this scheme can further reduce the required time slot overhead.

- **Non-orthogonal downlink pilot design at BS**: i) We theoretically prove that the GMMV outperforms the conventional MMV on the sparse signal recovery performance. This enlightens us to design the non-orthogonal pilot for CS based adaptive CSI acquisition for improving channel estimation performance. ii) We derive the CRLB for the proposed scheme. In the stage of closed-loop channel tracking, the derived CRLB enlightens us to adaptively design the non-orthogonal pilot according to the previous channel estimation for further improving performance.

- **DSAMP algorithm**: This algorithm leverages the spatially common sparsity of massive MIMO channels to jointly estimate multiple channels associated with different subcarriers. Compared with the conventional algorithms, such as sparsity adaptive matching pursuit (SAMP), subspace pursuit (SP) and OMP, the proposed DSAMP substantially reduces the required time slot overhead with similar computational complexity.

Throughout our discussions, scalar variables are denoted by normal-face letters, while boldface lower and upper-case symbols denote column vectors and matrices, respectively, and $j = \sqrt{-1}$ is the imaginary axis. The Moore-Penrose inversion, transpose and conjugate transpose operators are given by $(\cdot)^\dagger$, $(\cdot)^T$ and $(\cdot)^*$, respectively, while $\lceil \cdot \rceil$ is the integer ceiling operator and $(\cdot)^{-1}$ is the inverse operator. The ℓ_0-norm and ℓ_2-norm are given by $\|\cdot\|_0$ and $\|\cdot\|_2$, respectively, and $|\Gamma|$ is the cardinality of the set Γ. The support set of the vector $\mathbf{a}$ is denoted by supp$\{\mathbf{a}\}$, $[\mathbf{a}]_i$ denotes the ith entry of the vector $\mathbf{x}$, and $[\mathbf{A}]_{i,j}$ denotes the ith-row and jth-column element of the matrix $\mathbf{A}$, while $\mathbf{I}_K$ is the $K \times K$ identity matrix. The rank of $\mathbf{A}$ is denoted by rank$\{\mathbf{A}\}$ and $\text{Tr}\{\cdot\}$ is the matrix trace operator, while $\text{E}\{\cdot\}$ is the expectation operator and var$\{\cdot\}$ is the variance of a random variable. Finally, $(\mathbf{a})_\Gamma$ denotes the entries of $\mathbf{a}$ whose indices are defined by Γ, while $(\mathbf{A})_\Gamma$ denotes a sub-matrix of $\mathbf{A}$ with column indices defined by Γ.

II. System Model

A. Massive MIMO in the Downlink

In a typical massive MIMO system, the BS employing M antennas simultaneously serves K single-antenna users [2], where
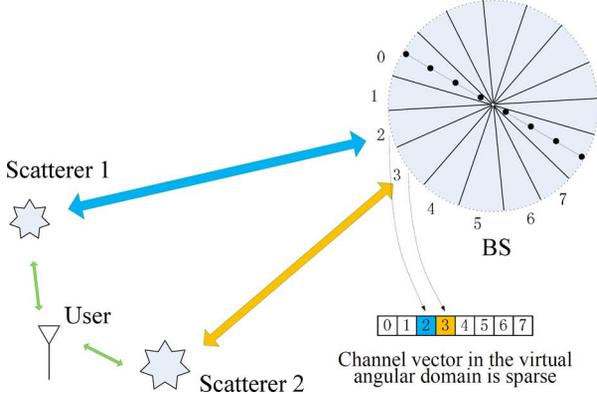
Fig. 1. Channel vector representation in the virtual angular domain, where the BS employs the ULA with half wave-length spacing, $M = 8$, and two clusters of scatterers are considered as an example.

$M \gg K$. For the subchannel at the nth subcarrier, where $1 \leq n \leq N$ and N is the size of the OFDM symbol, the received signal $y_{k,n}$ of the kth user can be expressed as

$$y_{k,n} = \mathbf{h}_{k,n}^{\mathrm{T}} \mathbf{x}_n + w_{k,n}, \qquad (1)$$

where $\mathbf{h}_{k,n} \in \mathbb{C}^{M \times 1}$ denotes the downlink channel between the kth user and the M antennas at the BS, $\mathbf{x}_n \in \mathbb{C}^{M \times 1}$ is the transmitted signal after precoding, and $w_{k,n}$ is the associated additive white Gaussian noise (AWGN). The received signal of the K users $\mathbf{y}_n = [y_{1,n} \ y_{2,n} \cdots y_{K,n}]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ can be collected together as

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{w}_n, \qquad (2)$$

in which $\mathbf{H}_n = [\mathbf{h}_{1,n} \ \mathbf{h}_{2,n} \cdots \mathbf{h}_{K,n}]^{\mathrm{T}} \in \mathbb{C}^{K \times M}$ is the downlink channel matrix, and $\mathbf{w}_n = [w_{1,n} \ w_{2,n} \cdots w_{K,n}]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ is the corresponding AWGN vector.

B. Massive MIMO Channels in Virtual Angular Domain

We model the channel vector $\mathbf{h}_{k,n}$ by using the virtual angular domain representation [18], [19]

$$y_n = \mathbf{h}_n^{\mathrm{T}} \mathbf{x}_n + w_n = \widetilde{\mathbf{h}}_n^{\mathrm{T}} \mathbf{A}_B^* \mathbf{x}_n + w_n, \qquad (3)$$

where the user index k in $y_{k,n}$, $\mathbf{h}_{k,n}$ and $w_{k,n}$ is dropped to simplify the notations, while $\mathbf{h}_n^{\mathrm{T}} = \widetilde{\mathbf{h}}_n^{\mathrm{T}} \mathbf{A}_B^*$ and $\mathbf{A}_B \in \mathbb{C}^{M \times M}$ is the unitary matrix representing the transformation matrix of the virtual angular domain at the BS side. $\mathbf{A}_B$ is determined by the geometrical structure of the BS's antenna array.

To intuitively explain the channel vector $\widetilde{\mathbf{h}}_n$, a simple example is illustrated in Fig. 1, where the BS employs the uniform linear array (ULA) with the antenna spacing of $d = \lambda/2$ and λ is the wave-length. In this case, $\mathbf{A}_B$ becomes the discrete Fourier transform (DFT) matrix [19]. The channel vector in the virtual angular domain then simply means to 'sample' the channel in the angular domain at equi-spaced angular intervals at the BS side, or equivalently to represent the channel in the virtual angular domain coordinates. More specifically, the mth element of $\widetilde{\mathbf{h}}_n$ is the channel gain consisting of the aggregation of all the paths, whose transmit/receive directions are within an angular window around the mth angular coordinate [19].
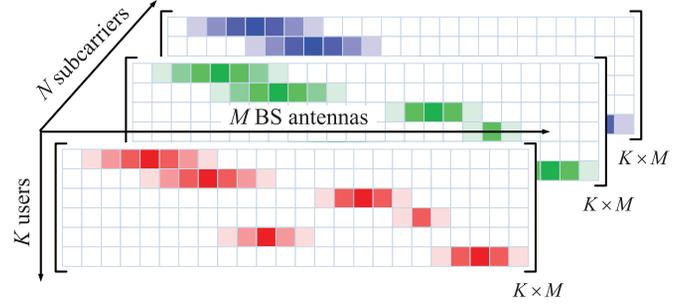
Fig. 2. The virtual angular-domain channel vectors within the system bandwidth exhibit the common sparsity.

As the BS is usually elevated high with few scatterers around, while users are located at low elevation with relatively rich local scatterers, the angle spread at the BS side is small [4], [16], [17]. Since the angle spread is limited at the BS, a small part of the elements in $\widetilde{\mathbf{h}}_n$ contain almost all the multipath signals reflected, diffracted, or refracted by scatterers around the user. If we take the typical angular-domain spread of $10°$ and the ULA with $M = 128$ as an example [4], the uniformly virtual angular domain sampling interval is $\varphi_s = 180°/M = 1.406°$ [18], and the vast majority of the channel energy is concentrated on around $8 = \lceil 10°/1.406° \rceil$ virtual angular domain coordinates, which is far smaller than the total dimension $M = 128$ of the channel vector. Consequently, $\widetilde{\mathbf{h}}_n$ exhibits the sparsity [19], namely,

$$|\Theta_n| = \left| \mathrm{supp}\left\{ \widetilde{\mathbf{h}}_n \right\} \right| = S_a \ll M, \qquad (4)$$

where Θ_n is the support set, and S_a is the sparsity level.

Moreover, since the spatial propagation characteristics of the channels within the system bandwidth are almost unchanged, the subchannels associated with different subcarriers share very similar scatterers in the propagation environment [19]. Hence the small angle spreads of the subchannels within the system bandwidth are very similar. Consequently, $\left\{ \widetilde{\mathbf{h}}_n \right\}_{n=1}^{N}$ have the common sparsity, namely,

$$\mathrm{supp}\left\{ \widetilde{\mathbf{h}}_1 \right\} = \mathrm{supp}\left\{ \widetilde{\mathbf{h}}_2 \right\} = \cdots = \mathrm{supp}\left\{ \widetilde{\mathbf{h}}_N \right\} = \Theta, \quad (5)$$

which is illustrated in Fig. 2.

C. Temporal Correlation of Wireless Channels

Since the user mobility is not very high in massive MIMO systems, the channels remain static for the duration of a block that consists of J consecutive time slots, while the channels change from block to block. Here, one time slot represents one OFDM symbol. This block fading implies that $\mathbf{h}_n^{(q,t)} = \mathbf{h}_n^{(q)}$ for $1 \leq t \leq J$, where $\mathbf{h}_n^{(q,t)}$ is the channel at the tth time slot of the qth block and $\mathbf{h}_n^{(q)}$ denotes the quasi-static channel in the qth block. Similarly, there exists the quasi-static relationship $\widetilde{\mathbf{h}}_n^{(q,t)} = \widetilde{\mathbf{h}}_n^{(q)}$ for $1 \leq t \leq J$, with $\widetilde{\mathbf{h}}_n^{(q,t)}$ and $\widetilde{\mathbf{h}}_n^{(q)}$ being the virtual angular representations of $\mathbf{h}_n^{(q,t)}$ and $\mathbf{h}_n^{(q)}$, respectively.

For massive MIMO channels, $J < M$ due to the limited coherence time and the large number of BS antennas. For example, consider massive MIMO systems with: the carrier frequency $f_c = 2$ GHz, the system bandwidth $B_s = 10$ MHz, the OFDM size $N = 2048$, the number of BS antennas $M = 128$, and the

maximum delay spread $\tau_{\max} = 6.4\ \mu s$ (need the guard interval $N_g = 64$) [2], [20]. Suppose that the maximum mobile velocity of the supported users is $v = 36$ km/h. Then the maximum Doppler frequency shift is $f_d = v f_c / c = 66.67$ Hz, where c is the velocity of electromagnetic wave. Hence the coherence time $T_c = \sqrt{9/(16\pi f_d^2)} \approx 6.3$ ms [21], or the coherence time slots $J = T_c B_s / (N + N_g) \approx 30$, which is much smaller than M.

Since the channels change from block to block, they must be estimated in every time block, which may impose very high complexity and overhead. Fortunately, experiments and theoretical analysis have shown that although the channels vary continuously from one block to another, the variation rate of the channel angle spread is much lower than that of the associated channel gains [18]. This implies that

$$\text{supp}\left\{\widetilde{\mathbf{h}}_n^{(q)}\right\} = \text{supp}\left\{\widetilde{\mathbf{h}}_n^{(q+1)}\right\} = \cdots = \text{supp}\left\{\widetilde{\mathbf{h}}_n^{(q+Q-1)}\right\}, \tag{6}$$

where Q is the number of consecutive time blocks over the duration of which the common support of virtual angular domain channels holds. For the example of Fig. 1, assume that the distance between the BS and the user is $L_{BU} = 250$ m and $v = 36$ km/h. Further assume the case of the mobile direction of the user being perpendicular to the direction connecting the BS and the user. Then, over the duration of $Q = 5$ successive time blocks, the maximum variance in the virtual angular domain is around $\theta_\Delta = \arctan(QT_c v/L_{BU}) \approx 0.072°$. Such a small variance of the angle spread is negligible, compared to the resolution of the virtual angular domain $\varphi_s = 1.406°$. If $v < 36$ km/h and/or the mobile direction of the user is not perpendicular to the direction connecting the BS and the user, Q can be larger than 5.

D. Challenges of Channel Estimation and Feedback

Consider the downlink channel estimation in the qth time block. To reliably estimate the channel of the nth subcarrier, the user should jointly utilize the received pilot signals over several successive time slots, say, G time slots, for channel estimation. Let $y_n^{(q,t)}$ be the received pilot of (3) at the nth subcarrier in the tth time slot, and $y_n^{(q,t)}$ for $1 \le t \le G$ can be collected together in the vector $\mathbf{y}_n^{[q,G]} = [y_n^{(q,1)}\ y_n^{(q,2)} \cdots y_n^{(q,G)}]^{\mathrm{T}} \in \mathbb{C}^{G \times 1}$. Then

$$\mathbf{y}_n^{[q,G]} = \mathbf{X}_n^{[q,G]} \mathbf{h}_n^{(q)} + \mathbf{w}_n^{[q,G]}, \tag{7}$$

where $\mathbf{X}_n^{[q,G]} = [\mathbf{x}_n^{(q,1)}\ \mathbf{x}_n^{(q,2)} \cdots \mathbf{x}_n^{(q,G)}]^{\mathrm{T}} \in \mathbb{C}^{G \times M}$ with $\mathbf{x}_n^{(q,t)} \in \mathbb{C}^{M \times 1}$ being the transmitted pilot signals in the tth time slot, and $\mathbf{w}_n^{[q,G]} = [w_n^{(q,1)}\ w_n^{(q,2)} \cdots w_n^{(q,G)}]^{\mathrm{T}} \in \mathbb{C}^{G \times 1}$ is the corresponding AWGN vector. To accurately estimate the channel from (7), the value of G used in conventional algorithms, such as the minimum mean square error (MMSE) algorithm, is heavily dependent on the value of M. Usually, G can be larger than J, which leads to the poor channel estimation performance [10]. Moreover, to minimize the mean square error (MSE) of the channel estimate, $\mathbf{X}_n^{[q,G]}$ should be a unitary matrix scaled by a transmit power factor [10]. Usually, $\mathbf{X}_n^{[q,G]}$ is a diagonal matrix with equal-power diagonal elements. Such a pilot design is illustrated in Fig. 3(a), which is called the time-domain orthogonal pilot. It should be pointed out that in MIMO-OFDM systems, to estimate the channel associated
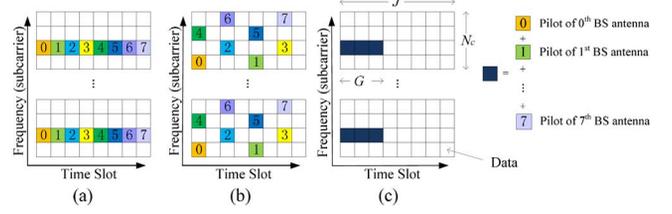
Fig. 3. (a) Time-domain orthogonal pilot [10], (b) time-frequency orthogonal pilot in LTE-A [9], and (c) proposed non-orthogonal pilot, assuming $M = 8$.

with one transmit antenna, P pilot subcarriers should be used, and usually $P = N_g$ is considered since $N_c = N/N_g$ adjacent subcarriers are correlated [10]. Hence the total pilot overhead to estimate the complete MIMO channel is $P_{\text{total}} = PG = N_g M$. Similarly, LTE-A adopts the time-frequency orthogonal pilots as shown in Fig. 3(b), which also needs $P_{\text{total}} = PM = N_g M$. These two kinds of orthogonal pilots are equivalent, since both of them are based on the framework of Nyquist sampling theorem and have the same pilot overhead. Hence we only consider the time-domain orthogonal pilot in this paper, and we will propose an efficient non-orthogonal pilot scheme.

Codebook based channel feedback schemes are widely adopted in small-scale MIMO systems. However, to obtain the fine-grain spatial channel structures in massive MIMO systems, the codebook size can be huge. Moreover, the storage and encoding of large dimension codebook at the user is challenging. To overcome this difficulty, we combine the channel estimation and feedback, whereby the CSI acquisition is mainly realized at the BS which has sufficient computation capability. By exploiting the spatially common sparsity and temporal correlation of massive MIMO channels, the proposed scheme can significantly reduce the required overhead and complexity for channel estimation and feedback.

III. SPATIALLY COMMON SPARSITY BASED ADAPTIVE CHANNEL ESTIMATION AND FEEDBACK SCHEME

The procedure of the proposed adaptive channel estimation and feedback scheme is first summarized.

Step 1: In each time slot, the BS transmits a non-orthogonal pilot to the user, and the user directly feeds back the received pilot signal to the BS. Except for *Step 4*, the pilot signal is designed in advance.

Step 2: The BS uses the proposed DSAMP algorithm to jointly reconstruct multiple sparse virtual angular domain channels of high dimension from the feedback signals of low dimension collected in multiple time slots.

Step 3: The BS judges the reliability of the estimated sparse channels according to a pre-specified criterion. If the given criterion is met, the BS stops transmitting pilot in the following time slots, and the acquired CSI at the BS is used for precoding and user scheduling in the current time block. Otherwise, the BS goes back to *Step 1* until the feedback signals are sufficient for acquiring the reliable CSI.

Step 4: Since the BS has acquired the estimated support set $\widehat{\Theta}$ and the estimated sparsity level $\widehat{S}_a$, it can directly use the LS algorithm to estimate the channels in

every time block of the following $Q - 1$ blocks. Here, the time slot overhead required in *Step 1* can be reduced to $G = \widehat{S}_a$, and the pilot signals can be adaptively adjusted according to $\widehat{\Theta}$ for further improving performance.

It is seen that the proposed adaptive channel estimation and feedback scheme consists of two stages: the CS based adaptive CSI acquisition in the qth time block, which includes *Step 1* to *Step 3*, and the following closed-loop channel tracking in the following $Q - 1$ time blocks, which includes *Step 1* and *Step 4*. We now detail all the key technical components.

A. Non-Orthogonal Pilot for Downlink Channel Estimation

The proposed non-orthogonal pilot scheme is illustrated in Fig. 3(c). Similar to the time-domain orthogonal pilot scheme, P subcarriers are dedicated to pilots in each OFDM symbol. However, the proposed scheme allows the non-orthogonal pilot signals associated with different BS antennas to occupy the completely identical frequency-domain subcarriers.

The orthogonal pilot based conventional designs usually require $G \geq M$. By contrast, the proposed non-orthogonal pilot for CS based adaptive CSI acquisition is capable of providing the efficient compression and reliable recovery of sparse signals. Therefore, G is mainly determined by $S_a \ll M$. The non-orthogonal pilot of the first stage is designed in advance, which will be discussed in Section IV-A. According to the CSI acquired in the first stage, the non-orthogonal pilot used for closed-loop channel tracking is adaptively designed to minimize both G and the MSE performance of CSI acquisition, which will be illustrated in Section III-D.

For the placement of pilot subcarriers, the widely used equi-spaced pilot is considered, and the specific reason is given in Section IV-C. For convenience, we denote $\Omega_\xi = \{\xi_1, \xi_2, \ldots, \xi_P\}$ as the index set of the pilot subcarriers, where ξ_p for $1 \leq p \leq P$ denotes the subcarrier index dedicated to the pth pilot subcarrier. It is worth pointing out that the pth pilot subcarrier is shared by the pilot signals of the M transmit antennas as illustrated in Fig. 3(c).

B. CS Based Adaptive CSI Acquisition Scheme

In the qth time block, as indicated in *Step 1*, the user directly feeds back the received pilot signals to the BS without performing downlink channel estimation where the feedback channel can be considered as the AWGN channel [12]–[15]. According to (7), at the BS, the fed back signal[1] (at the ξ_pth subcarrier) in the tth time slot can be expressed as

$$r_p^{(q,t)} = (\bar{\mathbf{h}}_p^{(q)})^{\mathrm{T}} \mathbf{A}_B^* \mathbf{s}_p^{(q,t)} + v_p^{(q,t)}, 1 \leq p \leq P, \quad (8)$$

where $r_p^{(q,t)} = y_{\xi_p}^{(q,t)}$ is the pth feedback pilot signal in the tth time slot, $\bar{\mathbf{h}}_p^{(q)} = \widetilde{\mathbf{h}}_{\xi_p}^{(q)}$ is the virtual angular domain channel vector associated with the pth pilot subcarrier, $\mathbf{s}_p^{(q,t)} = \mathbf{x}_{\xi_p}^{(q,t)}$ is the pilot signal vector transmitted by the M BS antennas, and

[1] The delay of the feedback signal is negligible, compared with the relatively long channel coherence time.

$v_p^{(q,t)} = w_{\xi_p}^{(q,t)}$ is the effective noise which aggregates both the downlink channel's AWGN and feedback channel's AWGN.

Due to the quasi-static property of the channel during one time block, the feedback signals in G successive time slots can be jointly exploited to acquire the downlink CSI at the BS, which can be expressed as

$$\mathbf{r}_p^{[q,G]} = \mathbf{S}_p^{[q,G]}(\mathbf{A}_B^*)^{\mathrm{T}}\bar{\mathbf{h}}_p^{(q)} + \mathbf{v}_p^{[q,G]} = \mathbf{\Phi}_p^{[q,G]}\bar{\mathbf{h}}_p^{(q)} + \mathbf{v}_p^{[q,G]}, \quad (9)$$

for $1 \leq p \leq P$, where $\mathbf{r}_p^{[q,G]} = [r_p^{(q,1)} \ r_p^{(q,2)} \cdots r_p^{(q,G)}]^{\mathrm{T}}$, $\mathbf{S}_p^{[q,G]} = [\mathbf{s}_p^{(q,1)} \ \mathbf{s}_p^{(q,2)} \cdots \mathbf{s}_p^{(q,G)}]^{\mathrm{T}} \in \mathbb{C}^{G \times M}$, $\mathbf{v}_p^{[q,G]} = [v_p^{(q,1)} \ v_p^{(q,2)} \cdots v_p^{(q,G)}]^{\mathrm{T}}$ and $\mathbf{\Phi}_p^{[q,G]} = \mathbf{S}_p^{[q,G]}(\mathbf{A}_B^*)^{\mathrm{T}} \in \mathbb{C}^{G \times M}$. The system's signal to noise ratio (SNR) is defined as $\mathrm{SNR} = \mathrm{E}\left\{\left\|\mathbf{\Phi}_p^{[q,G]}\bar{\mathbf{h}}_p^{(q)}\right\|_2^2\right\} \Big/ \mathrm{E}\left\{\left\|\mathbf{v}_p^{[q,G]}\right\|_2^2\right\}$, according to (9). By exploiting the spatially common sparsity within the system bandwidth, the proposed DSAMP algorithm can reconstruct the sparse angular domain channels of multiple pilot subcarriers, as will be detailed in Section III-C.

Algorithm 1: CS Based Adaptive CSI Acquisition Scheme

1: Determine the initial time slot overhead G_0, and set the iteration index $i = 0$.

2: **repeat**

3:     Collect $\mathbf{r}_p^{[q,G_i]}$ and $\mathbf{\Phi}_p^{[q,G_i]}$ in (9) for given G_i, $1 \leq p \leq P$. $\% G_i$ is the required overhead at the ith iteration.

4:     Acquire the channel vectors $\widehat{\bar{\mathbf{h}}}_p^{(q)} \ \forall p$ by using the proposed DSAMP algorithm (Algorithm 2).

5:     $G_{i+1} = G_i + 1; i = i + 1$.

6: **until** $\sum_{p=1}^{P} \left\|\mathbf{r}_p^{[q,G_{i-1}]} - \mathbf{\Phi}_p^{[q,G_{i-1}]}\widehat{\bar{\mathbf{h}}}_p^{(q)}\right\|_2^2 / (PG_{i-1}) \leq \varepsilon$.
    $\%$ If the error is smaller than the threshold ε, end repeat; otherwise, continue transmitting the pilot in the next time slot.

7: $G_0 = G_i - 1$. $\%$ Optional, determine the initial time slot overhead for the next CS based adaptive CSI acquisition.

For practical massive MIMO systems, the sparsity level S_a of the channels in the virtual angular domain can be time-varying. If S_a is relatively small, a small time slot overhead G is sufficient to acquire an accurate CSI estimate, while if S_a is relatively large, a large G is required to guarantee the reliable sparse signal recovery. We propose the CS based adaptive CSI acquisition as presented in Algorithm 1, which can adaptively adjust G to acquire the reliable CSI at the BS efficiently.

At the first CS based adaptive CSI acquisition, we need to empirically determine the initial time slot overhead G_0. Since the typical angle spread is about $10°$ [4], for massive MIMO with $M = 128$, the effective sparsity level $S_a = 8$. Thus, we may set $G_0 = 8$ to start. Given G_i, the DSAMP algorithm (Algorithm 2) acquires the set of channel vectors $\widehat{\bar{\mathbf{h}}}_p^{(q)} \ \forall p$. If $\sum_{p=1}^{P} \left\|\mathbf{r}_p^{[q,G_i]} - \mathbf{\Phi}_p^{[q,G_i]}\widehat{\bar{\mathbf{h}}}_p^{(q)}\right\|_2^2 / (PG_i)$ is larger than the predefined threshold ε, the sparse signal recovery is judged to be unreliable. Hence, the training time slots increase by one, and a set of the feedback pilot signals and transmitted pilot signals, $r_p^{(q,G_{i+1})}$ and $\mathbf{s}_p^{(q,G_{i+1})} \ \forall p$, are collected in the (G_{i+1})th time

slot, which are combined with the previously collected $\mathbf{r}_p^{[q,G_i]}$ and $\mathbf{\Phi}_p^{[q,G_i]}$ to enlarge the dimension of the measurement vectors sequentially, yielding

$$\mathbf{r}_p^{[q,G_{i+1}]} = \begin{bmatrix} \mathbf{r}_p^{[q,G_i]} \\ r_p^{(q,G_{i+1})} \end{bmatrix} \text{ and } \mathbf{\Phi}_p^{[q,G_{i+1}]} = \begin{bmatrix} \mathbf{\Phi}_p^{[q,G_i]} \\ \left(\mathbf{s}_p^{(q,G_{i+1})}\right)^{\mathrm{T}} (\mathbf{A}_B^*)^{\mathrm{T}} \end{bmatrix}$$

to improve the channel estimation. Furthermore, an appropriate initial time slot overhead for the next CS based adaptive CSI acquisition is automatically determined at the end.

Algorithm 2: Proposed DSAMP Algorithm

Input: Noisy feedback signals $\mathbf{r}_p^{[q,G]}$ and sensing matrices $\mathbf{\Phi}_p^{[q,G]}$ in (10), $1 \le p \le P$; termination threshold p_{th}.

Output: Estimated channel vectors in the virtual angular domain at multiple pilot subcarriers $\widehat{\bar{\mathbf{h}}}_p^{(q)}$, $\forall p$.

1: $\mathcal{T} = 1$; $i = 1$; $j = 1$. % $\mathcal{T}$, i, j are the sparsity level of the current stage, iteration index, and stage index, respectively.
2: $\mathbf{c}_p = \mathbf{t}_p = \mathbf{c}_p^{\mathrm{last}} = \mathbf{0} \in \mathbb{C}^{M \times 1}$, $\forall p$. % $\mathbf{c}_p$ and $\mathbf{t}_p$ are intermediate variables, and $\mathbf{c}_p^{\mathrm{last}}$ is the channel estimation of the last stage.
3: $\Omega^0 = \Gamma = \widetilde{\Gamma} = \Omega = \widetilde{\Omega} = \emptyset$; $l_{\min} = \tilde{l} = 0$. % Ω^i is the estimated support set in the ith iteration, Γ, $\widetilde{\Gamma}$, Ω, and $\widetilde{\Omega}$ are sets, $l_{\min}$ and $\tilde{l}$ denote the support indexes.
4: $\mathbf{b}_p^0 = \mathbf{r}_p^{[q,G]} \in \mathbb{C}^{G \times 1}$, $\forall p$. % $\mathbf{b}_p^i$ is the residual of the ith iteration.
5: $\sum_{p=1}^{P} \left\| \mathbf{b}_p^{\mathrm{last}} \right\|_2^2 = +\infty$. % $\mathbf{b}_p^{\mathrm{last}}$ is the residual of the last stage.
6: **repeat**
7: $\quad \mathbf{a}_p = \left(\mathbf{\Phi}_p^{[q,G]} \right)^* \mathbf{b}_p^{i-1}$, $\forall p$. % Signal proxy is saved in $\mathbf{a}_p$.
8: $\quad \Gamma = \arg\max_{\widetilde{\Gamma}} \left\{ \sum_{p=1}^{P} \left\| (\mathbf{a}_p)_{\widetilde{\Gamma}} \right\|_2^2, \left| \widetilde{\Gamma} \right| = \mathcal{T} \right\}$. % Identify support.
9: $\quad (\mathbf{t}_p)_{\Omega^{i-1} \cup \Gamma} = \left(\left(\mathbf{\Phi}_p^{[q,G]} \right)_{\Omega^{i-1} \cup \Gamma} \right)^{\dagger} \mathbf{r}_p^{[q,G]}$, $\forall p$. % LS estimation.
10: $\quad \Omega = \arg\max_{\widetilde{\Omega}} \left\{ \sum_{p=1}^{P} \left\| (\mathbf{t}_p)_{\widetilde{\Omega}} \right\|_2^2, \left| \widetilde{\Omega} \right| = \mathcal{T} \right\}$. % Prune support.
11: $\quad (\mathbf{c}_p)_{\Omega} = \left(\left(\mathbf{\Phi}_p^{[q,G]} \right)_{\Omega} \right)^{\dagger} \mathbf{r}_p^{[q,G]}$, $\forall p$. % LS estimation.
12: $\quad \mathbf{b}_p = \mathbf{r}_p^{[q,G]} - \mathbf{\Phi}_p^{[q,G]} \mathbf{c}_p$, $\forall p$. % Compute the residual.
13: $\quad l_{\min} = \arg\min_{\tilde{l}} \left\{ \sum_{p=1}^{P} \left\| [\mathbf{c}_p]_{\tilde{l}} \right\|_2^2, \tilde{l} \in \Omega \right\}$. % Find the support of the minimum average energy according to the estimated $\mathbf{c}_p$.
14: $\quad$ **if** $\sum_{p=1}^{P} \left\| [\mathbf{c}_p]_{l_{\min}} \right\|_2^2 / P < p_{\mathrm{th}}$ **then**
15: $\quad\quad$ Quit iteration. % Support index associated with AWGN may be included in Ω.
16: $\quad$ **else if** $\sum_{p=1}^{P} \left\| \mathbf{b}_p^{\mathrm{last}} \right\|_2^2 < \sum_{p=1}^{P} \left\| \mathbf{b}_p \right\|_2^2$ **then**
17: Quit iteration. % Larger residual of the current stage than that of the last stage indicates that it is unnecessary to continue the iteration.
18: **else if** $\sum_{p=1}^{P} \left\| \mathbf{b}_p^{i-1} \right\|_2^2 \le \sum_{p=1}^{P} \left\| \mathbf{b}_p \right\|_2^2$ **then**
19: $\quad\quad$ $j = j + 1$; $\mathcal{T} = j$; $\mathbf{c}_p^{\mathrm{last}} = \mathbf{c}_p$, $\mathbf{b}_p^{\mathrm{last}} = \mathbf{b}_p$, $\forall p$. % Begin a new stage. The larger residual of current

iteration than that of last iteration indicates that iteration at current stage converges.
20: $\quad$ **else**
21: $\quad\quad$ $\Omega^i = \Omega$; $\mathbf{b}_p^i = \mathbf{b}_p$, $\forall p$; $i = i + 1$. % Continue the iteration at the current stage.
22: $\quad$ **end if**
23: **until** $\sum_{p=1}^{P} \| [\mathbf{c}_p]_{l_{\min}} \|_2^2 / P < p_{\mathrm{th}}$
24: $\widehat{\bar{\mathbf{h}}}_p^{(q)} = \mathbf{c}_p^{\mathrm{last}}$, $\forall p$. % Obtain the final channel estimation.

C. Proposed DSAMP Algorithm for Channel Estimation

Given the measurements (9), the CSI can be acquired by solving the following optimization

$$\min_{\bar{\mathbf{h}}_p^{(q)}, 1 \le p \le P} \left(\sum_{p=1}^{P} \left\| \bar{\mathbf{h}}_p^{(q)} \right\|_0^2 \right)^{1/2}$$

$$\mathrm{s.t.} \, \mathbf{r}_p^{[q,G]} = \mathbf{\Phi}_p^{[q,G]} \bar{\mathbf{h}}_p^{(q)}, \, \forall p \text{ and } \left\{ \bar{\mathbf{h}}_p^{(q)} \right\}_{p=1}^{P}$$

share the common sparse support set. $\quad$ (10)

The DSAMP algorithm, listed in Algorithm 2, is used to solve the optimization (10) to simultaneously acquire multiple sparse channel vectors at different pilot subcarriers. This algorithm is developed from the SAMP algorithm [22]. Specifically, for each stage with the fixed sparsity level $\mathcal{T}$, *line 8* selects the $\mathcal{T}$ potential non-zero elements; *line 9* estimates the values associated with the support set $\Omega^{i-1} \cup \Gamma$ using LS; *line 10* selects $\mathcal{T}$ most likely supports. *Lines 7 $\sim$ 12* and *line 21* together aim to find $\mathcal{T}$ virtual angular domain coordinates which contain most of the channel energy. In particular, *Lines 7 $\sim$ 12* remove wrong indices added in the previous iteration and add the indices associated with the potential true indices. If *line 18* is triggered, the algorithm updates $\mathcal{T}$ and begins a new stage. The algorithm is halted when the stopping criteria, indicated in *lines 14 $\sim$ 17* and *line 23*, are met.

Compared to the classical SAMP algorithm [22] which recovers one high-dimensional sparse signal from single low-dimensional received signal, the proposed DSAMP algorithm can simultaneously recover multiple high-dimensional sparse signals with the common support set by jointly processing multiple low-dimensional received signals. In terms of termination condition, the SAMP algorithm stops the iteration once the residual is smaller than a threshold p_{th}. By contrast, the proposed DSAMP algorithm has two halting criteria. Specifically, if the energy associated with one virtual angular coordinate in the estimated support set is smaller than p_{th} or the residual of the current stage is larger than that of the previous stage, the algorithm stops. The proposed halting criteria ensure the robust signal recovery performance, which will be discussed in Section IV-D2 and confirmed by simulations.

By using the DSAMP algorithm at the BS, we can acquire the estimates of the virtual angular domain channels at the pilot subcarriers, i.e., $\widehat{\bar{\mathbf{h}}}_p^{(q)}$ for $1 \le p \le P$. Consequently, the actual channel at the ξ_pth subcarrier dedicated to the pth pilot signal can be acquired according to (8), yielding

$$\widehat{\mathbf{h}}_{\xi_p}^{(q)} = (\mathbf{A}_B^*)^{\mathrm{T}} \widehat{\bar{\mathbf{h}}}_{\xi_p}^{(q)} = (\mathbf{A}_B^*)^{\mathrm{T}} \widehat{\bar{\mathbf{h}}}_p^{(q)}. \quad (11)$$

D. Closed-Loop Channel Tracking With Adaptive Pilot Design

Since the channels in Q successive time blocks share the spatially common sparsity, in the following $Q-1$ time blocks, we can use the simple LS algorithm to estimate the channels at the BS from the feedback pilot signals by utilizing the estimated support set $\widehat{\Theta} = \text{supp}(\widehat{\bar{\mathbf{h}}}_p^{(q)})$ and the sparsity level $\widehat{S}_a = |\widehat{\Theta}|$ acquired in the qth time block. Specifically, for the q_bth time block, where $q + 1 \leq q_b \leq q + Q - 1$, the BS first transmits a non-orthogonal pilot to the user, and the user again directly feeds back the received pilot signal to the BS. At the BS, similar to (9), the feedback pilot signal associated with the pth pilot subcarrier $\mathbf{r}_p^{[q_b, G]}$ can be expressed as

$$\mathbf{r}_p^{[q_b,G]} = \mathbf{S}_p^{[q_b,G]}(\mathbf{A}_B^*)^{\mathrm{T}}\bar{\mathbf{h}}_p^{(q_b)} + \mathbf{v}_p^{[q_b,G]} = \boldsymbol{\Phi}_p^{[q_b,G]}\bar{\mathbf{h}}_p^{(q_b)} + \mathbf{v}_p^{[q_b,G]}, \quad (12)$$

where $\mathbf{S}_p^{[q_b,G]}$, $\bar{\mathbf{h}}_p^{(q_b)}$ and $\mathbf{v}_p^{[q_b,G]}$ are the pilot signal matrix, virtual angular domain channel, and effective noise in the q_bth time block, respectively. If Θ and S_a are known, the CSI can be acquired using the LS algorithm as

$$\left(\widehat{\bar{\mathbf{h}}}_p^{(q_b)}\right)_\Theta = \left(\left(\boldsymbol{\Phi}_p^{[q_b,G]}\right)_\Theta\right)^\dagger \mathbf{r}_p^{[q_b,G]}, \quad (13)$$

which is an unbiased estimator for $\bar{\mathbf{h}}_p^{(q_b)}$ that is capable of approaching the CRLB [23]. The BS can use $\widehat{\Theta}$ and $\widehat{S}_a$, obtained in the qth time block, to calculate this LS estimate.

As will be shown in Section IV-F, to acquire the estimate of $\bar{\mathbf{h}}_p^{(q_b)}$, the required time slot overhead can be reduced to S_a. Moreover, the non-orthogonal pilot used for channel tracking (for the time blocks of $q + 1 \leq q_b \leq q + Q - 1$) is very different from that used in the qth time block. Specifically, to minimize the MSE performance of the channel estimation with $G = S_a$, $(\boldsymbol{\Phi}_p^{[q_b,G]})_\Theta$ should be a unitary matrix scaled by a power factor $\sqrt{G}$. Therefore, for the closed-loop channel tracking, we can design the non-orthogonal pilot signal to guarantee this condition, and reduce G to S_a while attaining the best MSE performance for the channel estimation. Specifically, let $G = S_a$ and $\mathbf{U}_{S_a} \in \mathbb{C}^{S_a \times S_a}$ be a unitary matrix. Then

$$\left(\boldsymbol{\Phi}_p^{[q_b,G]}\right)_\Theta = \left(\mathbf{S}_p^{[q_b,G]}(\mathbf{A}_B^*)^{\mathrm{T}}\right)_\Theta = \sqrt{G}\mathbf{U}_{S_a}, \quad (14)$$

which yields the required non-orthogonal pilot matrix $\mathbf{S}_p^{[q_b,G]} = \sqrt{G}\mathbf{U}_{S_a}\left(((\mathbf{A}_B^*)^{\mathrm{T}})_\Theta\right)^\dagger$.

IV. PERFORMANCE ANALYSIS

The performance analysis of the proposed scheme includes: 1) the non-orthogonal pilot design for the CS based adaptive CSI acquisition; 2) the theoretical limit of the required time slot overhead for the CS based adaptive CSI acquisition; 3) the placement of pilot subcarriers; 4) the computational complexity and convergence of the DSAMP algorithm; 5) the performance bound of the proposed scheme; 6) the required time slot overhead and the performance analysis for the adaptive non-orthogonal pilot based closed-loop channel tracking; and 7) the selection of thresholds for Algorithms 1 and 2.

A. Non-Orthogonal Pilot Design for CS Based Adaptive CSI Acquisition

In the qth time block, the measurement matrices $\boldsymbol{\Phi}_p^{[q,G]} \forall p$ in (9) are very important for guaranteeing the reliable CSI acquisition at the BS. Usually, $G \ll M$. Since $\boldsymbol{\Phi}_p^{[q,G]} = \mathbf{S}_p^{[q,G]}(\mathbf{A}_B^*)^{\mathrm{T}}$ and $\mathbf{A}_B$ is determined by the geometrical structure of the antenna array at the BS, the pilot signals $\mathbf{S}_p^{[q,G]} \forall p$ transmitted by the BS should be designed to guarantee the desired robust channel estimation and feedback.

1) Restricted Isometry Property (RIP): In CS theory, RIP is used to evaluate the quality of the measurement matrix, in terms of the reliable compression and reconstruction of sparse signals. It is proven in [24] that the measurement matrix with its elements following the independent and identically distributed (i.i.d.) complex Gaussian distributions satisfies the RIP and enjoys a satisfying performance in compressing and recovering sparse signals.

2) Processing Multiple Measurement Vectors in Parallel: The optimization problem (10) is essentially different from the single-measurement-vector (SMV) and MMV problems in CS.

The SMV recovers single high-dimensional sparse signal $\mathbf{f}$ from its low-dimensional measurement signal $\mathbf{d}$, which may be formulated as $\mathbf{d} = \boldsymbol{\Phi}\mathbf{f}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{D \times F}$, $D < F$, and the support set $\Xi = \text{supp}\{\mathbf{f}\}$ with the sparsity level $|\Xi| = S \ll F$.

On the other hand, the MMV can simultaneously recover multiple high-dimensional sparse signals with the common support set and common measurement matrix from multiple low-dimensional measurement signals, which may be formulated as $\mathbf{D} = \boldsymbol{\Phi}\mathbf{F}$, with $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \cdots \mathbf{d}_L]$, $\mathbf{F} = [\mathbf{f}_1 \ \mathbf{f}_2 \cdots \mathbf{f}_L]$, $\text{supp}\{\mathbf{f}_1\} = \text{supp}\{\mathbf{f}_2\} = \cdots = \text{supp}\{\mathbf{f}_L\} = \Xi$, and the sparsity level $|\Xi| = S$.

By contrast, our problem (10) can jointly reconstruct multiple high-dimensional sparse signals with the common support set but having different measurement matrices, i.e.,

$$\mathbf{d}_l = \boldsymbol{\Phi}_l\mathbf{f}_l, 1 \leq l \leq L, \quad (15)$$

where $\boldsymbol{\Phi}_l \in \mathbb{C}^{D \times F}, \forall l$. Therefore, our problem can be regarded as the GMMV problem, which includes the SMV and MMV problems as its special cases. Specifically, if the multiple measurement matrices are identical, our GMMV becomes the conventional MMV, and furthermore if $L = 1$, it reduces to the conventional SMV.

Typically, the MMV has the better recovery performance than the SMV, due to the potential diversity from multiple sparse signals [24]. Intuitively, the recovery performance of multiple sparse signals with different measurement matrices, as defined in the GMMV, should be better than that with the common measurement matrix as given in the MMV. This is because the further potential diversity can benefit from different measurement matrices for the GMMV. To prove this, we investigate the uniqueness of the solution to the GMMV problem. First, we introduce the concept of 'spark' and the ℓ_0-minimization based GMMV problem associated with (15).

Definition 1 [24]: The smallest number of columns of $\boldsymbol{\Phi}$ which are linearly dependent is the spark of the given matrix $\boldsymbol{\Phi}$, denoted by $\text{spark}(\boldsymbol{\Phi})$.

Problem 1: $\min_{\mathbf{f}_l, \forall l} \sum_{l=1}^{L} \|\mathbf{f}_l\|_0^2$, s.t. $\mathbf{d}_l = \mathbf{\Phi}_l \mathbf{f}_l$, $\text{supp}\{\mathbf{f}_l\} = \Xi$, $\forall l$.

For the above ℓ_0-minimization based GMMV problem, the following result can be obtained.

Theorem 1: For $\mathbf{\Phi}_l$, $1 \leq l \leq L$, whose elements obey an i.i.d. continuous distribution, there exist full rank matrices $\mathbf{\Psi}_l$ for $2 \leq l \leq L$ satisfying $(\mathbf{\Phi}_l)_\Xi = \mathbf{\Psi}_l (\mathbf{\Phi}_1)_\Xi$ if we select $(\mathbf{\Phi}_1)_\Xi$ as the bridge, where Ξ is the common support set. Consequently, $\mathbf{f}_l$ for $1 \leq l \leq L$ will be the unique solution to Problem 1 if

$$2S < \text{spark}(\mathbf{\Phi}_1) - 1 + \text{rank}\{\widetilde{\mathbf{D}}\}, \tag{16}$$

where $\widetilde{\mathbf{D}} = \left[\mathbf{d}_1 \, \mathbf{\Psi}_2^{-1} \mathbf{d}_2 \cdots \mathbf{\Psi}_L^{-1} \mathbf{d}_L \right]$.

Proof: Consider (15) with $\mathbf{\Phi}_l \in \mathbb{C}^{D \times F}$ for $1 \leq l \leq L$, whose elements follow an i.i.d. continuous distribution. The common support set is $\Xi = \text{supp}\{\mathbf{f}_l\}$ with the sparsity level $|\Xi| = S$. This GMMV can be expressed as

$$\mathbf{d}_l = (\mathbf{\Phi}_l)_\Xi (\mathbf{f}_l)_\Xi = \mathbf{Z}_l (\mathbf{f}_l)_\Xi, 1 \leq l \leq L. \tag{17}$$

The random matrix $\mathbf{Z}_l = (\mathbf{\Phi}_l)_\Xi \in \mathbb{C}^{D \times S}$ is a tall matrix, as $D > S$. Clearly, $\text{rank}\{\mathbf{Z}_l\} = S$ with high probability, since the measure of the set $\{\mathbf{Z}_l \in \Omega_{\mathbf{Z}} : \text{rank}\{\mathbf{Z}_l\} < S\}$ is zero [25]. If we take $(\mathbf{\Phi}_1)_\Xi$ as the bridge, then there exist the full rank matrices $\mathbf{\Psi}_l$, $2 \leq l \leq L$, satisfying $(\mathbf{\Phi}_l)_\Xi = \mathbf{\Psi}_l (\mathbf{\Phi}_1)_\Xi$, and thus

$$\mathbf{\Psi}_l^{-1} \mathbf{d}_l = (\mathbf{\Phi}_1)_\Xi (\mathbf{f}_l)_\Xi = \mathbf{\Phi}_1 \mathbf{f}_l. \tag{18}$$

In this way, the GMMV is converted to the 'equivalent' MMV

$$\widetilde{\mathbf{D}} = \mathbf{\Phi}_1 \mathbf{F}, \tag{19}$$

where $\mathbf{F} = [\mathbf{f}_1 \, \mathbf{f}_2 \cdots \mathbf{f}_L]$. Applying the existing result for the MMV given in [26], (16) can be directly obtained. ∎

From Theorem 1, it is clear that the achievable diversity gain introduced by diversifying measurement matrices and sparse vectors is determined by $\text{rank}\{\widetilde{\mathbf{D}}\}$. The larger $\text{rank}\{\widetilde{\mathbf{D}}\}$ is, the more reliable recovery of sparse signals can be achieved. Hence, compared to the SMV and MMV, more reliable recovery performance can be achieved by the proposed GMMV. For the special case that multiple sparse signals are identical, the MMV reduces to the SMV since $\text{rank}(\mathbf{D}) = 1$, and there is no diversity gain by introducing multiple identical sparse signals. However, the GMMV in this case can still achieve diversity gain which comes from diversifying measurement matrices.

Pilot Design for CS Based Adaptive CSI Acquisition: According to the above discussions, a measurement matrix whose elements follow an i.i.d. Gaussian distribution satisfies the RIP. Furthermore, diversifying measurement matrices can further improve the recovery performance of sparse signals. This enlightens us to appropriately design pilot signals.

Specifically, each element of pilot signals is given by

$$[\mathbf{S}_p^{[q,G]}]_{t,m} = e^{j\theta_{t,m,p}}, 1 \leq t \leq G, 1 \leq m \leq M, \tag{20}$$

where $\mathbf{S}_p^{[q,G]} \in \mathbb{C}^{G \times M}$, and each $\theta_{t,m,p}$ has the i.i.d. uniform distribution in $[0, 2\pi)$, namely, the i.i.d. $\mathcal{U}[0, 2\pi)$. Note that the pilot signals for the CS based adaptive CSI acquisition are fixed

once they have been designed. Moreover, when designing the pilot signals, the worst case of $G = M$ has to be considered. It is readily seen that the designed pilot signals (20) guarantee that the elements of $\mathbf{\Phi}_p^{[q,G]}$ of (10), $\forall p$, obey the i.i.d. complex Gaussian distribution with zero mean and unit variance, i.e., the i.i.d. $\mathcal{CN}(0,1)$. Hence, the proposed pilot signal design is 'optimal', in terms of the reliable compression and recovery of sparse angular domain channels.

B. Time Slot Overhead for CS Based Adaptive CSI Acquisition

According to Theorem 1, for the optimization problem (10), $\widetilde{\mathbf{D}} = \mathbf{\Phi}_1^{[q,G]} \mathbf{F}$ with $\widetilde{\mathbf{D}} = [\mathbf{r}_1^{[q,G]} \, \mathbf{\Psi}_2^{-1} \mathbf{r}_2^{[q,G]} \cdots \mathbf{\Psi}_P^{-1} \mathbf{r}_P^{[q,G]}]$ and $\mathbf{F} = [\bar{\mathbf{h}}_1^{(q)} \, \bar{\mathbf{h}}_2^{(q)} \cdots \bar{\mathbf{h}}_P^{(q)}]$. Since $|\text{supp}\{\bar{\mathbf{h}}_p^{(q)}\}| = S_a$, it is clear that

$$\text{rank}\{\widetilde{\mathbf{D}}\} \leq \text{rank}\{\mathbf{F}\} \leq S_a. \tag{21}$$

Moreover, as $\mathbf{\Phi}_1^{[q,G]} \in \mathbb{C}^{G \times M}$,

$$\text{spark}(\mathbf{\Phi}_1^{[q,G]}) \in \{2, 3, \ldots, G+1\}. \tag{22}$$

Substituting (21) and (22) into (16) yields $G \geq S_a + 1$. Therefore, the smallest required time slot overhead is $G = S_a + 1$. As discussed in Section III-B, an appropriate value of G that ensures the reliable CSI acquisition is adaptively determined by Algorithm 1. By increasing the number of measurement vectors P, the required time slot overhead for reliable channel estimation can be reduced, since more measurement matrices and sparse signals can increase $\text{rank}\{\widetilde{\mathbf{D}}\}$.

C. Frequency-Domain Placement of Pilot Signals

Like any OFDM channel estimator, the proposed adaptive channel estimation and feedback scheme only estimates the channels at pilot subcarriers. Channels at data subcarriers are usually acquired based on the estimated channels at pilot subcarriers by using the off-the-shelf interpolation algorithms [21]. Clearly the frequency-domain placement of pilot signals Ω_ξ significantly influences the achievable performance of an interpolation algorithm. Additionally, due to the frequency-domain correlation of wireless channels, the channels of adjacent subcarriers exhibit strong correlation. Hence, two adjacent subcarriers both dedicated to the pilot may result in $\widetilde{\mathbf{D}}$ to be rank deficient. We adopt the widely used uniformly-spacing pilot placement with the spacing equal to the coherence bandwidth [21], which can reduce the correlation between different virtual angular domain channels, so that more diversity gain from the multiple sparse channels can be achieved.

D. Performance Analysis of Proposed DSAMP Algorithm

1) Complexity: The computational complexity of the proposed DSAMP algorithm (Algorithm 2) in each iteration mainly depends on the following operations.

Signal proxy (*line 7*): The matrix-vector multiplication involved has the complexity on the order of $\mathsf{O}(PMG)$.

ℓ_2**-norm operation** (*lines 8, 10, 13, 14, 16, 18, and 23*): The computational complexity is $\mathsf{O}(P)$.

Identifying or Pruning (*lines 8* and *10*): The cost to locate the largest $\mathcal{T}$ entries of a size-M vector is $\mathsf{O}(M)$ [24].

TABLE I
COMPUTATIONAL COMPLEXITY TO ESTIMATE ONE SPARSE SIGNAL

Algorithm	Number of complex multiplications in each iteration
OMP	$2GM + M + 2Gi^2 + i^3$
SP	$2GM + G + M + 2S_a + 2(2GS_a^2 + S_a^3)$
SAMP	$2GM + G + M + 3S_a + 2(2Gj^2 + j^3)$
DSAMP	$2GM + G + M + 3S_a + 2(2Gj^2 + j^3)$

Note: i denotes the iteration index, and j denotes the stage index.

LS operation (*lines 9* and *11*): LS solution has the computational complexity on the order of $\mathsf{O}(P(2G\mathcal{T}^2 + \mathcal{T}^3))$ due to the joint recovery of P sparse signals [27].

Residual computation (*line 12*): The complexity of computing the residual is $\mathsf{O}(PMG)$.

Obviously the matrix inversion implemented in Algorithm 2 for LS operation contributes to most of the computational complexity. Table I compares the complexity of the proposed DSAMP algorithm, classical OMP algorithm, SP algorithm [24], and SAMP algorithm, in terms of the number of required complex multiplications in each iteration to estimate one sparse signal. It is clear that the four algorithms have the same order of computational complexity.

2) Stopping Criteria: For the conventional SAMP algorithm, the iterative procedure stops when the residual is less than a given threshold. By contrast, the proposed DSAMP algorithm has two halting criteria, and meeting either of them will trigger the termination of the iterative procedure. Regarding the first halting criterion, when the average energy of the wireless channels at a certain virtual angular coordinate is lower than the noise floor (*lines 14* and *23*), the iterative procedure stops. When the residual of the current stage becomes larger than that of the previous stage (*line 16*), the second halting criterion is met and the algorithm also terminates.

Due to $S_a \ll M$, after coordinates accounting for the majority of the channel energy is achieved, the next iteration will include a virtual angular coordinate that is dominated by the AWGN. The energy of such a new coordinate is usually lower than the noise floor. The first stopping criterion is designed to detect this situation and to terminate the algorithm when an appropriate number of virtual angular domain coordinates have been tracked.

As for the second halting criterion, the DSAMP algorithm is similar to the conventional SP algorithm in each stage with the fixed sparsity level, which can guarantee the sparse signal recovery with the exact sparsity level. The residual of the stage with the exact sparsity level is usually smaller than that with the incorrect sparsity level. Therefore, the DSAMP algorithm stops at the stage when the smallest residual is reached, which tends to be the stage associated with the exact sparsity level of the channels in the virtual angular domain.

E. Performance Bound of Channel Estimation

By omitting q, p, and ξ_p in (11) for simplicity, the variance of the channel estimation can be expressed as

$$\mathrm{var}\left\{\widehat{\mathbf{h}}\right\} = \mathrm{E}\left\{\left\|\widehat{\mathbf{h}} - \mathbf{h}\right\|_2^2\right\} = \mathrm{E}\left\{\left\|(\mathbf{A}_B^*)^{\mathrm{T}}\widehat{\bar{\mathbf{h}}} - (\mathbf{A}_B^*)^{\mathrm{T}}\bar{\mathbf{h}}\right\|_2^2\right\}$$
$$= \mathrm{E}\left\{\left\|\widehat{\bar{\mathbf{h}}} - \bar{\mathbf{h}}\right\|_2^2\right\} = \mathrm{var}\left\{\widehat{\bar{\mathbf{h}}}\right\}. \quad (23)$$

Consider the CRLB for the estimation problem associated with (9) given the true channel $\bar{\mathbf{h}}$ and the support set Θ. Again for notational simplicity, q, G, and p in $\mathbf{r}_p^{[q,G]}$, $\boldsymbol{\Phi}_p^{[q,G]}$, and $\mathbf{v}_p^{[q,G]}$ are omitted. Since the distribution of $\mathbf{v}$ is $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_G)$, the conditional probability density function (PDF) of $\mathbf{r}$ given $\bar{\mathbf{h}}$ is

$$p_{\mathbf{r}|\bar{\mathbf{h}}}\left(\mathbf{r}|\bar{\mathbf{h}}\right) = \frac{1}{(\pi\sigma^2)^G}e^{-\frac{\|\mathbf{r}-(\boldsymbol{\Phi})_\Theta(\bar{\mathbf{h}})_\Theta\|_2^2}{\sigma^2}}, \quad (24)$$

where σ^2 is the power of the effective noise. The element at the s_ith-row and s_jth-column of the Fisher information matrix $\mathcal{I}\left((\bar{\mathbf{h}})_\Theta\right)$ associated with this estimation problem is

$$\left[\mathcal{I}\left((\bar{\mathbf{h}})_\Theta\right)\right]_{s_i,s_j} = \frac{1}{\sigma^2}\left[((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta\right]_{s_i,s_j}, \quad (25)$$

where $1 \leq s_i, s_j \leq |\Theta|$. Therefore, we have

$$\mathrm{var}\left\{\widehat{\bar{\mathbf{h}}}\right\} \geq \mathrm{Tr}\left\{(\mathcal{I}((\bar{\mathbf{h}})_\Theta))^{-1}\right\} = \sigma^2\mathrm{Tr}\left\{(((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta)^{-1}\right\}. \quad (26)$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_{S_a}$ be the S_a eigenvalues of the matrix $((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta \in \mathbb{C}^{S_a \times S_a}$. It is clear that

$$\mathrm{Tr}\left\{(((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta)^{-1}\right\} = \sum_{i=1}^{S_a}\lambda_i^{-1}, \quad (27)$$

which can be calculated after the pilot signals, the geometrical structure of the BS antenna array, and the support set of the channel vectors in the virtual angular domain are given.

However, the support set Θ is 'random' since the channel vectors in practice are random and the elements of the measurement matrix $\boldsymbol{\Phi}$ obey the i.i.d. $\mathcal{CN}(0,1)$. Thus we should consider the 'expectation' of the CRLB defined by

$$\mathrm{E}\left\{\mathrm{var}\left\{\widehat{\bar{\mathbf{h}}}\right\}\right\} \geq \mathrm{E}\left\{\sigma^2\sum_{i=1}^{S_a}\lambda_i^{-1}\right\}. \quad (28)$$

For the matrix $((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta$ with the elements of $\boldsymbol{\Phi}$ obeying the i.i.d. $\mathcal{CN}(0,1)$, its eigenvalues $\{\lambda_i\}_{i=1}^{S_a}$ obey the following joint distribution [28]

$$p_{\vec{\lambda}}(\lambda_1, \lambda_2, \ldots, \lambda_{S_a}) = e^{-\sum_{i=1}^{S_a}\lambda_i}\prod_{i=1}^{S_a}\left(\frac{\lambda_i^{G-S_a}}{(S_a-i)!,\, i!}\prod_{j>i}^{S_a}(\lambda_j-\lambda_i)^2\right). \quad (29)$$

Consequently, the expectation of the CRLB can be written as

$$\mathrm{E}\left\{\mathrm{var}\left\{\widehat{\bar{\mathbf{h}}}\right\}\right\} \geq \int_0^\infty \cdots \int_0^\infty \sigma^2\sum_{i=1}^{S_a}\lambda_i^{-1}p_{\vec{\lambda}}(\lambda_1, \ldots, \lambda_{S_a})d\lambda_1\cdots d\lambda_{S_a}. \quad (30)$$

Since the computation of (30) can be highly complex, in practice we adopt the performance of the oracle LS estimator as the performance bound in the simulation study.

F. Adaptive Pilot Design and Required Time Slot Overhead for Closed-Loop Channel Tracking

For the simplicity of analysis, the true support set Θ and the sparsity level S_a of the virtual angular domain channels are assumed to have been acquired by the CS based adaptive CSI acquisition. Clearly, if S_a is known, the smallest time slot overhead for CSI acquisition can be reduced to $G = S_a$.
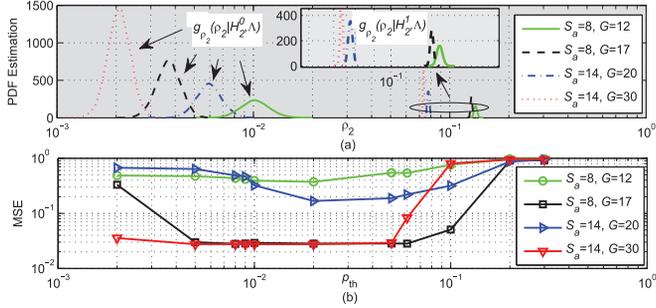
Fig. 4. The selection of threshold p_{th} in Algorithm 2 given SNR = 15 dB and $P = 64$: (a) estimated PDFs of $g_{\rho_2}(\rho_2|\mathcal{H}_2^0, \Lambda)$ and $g_{\rho_2}(\rho_2|\mathcal{H}_2^1, \Lambda)$; and (b) MSE performance of the DSAMP algorithm as the function of p_{th}.

Fig. 5. The selection of threshold ε in Algorithm 1 given SNR = 15 dB, $G_0 = 10$, and $P = 64$: (a) estimated PDF of $g_{\rho_1}(\rho_1|\mathcal{H}_1^0)$; and (b) $\Pr(\mathcal{H}_1^1|\mathcal{H}_1^0)$ and $\Pr(\mathcal{H}_1^0|\mathcal{H}_1^1)$ as the functions of ε.

With the known Θ, by exploiting the arithmetic-harmonic means inequality [27], (27) can be further expressed as

$$\text{Tr}\left\{(((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta)^{-1}\right\} \geq \frac{S_a^2}{\sum_{i=1}^{S_a} \lambda_i} = \frac{S_a^2}{\text{Tr}\left\{((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta\right\}}, \tag{31}$$

where the equality holds if and only if $\lambda_1 = \lambda_2 = \cdots = \lambda_{S_a}$. This indicates that $((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta$ should be a diagonal matrix with identical diagonal elements to approach the lower bound. In particular, for $(\boldsymbol{\Phi})_\Theta$ with $\text{Tr}\left\{((\boldsymbol{\Phi})_\Theta)^*(\boldsymbol{\Phi})_\Theta\right\} = S_a G$,

$$\text{var}\left\{\widehat{\mathbf{h}}\right\} \geq \sigma^2 S_a / G, \tag{32}$$

and the lower bound of (32) is attained if $(\boldsymbol{\Phi})_\Theta$ is a unitary matrix scaled by the factor $\sqrt{G}$. This has inspired us to design the pilot signal matrix as $\mathbf{S} = \sqrt{G} \mathbf{U}_{S_a}\left(((\mathbf{A}_B^*)^{\mathrm{T}})_\Theta\right)^\dagger$, where $\mathbf{U}_{S_a} \in \mathbb{C}^{S_a \times S_a}$ is a unitary matrix. With this non-orthogonal pilot matrix, the lower bound of (32) is attained, i.e., $\text{var}\left\{\widehat{\bar{\mathbf{h}}}\right\} = \sigma^2 S_a / G = \sigma^2$.

G. Selection of Thresholds for Algorithms 1 and 2

1) p_{th} in Algorithm 2: Consider the case Λ that for the stage of $\mathcal{T} = S_a + 1$ in Algorithm 2, the final estimated support set, denoted as $\widehat{\Omega}_\mathcal{T}$, is the proper superset of the true support set Θ, i.e., $\widehat{\Omega}_\mathcal{T} \supsetneq \Theta$. This case implies that $\widehat{\Omega}_\mathcal{T}$ includes the support index associated with the noise. Define the test statistic as $\rho_2 = \sum_{p=1}^{P} \left\|[\mathbf{c}_p]_{\widetilde{l}}\right\|_2^2 / P$ with $\widetilde{l} \in \widehat{\Omega}_\mathcal{T}$ (lines 13 and 14 in Algorithm 2). Two complete hypotheses for the case Λ are defined as: $\mathcal{H}_2^0$, indicating $\widetilde{l} \in \Theta$, and $\mathcal{H}_2^1$, indicating $\widetilde{l} \notin \Theta$. Furthermore, denote the PDFs of ρ_2 under $\mathcal{H}_2^0$ and $\mathcal{H}_2^1$ as $g_{\rho_2}(\rho_2|\mathcal{H}_2^0, \Lambda)$ and $g_{\rho_2}(\rho_2|\mathcal{H}_2^1, \Lambda)$, respectively. By using the *ksdensity* function of MATLAB, we can obtain the estimated PDFs according to Monte-Carlo simulations, since the closed-form expressions are difficult to obtain.

Fig. 4(a) depicts the estimated PDFs of $g_{\rho_2}(\rho_2|\mathcal{H}_2^0, \Lambda)$ and $g_{\rho_2}(\rho_2|\mathcal{H}_2^1, \Lambda)$ with typical values of S_a and G, given SNR = 15 dB and $P = 64$. Fig. 4(b) provides the MSE performance of Algorithm 2 as the function of p_{th}, which indicates that $p_{\text{th}} = 0.02$ achieves good MSE performance given typical values of S_a and G. Following a similar procedure, suitable values of p_{th} for different SNRs can be obtained.

2) ε in Algorithm 1: Consider the test statistic $\rho_1 = \sum_{p=1}^{P} \left\|\mathbf{r}_p^{[q,G]} - \boldsymbol{\Phi}_p^{[q,G]} \widehat{\mathbf{h}}_p^{(q)}\right\|_2^2 / (GP)$ (line 6 in Algorithm
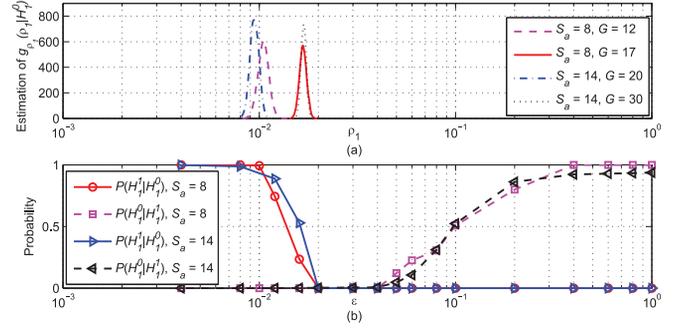
1 with iteration index i omitted) and the two complete hypotheses $\mathcal{H}_1^0$ and $\mathcal{H}_1^1$, where $\mathcal{H}_1^0$ indicates that the support set of $\{\widehat{\mathbf{h}}_p^{(q)}\}_{p=1}^{P}$ is correct, and $\mathcal{H}_1^1$ is complementary to $\mathcal{H}_1^0$. Under $\mathcal{H}_1^0$, $\rho_1 = \sum_{p=1}^{P}\left\|(\mathbf{I} - \boldsymbol{\Phi})_\Theta((\boldsymbol{\Phi})_\Theta)^\dagger) \mathbf{v}_p^{[q,G]}\right\|_2^2 / (GP)$. However, under $\mathcal{H}_1^1$, the closed-form expression of ρ_1 is difficult to derive. Similar to Fig. 4, Fig. 5(a) provides the estimates of the PDF $g_{\rho_1}(\rho_1|\mathcal{H}_1^0)$ with typical values of S_a and G, given SNR = 15 dB, $G_0 = 10$ and $P = 64$. According to Neyman-Pearson criterion [23], an appropriate threshold ε should minimize the probability of false alarm given the probability of miss. Fig. 5(b) depicts the simulated probability of false alarm $\Pr(\mathcal{H}_1^1|\mathcal{H}_1^0)$ and the miss probability $\Pr(\mathcal{H}_1^0|\mathcal{H}_1^1)$ of Algorithm 1 as the functions of ε, where $p_{\text{th}} = 0.02$ is used in the simulation. The results of Fig. 5(b) indicate that $\varepsilon = 0.03$ minimizes both $\Pr(\mathcal{H}_1^1|\mathcal{H}_1^0)$ and $\Pr(\mathcal{H}_1^0|\mathcal{H}_1^1)$ given typical values of S_a and G. Similarly, appropriate values of ε for different SNRs can be obtained.

V. SIMULATION RESULTS

Massive MIMO system with the ULA of $M = 128$ antennas and $d = \lambda/2$ was considered. The spatial angle spread varied from $10°$ to $20°$ [4], [16], and thus the effective sparsity level in the virtual angular domain S_a was in the range of 8 to 14. In the simulations, $f_c = 2$ GHz, $B_s = 10$ MHz, $N = 2048$, and $v = 36$ km/h, while the channels in the virtual angular domain exhibited the spatially common sparsity over $Q = 5$ time blocks. The length of the guard interval was 64, which indicates that the system can combat the maximum delay spread of 6.4 μs [20], and thus we adopted $P = 64$ [10]. The threshold parameters, ε in Algorithm 1 and p_{th} in Algorithm 2, were selected according to Section IV-G. Specifically, we set p_{th} to 0.06, 0.02, 0.01, 0.008, and 0.005, while ε to 0.08, 0.03, 0.0.09, 0.003, and 0.001, respectively, at the SNR of 10dB, 15dB, 20dB, 25dB and $\geq$ 30 dB. The oracle LS estimator and the CRLB were used as the benchmarks for the CS based adaptive CSI acquisition and the following closed-loop channel tracking, respectively. The time slot overhead G employed in the closed-loop channel tracking scheme was set to the estimated sparsity level obtained by the CS based adaptive CSI acquisition stage. The joint OMP (J-OMP) based CSI acquisition scheme [15] was also adopted for comparison.
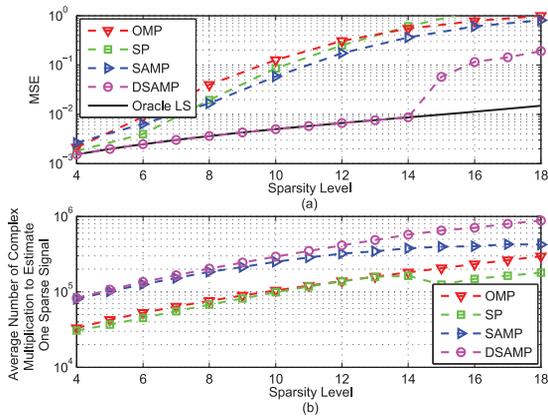
Fig. 6. Performance comparison of different CS algorithms as functions of the sparsity level S_a given $G = 30$, $P = 64$ and SNR $= 20$ dB: (a) MSE performance, and (b) computational complexity.

Fig. 6 compares the MSE performance and complexity of four CS algorithms under various sparsity levels S_a. In the simulations, $P = 64$ sparse signals with the length of $M = 128$ had the common sparsity, the measurement dimension was $G = 30$, and SNR $= 20$ dB, while P measurement matrices were mutually independent with elements obeying the i.i.d $\mathcal{CN}(0,1)$. Note that the conventional OMP and SP algorithms require S_a as the priori information. Fig. 6(a) shows that the DSAMP algorithm achieves the best MSE performance and it approaches the oracle LS estimator for $S_a \leq 14^2$. This is because the DSAMP algorithm jointly estimates P sparse signals by exploiting the common sparsity. Moreover, Fig. 6(b) shows that the complexity of the DSAMP algorithm is slightly higher than those of its counterparts, but all the four CS algorithms have the same order of complexity.

We defined the sparse signal detection probability as the probability of correctly acquiring the support set of sparse signal (channel). Fig. 7 compares the detection probabilities as functions of the measurement dimension G achieved by the SMV, MMV, and GMMV in noiseless scenario. In the simulation, the length of multiple sparse signals was $M = 128$ with the common sparsity level $S_a = 8$, and the DSAMP algorithm was employed to recover sparse signals. In particular, the SMV recovers single sparse signal from single measurement matrix, and the MMV jointly recovers P sparse signals with the multiple identical measurement matrices, where the elements of the measurement matrix obey the i.i.d. $\mathcal{CN}(0,1)$. By contrast, the GMMV recovers P sparse signals with mutually independent measurement matrices in parallel, where the elements of the measurement matrices also obey the i.i.d. $\mathcal{CN}(0,1)$. From Fig. 7, it is clear that the joint processing of multiple sparse signals with the common support set and diversifying measurement matrices significantly enhance the performance of sparse signal recovery. For example, to obtain the detection probability of one

---

²The DSAMP algorithm suffers from certain performance loss, compared to the oracle LS estimator in the noisy scenario with $G \leq 2S_a$. For $G = 2S_a$, the case of $\Omega^{i-1} \cap \Gamma = \emptyset$ (*line 9*) and $\Omega = \Omega^{i-1}$ (*line 13*) may repeatedly appear due to noise, resulting in the failure of the backtracking function of *lines 7 ~ 12*. For $G < 2S_a$, the case of $\Omega^{i-1} \cap \Gamma = \emptyset$ can lead to a poor LS estimation (*line 9*) due to $|\Omega| > G$. Also see [22].
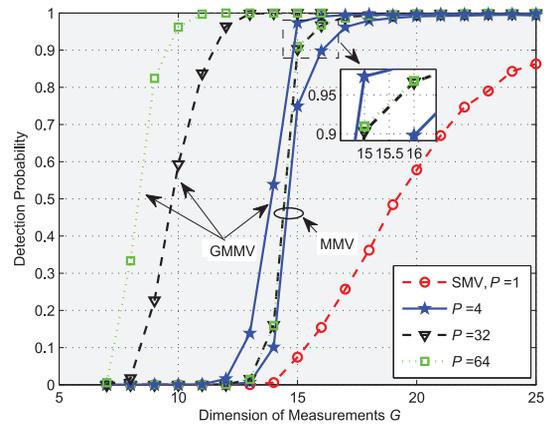
Fig. 7. Comparison of the sparse signal detection probabilities of the SMV, MMV and the proposed GMMV as functions of G.
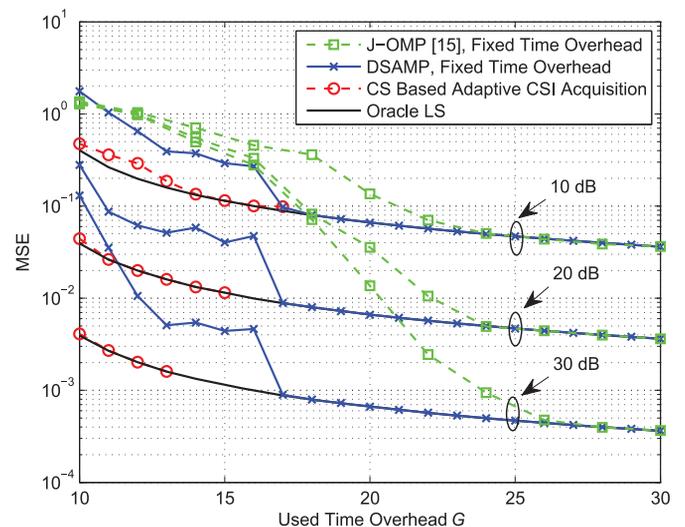
Fig. 8. MSE performance of different channel estimation and feedback schemes as functions of the time overhead G and SNR.

with $P = 64$, the MMV requires $G = 17$, but the proposed GMMV only needs $G = 11$, which indicates a reduction of approximately 35% in the required time slot overhead. Even the GMMV with $P = 4$ outperforms the MMV with $P = 64$.

Fig. 8 compares the MSE performance of the J-OMP scheme [15] with fixed G, the DSAMP algorithm with fixed G, and the CS based adaptive CSI acquisition scheme (Algorithm 1), where $S_a = 8$ was considered. The oracle LS estimator with the known support set of the sparse channel vectors was adopted as the performance bound. From Fig. 8, it can be seen that the J-OMP based CSI acquisition scheme performs poorly. By contrast, the proposed DSAMP algorithm is capable of approaching the oracle LS performance bound when $G > 2S_a$. However, there still exists a significant performance gap between the DSAMP algorithm and the oracle LS estimator for $G \leq 2S_a$. This is because the unreliable sparse signal recovery may occur when the time slot overhead G is insufficient, which degrades the MSE performance. Fortunately, the proposed CS based adaptive CSI acquisition scheme can adaptively adjust G to acquire the robust channel estimation. Observe from Fig. 8 that the proposed CS based adaptive CSI acquisition scheme approaches the oracle

Fig. 9. Distributions of adaptively selected time slot overhead by the CS based adaptive CSI acquisition scheme for different sparsity levels and SNRs.

LS performance bound even for $G \leq 2S_a$. Note that for Algorithm 1, we only plot the MSE associated with $G \leq 2S_a$, because Algorithm 1 actually determines an appropriate $G \leq 2S_a$ adaptively.

Fig. 9 depicts the distributions of the adaptively determined time slot overhead G by the CS based adaptive CSI acquisition, given different sparsity level S_a and SNRs. In Algorithm 1, G_0 was set to 8. The results of Fig. 9 show that the proposed scheme can adaptively determine an appropriate G according to S_a. As pointed out in Section II-D, to reliably acquire CSI, the required G in conventional schemes can be as large as $G = M = 128$. By exploiting the spatially common sparsity and temporal correlation of massive MIMO channels, the proposed scheme can effectively estimate the channels associated with hundreds of antennas at the BS with a dramatically reduced time slot overhead. Considering $S_a = 8$ at SNR = 30 dB for example, our scheme only uses a time slot overhead of $G \approx 10$ to acquire CSI at the BS, which represents a reduction in the required G by about 92%, compared to conventional schemes.

Fig. 10 plots the distributions of the acquired sparsity level $\widehat{S}_a$ by the proposed CS based adaptive CSI acquisition scheme, under the same settings of Fig. 9. The results of Fig. 10 show that the proposed scheme can accurately acquire the true sparsity level S_a. Note that the acquired $\widehat{S}_a$ may be smaller than S_a at low SNR. This is because some virtual angular domain coordinates whose channel energy is lower than the noise floor may be discarded by the DSAMP algorithm. Because we set the time slot overhead G to $\widehat{S}_a$ in the closed-loop channel tracking, Fig. 10 also provides the probability distributions of the time slot overhead used in the closed-loop channel tracking stage. As expected, the required time slot overhead in this stage is smaller than the time slot overhead actually used in the CS based adaptive CSI acquisition stage, which is confirmed by comparing Fig. 10 to Fig. 9.

Fig. 11 compares the MSE and required average time slot overhead $\bar{G}$ of the CS based adaptive CSI acquisition with those of the closed-loop channel tracking for different S_a at SNR = 20 dB. The initial overhead $G_0 = 10$ was set for the CS based adaptive CSI acquisition. It is clear that benefiting from the accurately estimated sparsity level information provided by the CS based adaptive CSI acquisition, the closed-loop channel tracking enjoys the better MSE performance with a smaller required time slot overhead. For $S_a = 14$, the required $\bar{G}$ by the CS based adaptive CSI acquisition and following closed-loop channel tracking are 20.43 and 14.02, respectively. Since the acquired CSI by the CS based adaptive CSI acquisition is utilized to adaptively adjust the pilot signal for enhancing performance, the closed-loop channel tracking approaches the CRLB, as can be seen in Fig. 11. Also note that for the CS based adaptive CSI acquisition, the ratio $\bar{G}/S_a$ increases slightly as S_a increases. Hence the MSE performance of the CS based adaptive CSI acquisition improves slightly as the true sparsity level S_a increases.

Fig. 12 provides the MSE performance comparison for different channel estimation and feedback schemes, given $S_a = 8$ and various SNRs. Both the J-OMP based CSI acquisition scheme [15] and the DSAMP algorithm used the fixed $G = 15$. For the CS based adaptive CSI acquisition scheme, $G_0 = 13$ was considered. The required average time slot overheads for the proposed scheme are also marked in Fig. 12. Again, it is clear that the proposed CS based adaptive CSI acquisition stage (Algorithm 1), which uses the DSAMP algorithm with fixed G to adaptively determine an appropriate time slot overhead, outperforms the J-OMP based CSI acquisition scheme and DSAMP algorithm with a reduced time slot overhead requirement. By utilizing the accurately estimated channel sparsity information provided by the CS based adaptive CSI acquisition scheme, the closed-loop channel tracking stage can adaptively adjust the pilot signal to approach the CRLB with a further reduced time slot overhead. Specifically, the proposed scheme can reliably acquire the CSI of this massive MIMO system, approaching the CRLB, with an average time slot overhead $\bar{G} < 2S_a$.
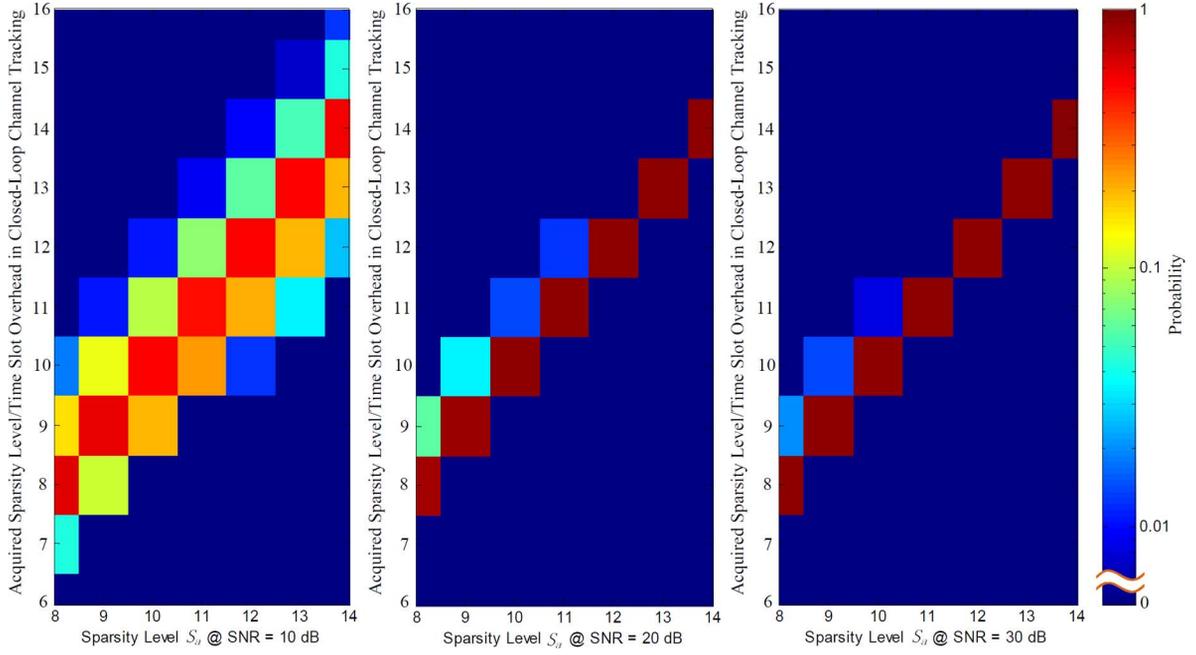
Fig. 10. Distributions of the acquired sparsity level $\widehat{S}_a$ by the CS based adaptive CSI acquisition scheme (which is then used as the time slot overhead for the proposed closed-loop channel tracking scheme) for different sparsity levels and SNRs.
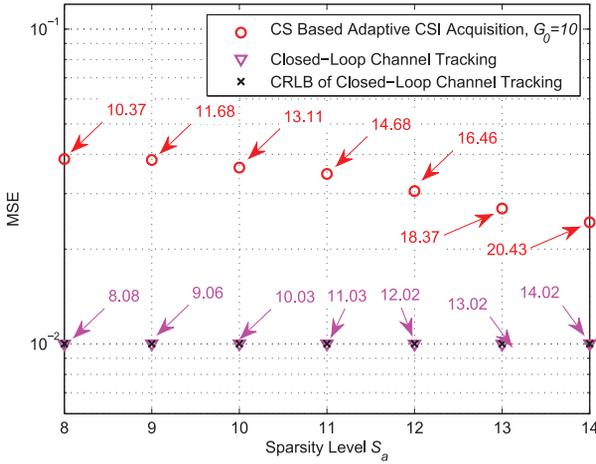
Fig. 11. MSE performance comparison of the CS based adaptive CSI acquisition stage and closed-loop channel tracking stage for different sparsity levels S_a at SNR = 20 dB, where the required $\bar{G}$ for each case is indicated.
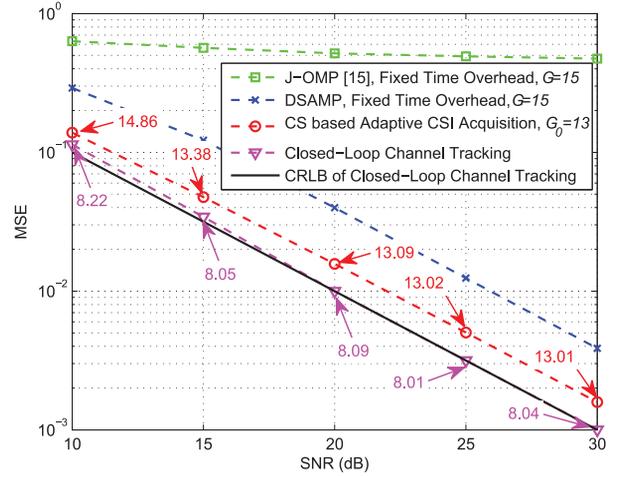
Fig. 12. MSE performance comparison of different channel estimation and feedback schemes for various SNRs and true sparsity level $S_a = 8$, where the required average time slot overhead $\bar{G}$ for each case is indicated.

Fig. 13 compares the downlink bit error rate (BER) performance with zero-forcing (ZF) precoding, where the precoding is based on the estimated CSI corresponding to Fig. 12 under the same setup. In the simulations, the BS simultaneously served 16 users using 16-quadrature amplitude modulation signaling, and the effective noise in CSI acquisition was only introduced in the downlink channel. It can be observed that the proposed channel estimation and feedback scheme outperforms its counterparts, and its BER performance is capable of approaching that of the CRLB.

VI. CONCLUSIONS

An adaptive channel estimation and feedback scheme has been proposed for FDD massive MIMO, which achieves robust and accurate CSI acquisition at the BS, while dramatically reducing the overhead for channel estimation and feedback. The
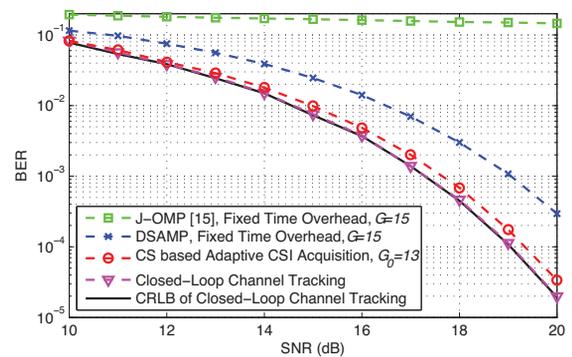
Fig. 13. Downlink BER performance with ZF precoding, where the CSI at the BS is acquired by different channel estimation and feedback schemes.

proposed scheme consists of two stages, the CS based adaptive CSI acquisition and the following closed-loop channel tracking. By exploiting the spatially common sparsity of massive MIMO channels within the system bandwidth, the CS based adaptive CSI acquisition can acquire the high-dimensional CSI from a small number of non-orthogonal pilots. The closed-loop channel tracking, which exploits the spatially common sparsity of massive MIMO channels over multiple consecutive time blocks, can effectively utilize the acquired CSI in the first stage to approach the CRLB. We have generalized the MMV to the GMMV in CS theory and provided the CRLB of the proposed scheme, which enlightens us to design the non-orthogonal pilot for different stages of the proposed scheme. Simulation results have confirmed that our scheme can reliably acquire the CSI of massive MIMO systems, specifically, approaching the performance bound with an adaptively determined time slot overhead.

References

[1] 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.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] J. Hoydis, S. t. 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.

[4] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[5] 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.

[6] 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.

[7] D. Angelosante, E. Biglieri, and M. Lops, "Sequential estimation of multipath MIMO-OFDM channels," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3167–3181, Aug. 2009.

[8] 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.

[9] Y.-H. Nam, Y. Akimoto, Y. Kim, M.-L. 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] 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.

[12] P. Cheng and Z. Chen, "Multidimensional compressive sensing based analog CSI feedback for massive MIMO-OFDM systems," in *Proc. Veh. Technol. Conf. (VTC)—Fall 2014*, Vancouver, Canada, Sep. 14–17, 2014, pp. 1–6.

[13] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. WCNC*, Shanghai, China, Apr. 1–4, 2012, pp. 492–497.

[14] X. Rao, V. K. N. Lau, and X. Kong, "CSIT estimation and feedback for FDD multi-user massive MIMO systems," in *Proc. ICASSP*, Florence, Italy, May 4–9, 2014, pp. 3157–3161.

[15] 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.

[16] 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.

[17] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1101, Oct. 2014.

[18] 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, 2007.

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

[20] *Mobile Broadband Multimedia Networks, Techniques, Models and Tools for 4G*, L. M. Correia, Ed. San Diego, CA, USA: Academic, 2006.

[21] Y. S. Cho, J. Kim, W. Y. Yang, and C.-G Kang, *MIMO-OFDM Wireless Communications With MATLAB*. Singapore: Wiley (Asia) Pte. Ltd., 2010.

[22] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. 42nd Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 26–29, 2008, pp. 581–587.

[23] S. M. Kay, *Fundamentals of Statistical Signal processing, Volume II: Detection Theory*. Englwood Cliffs, NJ, USA: Prentice-Hall, 1998.

[24] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2009.

[25] P. Billingsley, *Probability and Measure*. New York, NY, USA: Wiley, 1979.

[26] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[27] A. Björck, *Numerical Methods for Matrix Computations*. New York, NY, USA: Springe International Publishing AG, 2014.

[28] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Zhen Gao (S'14) received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 2011.

He is currently working towards Ph.D. degree in Electronic Engineering of Tsinghua University, Beijing, China. His research interests include sparse signal processing, massive MIMO, millimeter-wave communications, etc.
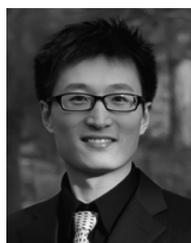
Linglong Dai (M'11–SM'14) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003; the M.S. degree (with the highest honor) from the China Academy of Telecommunications Technology, Beijing, China, in 2006; and the Ph.D. degree (with the highest honors) from Tsinghua University, Beijing, in 2011.

From 2011 to 2013, he was a Postdoctoral Fellow with the Department of Electronic Engineering, Tsinghua University, where he has been an Assistant Professor since July 2013. He has authored or coauthored over 60 journal and conference papers. His research interests are in wireless communications, with a focus on multi-carrier modulations, multiple antenna systems, multiple access techniques.

Dr. Dai received the Outstanding Ph.D. Graduate Award of Tsinghua University in 2011, the Excellent Doctoral Dissertation Award of Beijing in 2012, the IEEE ICC Best Paper Award in 2013, the National Excellent Doctoral Dissertation Nomination Award in 2013, the IEEE ICC Best Paper Award in 2014, the URSI Young Scientists Award in 2014, and the IEEE Scott Helt Memorial Award in 2015 (IEEE Transactions on Broadcasting Best Paper Award). He currently serves as a co-chair of the IEEE ComSoc Special Interest Group (SIG) "Signal processing techniques in 5G communication systems".

Zhaocheng Wang (M'09–SM'11) received his B.S., M.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was a Postdoctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was with OKI Techno Centre (Singapore) Pte. Ltd., Singapore, where he was first a Research Engineer and later became a Senior Engineer. From 1999 to 2009, he was with Sony Deutschland GmbH, where he was first a Senior Engineer and later became a Principal Engineer. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University and serves as the Director of Broadband Communication Key Laboratory, Tsinghua National Laboratory for Information Science and Technology (TNlist). He has authored or coauthored around 90 journal papers (SCI indexed). He is the holder of 34 granted U.S./EU patents. He co-authored two books, one of them with the title of *Millimeter Wave Communication Systems*, was selected by IEEE Series on Digital & Mobile Communication and published by Wiley-IEEE Press. His research areas include wireless communications, visible light communications, millimeter-wave communications, and digital broadcasting.

Professor Wang is a Fellow of the Institution of Engineering and Technology. Currently he serves as an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Associate Editor of IEEE COMMUNICATIONS LETTERS, and has also served as Technical Program Committee Co-chairs of various international conferences.

Sheng Chen (M'90–SM'97–F'08) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with Electronics and Computer Science, the University of Southampton, UK, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr. Chen's research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 550 research papers.

Dr. Chen is a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an ISI highly cited researcher in engineering (March 2004). In 2014, he was elected as a Fellow of the United Kingdom Royal Academy of Engineering.