Joint Channel Estimation and Feedback with Low Overhead for FDD Massive MIMO Systems

Linglong Dai, Zhen Gao, and Zhaocheng Wang
Tsinghua National Laboratory for Information Science and Technology (TNList),
Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China
Emails: daill@tsinghua.edu.cn, gao-z11@mails.tsinghua.edu.cn, zcwang@tsinghua.edu.cn

***Abstract*—Accurate channel state information (CSI) is essential to realize the potential advantages of massive MIMO. However, the overhead required by conventional channel estimation and feedback schemes will be unaffordable, especially for frequency division duplex (FDD) massive MIMO. To solve this problem, we propose a structured compressive sensing (SCS) based spatio-temporal joint channel estimation and feedback scheme to reduce the required overhead. Particularly, we first propose the non-orthogonal pilots at the base station (BS) under the framework of CS theory. Then, an adaptive structured subspace pursuit (ASSP) algorithm is proposed to jointly estimate channels associated with multiple OFDM symbols at the receiver, whereby the spatio-temporal common sparsity of massive MIMO channels is exploited to improve the channel estimation accuracy. Moreover, we propose a parametric channel feedback scheme, which exploits the sparsity of channels to acquire accurate CSI at the BS with reduced feedback overhead. Simulation results show that the channel estimation performance approaches that of the oracle least squares (LS) channel estimator, and the parametric channel feedback scheme only suffers from a negligible performance loss compared with the complete channel feedback scheme.**

***Index Terms*—Massive MIMO, structured compressive sensing (SCS), channel estimation, channel feedback.**

I. INTRODUCTION

By exploiting the large number of degrees of spatial freedom, massive MIMO can boost the system capacity and energy efficiency by orders of magnitude. Therefore, massive MIMO has been widely recognized as a key enabling technique for future spectrum and energy efficient 5G communications [1].

In massive MIMO systems, the accurate channel state information (CSI) is essential for signal detection, beamforming, resource allocation, etc. However, due to massive antennas at the base station (BS), each user has to accurately acquire and feed back channels associated with hundreds of transmit antennas, which results in the prohibitively high pilot and feedback overhead. Hence, how to realize the accurate channel estimation and reliable channel feedback with affordable overhead becomes a challenging problem, especially for frequency division duplex (FDD) massive MIMO systems [1]. By far, most of studies on massive MIMO sidesteps this challenge by assuming the time division duplex (TDD) protocol, where the CSI in the uplink can be more easily acquired at the BS due to the small number of single-antenna users and the powerful signal processing capability of the BS, and then the channel reciprocity property can be leveraged to directly obtain the CSI in the downlink. However, since we have to reuse the limited number of orthogonal pilots in adjacent cells, TDD massive MIMO suffers from the well-known problem of pilot contamination. Moreover, due to the calibration error of radio frequency chains and limited coherence time, the CSI acquired in the uplink may be inaccurate for the downlink [2]. Finally, compared with TDD, FDD can provide more efficient communications with symmetric traffic and low latency [2], thus it has dominated current cellular systems [12]. Therefore, this paper focuses on the more challenging problem of channel estimation and feedback for FDD massive MIMO systems.

It has been proven that the equi-spaced and equi-power orthogonal pilots are optimal to estimate the non-correlated Rayleigh fading MIMO channels for one OFDM symbol, where the required pilot overhead increases with the number of transmit antennas [3]. Currently, orthogonal pilots have been widely used in standardized MIMO systems, where the pilot overhead is not a big issue due to the small number of transmit antennas in existing MIMO systems (e.g., up to eight antennas in LTE-Advanced system) [4]. However, this issue can be critical in massive MIMO systems due to massive antennas at the BS (e.g., 128 antennas or even more at the BS [2]). For FDD massive MIMO systems, we have proposed to exploit the temporal correlation and sparsity of time-domain channels to reduce the pilot overhead [5], but the interference cancellation of training sequences of different transmit antennas will be difficult when the number of transmit antennas is large. [6] leveraged the spatial correlation and sparsity of time-domain MIMO channels to acquire CSI with reduced pilot overhead, but the assumption of known channel sparsity level at the user is unrealistic. By exploiting the spatial channel correlation, a compressive sensing based channel estimation was proposed [7], but it considers the estimation of flat-fading MIMO channels, where the leveraged spatial correlation can be impaired due to the non-ideal antenna array [2]. Additionally, [8] proposed an open-loop and closed-loop channel estimation scheme for massive MIMO, but the requirement that the long-term channel statistics perfectly known at the user can be difficult.

Moreover, CSI at the transmitter (CSIT) in massive MIMO systems is also necessary for signal processing at the BS, which indicates that the acquired CSI at the user should be accurately fed back to the BS. For conventional quantized channel feedback strategy, the design, storage, and encoding of large Grassmannian codebooks can be challenging in massive MIMO systems. Moreover, the quantization error will reduce the beamforming gain. By leveraging the spatial channel correlation, the analog channel feedback schemes have been proposed to reduce the required feedback overhead, where the high-dimensional CSI is first compressed at the user, and then reconstructed at the BS from the feedback signal [9]–[11]. However, these schemes usually assume the perfect CSI at the user, which is not realistic.

In this paper, by exploiting the spatio-temporal common sparsity of time-domain MIMO channels, we propose a structured compressive sensing (SCS) based spatio-temporal joint channel estimation and feedback scheme with low overhead for FDD massive MIMO systems. Specifically, at the BS, we first propose a non-orthogonal pilot scheme under the framework of compressive sensing (CS) theory, which is essentially different from the widely used orthogonal pilots under the framework of classical Nyquist sampling theorem. Compared with conventional orthogonal pilots, the proposed non-orthogonal pilot scheme can substantially reduce the required pilot overhead for channel estimation. At the user side, we propose an adaptive structured subspace pursuit (ASSP) algorithm for channel estimation, whereby the spatio-temporal common sparsity of time-domain MIMO channels is leveraged to improve the channel estimation performance from the limited number of pilots. Moreover, to reduce the feedback overhead for CSIT, we propose a parametric channel feedback scheme, which can achieve accurate CSIT by only feeding back the small number of dominated channel parameters to the BS. Finally, simulation results verify that the proposed scheme outperforms its conventional counterparts with reduced overhead, where the performance of the SCS based channel estimation scheme approaches that of the oracle least squares (LS) channel estimator, and the parametric channel feedback scheme only suffers from a negligible performance loss compared with the complete channel feedback scheme.

Notation: Boldface lower and upper-case symbols represent column vectors and matrices, respectively. The operator $\circ$ represents the Hadamard product, $\lfloor \cdot \rfloor$ denotes the integer floor operator, and $\mathrm{diag}\{\mathbf{x}\}$ is a diagonal matrix with elements of the vector $\mathbf{x}$ on its diagonal. The matrix inversion, transpose, and Hermitian transpose operations are denoted by $(\cdot)^{-1}$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$, respectively. $(\cdot)^{\dagger}$ denotes the Moore-Penrose matrix inversion, $|\cdot|_c$ denotes the cardinality of a set. The l_2-norm operation and Frobenius-norm operation are given by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. Finally, $\mathbf{\Phi}^{(l)}$ denotes the lth column vector of the matrix $\mathbf{\Phi}$.

II. Spatio-Temporal Common Sparsity of Time-Domain MIMO Channels

Extensive experimental studies have shown that wireless broadband channels appear the *sparsity* in the time domain [5]. This is caused by the fact that the number of multipaths that dominate the majority of channel energy is small due to the limited number of significant scatterers in the wireless signal propagation environments, while the channel delay spread can be large due to the significant difference between the time of arrival (ToA) of the earliest multipath and the ToA of the latest multipath [5]. Specifically, in the downlink, the time-domain channel impulse response (CIR) between the mth BS transmit antenna and one user can be expressed as

$$\mathbf{h}_{m,r} = \left[h_{m,r}[1], h_{m,r}[2], \cdots, h_{m,r}[L]\right]^{\mathrm{T}}, 1 \le m \le M, \quad (1)$$

where r is the index of the OFDM symbol in the time domain, L is the maximum channel delay spread, $D_{m,r} = \mathrm{supp}\{\mathbf{h}_{m,r}\} = \{l : |h_{m,r}[l]| > p_{\mathrm{th}}, 1 \le l \le L\}$ is the support set of $\mathbf{h}_{m,r}$, p_{th} is the noise floor according to [5], and M is the number of antennas at the BS. The sparsity level of wireless channels is denoted as $P_{m,r} = |D_{m,r}|_c$, and we have $P_{m,r} \ll L$ due to the sparse nature of time-domain channels [5].

Moreover, there are measurements showing that CIRs between different transmit antennas and one user appear very similar path delays [2], [14]. The reason is that, in typical massive MIMO geometry, the scale of the compact antenna array at the BS is relatively small compared with the long signal transmission distance, and channels associated with different transmit-receive antenna pairs share the common scatterers. Therefore, CIRs of different transmit-receive antenna pairs share a common sparse pattern [2], [6], [14], i.e.,

$$D_{1,r} = D_{2,r} = \cdots = D_{M,r}. \quad (2)$$

For example, we consider the LTE-Advanced system working at a carrier frequency of $f_c = 2$ GHz with a signal bandwidth of $f_s = 10$ MHz, and the uniform linear array (ULA) with the antenna spacing of half-wavelength. For two transmit antennas with the distance of 8 half-wavelengths, their maximum difference of path delays from the common scatterer is $\frac{f_s}{2f_c} \times 8 = 0.002\ \mu s$, which is negligible compared with the system sample period $T_s = 1/f_s = 0.1\ \mu s$.

Finally, practical wireless channels also appear the temporal correlation even in fast time-varying scenarios [5]. It has been demonstrated that the path delays usually vary much slower than the path gains [5]. This is due to the fact that the coherence time of path gains over time-varying channels is inversely proportional to the system carrier frequency, while the duration for path delay variation is inversely proportional to the system bandwidth [5]. For example, in the LTE-Advanced system with $f_c = 2$ GHz and $f_s = 10$ MHz, the path delays vary at a rate that is about several hundred times slower than that of the path gains [5]. That is to say, during the coherence time of path delays, CIRs associated with R successive OFDM symbols have the common sparsity due to the almost unchanged path delays, i.e.,

$$D_{m,r} = D_{m,r+1} = \cdots = D_{m,r+R-1}, 1 \le m \le M. \quad (3)$$

The spatial and temporal channel correlations shown in (2) and (3) are jointly referred to as the *spatio-temporal common sparsity* of time-domain MIMO channels. This channel property is usually not considered in existing channel estimation and feedback schemes. In this paper, we will exploit this channel property to overcome the challenging problem of channel estimation and feedback for FDD massive MIMO.

III. Proposed SCS Based Spatio-Temporal Joint Channel Estimation and Feedback Scheme

In this section, the SCS based spatio-temporal joint channel estimation and feedback scheme is proposed for FDD massive

MIMO. First, we propose the non-orthogonal pilot scheme at the BS to reduce the pilot overhead. Moreover, we propose the ASSP algorithm at the user for reliable channel estimation. Finally, we propose a low-overhead parametric channel feedback scheme to achieve the accurate CSIT.

A. Non-Orthogonal Pilot Scheme at the BS

The design of conventional orthogonal pilots is based on the framework of classical Nyquist sampling theorem, and this design has been widely used in standardized MIMO systems. For orthogonal pilots, pilots associated with different transmit antennas occupy the different subcarriers. For massive MIMO systems with hundreds of transmit antennas, such orthogonal pilots will suffer from the prohibitively high pilot overhead. In contrast, the design of the proposed non-orthogonal pilot scheme is based on CS theory, and it allows pilots of different transmit antennas to occupy the completely same subcarriers. By leveraging the sparse nature of channels, the pilots used for channel estimation can be reduced substantially. Particularly, we denote the index set of subcarriers allocated to pilots as ξ, which is uniquely selected from the set of $\{1, 2, \cdots, N\}$ and identical for all transmit antennas. Here $N_p = |\xi|_c$ is the number of pilot subcarriers in one OFDM symbol, and N is the size of the OFDM symbol. Moreover, we denote the pilot sequence of the mth transmit antenna as $\mathbf{p}_m \in \mathbb{C}^{N_p \times 1}$. In this paper, we propose that ξ is equi-spaced and elements of $\{\mathbf{p}_m\}_{m=1}^{M}$ are constant module with the phases following the mutually independent uniform distribution $\mathcal{U}[0, 2\pi)$.

B. SCS Based Channel Estimation at the User

At the user, after the removal of the guard interval and discrete Fourier transformation (DFT), the received pilot sequence $\mathbf{y}_r \in \mathbb{C}^{N_p \times 1}$ of the rth OFDM symbol can be expressed as

$$\begin{aligned} \mathbf{y}_r &= \sum_{m=1}^{M} \mathrm{diag}\{\mathbf{p}_m\} \mathbf{F}|_\xi \begin{bmatrix} \mathbf{h}_{m,r} \\ \mathbf{0}_{(N-L)\times 1} \end{bmatrix} + \mathbf{w}_r \\ &= \sum_{m=1}^{M} \mathbf{P}_m \mathbf{F}_L|_\xi \mathbf{h}_{m,r} + \mathbf{w}_r = \sum_{m=1}^{M} \mathbf{\Phi}_m \mathbf{h}_{m,r} + \mathbf{w}_r, \end{aligned} \tag{4}$$

where $\mathbf{P}_m = \mathrm{diag}\{\mathbf{p}_m\}$, $\mathbf{F} \in \mathbb{C}^{N \times N}$ is a DFT matrix, $\mathbf{F}_L \in \mathbb{C}^{N \times L}$ is a partial DFT matrix consisted of the first L columns of $\mathbf{F}$, $\mathbf{F}_L|_\xi \in \mathbb{C}^{N_p \times L}$ denotes the sub-matrix by selecting the rows of $\mathbf{F}_L$ according to ξ, $\mathbf{w}_r \in \mathbb{C}^{N_p \times 1}$ is the additive white Gaussian noise (AWGN) vector in the rth OFDM symbol, and $\mathbf{\Phi}_m = \mathbf{P}_m \mathbf{F}_L|_\xi$. Moreover, (4) can be rewritten in a more compact form as

$$\mathbf{y}_r = \mathbf{\Phi} \tilde{\mathbf{h}}_r + \mathbf{w}_r, \tag{5}$$

where $\mathbf{\Phi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_M] \in \mathbb{C}^{N_p \times ML}$, and $\tilde{\mathbf{h}}_r = [\mathbf{h}_{1,r}^{\mathrm{T}}, \mathbf{h}_{2,r}^{\mathrm{T}}, \cdots, \mathbf{h}_{M,r}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML \times 1}$ is an aggregate CIR vector.

For massive MIMO systems, we usually have $N_p \ll ML$ due to the large number of transmit antennas M and the limited number of pilots N_p. This indicates that we cannot reliably estimate $\tilde{\mathbf{h}}_r$ from $\mathbf{y}_r$ using conventional channel estimation schemes, since (5) is an under-determined system. However, the observation that $\tilde{\mathbf{h}}_r$ is a sparse signal due to the sparsity of $\{\mathbf{h}_{m,r}\}_{m=1}^{M}$ inspires us to estimate the sparse signal $\tilde{\mathbf{h}}_r$ of high dimension from the received pilot sequence $\mathbf{y}_r$ of low dimension under the framework of CS theory [15]. Moreover, the inherent spatial common sparsity of wireless MIMO channels can be also exploited for performance enhancement. Specifically, we rearrange the aggregate CIR vector $\tilde{\mathbf{h}}_r$ to obtain the equivalent CIR vector $\tilde{\mathbf{d}}_r$ as

$$\tilde{\mathbf{d}}_r = [\mathbf{d}_{1,r}^{\mathrm{T}}, \mathbf{d}_{2,r}^{\mathrm{T}}, \cdots, \mathbf{d}_{L,r}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML \times 1}, \tag{6}$$

where $\mathbf{d}_{l,r} = [h_{1,r}[l], h_{2,r}[l], \cdots, h_{M,r}[l]]^{\mathrm{T}}$ for $1 \le l \le L$. Similarly, $\mathbf{\Phi}$ can be rearranged as $\mathbf{\Psi}$, i.e.,

$$\mathbf{\Psi} = [\mathbf{\Psi}_1, \mathbf{\Psi}_2, \cdots, \mathbf{\Psi}_L] \in \mathbb{C}^{N_p \times ML}, \tag{7}$$

where $\mathbf{\Psi}_l = \left[\mathbf{\Phi}_1^{(l)}, \mathbf{\Phi}_2^{(l)}, \cdots, \mathbf{\Phi}_M^{(l)}\right] = [\psi_{1,l}, \psi_{2,l}, \cdots, \psi_{M,l}] \in \mathbb{C}^{N_p \times M}$. In this way, (5) can be reformulated as

$$\mathbf{y}_r = \mathbf{\Psi} \tilde{\mathbf{d}}_r + \mathbf{w}_r. \tag{8}$$

From (8), it can be observed that due to the spatial common sparsity of wireless MIMO channels, the equivalent CIR vector $\tilde{\mathbf{d}}_r$ appears the structured sparsity [15].

Furthermore, the temporal correlation of wireless channels indicates that such spatial common sparsity in MIMO systems remains virtually unchanged over R successive OFDM symbols, where R is determined by the coherence time of the path delays [5]. Hence, wireless MIMO channels appear the spatio-temporal common sparsity during R successive OFDM symbols. Considering (8) during R adjacent OFDM symbols with the same pilot pattern, we have

$$\mathbf{Y} = \mathbf{\Psi} \mathbf{D} + \mathbf{W}, \tag{9}$$

where $\mathbf{Y} = [\mathbf{y}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_{r+R-1}] \in \mathbb{C}^{N_p \times R}$ is the measurement matrix, $\mathbf{D} = \left[\tilde{\mathbf{d}}_r, \tilde{\mathbf{d}}_{r+1}, \cdots, \tilde{\mathbf{d}}_{r+R-1}\right] \in \mathbb{C}^{ML \times R}$ is the equivalent CIR matrix, and $\mathbf{W} = [\mathbf{w}_r, \mathbf{w}_{r+1}, \cdots, \mathbf{w}_{r+R-1}] \in \mathbb{C}^{N_p \times R}$ is the AWGN matrix. It should be pointed out that $\mathbf{D}$ can be expressed as

$$\mathbf{D} = [\mathbf{D}_1^{\mathrm{T}}, \mathbf{D}_2^{\mathrm{T}}, \cdots, \mathbf{D}_L^{\mathrm{T}}]^{\mathrm{T}}, \tag{10}$$

where $\mathbf{D}_l$ for $1 \le l \le L$ has the size of $M \times R$, and the mth row and rth column element of $\mathbf{D}_l$ is the channel gain of the lth path delay associated with the mth transmit antenna in the rth OFDM symbol.

It is clear that the equivalent CIR matrix $\mathbf{D}$ in (10) appears the structured sparsity due to the spatio-temporal common sparsity of wireless MIMO channels, and this intrinsic sparsity in $\mathbf{D}$ can be exploited for better estimation performance. In this way, we can jointly estimate channels associated with M transmit antennas in R OFDM symbols by jointly processing the received pilots of R OFDM symbols.

By exploiting the structured sparsity of $\mathbf{D}$ in (9), we propose the ASSP algorithm as described in Algorithm 1 to estimate channels for massive MIMO systems. Developed from the classical subspace pursuit (SP) algorithm [16], the proposed ASSP algorithm exploits the structured sparsity of $\mathbf{D}$ for further improved sparse signal recovery performance.

For Algorithm 1, some notations should be further detailed. First, both $\mathbf{Z} \in \mathbb{C}^{ML \times R}$ and $\tilde{\mathbf{D}} \in \mathbb{C}^{ML \times R}$ are consisted of L sub-matrices with the equal size of $M \times R$, i.e., $\mathbf{Z} = [\mathbf{Z}_1^{\mathrm{T}}, \mathbf{Z}_2^{\mathrm{T}}, \cdots, \mathbf{Z}_L^{\mathrm{T}}]^{\mathrm{T}}$ and $\tilde{\mathbf{D}} = [\tilde{\mathbf{D}}_1^{\mathrm{T}}, \tilde{\mathbf{D}}_2^{\mathrm{T}}, \cdots, \tilde{\mathbf{D}}_L^{\mathrm{T}}]^{\mathrm{T}}$.

Algorithm 1 Proposed ASSP Algorithm.

Input: Noisy measurement matrix $\mathbf{Y}$ and sensing matrix $\mathbf{\Psi}$.
Output: The estimation of channels $\{\mathbf{h}_{m,t}\}_{m=1,t=r}^{m=M,t=r+R-1}$.

- **Step 1** (*Initialization*) The initial channel sparsity level $s=1$, the iterative index $k=1$, the support set $\Omega^{k-1}=\emptyset$, and the residual matrices $\mathbf{R}^{k-1}=\mathbf{Y}$ and $\|\mathbf{R}_{s-1}\|_F=+\inf$.
- **Step 2** (*Solve the Structured Sparse Matrix* $\mathbf{D}$ *to (9)*)
  repeat
  1. (*Correlation*) $\mathbf{Z}=\mathbf{\Psi}^{\mathrm{H}}\mathbf{R}^{k-1}$;
  2. (*Support Estimate*) $\tilde{\Omega}'^{k}=\Omega^{k-1}\cup\Pi^{s}\left(\left\{\|\mathbf{Z}_l\|_F\right\}_{l=1}^{L}\right)$;
  3. (*Support Pruning*) $\tilde{\mathbf{D}}_{\tilde{\Omega}'^{k}}=\mathbf{\Psi}_{\tilde{\Omega}'^{k}}^{\dagger}\mathbf{Y}$;
     $\tilde{\Omega}^{k}=\Pi^{s}\left(\left\{\left\|\tilde{\mathbf{D}}_l\right\|_F\right\}_{l=1}^{L}\right)$;
  4. (*Matrix Estimate*) $\tilde{\mathbf{D}}_{\tilde{\Omega}^{k}}=\mathbf{\Psi}_{\tilde{\Omega}^{k}}^{\dagger}\mathbf{Y}$;
  5. (*Residue Update*) $\mathbf{R}^{k}=\mathbf{Y}-\mathbf{\Psi}\tilde{\mathbf{D}}$;
  6. (*Matrix Update*) $\tilde{\mathbf{D}}^{k}=\tilde{\mathbf{D}}$;
  if $\left\|\mathbf{R}^{k-1}\right\|_F>\left\|\mathbf{R}^{k}\right\|_F$
  7. (*Iteration with Fixed Sparsity Level*) $\Omega^{k}=\tilde{\Omega}^{k}$; $k=k+1$;
  else
  8. (*Update Sparsity Level*) $\tilde{\mathbf{D}}_s=\tilde{\mathbf{D}}^{k-1}$; $\mathbf{R}_s=\mathbf{R}^{k-1}$;
     $\Omega_s=\Omega^{k-1}$; $s=s+1$;
  end if
  until stopping criteria are met
- **Step 3** (*Obtain Channels*) $\widehat{\mathbf{D}}=\tilde{\mathbf{D}}_{s-1}$ and obtain the estimation of channels $\{\mathbf{h}_{m,t}\}_{m=1,t=r}^{m=M,t=r+R-1}$ according to (4)-(9).

Second, $\tilde{\mathbf{D}}_{\tilde{\Omega}}=\left[\tilde{\mathbf{D}}_{\tilde{\Omega}(1)}^{\mathrm{T}},\tilde{\mathbf{D}}_{\tilde{\Omega}(2)}^{\mathrm{T}},\cdots,\tilde{\mathbf{D}}_{\tilde{\Omega}(|\tilde{\Omega}|_c)}^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\mathbf{\Psi}_{\tilde{\Omega}}=\left[\mathbf{\Psi}_{\tilde{\Omega}(1)}^{\mathrm{T}},\mathbf{\Psi}_{\tilde{\Omega}(2)}^{\mathrm{T}},\cdots,\mathbf{\Psi}_{\tilde{\Omega}(|\tilde{\Omega}|_c)}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\tilde{\Omega}(1)<\tilde{\Omega}(2)<\cdots<\tilde{\Omega}(|\tilde{\Omega}|_c)$ are elements in the set $\tilde{\Omega}$. Third, $\Pi^{s}(\cdot)$ is a set, whose elements are the indices of the largest s elements of its argument. Finally, to reliably acquire the channel sparsity level, we stop the iteration if $\left\|\mathbf{R}^{k}\right\|_F>\|\mathbf{R}_{s-1}\|_F$ or $\left\|\tilde{\mathbf{D}}_{\tilde{l}}\right\|_F\le\sqrt{MR}p_{\mathrm{th}}$, where $\left\|\tilde{\mathbf{D}}_{\tilde{l}}\right\|_F$ is the smallest $\left\|\tilde{\mathbf{D}}_{l}\right\|_F$ for $l\in\tilde{\Omega}^{k}$, and p_{th} is the noise floor according to [5].

Here we further explain the main steps in Algorithm 1 as follows. First, for **step 2.1**$\sim$**2.7**, the ASSP algorithm aims to acquire the solution $\mathbf{D}$ to (9) with the fixed sparsity level s in a greedy way, which is similar to the classical SP algorithm. Second, $\left\|\mathbf{R}^{k-1}\right\|_F\le\left\|\mathbf{R}^{k}\right\|_F$ indicates that the s-sparse solution $\mathbf{D}$ to (9) has been obtained, and then the sparsity level is updated to find the $(s+1)$-sparse solution $\mathbf{D}$. Finally, if the stopping criteria are met, the iteration quits, and we consider the estimated solution to (9) with the last sparsity level as the estimated channels, i.e., $\widehat{\mathbf{D}}=\tilde{\mathbf{D}}_{s-1}$.

Compared to the SP algorithm, the proposed ASSP algorithm has the following distinctive features:

- The classical SP algorithm reconstructs one high-dimensional sparse vector from one low-dimensional measurement vector. In contrast, the proposed ASSP algorithm recovers the high-dimensional sparse matrix with the inherent structured sparsity from the low-dimensional measurement matrix, whereby the inherent structured sparsity of the sparse matrix is exploited for the improved matrix reconstruction performance.
- The classical SP algorithm requires the sparsity level as the priori information for reliable sparse signal reconstruction. In contrast, the proposed ASSP algorithm can adaptively acquire the sparsity level of the structured sparse matrix. By exploiting the practical physical property of wireless channels, the proposed stopping criteria enable ASSP algorithm to estimate channels with good mean square error (MSE) performance. Moreover, simulation results in Section IV also verify its accurate acquisition of channel sparsity level.

C. Parametric Channel Feedback Scheme

The proposed parametric channel feedback scheme enables the BS to acquire the accurate CSI by only feeding back the small number of dominated channel parameters thanks to the structured sparse nature of wireless MIMO channels. Particularly, after Section III-A and III-B, the user has estimated the common path delays of CIRs and corresponding path gains, where the small number of parameters can dominate the fine grain spatial structure of large-dimensional massive MIMO channel matrix. Therefore, we can directly feed back the limited number of path delays and path gains instead of the complete CSI to reduce the required feedback overhead.

Compared to the conventional CS based channel feedback schemes, the proposed parametric channel feedback scheme enjoys more reliable CSIT when the feedback overhead is limited. It has been shown that for the high-dimensional sparse signal, the compression and reconstruction performance by directly preserving its support set and corresponding non-zero values is superior to that by first compressing the sparse signal via projection matrix and then reconstructing the sparse signal via CS algorithms [13]. Moreover, in the proposed scheme, we do not consider the spatial correlation of path gains over different transmit antennas, since the spatial correlation may be impaired by practical electronic components, and it is expected to reduce the spatial correlation of path gains for capacity improvement in practical massive MIMO systems [2]. Therefore, compared to the conventional CS based channel feedback schemes exploiting such spatial channel correlation of ideal antenna array, the proposed scheme is more realistic.

IV. Simulation Results

In this section, a simulation study was carried out to investigate the performance of the proposed channel estimation and feedback scheme for FDD massive MIMO systems. To provide a benchmark for performance comparison, we consider the oracle LS algorithm by assuming the true channel support set known at the user and the oracle ASSP algorithm[1] by assuming the true channel sparsity level known at the user. Moreover, to investigate the performance gain from the exploitation of the spatial common sparsity of CIRs, we provide the MSE performance of adaptive subspace pursuit (ASP) algorithm, which is a special case of the proposed ASSP algorithm without leveraging such spatial common sparsity of CIRs. Simulation system parameters were set as: system carrier was $f_c=2$ GHz, system bandwidth was $f_s=10$ MHz, DFT size was $N=4096$, and the length of the guard interval was

[1]The oracle ASSP algorithm is a special case of the proposed ASSP algorithm, where the initial channel sparsity level s is set to the true channel sparsity level, **Step 2.8** is not performed, and the stopping criterion is $\left\|\mathbf{R}^{k-1}\right\|_F\le\left\|\mathbf{R}^{k}\right\|_F$.

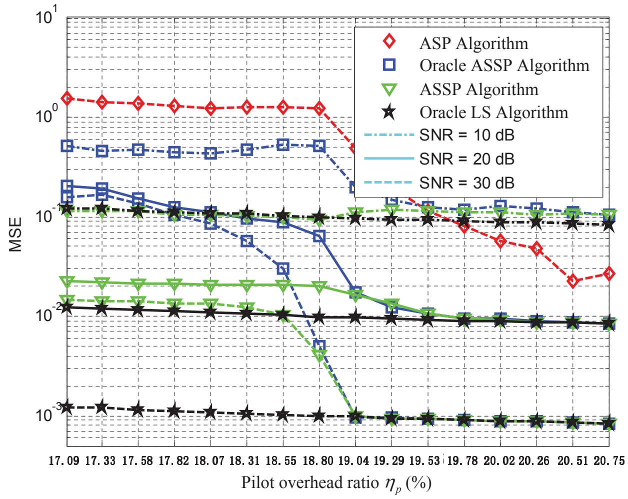

Fig. 1. MSE performance comparison of different channel estimation algorithms against pilot overhead ratio and SNR.

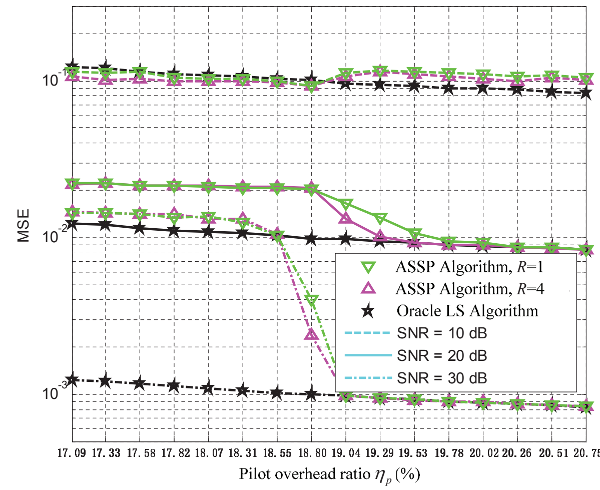

Fig. 2. MSE performance comparison of the ASSP algorithm with different R's over time-varying ITU-VA channel with the mobile speed of 60 km/h.

$N_g = 64$, which could combat the maximum delay spread of 6.4 μs [4]. We consider $M = 64$, $M_G = 32$, the number of pilots to estimate channels for one antenna group is N_p, and the pilot overhead ratio is $\eta_p = (N_p M)/(N M_G)$. The International Telecommunications Union Vehicular-A (ITU-VA) channel model with $P = 6$ paths was adopted [4]. Finally, p_{th} was set as 0.1, 0.08, 0.06, 0.05, and 0.04 for $\text{SNR} = 10$ dB, 15 dB, 20 dB, 25 dB, and 30 dB, respectively.

Fig. 1 compares the MSE performance of the ASSP algorithm, the oracle ASSP algorithm, the ASP algorithm, and the oracle LS algorithm over static ITU-VA channel. In the simulation, we only consider the channel estimation for one OFDM symbol with $R = 1$. From Fig. 1, it can be observed that the ASP algorithm performs poorly. The proposed ASSP algorithm outperforms the ASP algorithm, since the spatial common sparsity of MIMO channels is leveraged for the enhanced channel estimation performance. Moreover, for $\eta_p \geq 19.04\%$, the ASSP algorithm and the oracle ASSP algorithm have the similar MSE performance, and their performance approaches that of the oracle LS algorithm. This indicates that the proposed ASSP algorithm can reliably acquire the channel sparsity level and the support set for $\eta_p \geq 19.04\%$. Moreover, the low pilot overhead implies that the average pilot overhead to estimate the channel associated with one transmit antenna is $N_{p\_\text{avg}} = N_p/M_G = 12.18$, which approaches $2P = 12$, the minimum number of observations to reliably recover a P-sparse signal [17]. Therefore, the good sparse signal recovery performance of the proposed non-orthogonal pilot scheme and the near-optimal channel estimation performance of the proposed ASSP algorithm are confirmed.

Fig. 2 provides the MSE performance comparison of the proposed ASSP algorithm with ($R = 4$) and without ($R = 1$) exploiting the temporal common support of wireless channels, where the time-varying ITU-VA channel with the user's mobile speed of 60 km/h is considered. In the simulation, $R = 1$ or 4 denotes the joint processing of the received pilot signals in R successive OFDM symbols. It is clear that the channel estimation performance by exploiting the temporal channel correlation is better than that without considering this channel property, since more measurements can be used for the improved channel estimation performance.

Fig. 3 provides the MSE performance comparison of several channel estimation schemes for massive MIMO, where we consider the channel estimation for one OFDM symbol with $R = 1$. The Cramer-Rao lower bound (CRLB) of conventional linear channel estimation schemes (e.g., minimum mean square error (MMSE) algorithm and LS algorithm) is also plotted as the performance benchmark, where $\text{CRLB} = 1/\text{SNR}$ [5]. The ASP algorithm does not perform well due to the insufficient pilots. The time-frequency joint training based scheme [5] works poorly since the mutual interferences of time-domain training sequences of different transmit antennas degrade the channel estimation performance when M is large. Both the MMSE algorithm [3] and the proposed ASSP algorithm have the 9 dB gain than that proposed in [5], and both of them approach the CRLB of conventional linear algorithms. It is worth mentioning that the proposed scheme enjoys the significantly reduced pilot overhead compared with the MMSE algorithm, since the MMSE algorithm work well only when (8) is well-determined or over-determined. Finally, since the proposed ASSP algorithm can adaptively acquire the channel sparsity level and discards the multipath components buried by the noise at low SNR for improved channel estimation, we can find the proposed scheme even works better than the oracle ASSP algorithm at low SNR.

Fig. 4 compares the average capacity per user in the downlink massive MIMO by using different channel estimation and feedback schemes, where $R = 1$ is considered. In the simulation, we set the number of users to 8, and the zero forcing (ZF) precoding is adopted at the BS. The acquired CSI at the user is assumed to be fed back to the BS without noise. The specific channel estimation and feedback schemes for comparison are listed as follows. The complete CSI feedback scheme feeds back the complete CSI acquired by the MMSE algorithm [3] to the BS, where the CSI compression ratio is η_f=100%. The CS based channel feedback scheme [9] feeds back the compressed CSI to the BS with $\eta_f = 17.09\%$, where the ideal CSI is assumed at the user side [18]. The proposed parametric channel feedback scheme feeds back the oracle path delays and path gains estimated by the oracle LS algorithm with $\eta_f = 9.52\%$ (The common path delays and different path gains are fed back with the compression ratio

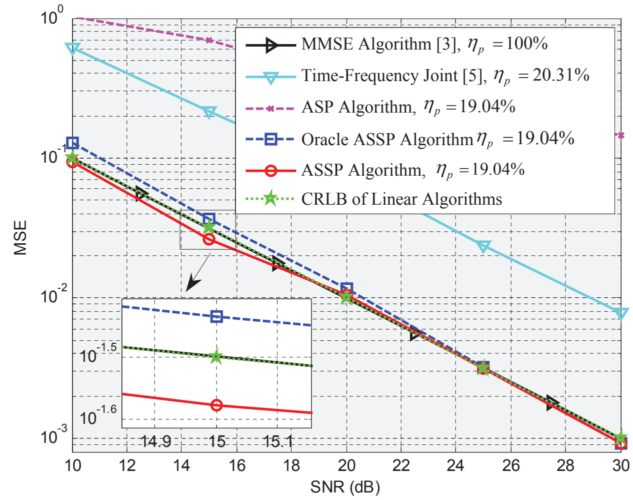

Fig. 3. MSE performance comparison of different channel estimation schemes for FDD massive MIMO systems.

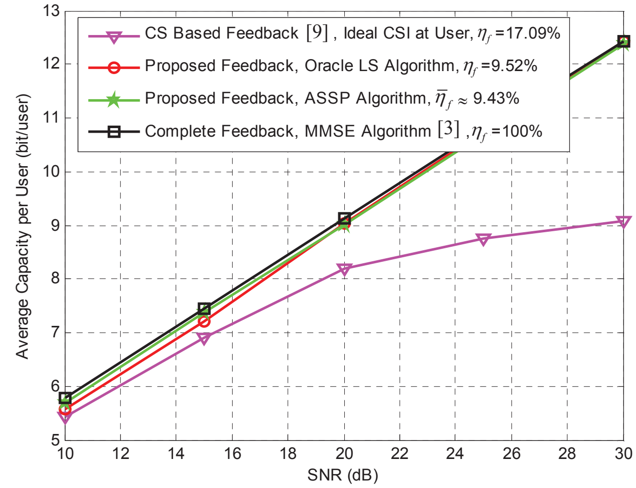

Fig. 4. Comparison of average capacity per user in the downlink massive MIMO by using different channel estimation and feedback schemes, where ZF precoding is adopted.

$\eta_f = (PM+P)/(LM) = 9.52\%$). Finally, the proposed parametric channel feedback scheme feeds back both the path delays and path gains estimated by the proposed ASSP algorithm with the average compression ratio $\bar{\eta}_f \approx 9.43\%$[2]. Clearly, the CS based channel feedback scheme works poorly, and it performs worse than the proposed parametric channel feedback scheme. This is consistent with the conclusion in [13], i.e., the compression/reconstruction performance by preserving the values and positions of non-zero elements of the sparse signal is superior to that by compressing/recovering the sparse signal with CS algorithms. Moreover, both the proposed parametric channel feedback scheme with the oracle LS algorithm and the proposed parametric channel feedback scheme with the ASSP algorithm have the similar performance. Moreover, their performance approaches the performance bound obtained by the complete channel feedback scheme, which confirms the near-optimal performance of the proposed scheme.

V. Conclusions

In this paper, we have proposed an SCS based spatio-temporal joint channel estimation and feedback scheme for FDD massive MIMO systems, whereby the intrinsic spatio-temporal common sparsity of wireless MIMO channels is exploited to reduce the pilot and channel feedback overhead. First, the non-orthogonal pilot scheme at the BS and the ASSP algorithm at the user can reliably estimate channels with significantly reduced pilot overhead. Moreover, by leveraging the sparse nature of channels, the parametric channel feedback scheme can achieve the accurate CSIT with reduced feedback overhead. Simulation results have shown that the proposed channel estimation and feedback scheme can achieve much better performance than its counterparts, and it only suffers from a negligible performance loss compared with the performance bound.

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]The fed back path delays and path gains are estimated by the proposed ASSP algorithm, and the estimated channel sparsity level in different Monte Carlo trial can be different.

[2] 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.

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

[4] 3GPP Technical Specification 36.211. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation. www.3gpp.org.

[5] 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.

[6] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Super-resolution sparse MIMO-OFDM channel estimation based on spatial and temporal correlations," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1266-1269, Jul. 2014.

[7] S. L. H. Nguyen, and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE WCNC'13*, Shanghai, China, Apr. 2013.

[8] 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.

[9] P. Cheng and Z. Chen, "Multidimensional compressive sensing based analog CSI feedback for massive MIMO-OFDM systems," in *Proc. IEEE VTC'14-Fall*, Vancouver, Canada, Sep. 2014.

[10] J. Lee and S. H. Lee, "A compressed analog feedback strategy for spatially correlated massive MIMO systems," in *Proc. IEEE VTC'12 Fall*, Quebec City, Canada, Sep. 2012.

[11] M. S. Sim and C. B. Chae, "Compressed channel feedback for correlated massive MIMO systems," in *Proc. IEEE Globecom'14*, Austin, USA, Dec. 2014.

[12] M. Dai and B. Clerckx, "Transmit beamforming for MISO broadcast channels with statistical and delayed CSIT," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1202-1215, Apr. 2015.

[13] V. K Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48-56, Mar. 2008.

[14] T. Santos, J. Kredal, P. Almers, F. Tufvesson, and A. Molisch, "Modeling the ultra-wideband outdoor channel: Measurements and parameter extraction method," *IEEE Trans. Wireless. Commun.*, vol. 9, no. 1, pp. 282-290, Jan. 2010.

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

[16] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.

[17] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197-2202, Mar. 2003.

[18] Y. Huang and B. Clerckx, "Joint wireless information and power transfer for an autonomous multiple antenna relay system," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1113-1116, Jul. 2015.