Massive MIMO Channel Estimation Based on Block Iterative Support Detection

Wenqian Shen[1], Linglong Dai[1], Yi Shi[2], Zhen Gao[1], and Zhaocheng Wang[1]
[1]Tsinghua National Laboratory for Information Science and Technology (TNlist)
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Huawei Technologies, Beijing 100095, China
Email: swq13@mails.tsinghua.edu.cn

***Abstract*—Massive MIMO has become a promising key technology for future 5G wireless communications to increase the channel capacity and link reliability. However, with greatly increased number of transmit antennas at the base station (BS) in massive MIMO systems, the pilot overhead for accurate acquisition of channel state information (CSI) will be prohibitively high. To address this issue, we propose a block iterative support detection (block-ISD) based algorithm for channel estimation to reduce the pilot overhead. The proposed block-ISD algorithm fully exploits the block sparsity inherent in the block-sparse equivalent channel impulse response (CIR) generated by considering the spatial correlations of MIMO channels. Furthermore, unlike conventional greedy compressive sensing (CS) algorithms that rely on prior knowledge of the channel sparsity level, block-ISD relaxes this demanding requirement and is thus more practically appealing. Simulation results demonstrate that block-ISD yields better normalized mean square error (NMSE) performance than classical CS algorithms, and achieve a reduction of 87.5% pilot overhead than conventional channel estimation techniques.**

I. Introduction

Multiple-input multiple-output (MIMO) technology has been adopted by numerous wireless communication standards like long term evolution-advanced (LTE-A) [1] due to its attractive potential gains both in system capacity and link reliability. Recently, massive MIMO using substantially more antennas than traditional MIMO at the base station (BS) has emerged as a key promising technology for future 5G wireless communications. It is proved that massive MIMO can reduce transmit power as well as increase spectrum efficiency by orders of magnitude [2]. In massive MIMO systems, accurate downlink channel state information (CSI) is essential for channel adaptive techniques such as water-filling, beamforming, etc. [3]-[5]. The pilot overhead for downlink channel estimation in LTE-A standard with 8 antennas has already exceeded 25% [6]. As the number of BS antennas keeps increasing in massive MIMO systems, efficient low-overhead channel estimation will be an increasingly important and challenging problem.

Basically, the number of orthogonal pilots required for downlink channel estimation is proportional to the number of BS antennas, while the number of orthogonal pilots required for uplink channel estimation is proportional to the number of scheduled users [2]. Therefore, channel estimation in the downlink is more challenging than that in the uplink, since the number of BS antennas is usually much larger than the number of scheduled users in massive MIMO systems. By exploiting channel reciprocity in time division duplexing (TDD) systems, the estimated uplink CSI can be directly applied in the downlink, thereby alleviating the need of downlink channel estimation. Nonetheless, frequency division duplexing (FDD) generally outperforms TDD in terms of transmission delay, communication range, mobility support, etc. [3]-[5]. That's why FDD still dominates the current wireless communication systems, where the channel reciprocity does not exist. Thus, it is of great importance to study the challenging problem of downlink channel estimation, especially in FDD massive MIMO systems.

Conventional downlink channel estimation techniques include least square (LS) and minimum mean square error (MMSE) [7]. However, they are not suitable for massive MIMO systems due to the number of required orthogonal pilots scales linearly with the number of antennas at the BS, which results in prohibitively high pilot overhead. Recently, several efficient channel estimation schemes based on compressive sensing (CS) have been proposed to reduce pilot overhead by taking channel sparsity into account [8]-[12]. CS is a new signal processing theory that can recover high-dimensional sparse signals from low-dimensional measurements with an overwhelming probability at a sampling rate much lower than the classical Nyquist sampling rate [13]. The utilization of CS enables accurate CSI acquisition with acceptable pilot overhead. Conventional CS-based channel estimation schemes usually assume prior knowledge of the channel sparsity level, i.e., the number of non-zero elements of channel impulse response (CIR). However, in practice, channel sparsity level is usually unknown and difficult to be accurately estimated. In addition, due to the physical propagation characteristics of multiple antennas and close antenna spacing at the BS, CIRs associated with different antennas inevitably share some correlations [14], [15], which have not been considered by the existing CS-based channel estimation schemes to further reduce the pilot overhead[1].

[1]We have proposed a block-ISD algorithm to solve this problem in our previously published paper [16], which only briefly discussed the algorithm within two pages. In this paper, we will discuss this problem in detail with more analysis of the pilot design problem and the computational complexity of block-ISD. We also provide more numerical results to verify the performance of the proposed block-ISD algorithm.

In this paper, building upon the iterative support detection (ISD) algorithm, we propose an improved block-ISD based algorithm for downlink channel estimation in FDD massive MIMO systems to reduce the pilot overhead. Specifically, by considering the spatial correlations of MIMO channels caused by the physical propagation characteristics of multiple antennas and close antenna spacing at the BS, we generate the block-sparse equivalent CIR with the preferred block sparsity. Accordingly, we propose a block-ISD based algorithm to significantly reduce pilot overhead required for accurate channel estimation. Finally, we propose a pilot pattern and theoretically show its capacity of achieving small maximal mutual coherence (MMC), which ensures reliable channel estimation performance. Simulation results show that block-ISD yields better channel estimation performance and significantly reduces pilot overhead. Additionally, block-ISD requires no prior knowledge of the channel sparsity level, and is thus more practically appealing than conventional CS-based algorithms.

The remainder of this paper is organized as follows. The system model of massive MIMO is introduced in section II. Section III presents the downlink channel estimation based on the proposed block-ISD algorithm. In Section IV, simulation results are provided to illustrate the performance of block-ISD. Finally, conclusions are drawn in Section V.

Notation: Lower-case and upper-case boldface letters denote vectors and matrices, respectively; $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote the transpose, conjugate and conjugate transpose of a matrix, respectively; diag$\{\mathbf{c}\}$ denotes the diagonal matrix with the diagonal vector $\mathbf{c}$. $\mathbf{y}_\Omega$ denotes the sub-vector consisting of the elements with indexes from Ω. $\|\cdot\|_0$ is the l_0-norm denoting the number of non-zero elements in a vector, and $\|\cdot\|_p$ is the l_p-norm. Ω^c denotes the complementary set of Ω. Card(Γ) denotes the number of elements in set Γ.

II. System Model

We consider a massive MIMO system with N_t antennas at the BS and K scheduled single-antenna users ($N_t >> K$) with the commonly used orthogonal frequency division multiplexing (OFDM) modulation at the BS. Normally, frequency-domain pilots are used for channel estimation in OFDM-based systems [8]. At the user side, the received signal (including data and pilots) in the frequency domain can be expressed as

$$\mathbf{y} = \sum_{i=1}^{N_t} \mathbf{X}_i \mathbf{F}_L \mathbf{h}_i + \mathbf{n}, \tag{1}$$

where $\mathbf{X}_i = \text{diag}\{\mathbf{x}_i\}$ with $\mathbf{x}_i \in \mathcal{C}^{N\times 1}$ denotes the transmitted signal (including data and pilots) from the ith transmit antenna, where N is the OFDM symbol length. $\mathbf{F}_L \in \mathcal{C}^{N\times L}$ is a sub-matrix consisting of the first L columns of the normalized discrete fourier transform (DFT) matrix of size $N \times N$. $\mathbf{h}_i = [\mathbf{h}_i(1), \mathbf{h}_i(2), \cdots, \mathbf{h}_i(L)]^T$ denotes the CIR between the ith transmit antenna of the BS and the single receive antenna of the user with the maximal channel length L. $\mathbf{n} = [n_1, \cdots, n_N]^T$ represents the noise vector consisting of independent and identically distributed (i.i.d.) additive white complex Gaussian noise (AWGN) variables with zero mean and unit variance.

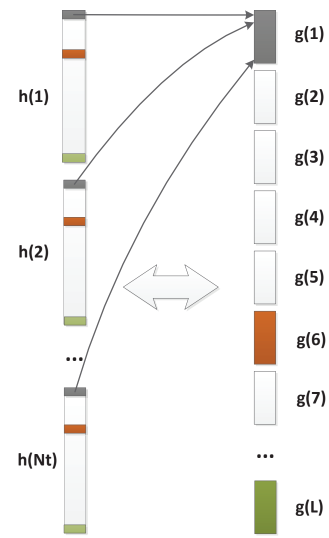

Fig. 1. Generation of the block-sparse equivalent CIR.

In order to estimate CIR $\{\mathbf{h}_i\}_{i=1}^{N_t}$, the received pilots $\mathbf{y}_\Omega$ can be extracted from the received signal $\mathbf{y}$ as

$$\mathbf{y}_\Omega = \sum_{i=1}^{N_t} \mathbf{C}_i (\mathbf{F}_L)_\Omega \mathbf{h}_i + \mathbf{n}_\Omega, \tag{2}$$

where Ω is the index set of subcarriers assigned to pilots, which can be randomly selected from the subcarrier set $[1, 2, \cdots, N]$. $(\mathbf{F}_L)_\Omega$ is the sub-matrix consisting of the rows with indexes from Ω, $\mathbf{C}_i = \text{diag}\{\mathbf{c}_i\}$ with $\mathbf{c}_i \in \mathcal{C}^{p\times 1}$ being the pilot vector for the ith transmit antenna, and the number of pilots is p. For simplicity, (2) can also be written as

$$\mathbf{y}_\Omega = \mathbf{P}\mathbf{h} + \mathbf{n}_\Omega, \tag{3}$$

where $\mathbf{P} = [\mathbf{C}_1(\mathbf{F}_L)_\Omega, \mathbf{C}_2(\mathbf{F}_L)_\Omega, \cdots, \mathbf{C}_{N_t}(\mathbf{F}_L)_\Omega]$ can be regarded as the sensing matrix, $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_{N_t}^T]^T$ is the aggregate CIR from N_t antennas to be estimated in massive MIMO systems.

III. Downlink Channel Estimation Based on Block-ISD

In this section, we firstly generate the block-sparse equivalent CIR by considering the correlations of CIRs associated with different antennas. Then, we propose an improved block-ISD algorithm for channel estimation to reduce the pilot overhead. Finally, we further discuss the pilot design problem which has a substantial impact on the channel estimation performance.

A. Generation of the block-sparse equivalent CIR

The massive MIMO channel to be recovered is $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_{N_t}^T]^T$. Due to the physical propagation characteristics of multiple antennas and close antenna spacing at the BS, CIRs associated with different antennas have similar path arrival times, and thus they share a common support [14], i.e.,

$$\mathbf{\Gamma}_{\mathbf{h}_1} = \mathbf{\Gamma}_{\mathbf{h}_2} = \cdots = \mathbf{\Gamma}_{\mathbf{h}_{N_t}}, \tag{4}$$

where $\mathbf{\Gamma}_{\mathbf{h}_i} = \{k : \mathbf{h}_i(k) \neq 0\}$ denotes the support of $\mathbf{h}_i$. This special structure of $\mathbf{h}$ is illustrated in Fig. 1, where the colored blocks denote the non-zero elements of CIRs.

Since the CIRs associated with different transmit antennas share a common support, we can group the elements of $\mathbf{h}_i$ with the same indexes into non-zero blocks and zero blocks as shown in Fig. 1 to generate the block-sparse equivalent CIR $\mathbf{g} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_L]^T$. More specifically, the relationship between $\mathbf{h}$ and $\mathbf{g}$ can be expressed as

$$\mathbf{g}((l-1)N_t + n_t) = \mathbf{h}((n_t - 1)L + l), \tag{5}$$

where $l = 1, 2, \cdots, L$ and $n_t = 1, 2, \cdots, N_t$. It is important that if we equally divide $\mathbf{g}$ into L blocks with N_t elements in each block, these N_t continuous elements in the lth block $\mathbf{g}_l$ are all zeros or non-zeros. Thus, the generated block-sparse equivalent CIR $\mathbf{g}$ enjoys the preferred block sparsity. This implies that we can treat the N_t continuous elements of the support $\mathbf{\Gamma}_{\mathbf{g}}$ of $\mathbf{g}$ as a whole and update them simultaneously.

Accordingly, similar to (5), we can obtain a new sensing matrix $\mathbf{\Theta}$ by rearranging the columns of $\mathbf{P}$ in (3) as

$$\mathbf{\Theta}(:, (l-1)N_t + n_t) = \mathbf{P}(:, (n_t - 1)L + l). \tag{6}$$

Therefore, the channel estimation problem (3) can be reformulated as

$$\mathbf{y}_\Omega = \mathbf{\Theta}\mathbf{g} + \mathbf{n}_\Omega. \tag{7}$$

This is an underdetermined problem with $\mathbf{g}$ of size $N_T L \times 1$ and $\mathbf{y}_{\mathbf{\Omega}}$ of size $p \times 1$, where p is usually much smaller than $N_T L$ due to the large number of antennas and the limited pilot overhead. Traditional channel estimation techniques like LS and MMSE can not recover CSI with limited pilot overhead. In what follows we propose a block-ISD algorithm to solve this problem by fully exploiting the block sparsity inherent in the generated block-sparse equivalent CIR.

B. Downlink channel estimation based on block-ISD

The pseudocode of block-ISD is in **Algorithm 1**. Note that block-ISD updates the recovered signal $\mathbf{g}^{(s)}$ in the sth iteration through solving the truncated basic pursuit (BP) problem [17] in step 4:

$$\min_{\mathbf{g}^{(s)}} \|\mathbf{g}^{(s)}_{W^{(s)}}\|_1 \quad s.t.\ \mathbf{y}_\Omega = \mathbf{\Theta}\mathbf{g}^{(s)} + \mathbf{n}_\Omega, \tag{8}$$

where $\|\mathbf{g}^{(s)}_{W^{(s)}}\|_1 = \sum_{w \in W^{(s)}} |\mathbf{g}^{(s)}(w)|$. This problem can be efficiently solved by calling a BP algorithm such as YALL1 [18]. Then, the support $\mathbf{\Gamma}^{(s)}_{\mathbf{g}}$ is also updated in the sth iteration through the adjacent support detection in steps 5-9. In these five steps, we firstly sort $\mathbf{g}^{(s)}$ in an ascending order in step 5 to obtain $\mathbf{v}^{(s)}$. Then we detect the support of $\mathbf{v}^{(s)}$ based on the 'first significant jump' rule [10] in step 6, which searches the smallest i that satisfies:

$$|\mathbf{v}^{(s)}(i+1)| - |\mathbf{v}^{(s)}(i)| > |\tau^s|, \tag{9}$$

Algorithm 1 Block-ISD Algorithm

Input:

1) Measurements $\mathbf{y}_\Omega$,
2) Sensing matrix $\mathbf{\Theta}$.

1: Initialization:
$s = 0$ and $\mathbf{g} : \mathbf{\Gamma}^{(0)}_{\mathbf{g}} = \emptyset$.
2: **while** Card$(\mathbf{\Gamma}^{(s)}_{\mathbf{g}}) < N_T L - p$ **do**
3: $\quad W^{(s)} = (\mathbf{\Gamma}^{(\mathbf{s})}_{\mathbf{g}})^c$;
4: $\quad \mathbf{g}^{(s)} \leftarrow \min_{\mathbf{g}^{(s)}} \|\mathbf{g}^{(s)}_{W^{(s)}}\|_1 \quad s.t.\ \mathbf{y}_\Omega = \mathbf{\Theta}\mathbf{g}^{(s)} + \mathbf{n}_\Omega$;
5: $\quad \mathbf{v}^{(s)} = \text{Sort}(\mathbf{g}^{(s)})$;
6: $\quad i \leftarrow \min i \quad s.t.\ |\mathbf{v}^{(s)}(i+1)| - |\mathbf{v}^{(s)}(i)| > |\tau^{(s)}|$;
7: $\quad \epsilon^{(s)} = |\mathbf{v}^{(s)}(i)|$;
8: $\quad \mathbf{\Gamma}^{(s)}_{\mathbf{v}} = \{k \quad s.t.\ |\mathbf{v}^{(s)}(k)| > \epsilon^{(s)}\}$;
9: $\quad \mathbf{\Gamma}^{(s)}_{\mathbf{g}} = \{(l-1)N_t + 1 : 1 : lN_T \quad s.t.\ \text{Card}(\{(l-1)N_t + 1 : 1 : lN_T\} \cap \mathbf{\Gamma}^{(\mathbf{s})}_{\mathbf{v}}) > N_t/2\}$;
10: $\quad s = s + 1$.
11: **end while**
12: **return** $\hat{\mathbf{g}} = \mathbf{g}^{(s)}$

Output:

Recovered block-sparse equivalent CIR $\hat{\mathbf{g}}$.

where $|\mathbf{v}^{(s)}(i)|$ denotes the absolute value of ith element of $\mathbf{v}^{(s)}$, and $\tau^{(s)} = (LN_T)^{-1}\|\mathbf{v}^{(s)}\|_\infty$ [18]. The smallest i is the index where the 'first significant jump' occurs in an ascending ordered vector $\mathbf{v}^{(s)}$. Next we set the threshold $\epsilon^{(s)} = |\mathbf{v}^{(s)}(i)|$ in step 7, then the support of $\mathbf{v}^{(s)}$ can be updated based on the threshold $\epsilon^{(s)}$ in step 8 as

$$\mathbf{\Gamma}^{(\mathbf{s})}_{\mathbf{v}}) = \{k \quad s.t.\ |\mathbf{v}^{(s)}(k)| > \epsilon^{(s)}\}. \tag{10}$$

Finally, due to the block sparsity of $\mathbf{g}^{(s)}$, the support of $\mathbf{g}^{(s)}$ can be updated in step 9 as

$$\begin{aligned} &\mathbf{\Gamma}^{(\mathbf{s})}_{\mathbf{g}} = \{(l-1)N_t + 1 : 1 : lN_T \quad s.t. \\ &\text{Card}(\{(l-1)N_t + 1 : 1 : lN_T\} \cap \mathbf{\Gamma}^{(\mathbf{s})}_{\mathbf{v}}) > N_t/2\}. \end{aligned} \tag{11}$$

Note that the support $\mathbf{\Gamma}^{(s)}_{\mathbf{g}}$ is independent of $\mathbf{\Gamma}^{(s-1)}_{\mathbf{g}}$ in block-ISD, which is essentially different from the classical greedy algorithm orthogonal matching pursuit (OMP) [13]. In OMP, only one element of $\mathbf{\Gamma}^{(s)}_{\mathbf{g}}$ is updated in each iteration, and once an element is added to $\mathbf{\Gamma}^{(s)}_{\mathbf{g}}$, this element will not be removed in the following iterations. From this aspect, block-ISD is similar to subspace pursuit (SP) [19] and compressive sampling matching pursuit (CoSaMP) [13]. They update all elements of the recovered signal in every iteration, whereby the support detection not only selects desired elements but also removes undesired elements. However, the support detection of SP and CoSaMP is based on the signal sparsity level to be known as a priori, while the support detection of block-ISD is based on the sparsity-independent threshold $\epsilon^{(s)}$. Thus, block-ISD can recover the signal without prior knowledge of the channel sparsity level.

Compared with the classical ISD algorithm, the key difference of block-ISD is the consideration of the block sparsity of $\mathbf{g}^{(s)}$ in step 9. For a certain non-zero block of $\mathbf{g}$, continuous

N_t elements of this block are supposed to be non-zeros. Their indexes are supposed to be included in $\mathbf{\Gamma}_{\mathbf{g}}^{(s)}$. However, some indexes of the block may be detected incorrectly due to the impact of noise. Nevertheless, we can determine whether this block is a zero block or a non-zero block by comparing the number of indexes included in $\mathbf{\Gamma}_{\mathbf{g}}^{(s)}$ with $N_t/2$ (half of the block length). Only when more than half of the indexes of the block are included in $\mathbf{\Gamma}_{\mathbf{g}}^{(s)}$, then all N_t indexes of the block will be added in $\mathbf{\Gamma}_{\mathbf{g}}^{(s)}$. This mechanism considering the block sparsity is expected to increase the robustness of the support detection and thus improve the channel estimation performance as will be verified by simulation results in section IV. Moreover, compared with ISD, block-ISD only adds some comparison operations, and thus the overall complexity does not increase much.

C. Pilot design based on mutual coherence properrty

Note that the sensing matrix $\mathbf{P}$ in (3) plays an important role in the sparse signal recovery performance. There are various results regarding what conditions $\mathbf{P}$ should satisfy to guarantee a robust signal recovery. The widely used conditions include the restricted isometry property (RIP) [19] and mutual coherence property (MCP) [13]. In this paper we discuss the sensing matrix based on MCP.

The maximal mutual coherence (MMC) of a matrix $\mathbf{P} \in \mathcal{C}^{p\times LN_T}$ is defined as the largest normalized absolute inner product between any two columns of $\mathbf{P}$, i.e.,

$$\mu(\mathbf{P}) = \max_{1\leq i\neq j\leq LN_T} \frac{|<\mathbf{P}_i, \mathbf{P}_j>|}{\|\mathbf{P}_i\|_2\|\mathbf{P}_j\|_2}. \quad (12)$$

Let the signal $\mathbf{h}$ be K-sparse (i.e., the number of non-zero elements in $\mathbf{h}$ is K), and $\mathbf{y} = \mathbf{P}\mathbf{h} + \mathbf{n}$, where $\mathbf{n}$ is noise vector consisting of i.i.d. AWGN variables with zero mean and variance γ. Then, $\mathbf{h}$ can be recovered by

$$\hat{\mathbf{h}} = \min_{\mathbf{h}} \|\mathbf{h}\|_1 \quad s.t.\ \|\mathbf{y} - \mathbf{P}\mathbf{h}\|_2 \leq \epsilon. \quad (13)$$

Suppose that $K \leq \frac{\frac{1}{\mu(\mathbf{P})}+1}{4}$ and $\epsilon \geq \gamma$, then the signal recovery error is bounded by [13]

$$\|\hat{\mathbf{h}} - \mathbf{h}\|_2 \leq \frac{\gamma}{\sqrt{1-\mu(\mathbf{P})(4K-1)}}. \quad (14)$$

It's clear that a smaller MMC results in a better signal recovery performance.

Accordingly, the MMC of the matrix $\mathbf{P}$ can be expressed as the maximal absolute element of the Grammar matrix $\mathbf{G} = \tilde{\mathbf{P}}^H\tilde{\mathbf{P}}$ except for the diagonal elements:

$$\mu(\mathbf{P}) = \max_{i\neq j} |g_{ij}|, \quad (15)$$

where $\tilde{\mathbf{P}}$ denotes the column-normalized version of $\mathbf{P}$, i.e., $\tilde{\mathbf{P}}(:,i) = \frac{\mathbf{P}(:,i)}{\|\mathbf{P}(:,i)\|}$.

As $\mathbf{P} = [\mathbf{C}_1(\mathbf{F}_L)_\Omega, \mathbf{C}_2(\mathbf{F}_L)_\Omega, \cdots, \mathbf{C}_{N_t}(\mathbf{F}_L)_\Omega]$, the MMC $\mu(\mathbf{P})$ is determined by the pilots $\mathbf{c}_i$ for $1 \leq i \leq N_t$. To achieve a small MMC, we propose the following pilot pattern where the elements of the pilot sequences $\mathbf{c}_i$ have independent random phases but a unit amplitude. Thus, the l_2-norm of the columns of $\mathbf{P}$ are constant $\sqrt{p}$, and thus the Grammar matrix $\mathbf{G}$ can be written as

$$\mathbf{G} = \frac{1}{p}\mathbf{P}^H\mathbf{P}. \quad (16)$$

Then, the MMC $\mu(\mathbf{P})$ of the sensing matrix $\mathbf{P}$ is

$$\mu(\mathbf{P}) = \frac{1}{p} \max_{i\neq j \text{ or } n_1\neq n_2} \sum_{k=1}^{p} c_{i,k}^* c_{j,k} e^{-j\frac{2\pi}{N}m_k(n_1-n_2)}, \quad (17)$$

where $1 \leq i,j \leq N_t$, $0 \leq n_1, n_2 \leq L-1$, $c_{i,k}$ denotes the kth element of $\mathbf{c}_i$, and $\{m_k\}_{k=1}^p$ is the set of subcarrier indexes assigned to pilots. The conditions that $i \neq j$ or $n_1 \neq n_2$ ensures that the selected elements are not the diagonal elements of $\mathbf{G}$. To derive the expection of $\mu(\mathbf{P})$, we consider three cases as below:

(a) If $i = j$ and $n_1 \neq n_2$, (17) can be simplified as

$$\mu(\mathbf{P}) = \frac{1}{p} \max \sum_{k=1}^{p} e^{-j\frac{2\pi}{N}m_k(n_1-n_2)}. \quad (18)$$

The expectation of $\mu(\mathbf{P})$ is zero because $\frac{m_k}{N}$ follows the i.i.d. uniform distribution $\mathcal{U}[0,1)$.

(b) If $i \neq j$ and $n_1 = n_2$, (17) can be simplified as

$$\mu(\mathbf{P}) = \frac{1}{p} \max \sum_{k=1}^{p} c_{i,k}^* c_{j,k} = \frac{1}{p} \max \sum_{k=1}^{p} e^{j2\pi(\theta_{j,k}-\theta_{i,k})}. \quad (19)$$

The expectation of $\mu(\mathbf{P})$ is zero because $\theta_{i,k}$ and $\theta_{j,k}$ follows the i.i.d. uniform distribution $\mathcal{U}[0,1)$.

(c) If $i \neq j$ and $n_1 \neq n_2$, (17) can be rewritten as

$$\mu(\mathbf{P}) = \frac{1}{p} \max_{i\neq j \text{ and } n_1\neq n_2} \sum_{k=1}^{p} e^{j2\pi(\frac{m_k}{N}(n_2-n_1)+(\theta_{j,k}-\theta_{i,k}))}. \quad (20)$$

Clearly, the expectation of $\mu(\mathbf{P})$ is zero because $\frac{m_k}{N}$, $\theta_{i,k}$ and $\theta_{j,k}$ follow the i.i.d. uniform distribution $\mathcal{U}[0,1)$.

To sum up, the proposed pilot pattern results in a small MMC, which ensures reliable channel estimation performance of block-ISD.

IV. Simulation Results

Simulations have been conducted to validate the performance of block-ISD. We consider a $N_t = 32$ massive MIMO system with the system bandwidth of 50 MHz and the OFDM symbol length $N = 4096$. We adopt the ITU Vehicular B channel model [8] with the maximal channel length $L = 128$.

Fig. 2 shows the normalized mean square error (NMSE) performance comparison between block-ISD and the classical ISD and BP algorithms, when the number of pilots is $p = 640$. In addition, the performance of the exact least square (LS) assuming the exact knowledge of the signal support is also presented as the lower bound of NMSE. It can be observed that block-ISD outperforms both classical ISD and BP algorithms. Specifically, block-ISD achieves over 4 dB SNR gain than ISD when the target NMSE of 10^{-1} is considered. The performance gain is mainly attributed to the exploration of the block sparsity of the block-sparse equivalent CIR. Note that

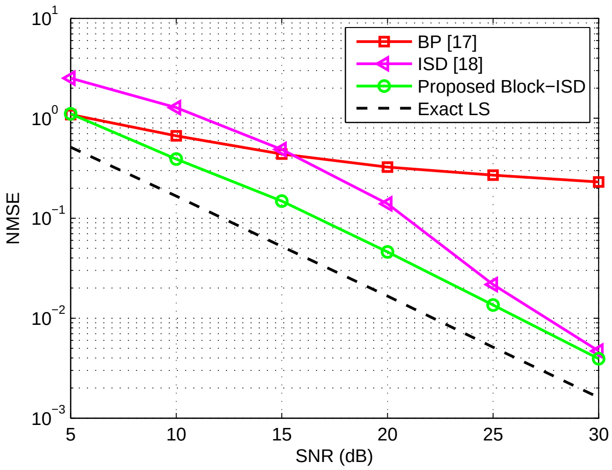

Fig. 2. The comparison of NMSE performance among block-ISD, ISD, and BP.

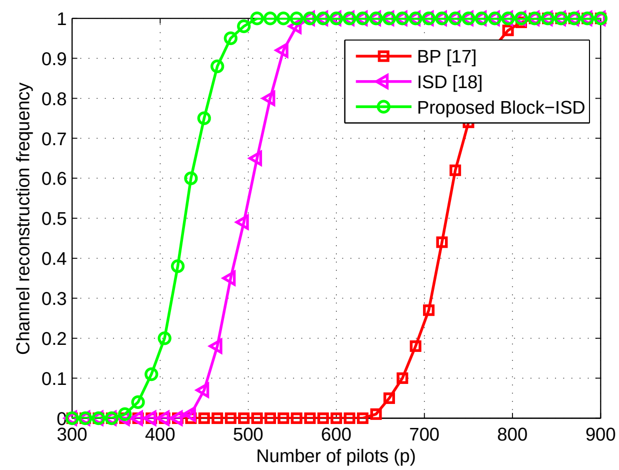

Fig. 3. The comparison of channel reconstruction frequency performance among block-ISD, ISD, and BP.

block-ISD obviously outperforms ISD when SNR is not very high. This is due to the fact that block-ISD is more capable of correcting the support detection error caused by the additive noise than ISD when SNR is not very high (e.g., SNR $<$ 20 dB), which therefore enhances the support detection and ultimately leads to a lower NMSE.

Fig. 3 illustrates the performance of the channel reconstruction frequency against the number of pilots p ranging from 300 to 900. The channel reconstruction frequency is defined as the ratio between the times of reliable channel estimation and the total times of simulation, where reliable channel estimation refers to the case that the NMSE is smaller than the NMSE threshold $\tau = 0.08$ [18]. It is evident that block-ISD significantly outperforms conventional ISD and BP algorithms. The number of pilots required for reliable channel estimation for block-ISD, ISD, and BP is 510, 585, and 840, respectively. For conventional channel estimation techniques such as LS and MMSE, the number of pilots should be as large as $N_T L = 32 \times 128 = 4096$ to ensure (7) as an overdetermined problem. That is to say, block-ISD achieves a substantial reduction of $(4096 - 510)/4096 = 87.5\%$ pilot overhead compared with these conventional channel estimation techniques without considering the channel sparsity.

V. Conclusions

In this paper, we have proposed a block-ISD based algorithm for downlink channel estimation with low pilot overhead for FDD massive MIMO systems. It is found that by exploring the block sparsity inherent in the block-sparse equivalent CIR, which is generated by considering the spatial correlations of CIRs sharing a common support, the proposed block-ISD algorithm could improve the channel estimation performance by over 4 dB than classical ISD and BP algorithms. In addition, we have shown that block-ISD requires no prior knowledge of the channel sparsity level, thereby making an important step toward practical implementation. Simulation results have demonstrated that block-ISD can achieve a reduction of 87.5% pilot overhead than conventional channel estimation techniques.

Acknowledgment

This work was supported by the National Key Basic Research Program of China (Grant No. 2013CB329203), the National Natural Science Foundation of China (Grant Nos. 61571270 and 61271266), the Beijing Natural Science Foundation (Grant No. 4142027), and the Foundation of Shenzhen government.

References

[1] E. Dahlman, S. Parkvall, J. Skold, et al., *3G evolution: HSPA and LTE for mobile broadband[M],* Academic press, 2010.

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

[3] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169-6183, Dec. 2015.

[4] B. Lee, J. Choi, J. Seol, D. Love, and B. Shim, "Antenna grouping based feedback reduction for FDD-based massive MIMO systems," in *Proc. IEEE Int. Conf. on Commun. (IEEE ICC'14)*, Jun. 2014, pp. 4477-4482.

[5] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing based spatial-temporal channel estimation for FDD massive MIMO," to appear in *IEEE Trans. Commun.*

[6] L. Correia *Moblie broadband multimedia networks, techniques, models and tools for 4G,* SanDiego, CA: Academic, 2006.

[7] M. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surv. Tutor.*, vol. 9, no. 2, pp. 18-48, 2nd Quarter 2007.

[8] L. Dai, J. Wang, Z. Wang, P. Tsiaflakis, and M. Moonen, "Spectrum- and energy-efficient OFDM based on simultaneous multi-channel reconstruction," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 6047-6059, Dec. 2013.

[9] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178-2181, Dec. 2015.

[10] 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.

[11] G. Gui and F. Adachi, "Stable adaptive sparse filtering algorithms for estimating multiple-inputCmultiple-output channels," *IET Commun.*, vol. 8, no. 7, pp. 1032-1040, May 2014.

[12] C. Qi, G. Yue, L. Wu, and N. Arumugam, "Pilot design for sparse channel estimation in OFDM-based cognitive radio systems," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 982-987, Feb. 2014.

[13] M. Duarte and Y. Eldar, “Structured compressed sensing: From theory to applications,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053-4085, Sep. 2011.
[14] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, “Estimation of sparse MIMO channels with common support,” *IEEE J. Sel. Areas Commun.*, vol. 60, no. 12, pp. 3705-3716, Dec. 2012.
[15] Z. Gao, L. Dai, C. Yuen, and Z. Wang, “Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels,” *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1826-1829, Oct. 2015.
[16] W. Shen, L. Dai, Z. Gao, and Z. Wang, “Spatially Correlated Channel Estimation Based on Block Iterative Support Detection for Massive MIMO,” *Electron. Lett.*, vol. 51, no. 7, pp. 587-588, Apr. 2015.
[17] D. Donoho and X. Huo,“Uncertainty principles and ideal atomic decomposition,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845-2862, Nov. 2001.
[18] Y. Wang and W. Yin, “Sparse signal reconstruction via iterative support detection,” *SIAM Journal on Imaging Sciences*, vol. 49, no. 6, pp. 2543-2563, Jun. 2011.
[19] 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.