Sparse Channel Estimation Based on Compressed Sensing for Massive MIMO Systems

Chenhao Qi, Yongming Huang, Shi Jin and Lenan Wu
School of Information Science and Engineering, Southeast University, Nanjing 210096, China
Email: qch@seu.edu.cn

***Abstract*—The sparse channel estimation which sufficiently exploits the inherent sparsity of wireless channels, is capable of improving the channel estimation performance with less pilot overhead. To reduce the pilot overhead in massive MIMO systems, sparse channel estimation exploring the joint channel sparsity is first proposed, where the channel estimation is modeled as a joint sparse recovery problem. Then the block coherence of MIMO channels is analyzed for the proposed model, which shows that as the number of antennas at the base station grows, the probability of joint recovery of the positions of nonzero channel entries will increase. Furthermore, an improved algorithm named block optimized orthogonal matching pursuit (BOOMP) is also proposed to obtain an accurate channel estimate for the model. Simulation results verify our analysis and show that the proposed scheme exploring joint channel sparsity substantially outperforms the existing methods using individual sparse channel estimation.**

***Index Terms*—Compressed sensing (CS); sparse channel estimation; massive MIMO; large-scale MIMO;**

I. Introduction

In order to improve the data rate as well as the reliability of wireless systems, the multi-antenna technology, termed as multiple-input multiple-output (MIMO), has been extensively investigated in the last two decades. In a typical multi-user MIMO system, a base station (BS) equipped with some antennas simultaneously communicates with several users each equipped with a single antenna. Recently, it has been shown that as the number of BS antennas grows to be infinity, the effect of additive noise and rayleigh fading is negligible and a very high data rate can be achieved [1]. Therefore, we may construct a massive MIMO or large-scale MIMO system by equipping the BS with orders of magnitude more antennas, e.g., 128, which is even larger than the number of users that the BS serves [2]. In this way, the BS can sufficiently exploit the spatial degree of freedom to simultaneously communicate with several users using the same temporal and frequency resource [3]. The researchers from Rice University establish a BS equipped with 64 antennas serving 15 users, which is demonstrated to achieve up to 6.7 fold capacity gains while using a mere 1/64th of transmission power [4].

One of the challenges in massive MIMO systems is the pilot overhead that grows linearly with the number of channels to be estimated. In massive MIMO systems, the number of wireless links and channels is very large, leading to the proliferation of pilot overhead and thus the reduced resource for data. To reduce the pilot overhead, one potential choice is to explore the inherent sparsity of wireless channels and to use the sparse channel estimation [5], [6], [7]. Wireless channel is essentially sparse, where only a small number of channel coefficients are nonzero. By applying recently emerged compressed sensing (CS) technique, sparse channel estimation can be used to estimate the channel impulse response (CIR) based on the received and transmitted pilot symbols. Compared to the least squares (LS) and minimum mean square error (MMSE) methods, sparse channel estimation is capable of improving the channel estimation performance and reducing the pilot overhead [8], [9], [10], [11]. In [5], distributed compressive channel estimation and feedback schemes are considered for frequency-division duplex (FDD) massive MIMO Systems. In [12], superimposed pilot design for downlink FDD massive MIMO systems is proposed based on structured CS. In [13], sparse channel estimation with structured CS is proposed for multi-input single-output (MISO) systems. In [14], based on the idea that the degree of freedom of the channel matrix is smaller than the number of free parameters, a low-rank matrix approximation is proposed and solved via semidefinite programming (SDP). In [15], uplink channel estimation exploring joint channel sparsity is investigated for massive MIMO systems.

It is shown in [16] that the CIR from different BS antennas to the same user antenna shares a common support, because the time of arrival (ToA) is similar while the paths amplitudes and phases are distinct. In other words, the nonzero positions of different CIRs are the same, exhibiting the joint sparsity. So it is beneficial to exploit the joint sparsity so that the number of pilots for channel estimation can be substantially reduced. In this paper, we first consider the sparse channel estimation exploring the joint channel sparsity, where the channel estimation is modeled as a joint sparse recovery problem. Then the block coherence of MIMO channels is analyzed for the proposed model, which shows that as the number of BS antennas grows, the probability of joint recovery of the positions of nonzero channel entries will increase. Furthermore, an algorithm named block optimized orthogonal matching pursuit (BOOMP) is also proposed to obtain a reliable solution to this model.

The remainder of this paper is organized as follows. Section II provides the system model for sparse channel estimation exploiting joint sparsity, which is then formulated as a joint sparse recovery problem. Section III analyzes the block coherence for the proposed model. In Section IV, an algorithm named BOOMP is proposed to get a solution to the model. Simulation results are provided in Section V. Finally, Section VI concludes this paper.

The notations used in this paper are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $(\cdot)^T$, $(\cdot)^H$, diag$\{\cdot\}$, $\boldsymbol{I}_L$, $\|\boldsymbol{a}\|_2$, $\mathcal{CN}$ and $\emptyset$, denote the matrix transpose, conjugate transpose (Hermitian), the diagonal matrix, the identity matrix of size L, ℓ_2-norm of a vector $\boldsymbol{a}$, the complex Gaussian distribution and the empty set, respectively.

II. System Model

As shown in Figure 1, we give a three-cell massive MIMO system. We illustrate three different configuration of BS antennas, e.g., linear antenna configuration on an edge of a building, rectangular antenna configuration at a wall of a building, and cylindrical antenna configuration on a tower. Therefore, it is flexible to deploy BS for different scenarios in practice. With the large number of antennas, the energy can be focused on extremely sharp beams, where the beamforming will be more efficient and the spatial degree of freedom can be fully exploited.

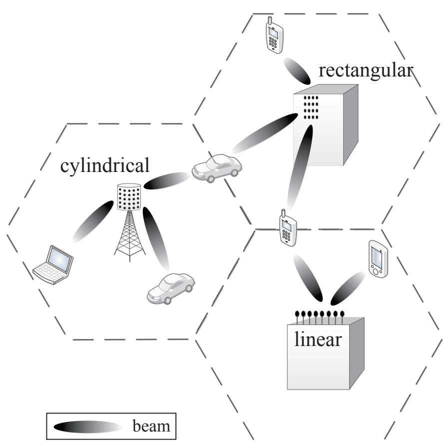

Fig. 1. Three different configuration of BS antennas in massive MIMO systems.

To overcome the frequency-selective fading and improve the spectral efficiency, orthogonal frequency division multiplexing (OFDM) is usually adopted for downlink transmission in current wireless standards, e.g., LTE-A [17]. OFDM transforms the frequency-selective wireless channel into several parallel flat-fading narrowband subchannels. Each subchannel only needs a single-tap equalizer, and therefore the high complexity associated with the long equalizer to combat inter-symbol interference (ISI) is substantially mitigated.

We consider a massive MIMO system including a BS equipped with M antennas and several users each equipped with a single antenna. We use OFDM for downlink transmission. Suppose the total number of OFDM subcarriers is N. From N subcarriers, $K(0 < K \leq N)$ subcarriers are selected to transmit pilot symbols for pilot-assisted channel estimation. In FDD systems, each user first performs channel estimation for each downlink channel and then feeds back the quantized CSI to the BS. With the downlink CSI, the BS designs the beamforming vector for each user so that the spatial degree of freedom in the massive MIMO system can be fully exploited. In order to distinguish M different downlink channels, the BS has to use M orthogonal pilots, either in time domain, frequency domain, or sequence domain. According to current wireless standards [17], frequency-orthogonal pilots are usually employed. With the increased number of BS antennas, i.e., growing M, these orthogonal pilots will occupy increased resource, resulting in reduced resource for data transfer. So in the following, we will explore the joint sparsity of downlink channels and reduce the pilot overhead.

The positions of pilot subcarriers make up a pilot pattern, which is a positive integer vector. Suppose the pilot pattern used by the ith BS antenna is $\boldsymbol{p}^{(i)}$, $i = 1, 2, \ldots, M$. The pilot patterns used by different antennas are frequency-orthogonal to each other, i.e., $\boldsymbol{p}^{(i)} \cap \boldsymbol{p}^{(j)} = \emptyset$ if $i \neq j$, where $\cap$ represents the intersection of two sets. Suppose the OFDM symbol transmitted by the ith BS antenna is $\boldsymbol{x}^{(i)}$, $i = 1, 2, \ldots, M$. The pilot vector transmitted by the ith BS antenna can be denoted as $\boldsymbol{x}^{(i)}(\boldsymbol{p}^{(i)})$, $i = 1, 2, \ldots, M$. After the user receives an OFDM symbols $\boldsymbol{y}$, it can extract the received pilot vectors $\boldsymbol{y}(\boldsymbol{p}^{(i)})$, $i = 1, 2, \ldots, M$, corresponding to different transmit pilot vectors, because the transmit pilot vectors are orthogonal in the frequency domain. To ease the notation, we define $\boldsymbol{y}^{(i)} \triangleq \boldsymbol{y}(\boldsymbol{p}^{(i)})$, $i = 1, 2, \ldots, M$. We then formulate the channel estimation problem as

$$\boldsymbol{y}^{(i)} = \boldsymbol{D}^{(i)} \boldsymbol{F}^{(i)} \boldsymbol{h}^{(i)} + \boldsymbol{\eta}^{(i)}, \quad i = 1, 2, \ldots, M \tag{1}$$

where $\boldsymbol{D}^{(i)} \triangleq \mathrm{diag}\{\boldsymbol{x}^{(i)}(\boldsymbol{p}^{(i)})\}$ denotes a diagonal square matrix, with the diagonal entries being the entries of $\boldsymbol{x}^{(i)}(\boldsymbol{p}^{(i)})$; $\boldsymbol{\eta}^{(i)} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_K)$ denotes the noise term of the ith downlink channel; $\boldsymbol{F}^{(i)}$ is a K by L submatrix indexed by $\boldsymbol{p}^{(i)}$ in row and $[1, 2, \ldots, L]$ in column from a standard N by N DFT matrix; and $\boldsymbol{h}^{(i)} = [h^{(i)}(1), h^{(i)}(2), \ldots, h^{(i)}(L)]^T$ denotes the CIR of the ith downlink channel. Due to the inherent sparsity of wireless channels, most entries of $\boldsymbol{h}^{(i)}$ are zero, and the number of the nonzero entries of $\boldsymbol{h}^{(i)}$ equals the number of multipath in the ith downlink channel. It is shown in [16] that the CIR of different downlink channels shares a common support, because the ToA from different transmit antennas to the same receive antenna is similar while the path amplitudes and phases are distinct. In other words, the nonzero positions of $\boldsymbol{h}^{(i)}$ are the same for $i = 1, 2, \ldots, M$, while their nonzero coefficients are different. We define the measurement matrix $\boldsymbol{A}^{(i)} \triangleq \boldsymbol{D}^{(i)} \boldsymbol{F}^{(i)}$, then (1) can be rewritten as

$$\boldsymbol{y}^{(i)} = \boldsymbol{A}^{(i)} \boldsymbol{h}^{(i)} + \boldsymbol{\eta}^{(i)}, \quad i = 1, 2, \ldots, M \tag{2}$$

which is essentially to use $\boldsymbol{y}^{(i)}$ and $\boldsymbol{A}^{(i)}$ to estimate $\boldsymbol{h}^{(i)}$ under the perturbation of $\boldsymbol{\eta}^{(i)}$. In order to explore the joint sparsity of MIMO downlink channels, we define $\boldsymbol{w}$ as a stack vector

$$\boldsymbol{w} \triangleq [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_L^T]^T \tag{3}$$

where $\boldsymbol{w}_l \triangleq [h^{(1)}(l), h^{(2)}(l), \ldots, h^{(M)}(l)]^T$ denotes the lth block of $\boldsymbol{w}$, $l = 1, 2, \ldots, L$. Since the nonzero positions of $\boldsymbol{h}^{(i)}$ are the same for $i = 1, 2, \ldots, M$, the entries of $\boldsymbol{w}_i$ will be either all zero or all nonzero, exhibiting the *block sparsity*. Correspondingly, we define a stack vector of received pilots as

$$\boldsymbol{z} \triangleq [\boldsymbol{z}_1^T, \boldsymbol{z}_2^T, \ldots, \boldsymbol{z}_K^T]^T \tag{4}$$

where

$$\boldsymbol{z}_l \triangleq [y^{(1)}(l), y^{(2)}(l), \ldots, y^{(M)}(l)]^T \tag{5}$$

denotes the lth block of $\boldsymbol{z}$, $l = 1, 2, \ldots, K$. In the same way, we define a stack vector of noise terms as

$$\boldsymbol{n} \triangleq [\boldsymbol{n}_1^T, \boldsymbol{n}_2^T, \ldots, \boldsymbol{n}_K^T]^T \tag{6}$$

where

$$\boldsymbol{n}_l \triangleq [\eta^{(1)}(l), \eta^{(2)}(l), \ldots, \eta^{(M)}(l)]^T \tag{7}$$

denotes the lth block of $\boldsymbol{n}$, $l = 1, 2, \ldots, K$. We generate a new measurement matrix $\boldsymbol{B}$ based on a matrix $\boldsymbol{E}$. Given any matrix $\boldsymbol{E}$ with K rows and L columns, we substitute the lth-row jth-column entry of $\boldsymbol{E}$, denoted as $E(l, j)$, by a diagonal matrix $\text{diag}\{A^{(1)}(l, j), A^{(2)}(l, j), \ldots, A^{(M)}(l, j)\}$, $l = 1, 2, \ldots, K$, $j = 1, 2, \ldots, L$. We thus construct a block-diagonal matrix $\boldsymbol{B}$ with MK rows and ML columns. The sparse channel estimation exploring joint sparsity can be finally formulated as

$$\boldsymbol{z} = \boldsymbol{B}\boldsymbol{w} + \boldsymbol{n}. \tag{8}$$

Recent works in CS show that the sparse recovery performance of (8) is determined by two factors, including the structure of $\boldsymbol{B}$ and the sparse recovery algorithm, which will be discussed in Section III and Section IV, respectively.

III. Analysis of Block Coherence

Assuming that each column of $\boldsymbol{B}$ in (8) is normalized. This assumption is reasonable because we can normalize $\boldsymbol{B}$ by simply decomposing it into a normalized matrix $\boldsymbol{Q}$ and a diagonal matrix $\boldsymbol{G}$ so that $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{G}$. And after the sparse recovery, we can obtain the solution to the original problem by multiplying the results with $\boldsymbol{G}^{-1}$.

We define the *coherence* of $\boldsymbol{A}^{(i)}$, $i = 1, 2, \ldots, M$ for (2) as

$$\mu(\boldsymbol{A}^{(i)}) = \max_{l \neq k} \left|(\boldsymbol{a}_l^{(i)})^H \boldsymbol{a}_k^{(i)}\right| \tag{9}$$

where $\boldsymbol{a}_l^{(i)}$ denotes the lth column of $\boldsymbol{A}^{(i)}$, $l = 1, 2, \ldots, L$. To improve the sparse recovery performance of (2), it is better to minimize $\mu(\boldsymbol{A}^{(i)})$ [8].

We represent $\boldsymbol{B}$ in (8) as a concatenation of blocks $\boldsymbol{B}_l$, $l = 1, 2, \ldots, L$, as

$$\boldsymbol{B} \triangleq [\underbrace{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_M}_{\boldsymbol{B}_1}, \underbrace{\boldsymbol{b}_{M+1}, \boldsymbol{b}_{M+2}, \ldots, \boldsymbol{b}_{2M}}_{\boldsymbol{B}_2}, \ldots, \underbrace{\boldsymbol{b}_{LM-M+1}, \boldsymbol{b}_{LM-M+2}, \ldots, \boldsymbol{b}_{LM}}_{\boldsymbol{B}_L}] \tag{10}$$

where $\boldsymbol{b}_j$ denotes the jth column of $\boldsymbol{B}$, $j = 1, 2, \ldots, LM$. It's observed that the columns within each block of $\boldsymbol{B}$ are orthogonal to each other, meaning that the rank of each block is M.

Similarly, we define the *coherence* of $\boldsymbol{B}$ as

$$\mu(\boldsymbol{B}) = \max_{l \neq k} |\boldsymbol{b}_l^H \boldsymbol{b}_k|. \tag{11}$$

Considering that the sparse $\boldsymbol{w}$ in (8) exhibits block sparsity, we further define the *block coherence* of $\boldsymbol{B}$ according to [18] as

$$\mu_B(\boldsymbol{B}) = \frac{1}{M} \max_{l \neq k} \rho(\boldsymbol{B}_l^H \boldsymbol{B}_k) \tag{12}$$

where we denote the spectrum norm of a given matrix $\boldsymbol{R}$ as

$$\rho(\boldsymbol{R}) \triangleq \lambda_{max}^{1/2}(\boldsymbol{R}^H \boldsymbol{R}), \tag{13}$$

with $\lambda_{max}(\boldsymbol{R}^H \boldsymbol{R})$ representing the largest eigenvalue of the positive-semidefinite matrix $\boldsymbol{R}^H \boldsymbol{R}$.

Theorem 1: For sparse channel estimation exploring the joint sparsity which is formulated in (8), we have

$$\mu_B(\boldsymbol{B}) = \frac{1}{M} \mu(\boldsymbol{B}). \tag{14}$$

Proof:

From (8) and (10), we observe that

$$\mu(\boldsymbol{B}) = \max_{i=1,2,\ldots,M} \mu(\boldsymbol{A}^{(i)}). \tag{15}$$

According to (12), we have

$$\begin{aligned}
\mu_B(\boldsymbol{B}) &= \frac{1}{M} \max_{l \neq k} \rho(\boldsymbol{B}_l^H \boldsymbol{B}_k) \\
&= \frac{1}{M} \max_{l \neq k} \rho\Big(\text{diag}\Big\{(\boldsymbol{a}_k^{(1)})^H \boldsymbol{a}_l^{(1)}, (\boldsymbol{a}_k^{(2)})^H \boldsymbol{a}_l^{(2)}, \ldots, (\boldsymbol{a}_k^{(M)})^H \boldsymbol{a}_l^{(M)}\Big\}\Big) \\
&= \frac{1}{M} \max_{l \neq k} \lambda_{max}^{1/2}\Big(\text{diag}\Big\{\Big|(\boldsymbol{a}_k^{(1)})^H \boldsymbol{a}_l^{(1)}\Big|^2, \Big|(\boldsymbol{a}_k^{(2)})^H \boldsymbol{a}_l^{(2)}\Big|^2, \ldots, \\
&\qquad \Big|(\boldsymbol{a}_k^{(M)})^H \boldsymbol{a}_l^{(M)}\Big|^2\Big\}\Big) \\
&= \frac{1}{M} \max_{l \neq k} \max_{i=1,2,\ldots,M} \Big|(\boldsymbol{a}_k^{(i)})^H \boldsymbol{a}_l^{(i)}\Big| \\
&= \frac{1}{M} \max_{i=1,2,\ldots,M} \max_{l \neq k} \Big|(\boldsymbol{a}_k^{(i)})^H \boldsymbol{a}_l^{(i)}\Big| \\
&= \frac{1}{M} \max_{i=1,2,\ldots,M} \mu(\boldsymbol{A}^{(i)}) = \frac{1}{M} \mu(\boldsymbol{B}).
\end{aligned} \tag{16}$$

■

If M grows to be infinity, $\mu_B(\boldsymbol{B})$ will be zero, which means that the blocks $\boldsymbol{B}_l, l = 1, 2, \ldots, L$, in (10) will be orthogonal to each other, leading to the unique recovery of blocks. So as the number of BS antennas grows, the probability of joint recovery of the positions of nonzero channel entries will increase. In this way, we can reduce the pilot overhead and therefore leave more resource for data transfer in the massive MIMO system.

IV. Block Optimized Orthogonal Matching Pursuit (BOOMP)

Existing methods for solving (8) can be roughly divided into two classes, including convex optimization algorithms and greedy algorithms. The convex optimization algorithms include BP algorithms such as ℓ_1-LS, YALL1, SpaRSA and other optimization solvers. The greedy algorithms construct a sparse solution by iteratively selecting the matrix columns and eventually forming a linear combination of them closest to the original signal, and they include methods such as orthogonal matching pursuit (OMP), CoSaMP, subspace pursuit and Homotopy. However, all of these algorithms did not exploit the joint sparsity.

Now we propose a BOOMP algorithm exploring the joint sparsity for the proposed model in (8). Note that the BOOMP algorithm presented in this work is based on the optimized OMP algorithm (OOMP) [19] instead of the basic OMP algorithm.

Algorithm 1-Block Optimized Orthogonal Matching Pursuit

1: Input: $\boldsymbol{B}$, z, M, L, σ.
2: Initializations: $\boldsymbol{r} \Leftarrow z$. $T \Leftarrow 0$. $\boldsymbol{\Lambda} \Leftarrow \emptyset$
3: **while** $\|\boldsymbol{r}\|_2 > M\sigma$ and $T \leq L$
4: $\quad T \Leftarrow T + 1$.
5: $\quad$ Obtain J via (17).
6: $\quad \boldsymbol{\Lambda} \Leftarrow \boldsymbol{\Lambda} \cup \{J\}$.
7: $\quad \boldsymbol{r} \Leftarrow z - \boldsymbol{B}_{\boldsymbol{\Lambda}}(\boldsymbol{B}_{\boldsymbol{\Lambda}}^H \boldsymbol{B}_{\boldsymbol{\Lambda}})^{-1} \boldsymbol{B}_{\boldsymbol{\Lambda}}^H z$.
8: **end**
9: Output: $\hat{\boldsymbol{h}}_{\boldsymbol{\Lambda}}^{(i)} \Leftarrow (\boldsymbol{A}_{\boldsymbol{\Lambda}}^H \boldsymbol{A}_{\boldsymbol{\Lambda}})^{-1} \boldsymbol{A}_{\boldsymbol{\Lambda}}^H \boldsymbol{y}^{(i)},\ i = 1, 2, \ldots, M$.

At first, we initialize a residue vector $\boldsymbol{r} \Leftarrow z$ and a loop counter $T \Leftarrow 0$. At each iteration, we obtain an index of the nonzero entry of $\boldsymbol{h}^{(i)}$ by

$$J = \arg \max_{j \in \{1,2,\ldots,L\} \setminus \boldsymbol{\Lambda}} \left\| (\boldsymbol{B}_j^H \boldsymbol{B}_j)^{-1} \boldsymbol{B}_j^H \boldsymbol{r} \right\|_2 \tag{17}$$

and keep J in an active set $\boldsymbol{\Lambda}$. Since $\boldsymbol{h}^{(i)}$ shares a common support for $i = 1, 2, \ldots, M$, we only need one active set to keep the common support. We denote the submatrix indexed by $\boldsymbol{\Lambda}$ in blocks from $\boldsymbol{B}$ and the submatrix indexed by $\boldsymbol{\Lambda}$ in columns from $\boldsymbol{A}$ as $\boldsymbol{B}_{\boldsymbol{\Lambda}}$ and $\boldsymbol{A}_{\boldsymbol{\Lambda}}$, respectively. We iteratively update the residue $\boldsymbol{r}$ by the LS estimation in step 7 of Algorithm 1, where $(\boldsymbol{B}_{\boldsymbol{\Lambda}}^H \boldsymbol{B}_{\boldsymbol{\Lambda}})^{-1} \boldsymbol{B}_{\boldsymbol{\Lambda}}^H$ is the pseudo inverse of $\boldsymbol{B}_{\boldsymbol{\Lambda}}$. Once the power of residue is comparable to the noise or the number of iterations is greater than L, we stop the iterations. Meanwhile we output the estimated CIR as $\hat{\boldsymbol{h}}^{(i)}$, with the coefficients of nonzero entries denoted as $\hat{\boldsymbol{h}}_{\boldsymbol{\Lambda}}^{(i)},\ i = 1, 2, \ldots, M$.

V. Simulation Results

We consider a massive MIMO system including a BS equipped with $M = 8$ antennas. The BS uses $N = 256$ OFDM subcarriers for downlink transmission, where $K = 16$ subcarriers are selected to transmit pilot symbols. QPSK is employed for modulation. The length of the CIR vector is set to be $L = 60$. The number of channel multipath is set to be $S = 12$, which means that there are only $S = 12$ nonzero entries in the CIR vector. Since different channels share a common support [16], the positions of nonzero entries in the CIR vector are the same while the coefficients of these nonzero entries are different.

A. With the fixed positions of nonzero entries of CIR

We first consider the fixed positions of nonzero entries of CIR. The positions of nonzero entries are fixed to be [2, 13, 21, 24, 29, 33, 41, 42, 43, 53, 54, 60]. In order to estimate the downlink channel, each BS antenna transmits a pilot symbol and the user will simultaneously receive $M = 8$ different pilot symbols, meaning that the user has to estimate $M = 8$ channels. In order to distinguish different downlink channels, frequency-orthogonal pilots are used. In Table I, we provide $M = 8$ frequency-orthogonal pilots via pilot optimization [8]. Each pilot in Table I is used by a BS antenna for downlink sparse channel estimation.

TABLE I
Frequency-orthogonal pilots for sparse channel estimation in the massive MIMO system.

	Positions of pilot subcarriers
1st antenna	8, 40, 48, 52, 72, 82, 99, 142, 145, 154, 158, 161, 183, 209, 212, 230
2nd antenna	9, 41, 49, 53, 73, 83, 100, 143, 146, 155, 159, 162, 184, 210, 213, 231
3rd antenna	10, 42, 50, 54, 74, 84, 101, 144, 147, 156, 160, 163, 185, 211, 214, 232
4th antenna	17, 25, 47, 56, 59, 63, 75, 111, 115, 130, 141, 149, 153, 174, 200, 250
5th antenna	12, 34, 55, 64, 67, 109, 112, 148, 173, 215, 222, 233, 238, 241, 249, 252
6th antenna	2, 15, 45, 58, 62, 66, 96, 103, 107, 132, 165, 181, 186, 189, 204, 206
7th antenna	18, 22, 33, 68, 76, 80, 88, 91, 95, 116, 133, 167, 198, 205, 229, 246
8th antenna	7, 79, 92, 117, 120, 152, 168, 180, 187, 197, 219, 223, 239, 243, 251, 255

Now we compare the individual sparse channel estimation using OMP and the joint sparse channel estimation using BOOMP. Since $K \leq 2S$, the individual sparse channel estimation can not succeed, from information theoretical point of view, because $K = 16$ equations are not enough to solve 24 unknown variables including $S = 12$ unknown positions and $S = 12$ unknown coefficients of nonzero entries. As shown in Table II, the positions of nonzero entries individually estimated for $\boldsymbol{h}^{(i)},\ i = 1, 2, \ldots, 8$, named as *individual for ith antenna*, are all incorrect. Then we use 2 of 8 antennas, 4 of 8 antennas, 6 of 8 antennas and all 8 antennas, respectively, for joint sparse channel estimation, named as *joint for x antennas* with $x = 2, 4, 6,$ or 8. It is seen from Table II that we can not obtain the true positions exactly as those of the original CIR unless we use all $x = 8$ antennas for joint sparse recovery, which verifies the analysis of block coherence in Section III. Moreover, the estimation performance will further increase if we use more antennas and explore the joint sparsity.

We further compare the performance of individual sparse channel estimation and joint sparse channel estimation in terms of mean square error (MSE). We define the MSE as

$$MSE = \frac{1}{V} \sum_{i=1}^{V} \frac{\left\| \boldsymbol{h}^{(i)} - \hat{\boldsymbol{h}}^{(i)} \right\|_2^2}{\|\boldsymbol{h}^{(i)}\|_2^2}. \tag{18}$$

where $\hat{\boldsymbol{h}}^{(i)}$ is the estimate of $\boldsymbol{h}^{(i)}$ and V is the number of all possibilities for the averaging. For example, in individual sparse channel estimation, $V = 8$. In joint sparse channel

TABLE II
COMPARISONS OF INDIVIDUAL SPARSE RECOVERY AND JOINT SPARSE RECOVERY IN TERMS OF ESTIMATED NONZERO POSITIONS FOR DOWNLINK CHANNELS.

	Positions of nonzero entries
True positions	2,13,21,24,29,33,41,42,43,53,54,60
Individual for 1st antenna	2,8,15,21,24,33,41,42,47,53,54,60
Individual for 2nd antenna	1,2,13,24,35,40,44,46,50,53,54
Individual for 3rd antenna	2,5,11,13,20,24,33,37,42,53,54,55,60
Individual for 4th antenna	1,8,13,17,24,27,33,41,43,46,53,60
Individual for 5th antenna	5,6,13,15,20,21,24,29,31,32,38,41,51,60
Individual for 6th antenna	2,7,12,21,24,26,33,41,42,49,54,60
Individual for 7th antenna	3,8,10,17,21,26,36,41,42,43,50,55,59
Individual for 8th antenna	2,6,15,18,20,21,24,29,32,41,49,56,60
Joint for 2 antennas	8,9,10,12,13,15,21,25,36,43,44,50,56,60
Joint for 4 antennas	2,10,12,13,19,21,24,41,47,50,53,54,57,60
Joint for 6 antennas	3,6,7,13,14,23,29,33,40,41,42,43,51,53,60
Joint for 8 antennas	2,13,21,24,29,33,41,42,43,53,54,60

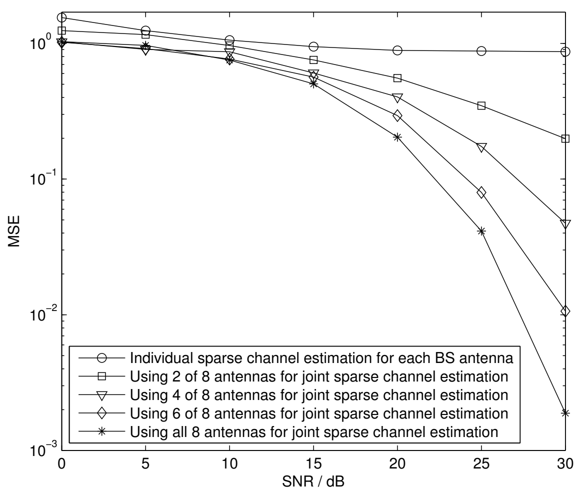

Fig. 2. Comparisons of individual sparse channel estimation and joint sparse channel estimation in terms of MSE with the fixed positions of nonzero entries of CIR.

estimation using 2 of 8 antennas, $V = \binom{8}{2} = 28$. It is seen from Figure 2 that the joint estimation using BOOMP notably outperforms the individual estimation using OMP. Moreover, we can further improve the MSE performance by employing more BS antennas for joint sparse channel estimation, which shows that our scheme is beneficial for massive MIMO systems.

We also evaluate the pilot reduction of joint sparse channel estimation, supposing that the individual sparse channel estimation and joint sparse channel estimation achieve the same MSE performance. The results show that when the number of pilots increases up to $K = 28$, the MSE performance of the individual sparse channel estimation and joint sparse channel estimation is the same. Therefore, the pilot reduction of $(28 - 16)/16 = 75\%$ can be achieved.

B. With random positions of nonzero entries of CIR

Now we consider the random positions of nonzero entries of CIR. For MIMO channel realization where the positions of nonzero entries of CIR are randomly generated, we execute the routine described in the previous subsection. In this way we repeat it 1000 times and make an average of them. As shown in Figure 3, the joint estimation using BOOMP outperforms the individual estimation using OMP. Moreover, as the number of BS antennas for joint sparse channel estimation increases, the MSE can be further reduced.

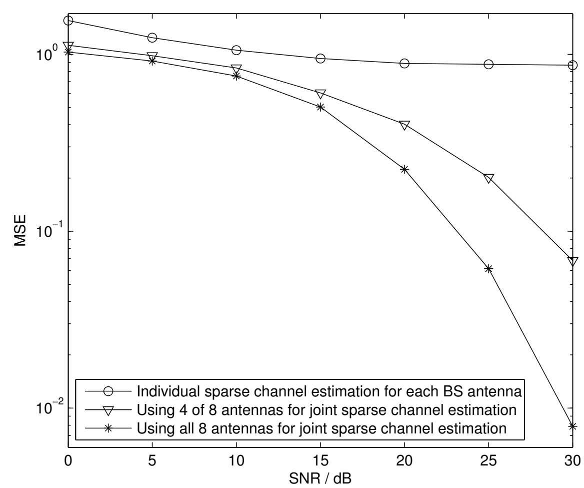

Fig. 3. Comparisons of individual sparse channel estimation and joint sparse channel estimation in terms of average MSE with random positions of nonzero entries of CIR.

VI. CONCLUSIONS

In this paper, we have investigated the sparse channel estimation based on CS for massive MIMO systems. We have proposed a system model for sparse channel estimation exploring the joint channel sparsity. We have analyzed the block coherence for the proposed model, which has shown that as the number of BS antennas grows, the probability of joint recovery of the positions of nonzero channel entries will increase. We have also proposed an algorithm named BOOMP to get a solution to the model. Simulation results have verified our analysis and shown that the proposed sparse channel estimation exploring joint sparsity substantially outperforms the existing methods using individual sparse channel estimation.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61302097 and 61201185, by the National Science and Technology Major Project of China under Grants 2013ZX03003006-002 and 2012ZX03001036-004, by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20120092120014, and by the Fundamental Research Funds for the Central Universities under Grant 2242013K30010.

References

[1] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. 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. 2012.

[3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[4] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Int. Conf. Mobile Computing and Networking (MobiCom)*, Istanbul, Turkey, Aug. 2012, pp. 53–64.

[5] X. Rao and V. 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, June 2014.

[6] 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.

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

[8] C. Qi and L. Wu, "A study of deterministic pilot allocation for sparse channel estimation in OFDM systems," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 742–744, May 2012.

[9] C. Qi, G. Yue, L. Wu, Y. Huang, and A. Nallanathan, "Pilot design schemes for sparse channel estimation in ofdm systems," *IEEE Trans. Veh. Technol.*, vol. 64, to appear in 2015.

[10] Z. Gao, L. Dai, Z. Lu, 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, July 2014.

[11] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems," *IEEE Signal Process. Lett.*, vol. 18, no. 12, pp. 749–752, Dec. 2011.

[12] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, June 2014.

[13] W. Hou and C. W. Lim, "Structured compressive channel estimation for large-scale MISO-OFDM systems," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 765–768, May 2014.

[14] S. L. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. 2013 IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[15] C. Qi and L. Wu, "Uplink channel estimation for massive MIMO systems exploring joint channel sparsity," *Electr. Lett.*, vol. 50, no. 23, pp. 1770–1772, Nov. 2014.

[16] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2012.

[17] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-Advanced: Next-generation wireless broadband technology," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10–22, June 2010.

[18] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, June 2010.

[19] L. R. Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.