Downlink Pilot Reduction for Massive MIMO Systems via Compressed Sensing

Jun Won Choi, *Member, IEEE*, Byonghyo Shim, *Member, IEEE*, and Seok-Ho Chang, *Member, IEEE*

***Abstract*—This letter addresses a problem of downlink pilot allocation for massive multiple-input multiple-output (MIMO) systems. When a massive MIMO is employed in frequency division duplex (FDD) systems, significant amount of radio resources are dedicated to the transmission of downlink pilots. Such huge pilot overhead leads to a substantial loss in the maximum data throughput, which motivates us to reduce the number of pilots. In this letter, we propose a pilot reduction strategy based on compressed sensing techniques for orthogonal frequency division multiplexing systems. The pilots are randomly located in a low density manner over the time and frequency domain. To estimate the channels with such low density pilots, we propose a novel sparse channel estimation technique that exploits the common support of the consecutive channel impulse responses over the certain time duration. The evaluation shows that for a massive MIMO with 128 antennas, the proposed scheme achieves significant reduction of pilot overhead, while maintaining good channel estimation performance.**

***Index Terms*—Channel estimation, compressed sensing, downlink pilot allocation, massive multiple-input multiple-output (MIMO), orthogonal frequency division multiplexing (OFDM).**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) is considered as a key technique to support high data throughput for the next generation wireless communications. By deploying a large number of antennas in the base station, massive MIMO systems can achieve drastic increase in the spectral efficiency [1]. In reality, however, there are many challenges that one needs to address to employ massive MIMO. One challenge is that downlink pilots sent from each of the base station antennas in an orthogonal manner consume the significant amount of radio resources such as time and bandwidth. Note that each user estimates the downlink channel using the downlink pilots and then sends it back to the base station for multi-user precoding or beamforming. Hence, it is crucial to reduce such huge pilot overhead without degrading channel estimation performance at each user.

One popular way to reduce the number of downlink pilots is to use a time division duplex (TDD) protocol. In the TDD systems, the state of downlink channels can be inferred from that of uplink channels using *channel reciprocity* [1]. Thus, the base station can estimate the downlink channels without any downlink pilots. While the TDD system provides a remedy for the pilot overhead issue, many current wireless systems adopt frequency division duplex (FDD), and hence it is important to employ the massive MIMO for the existing FDD systems.

In recent years, compressed sensing (CS) techniques have received attention since they can recover the unknown signals from only a small number of measurements [2]. The signal recovery schemes developed for CS exploit the *sparse nature* of signals (that is, only a small number of components in a signal vector are nonzero). In the scenarios where more than one measurement vectors are available and the unknown signal vectors have common support (i.e., the set of indices for nonzero elements), the performance of the recovery scheme can be improved by exploiting the joint sparsity of the source signals [3]. In particular, several recovery schemes that exploit the joint sparsity have been developed to estimate common support of the channel impulse responses (CIR) in the MIMO systems [3]–[6].

In this letter, we investigate the CS-based technique achieving the reduction of pilot overhead for massive MIMO-orthogonal frequency division multiplexing (OFDM) systems. In the base station, a small number of downlink pilots are placed at randomly chosen locations in time and frequency resource grid. As a result of the random pilot allocation, we can construct so called a random *sensing matrix*, which retains desirable properties in recovering the unknown signal vector from a small number of measurements [7]. In the user equipment, we propose a novel sparse signal recovery scheme. Based on the observation that the support of the CIR vector rarely changes during the time period of several OFDM symbols, the proposed scheme finds the common support of the several measured CIR vectors. Note that our channel estimation scheme differs from the existing methods exploiting common support [3]–[6] in that our approach is working for a system with time-varying sensing matrix and furthermore, our method exploits the temporal correlation structure underlying in sequentially acquired CIR vectors. Our numerical evaluation shows that the proposed technique achieves substantial reduction of pilot overhead while maintaining strong channel estimation performance.

II. Low Density Pilot Allocation for Massive MIMO-OFDM Systems

A. Pilot Processing in MIMO-OFDM Systems

We briefly describe the downlink pilot processing in a MIMO-OFDM system. Pilots are transmitted in an exclusive way so that only one antenna transmits pilots for any given time

Manuscript received April 25, 2015; revised August 12, 2015; accepted August 13, 2015. Date of publication August 28, 2015; date of current version November 9, 2015. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (NRF-2014R1A1A2055805) and the National Research Foundation of Korea (NRF) grant funded by Korean government (MSIP) (2014R1A5A1011478). The associate editor coordinating the review of this paper and approving it for publication was C. Zhong.
(Corresponding author:Seok-Ho Chang.)

J. W. Choi is with the Department of Electrical Engineering, Hanyang University, Seoul 133-791, Korea (e-mail: junwchoi@hanyang.ac.kr).

B. Shim is with the Institute of New Media and Communications and School of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, Korea (e-mail: bshim@snu.ac.kr).

S.-H. Chang is with the Department of Computer Science and Engineering, Dankook University, Yongin 448-701, Korea (e-mail: seokho@dankook.ac.kr).
Digital Object Identifier 10.1109/LCOMM.2015.2474398

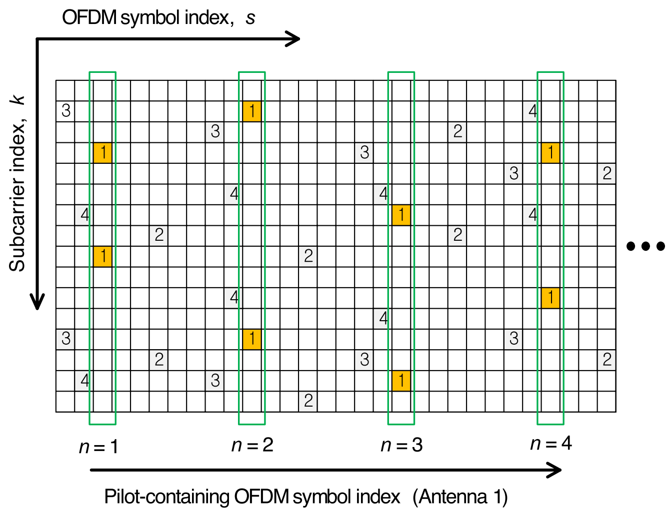

Fig. 1. The proposed low density pilot allocation in an example of OFDM systems with $N_{\text{FFT}} = 16$ and $N_{\text{ANT}} = 4$.

and frequency resource elements. With such pilot transmission, per antenna pilot processing is described in the following: i) generation of a sequence of pilots, ii) multiplexing of the pilots with the data in the frequency domain, iii) execution of the inverse discrete Fourier transform (IDFT), and iv) addition of cyclic prefix.

An OFDM resource grid is represented by the subcarrier index, $1 \le k \le N_{\text{FFT}}$, and the OFDM symbol index, $1 \le s \le N_{\text{FRAME}}$, where N_{FFT} is the number of subcarriers, and N_{FRAME} is the number of OFDM symbols. We let n be the index of the pilot containing OFDM symbols. Fig. 1 depicts an example of OFDM systems with $N_{\text{FFT}} = 16$ and $N_{\text{ANT}} = 4$, where N_{ANT} is the number of base station antennas. Each box represents one resource element in time-frequency domain and a downlink pilot is located at the box marked by some number, which indicates the index of the antenna from which the pilot is transmitted. Each pilot-containing OFDM symbol consists of $N_P = 2$ pilot subcarriers, and the remaining $(N_{\text{FFT}} - N_P) = 14$ subcarriers are used as data subcarriers or guard band.

The signal received by a user is transformed into the frequency-domain signal via DFT. Let $\mathbf{r}_n \in \mathcal{C}^{N_P \times 1}$ be the received frequency-domain pilots extracted from the nth pilot-containing OFDM symbol. Then, the received pilot vector, $\mathbf{r}_n$, can be expressed as

$$\mathbf{r}_n = \text{diag}(\mathbf{p}_n)\mathbf{h}_n + \mathbf{w}_n \tag{1}$$

where $\mathbf{p}_n \in \mathcal{C}^{N_P \times 1}$ is the transmitted pilot vector, $\mathbf{h}_n \in \mathcal{C}^{N_P \times 1}$ is the frequency-domain response of channels at the pilot subcarriers, and $\mathbf{w}_n \in \mathcal{C}^{N_P \times 1}$ is the additive white Gaussian noise $\sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I}_{N_P})$.

B. *Proposed Low Density Pilot Allocation*

In this subsection, we describe the proposed low density pilot allocation strategy for massive MIMO-OFDM systems. Fig. 1 depicts an example of the proposed allocation, where a small number of pilots are randomly located in the time and frequency resource grid. In the conventional systems, the pilots are uniformly located, under the constraint that the number of pilot subcarriers, N_P, is equal to the length of the time-domain CIR vector, denoted by N_{CIR} (i.e., $N_P = N_{\text{CIR}}$). Whereas, the proposed scheme employs much smaller number of pilots (i.e., $N_P \ll N_{\text{CIR}}$). Let $\xi_n = \{c_{n,1}, \ldots, c_{n,N_P}\}$ be the set of indices of N_P pilot subcarriers belonging to the nth pilot-containing OFDM symbol ($1 \le c_{n,i} \le N_{\text{FFT}}$). Due to our random pilot allocation, ξ_n changes with n.

Since perfectly random allocation requires that base station should transmit additional information on pilot locations to a user equipment, in this work, we instead use "pseudo-random" pilot locations that appear to be random but actually the location is chosen by the deterministic process. By doing so, the user equipment is able to see the pilot locations without receiving extra information from the base station.

Let $\mathbf{g}_n = [g_{n,1} \ldots g_{n,N_{\text{CIR}}}]^T$ be the time-domain CIR vector which is effective over the time duration of the nth pilot-containing OFDM symbol. Note that due to multipath fading of wireless channels, $\mathbf{g}_n$ changes with n. The received pilot vector $\mathbf{r}_n$, given by (1), can be rewritten as

$$\mathbf{r}_n = \text{diag}(\mathbf{p}_n)\boldsymbol{\Phi}_n \mathcal{F}_{N_{\text{FFT}}} \begin{bmatrix} \mathbf{g}_n \\ \mathbf{0}_{N_{\text{FFT}} - N_{\text{CIR}}} \end{bmatrix} + \mathbf{w}_n \tag{2}$$

$$= \underbrace{\text{diag}(\mathbf{p}_n)\boldsymbol{\Phi}_n \mathcal{F}_{N_{\text{FFT}}} \boldsymbol{\Pi}}_{=\mathbf{U}_n} \mathbf{g}_n + \mathbf{w}_n \tag{3}$$

where $\mathcal{F}_{N_{\text{FFT}}} \in \mathcal{C}^{N_{\text{FFT}} \times N_{\text{FFT}}}$ is the DFT matrix with (k, l)th entry given by $\exp(-j2\pi kl/N_{\text{FFT}})$, and $\boldsymbol{\Phi}_n \in \mathcal{R}^{N_P \times N_{\text{FFT}}}$ and $\boldsymbol{\Pi} \in \mathcal{R}^{N_{\text{FFT}} \times N_{\text{CIR}}}$ are the matrices consisting of coordinate vectors. Denoting the kth coordinate vector of length N_{FFT} as $\mathbf{e}_k$, we have $\boldsymbol{\Phi}_n = [\mathbf{e}_{c_{n,1}} \ldots \mathbf{e}_{c_{n,N_P}}]^T$ and $\boldsymbol{\Pi} = [\mathbf{e}_1 \ldots \mathbf{e}_{N_{\text{CIR}}}]$.

It is seen from (3) that $\mathbf{U}_n \in \mathcal{C}^{N_P \times N_{\text{CIR}}}$ becomes a composite sensing matrix that forms the measurement vector $\mathbf{r}_n$ from the sparse source vector $\mathbf{g}_n$. In view of this, the channel estimation can be interpreted as the recovery of the unknown time-domain CIR vector $\mathbf{g}_n$ with the knowledge of $\mathbf{U}_n$ and $\mathbf{r}_n$. Also note that the matrix $\boldsymbol{\Phi}_n \mathcal{F}_{N_{\text{FFT}}}$ in (3) corresponds to the random sensing matrix that consists of some specific rows of $\mathcal{F}_{N_{\text{FFT}}}$ selected by $\boldsymbol{\Phi}_n$, which is related to the random locations of pilots, $\xi_n = \{c_{n,1}, \ldots, c_{n,N_P}\}$. This random sensing matrix is identical to what is obtained from *random frequency measurement* [7].

III. Estimation of Channels With Locally Common Support

A. *Sparse Channels With Locally Common Support*

We now describe the model of sparse channels with common support. Due to the characteristic of multipath propagation in wireless channels, in most cases, relatively a small number of components in the CIR vector contain most of the energy. Further, the indices (locations) of those dominant components in the CIR vector are often varying slowly in time, and hence their locations are readily assumed to be constant over the certain time duration, which is the period of so called a *local block*. They henceforth will be referred to as *channels with locally common support*.

When compared to the location, the nonzero components of the CIR vector vary relatively faster with some temporal correlation. From the Jakes' fading model, the temporal correlation of the complex gains of the CIR vector is given by

$$E[g_{k,i}g_{l,j}^*] = \begin{cases} P_i J_0(2\pi f_d T_s(k-l)) & \text{for } i=j \\ 0 & \text{for } i \neq j \end{cases} \tag{4}$$

where $1 \leq i, j \leq N_{\text{CIR}}$, $(\cdot)^*$ denotes the complex conjugate operation, $P_i = E[|g_{k,i}|^2]$ is the variance of the ith component, f_d is the Doppler frequency, T_s is the time interval between adjacent pilot-containing OFDM symbols, and $J_0(x)$ is the zero-order Bessel function of the first kind. In (4), the distinct components of the CIR vector (i.e., the case of $i \neq j$) are assumed to be uncorrelated.

B. Locally Common Support-MMSE Pursuit Algorithm

For such sparse channels with locally common support, we propose a channel estimation algorithm referred to as *locally common support recovery-based channel estimator* (LCS-CE). Let L be the number of pilot-containing OFDM symbols per base station antenna in a local block. To estimate the CIR vector $\mathbf{g}_n$, (see (2)), the proposed LCS-CE uses L adjacent received pilot vectors which consist of $L_c + 1$ causal vectors, $\{\mathbf{r}_{n-L_c}, \ldots, \mathbf{r}_n\}$, and L_{nc} noncausal vectors, $\{\mathbf{r}_{n+1}, \ldots, \mathbf{r}_{n+L_{\text{nc}}}\}$. From (3), the stacked pilot vector $\bar{\mathbf{r}}_n = [\mathbf{r}_{n+L_{\text{nc}}}^T \ldots \mathbf{r}_{n-L_c}^T]^T$ can be expressed as

$$\bar{\mathbf{r}}_n = \underbrace{\text{diag}(\mathbf{U}_{n+L_{\text{nc}}}, \ldots, \mathbf{U}_{n-L_c})}_{=\text{sensing matrix}} \begin{bmatrix} \mathbf{g}_{n+L_{\text{nc}}} \\ \vdots \\ \mathbf{g}_{n-L_c} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{n+L_{\text{nc}}} \\ \vdots \\ \mathbf{w}_{n-L_c} \end{bmatrix} \tag{5}$$

We define a vector $\mathbf{d}_{n,i} \in \mathcal{C}^{L\times 1}$ as

$$\mathbf{d}_{n,i} = [(\mathbf{g}_{n+L_{\text{nc}}})_i \ldots (\mathbf{g}_{n-L_c})_i]^T, \quad i = 1, \ldots, N_{\text{CIR}} \tag{6}$$

where $(\cdot)_i$ denotes the ith component of the vector. Note that $\mathbf{d}_{n,i}$ contains the ith components of all the CIR vectors observed over a local block. Eq. (5) can be rewritten as

$$\bar{\mathbf{r}}_n = \underbrace{[\boldsymbol{\Sigma}_{n,1} \ldots \boldsymbol{\Sigma}_{n,N_{\text{CIR}}}]}_{=\text{sensing matrix}} \begin{bmatrix} \mathbf{d}_{n,1} \\ \vdots \\ \mathbf{d}_{n,N_{\text{CIR}}} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{n+L_{\text{nc}}} \\ \vdots \\ \mathbf{w}_{n-L_c} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\Sigma}_{n,i} \in \mathcal{C}^{LN_P\times L}$ forms a new sensing matrix that corresponds to the new unknown signal vectors $\mathbf{d}_{n,1}, \ldots, \mathbf{d}_{n,N_{\text{CIR}}}$, rather than $\mathbf{g}_{n+L_{\text{nc}}}, \ldots, \mathbf{g}_{n-L_c}$. From (4) and (6), we have

$$E\left[\mathbf{d}_{n,i}\mathbf{d}_{n,j}^H\right] = \begin{cases} P_i \, \mathbf{J}_{L\times L} & \text{for } i=j \\ \mathbf{0}_{L\times L} & \text{for } i \neq j \end{cases} \tag{8}$$

where $1 \leq i, j \leq N_{\text{CIR}}$, $(\cdot)^H$ denotes Hermitian operation, and $\mathbf{J}_{L\times L}$ is the $L \times L$ covariance matrix whose (k,l)th entry is given by $J_0(2\pi f_d T_s(k-l))$.

In order to check whether the location contains energy or not, we introduce a vector called the *support vector* $\mathbf{b} = [b_1 \ldots b_{N_{\text{CIR}}}]^T$. Note that $b_i = 1$ when the ith component of the CIR vector is in the support (i.e., dominant), and $b_i = 0$ otherwise. We assume that $\mathbf{b}$ does not vary over L pilot-containing OFDM symbols. Using the support vector $\mathbf{b}$, (7) can be rewritten as

$$\bar{\mathbf{r}}_n = [\boldsymbol{\Sigma}_{n,1} \ldots \boldsymbol{\Sigma}_{n,N_{\text{CIR}}}] \begin{bmatrix} b_1\mathbf{d}_{n,1} \\ \vdots \\ b_{N_{\text{CIR}}}\mathbf{d}_{n,N_{\text{CIR}}} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_{n+L_{\text{nc}}} \\ \vdots \\ \mathbf{w}_{n-L_c} \end{bmatrix} \tag{9}$$

where it is assumed that the non-dominant components of the CIR vector have zero energies.

Using the linear measurement model in (9), we derive the joint estimation of $\mathbf{b}$ and $\{\mathbf{d}_{n,1}, \ldots, \mathbf{d}_{n,N_{\text{CIR}}}\}$. We first describe the greedy search algorithm for the joint estimation, and then provide detailed mathematical steps. In the greedy search strategy [2], the index of the component with the largest energy in the CIR vector is chosen at each iteration. Based on the index of the strongest component, we update the support vector $\mathbf{b}$ and compute the LMMSE estimate of the unknown signal vectors $\{\mathbf{d}_{n,1}, \ldots, \mathbf{d}_{n,N_{\text{CIR}}}\}$. Then, we obtain the residual signal vector by subtracting some transformation of the support vector and the LMMSE estimates from the received pilot vector. From the residual signal vector, we search for the next strongest component of the CIR vector.

The detailed steps of the proposed algorithm are as follows; Let $\mathbf{a}_n^{(l)} \in \mathcal{C}^{LN_P\times 1}$ and $\mathbf{b}^{(l)} \in \mathcal{C}^{N_{\text{CIR}}\times 1}$ be the residual signal vector and the support vector at the lth iteration, respectively. At the first iteration, we initialize those vectors such that $\mathbf{a}_n^{(0)} = \bar{\mathbf{r}}_n$ and $\mathbf{b}^{(0)} = \mathbf{0}_{N_{\text{CIR}}}$. At the lth iteration, we find the components of the CIR vector maximizing the correlated energy metrics:

$$k_{\max} = \underset{1\leq k\leq N_{\text{CIR}}}{\arg\max} \left\| \boldsymbol{\Sigma}_{n,k}^H \mathbf{a}_n^{(l-1)} \right\|_2^2 \tag{10}$$

where $\|\cdot\|_2$ denotes ℓ_2 norm, and the cost function in (10) quantifies how much the unknown signal vector that corresponds to $\boldsymbol{\Sigma}_{n,k}$, given by (9), contributes to $\bar{\mathbf{r}}_n$. Once $k_{\max}$ is found, we update the support vector $\mathbf{b}^{(l-1)}$ for the next iteration, such that $\mathbf{b}^{(l)} = \mathbf{b}^{(l-1)}$ but the $k_{\max}$th component of $\mathbf{b}^{(l)}$ is set to one. Using the updated vector $\mathbf{b}^{(l)}$, we compute the LMMSE estimate of $\mathbf{d}_{n,i}$ at the lth iteration, denoted by $\hat{\mathbf{d}}_{n,i}^{(l)}$;

$$\hat{\mathbf{d}}_{n,i}^{(l)} = \mathbf{w}_{n,i}^{(l)}\bar{\mathbf{r}}_n, \quad i = 1, \ldots, N_{\text{CIR}} \tag{11}$$

where $\mathbf{w}_{n,i}^{(l)} \in \mathcal{C}^{L\times N_P L}$ is the LMMSE coefficient given by

$$\mathbf{w}_{n,i}^{(l)} = E\left[\mathbf{d}_{n,i}\,\bar{\mathbf{r}}_n^H\right] E^{-1}\left[\bar{\mathbf{r}}_n\bar{\mathbf{r}}_n^H\right] \tag{12}$$

From (8) and (9), it can be shown that

$$\begin{aligned} E\left[\bar{\mathbf{r}}_n\bar{\mathbf{r}}_n^H\right] &= \sum_{i=1}^{N_{\text{CIR}}} (\mathbf{b}^{(l)})_i \, P_i \boldsymbol{\Sigma}_{n,i}\, \mathbf{J}_{L\times L}\boldsymbol{\Sigma}_{n,i}^H + \sigma_w^2\, \mathbf{I}_{LN_P\times LN_P} \\ E\left[\mathbf{d}_{n,i}\,\bar{\mathbf{r}}_n^H\right] &= (\mathbf{b}^{(l)})_i \, P_i \, \mathbf{J}_{L\times L}\boldsymbol{\Sigma}_{n,i}^H \end{aligned} \tag{13}$$

The residual signal vector at the lth iteration is given by

$$\mathbf{a}_n^{(l)} = \bar{\mathbf{r}}_n - \sum_{i=1}^{N_{\text{CIR}}} \boldsymbol{\Sigma}_{n,i}\, (\mathbf{b}^{(l)})_i \, \hat{\mathbf{d}}_{n,i}^{(l)} \tag{14}$$

Let K be the number of dominant components of the CIR vector. Note that we can estimate K by modifying the timing

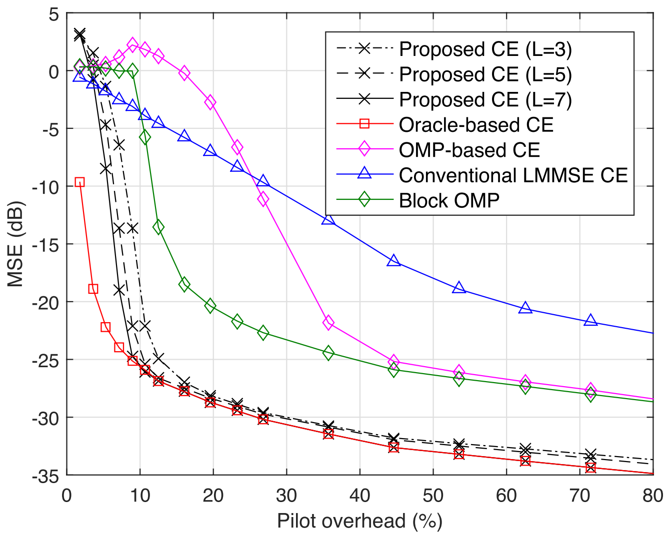

Fig. 2. The MSE performances as a function of pilot overhead (%).

acquisition process such that the number of dominant path delays is estimated using the downlink preamble. We repeat the above steps given by Eqs. (10)–(14), until all the K dominant components are identified in Eq. (10) (i.e., until the iteration index l equals K). Finally, from the LMMSE estimates, $\{\hat{\mathbf{d}}_{n,1}^{(K)}, \ldots, \hat{\mathbf{d}}_{n,N_{\mathrm{CIR}}}^{(K)}\}$, computed at the Kth iteration, we obtain the estimate of the CIR vector, $\hat{\mathbf{g}}_n$ as

$$\hat{\mathbf{g}}_n = \left[\left(\hat{\mathbf{d}}_{n,1}^{(K)} \right)_{L_{\mathrm{nc}}+1} \cdots \left(\hat{\mathbf{d}}_{n,N_{\mathrm{CIR}}}^{(K)} \right)_{L_{\mathrm{nc}}+1} \right]^T. \tag{15}$$

IV. Numerical Evaluations

In our simulations, we consider OFDM systems with 2048 subcarriers (i.e., $N_{\mathrm{FFT}} = 2048$), where adjacent subcarriers are spaced by 15 kHz. A base station is equipped with 128 antennas (i.e., $N_{\mathrm{ANT}} = 128$), which would require huge amount of pilots with the conventional pilot allocation. We use the finite-length pseudo-random sequences to determine the pilot locations in frequency domain. The maximum delay spread of the multipath channel is assumed to be 4.88 μs, which yields the CIR vector of length 150 (i.e., $N_{\mathrm{CIR}} = 4.88\ \mu\mathrm{s} \times 15\ \mathrm{kHz} \times 2048 \approx 150$). The time interval between adjacent pilot-containing OFDM symbols for each antenna is 0.5 ms (i.e., $T_{\mathrm{s}} = 0.5$ ms). We also assume that 10% of the components of the CIR vector are dominant (i.e., $K = 15$). The CIR energy is uniformly distributed over K dominant CIR taps, and the other $N_{\mathrm{CIR}} - K$ components have zero energies. The Doppler frequency is set to 70 Hz. The signal-to-noise ratio (SNR) for each subcarrier is set to 20 dB. We compare the MSE performance of the proposed scheme with that of the Oracle-based scheme which estimates channels with the perfect knowledge on the support vector $\mathbf{b}$. Note that the Oracle-based scheme provides the lower bound of the MSE for any practical channel estimators. Our comparison is also made with the existing channel estimators such as the conventional LMMSE estimator [8], the OMP [9] and block-OMP [5].

Fig. 2 shows the MSE performances of the channel estimators as a function of pilot overhead, which is the ratio of the pilot resources to the whole OFDM resources. Recall that the main goal of our pilot allocation is to minimize the pilot overhead, while not degrading the channel estimation performance. In Fig. 2, we observe that the MSE of the channel estimators increases as we reduce the pilot density. Note that the MSE goes up rapidly if we reduce pilot overhead below the certain level. With $L = 7$, the proposed channel estimator achieves 10% of pilot overhead achieving the MSE below -25 dB. The MSE performance of the proposed scheme is close to that of the Oracle-based scheme as long as the pilot overhead is larger than 10%. On the other hand, the conventional LMMSE estimator exhibits the poor performance because it lacks the mechanism which exploits the sparse nature of the CIR vector. Compared to the conventional LMMSE estimator, both the OMP and block-OMP schemes perform better, but they achieves only 45% and 38% of pilot overhead to provide the MSE lower than -25 dB. This is because the OMP fails to use the common sparsity while the block-OMP does not use the temporal correlation of the channel gains. From Fig. 2, we observe that if the number of available pilot-containing OFDM symbols increases in a local block (i.e., as L becomes larger), the proposed scheme performs better. Lastly, we note that any of the above channel estimators would not perform well if pilots are uniformly placed in the time and frequency resource grid. This is because, with the uniform allocation, the low pilot overhead leads to an ill-conditioned sensing matrix.

V. Conclusion

This letter studied the compressed sensing-based pilot reduction technique for massive MIMO-OFDM systems. We have demonstrated that the proposed pseudo-random pilot allocation combined with the new sparse channel estimator achieves much lower pilot overhead than the existing methods.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of BS antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Sep. 2006.

[3] 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.

[4] 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, Jun. 2014.

[5] 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.

[6] D. Eiwen, G. Taubock, F. Hlawatsch, H. Rauhut, and N. Czink, "Multichannel-compressive estimation of doubly selective channels in MIMO-OFDM systems: Exploiting and enhancing joint sparsity," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3082–3085

[7] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] M. K. Orzdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 2, pp. 18–48, 2nd Quart. 2007.

[9] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.