

Hybrid orthogonal and non-orthogonal pilot distribution based channel estimation in massive MIMO system

ZHANG Ruoyu, ZHAO Honglin*, ZHANG Jiayan, and JIA Shaobo

Communication Research Center, Harbin Institute of Technology, Harbin 150001, China

Abstract: How to obtain accurate channel state information (CSI) at the transmitter with less pilot overhead for frequency division duplexing (FDD) massive multiple-input multiple-output (MIMO) system is a challenging issue due to the large number of antennas. To reduce the overwhelming pilot overhead, a hybrid orthogonal and non-orthogonal pilot distribution at the base station (BS), which is a generalization of the existing pilot distribution scheme, is proposed by exploiting the common sparsity of channel due to the compact antenna arrangement. Then the block sparsity for antennas with hybrid pilot distribution is derived respectively and can be used to obtain channel impulse response. By employing the theoretical analysis of block sparse recovery, the total coherence criterion is proposed to optimize the sensing matrix composed by orthogonal pilots. Due to the huge complexity of optimal pilot acquisition, a genetic algorithm based pilot allocation (GAPA) algorithm is proposed to acquire optimal pilot distribution locations with fast convergence. Furthermore, the Cramer Rao lower bound is derived for non-orthogonal pilot-based channel estimation and can be asymptotically approached by the prior support set, especially when the optimized pilot is employed.

Keywords: massive multiple-input multiple-output (MIMO), frequency division duplexing (FDD), compressed sensing, hybrid pilot distribution, genetic algorithm based pilot allocation (GAPA).

DOI: 10.21629/JSEE.2018.05.01

1. Introduction

Massive multiple-input multiple-output (MIMO) can provide huge gain on system capacity and spectrum efficiency due to large number of antennas and has drawn great interest among academic and industrial community [1]. The large antenna arrays not only provide huge array gains, but also bring about the simplicity to perform signal detection and other signal processing compared with conventional MIMO system. Such system has the potential to average

the fast fading of channel coefficients and almost eliminate intra-cell interference [2]. Thus, massive MIMO system is emerged as promising technology and is recognized as a key technique for the next generation mobile cellular system [3].

In order to achieve the merits of massive MIMO system due to large spatial degrees of freedom, including precoding and beamforming technology, the base station (BS) is required to obtain accurate channel state information (CSI) [1–4]. However, with the increase of the number of BS antennas, the corresponding downlink pilot overhead increases linearly with the antenna number and occupies a large amount of radio resources when using classical least squares (LS), especially in frequency division duplexing (FDD) system where uplink and downlink utilizes different frequency bands [5]. Thus, to avoid the difficulty, a straightforward way is to employ time division duplexing (TDD) operation to acquire CSI by utilizing uplink pilots and channel reciprocity [6]. In spite of this, the FDD mode is more appropriate for symmetric traffic and low latency applications [7]. Considering the backward compatibility with cellular network nowadays [8], it is of importance to deal with channel estimation for FDD massive MIMO system with low pilot overhead.

Among the existing researches, channel sparsity [9–11], which is an important feature of wireless channel, has been spotted to deal with the above challenge under the framework of compressed sensing (CS) [12–14]. Under the condition that the channel is sparse, CS has been applied to enhance the estimate sparse channel in massive MIMO to reduce pilot overhead. In the existing literature, CS was naively used in CSI estimation by using the orthogonal matching pursuit (OMP) [15] algorithm and the OMP-based [16,17] algorithm, achieving better estimation performance than the conventional LS method. Due to the limited local scatterers around the BS, the channel matrices in angular domain tend to be sparse as the increasing of the number of antennas [18], a distributed CS recovery

Manuscript received May 27, 2017.

*Corresponding author.

This work was supported by the National Natural Science Foundation of China (61671176; 61671173) and the Fundamental Research Funds for the Center Universities (HIT.MKSTISP.2016 13).

exploiting common sparsity between users was proposed in [19], which leveraged the joint sparsity of the channel matrix and jointly recovered the CSI at the BS. In [20], an open-loop and closed-loop channel estimation scheme was proposed for massive MIMO by utilizing the long-term channel statistics. Shen et al. investigated the partial support information of sparse channels to achieve further overhead reduction and analyzed channel vector with a set of significant elements using the l_1 weighted minimization framework [21]. The aforementioned researches focus on flat fading channel, while for frequency selective fading channel, the sparse channel estimation is even more challenging.

In order to deal with frequency selective fading channel, the channel sparsity in delay-domain is fully exploited to improve channel estimation performance, where channel impulse response (CIR) only contains limited significant taps and accounts for dominant energy of wireless multipath channel. This characteristic was firstly exploited in [9], where the sparse underwater acoustic channel was analyzed and modelled. [10] was based on the framework of structured compressive sensing (SCS) to jointly exploit the sparsity of wireless channels and the property that path delays vary much slower than path attenuation. Barbotin et al. demonstrated that the CIR between transmit and receive link has similar path delays [11], which results in common support in MIMO channels under certain conditions. The common support in MIMO channels was also investigated in uplink channel. The estimation of CIR for uplink massive MIMO system was discussed in [22], where it is assumed that the antennas and the adjacent antennas share the same channel support. It can be true with respect to the number of antennas, time difference of arrival of signal to adjacent antennas, and the variation of operation frequency, etc. The antenna array at the BS is spatial invariant for the CIRs, which means the gains of the channel taps are distinct but the locations of the most significant amplitudes remain unchanged. This is referred to the common sparsity in spatial correlation among antenna arrays. Basing on the common sparsity of wireless channel in BS antennas, then the sparse channel estimation for each transmit-receive link is transmitted to 'block' sparse signal recovery [23–25], where the required number of pilot can be reduced significantly. In [26], the joint channel estimation was based on rearranging the CIRs at the receiver to a block sparse model and outperformed the individual recovery, in which different antennas at the BS utilize frequency orthogonal pilot. By this way, users can identify which antenna is the source of pilot sequence. Choi et al. also addressed downlink pilot reduction for massive MIMO system by exploiting the consecutive CIR during certain time duration [27], where the location of pilots is random in the

subcarriers of orthogonal frequency division multiplexing (OFDM) symbol and changes over the time domain. However, the aforementioned works fail to take the pilot allocation into consideration.

Pilot allocation is an important issue to the sparse channel estimation of MIMO channels due to the fact that the sensing matrix is consisted by pilot sequences. Unlike the conventional pilot-assisted channel estimation scheme using least square (LS) or minimum mean square error (MMSE) [28], in which equidistant distributed pilot is optimal, CS-based sparse channel estimation highlights the randomness [12,27]. Restricted isometry property (RIP) is the fundamental concept to evaluate the asymptotic orthogonal property of the sensing matrix, which is hard to be verified in polynomial time. An alternative evaluation is to minimize the mutual coherence of the sensing matrix. The cyclic different set theory [29] is introduced to construct optimal sensing matrices, but only matrices with specific rows and columns can satisfy the Welch bound. In order to adapt the general case of pilot design, pilot pattern optimization schemes based on discrete stochastic approximation [26], and cross-entropy optimization (CEO) [30] have been proposed, where the pilot positions are randomly searched. Recently in [31], a genetic algorithm (GA) based algorithm was proposed to design pilot locations and extended to MIMO-OFDM system to provide optimal pilot patterns under the framework of distributed CS [14]. However, due to the fact that each antenna is equipped with completely orthogonal pilot in the frequency domain, when the number of BS antennas tends to infinity, the available frequency resource becomes intractable. On the other hand, in [32], a non-orthogonal pilot distribution scheme was proposed to reduce the pilot overhead under the theory of structured CS, where spatio-temporal common sparsity of delay-domain massive MIMO channels were capitalized on using adaptive structured subspace pursuit (ASSP) algorithm. However, the number of pilot subcarriers required increases with the number of BS antennas.

In this paper, to reduce the overwhelming pilot overhead, a hybrid orthogonal and non-orthogonal pilot distribution at BS is proposed by exploiting the common sparsity of channel due to the compact antenna arrangement, which is a generalization of the existing pilot distribution scheme. Then the block sparsity for antennas with hybrid orthogonal and non-orthogonal pilot distribution is derived respectively and can be used to obtain CIRs. By employing the theoretical analysis of block sparse recovery, the total coherence criterion is proposed to optimize the sensing matrix composed by orthogonal pilots. Due to the huge complexity of optimal pilot acquisition, a genetic algorithm based pilot allocation (GAPA) algorithm is proposed to acquire optimal pilot distribution locations with

fast convergence. Furthermore, for the channel estimation with non-orthogonal pilots, the Cramer Rao lower bound is derived and can be asymptotically approached by the prior support set when the optimized pilot is employed.

Notations: Vectors and matrices are written in lower-case and upper-case boldface, respectively; $|\cdot|$ denotes the cardinality of a set, while $\|\cdot\|_p$ and $\|\cdot\|_F$ denote the l_p -norm and Frobenius-norm, respectively. The matrix transpose and conjugate transpose are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively, while $(\cdot)^\dagger$ is the Moore-Penrose inversion. For a given matrix \mathbf{A} and vector \mathbf{a} , $\mathbf{A}(l)$ and $\mathbf{a}(l)$ denote the l th column vector of \mathbf{A} and the l th entry of \mathbf{a} , respectively. $\text{supp}(\mathbf{a})$ denotes the index set of non-zero entries of vector \mathbf{a} . The spectral norm of matrix \mathbf{A} , $\rho(\mathbf{A}) = \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})}$ denotes the largest eigenvalue of positive-semidefinite matrix $\mathbf{A}^H \mathbf{A}$, $\text{Tr}(\cdot)$ denotes the trace of matrix.

2. Channel model

Consider an OFDM based massive MIMO system, which consists one BS with N_T transmit antennas. The frequency selective fading channel between the i th transmit antenna and the user can be represented by CIR:

$$h_i(n) = \sum_{l=1}^L a_i(l) \delta(n-l), \quad i = 1, 2, \dots, N_T \quad (1)$$

where L denotes the number of equivalent channel taps, $a_i(l)$ is the channel gain of the l th tap for the i th antenna. For OFDM-based signal transmission, wireless channel between the BS and the user has been illustrated with inherent sparsity in the time delay domain due to the fact that only a few significant paths account for dominant energy of channel with large delay spread. This leads to only K nonzero entries in channel vectors. The support set of non-zero elements for channel $\mathbf{h}_i = [h_i(0), h_i(1), \dots, h_i(L)]$ between the i th antenna at the BS and the user is denoted as

$$D_i = \text{supp}(\mathbf{h}_i) = \{l : |\mathbf{h}_i(l)| \neq 0, 1 \leq l \leq L\}. \quad (2)$$

The sparsity level for channel of the i th antenna is $K_i = |D_i| \ll L$. Moreover, for BS with large number of antennas, CIRs between different transmit antennas and users show correlation under certain conditions. Specifically, sparse channel among different antenna elements exhibits common support due to almost same scatterers shared by transmit-receive antenna pairs, which is mainly controlled by maximum distance of antenna array d_{\max} , signal bandwidth BW , carrier frequency f_c . In [11], it is suggested that the maximum time difference of arrival (TDOA) of signal for antennas at the BS be larger than $\frac{1}{10BW}$, we can say that these channel taps corresponding to antennas can be resolvable. Accordingly, for an antenna

array with maximum distance of antenna array d_{\max} , the maximum TDOA is d_{\max}/c , where c is the light speed, if

$$\frac{d_{\max}}{c} \leq \frac{1}{10BW} \quad (3)$$

holds, then the antenna array can be interpreted as spatially invariant for support of CIR [22].

Define $d_\Delta = \frac{c}{10BW}$ as the marginal spacing, which determines whether the antenna array is spatially invariant or not. Specifically, if $d_{\max} \leq d_\Delta$, among all the the CIRs \mathbf{h}_i ($i = 1, 2, \dots, N_T$) antennas at the BS share the same sparsity pattern, that is

$$D_1 = D_2 = \dots = D_i = \dots = D_{N_T} = D, \quad |D| = K \quad (4)$$

which indicates that the support sets of channel corresponding to transmit antennas remain unchanged, while the complex path gains are different for each transmission channel. The antenna array distribution at the BS side can be linear, planar and cylindrical, etc. It can be noticed that maximum distance of antenna array d_{\max} is relevant with antenna distribution and the number of BS antennas, for convenience, the planar (square or rectangular) antenna array with uniform linear array (ULA) is considered in this paper. Fig. 1 illustrates the distance of antenna for a square antenna array, which is the function of the carrier frequency and the number of antennas. The marginal plane perpendicular with the axis of distance of antenna stands for the marginal spacing of antenna array d_Δ when signal bandwidth $BW = 30$ MHz.

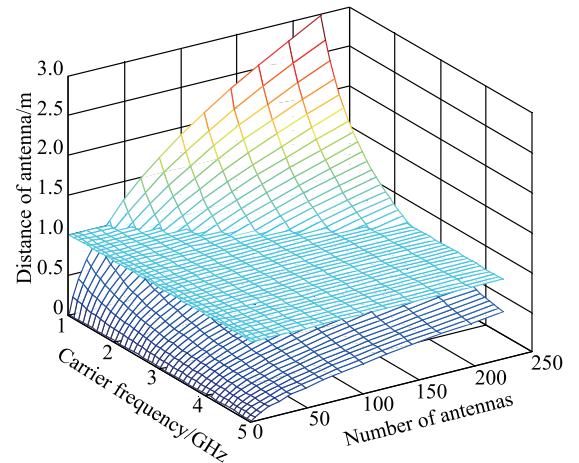


Fig. 1 Diagram of antenna distance of a square antenna array with carrier frequency and number of antennas

According to Fig. 1, the distance of antennas decreases with the carrier frequency and increases with the number of antennas. The marginal plane perpendicular with the axis of distance of antenna stands for the marginal spacing of antenna array d_Δ when signal bandwidth $BW = 30$ MHz. This means when the distance of antenna arrays is

below the plane, the support sets of channel among all the transmit antennas remain stable for a certain time. For instance, a system working on $f_c = 2$ GHz and signal bandwidth $BW = 30$ MHz, the wave length of signal is $\lambda = c/f_c = 0.15$ m and the marginal spacing $d_\Delta = \frac{c}{10BW} = 1$ m, which guarantees the invariability of channel support. For an antenna array of ULA with 64 antennas, which is shown in Fig. 2, the adjacent antenna spacing is $d = \lambda/2 = 0.075$ m, the farthest distance is $7\sqrt{2}d = 0.7425 < d_\Delta$. In this case, the antenna array is regarded as spatially invariant for support of CIR. When the antenna number increases to 144 with $f_c = 2$ GHz and $BW = 30$ MHz, then the farthest distance of antenna array is $10\sqrt{2}d = 1.0607 > d_\Delta$, which means the antenna arrays is spatially variant. This paper considers the channel estimation of spatially invariant antenna arrays at the BS, while Fig. 2 is an example of spatially invariant antenna arrays and corresponding CIRs from all the antennas to the user are shown in Fig. 3.

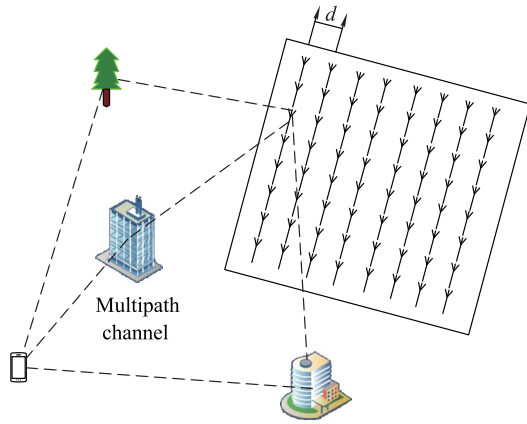


Fig. 2 Signal transmission between the BS and the user, which has limited scatterers

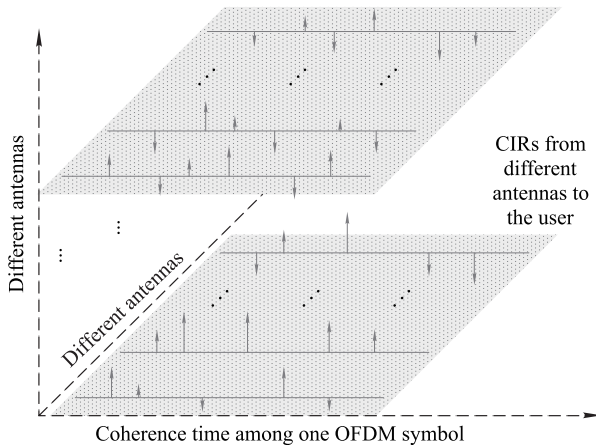
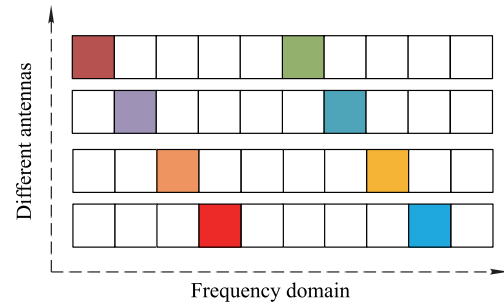


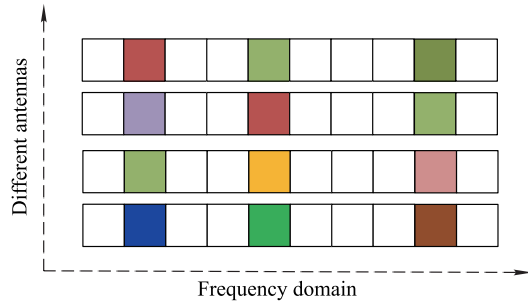
Fig. 3 Sparse channel with common support among the BS antennas, where CIRs among different BS antennas share the common sparsity

3. Hybrid pilot distribution at the BS

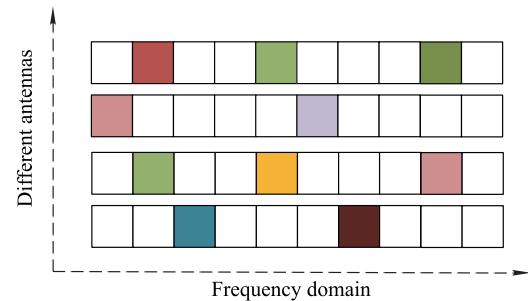
For sparse channel estimation, in order to guarantee the stable recovery of channel gains and support, a popular way is to obtain the sensing matrix by evaluating mutual coherence. The smaller mutual coherence indicates that the sensing matrix satisfies the RIP with higher probability. The orthogonal pilots mean the different transmit antennas are distributed to pilot carriers with totally different positions, which can be illustrated in Fig. 4(a). For massive MIMO system with a large number of antennas, it would be impossible to arrange such a large amount of orthogonal pilots. On the other hand, non-orthogonal pilot scheme, which allows different BS antennas to occupy the same pilot subcarriers, is proposed to accomplish the high spectral efficiency. This kind of pilot distribution is shown in Fig. 4(b).



(a) Traditional orthogonal pilot distribution in the frequency domain for different antennas



(b) Non-orthogonal pilot distribution in the frequency domain for different antennas



(c) Proposed hybrid pilot distribution for different antennas

Fig. 4 Pilot distribution

While in this paper, a hybrid pilot distribution with non-orthogonal and orthogonal pilot distribution in the frequency domain is proposed to achieve the balance between spectral efficiency and pilot overhead. This can be illustrated in Fig. 4(c). Particularly, part of BS antennas are allocated with frequency orthogonal pilot, while the other part of BS antennas are assigned with non-orthogonal pilot sequences among one OFDM symbol. For a system with N_T antennas at the BS, the number of antennas equipped with orthogonal pilot is M_O and the ratio of antennas with orthogonal pilot is defined as $\alpha = M_O/N_T$, $\alpha \in [0, 1]$, where the index of antennas with orthogonal pilot is $\Omega_O \subset \{1, 2, \dots, N_T\}$ and $M_O = |\Omega_O|$. Accordingly, the index of antennas with non-orthogonal pilot is $\Omega_N = \{1, 2, \dots, N_T\} \setminus \Omega_O$ and $M_N = |\Omega_N|$ satisfying $M_N + M_O = N_T$. For non-orthogonal pilot part, the index set of pilot subcarriers is uniquely selected from the carrier set of OFDM symbols and all the antennas in Ω_N share the same pilot position, i.e. $\zeta_{\Omega_N(1)} = \dots = \zeta_{\Omega_N(i)} = \dots = \zeta_{\Omega_N(M_N)} = \zeta_N$, $i \in \Omega_N$. For the part of orthogonal pilot, the index set should be selected from $\zeta_j \subset \{1, 2, 3, \dots, N\} \setminus \zeta_N$, which satisfies $\zeta_j \cap \zeta_{j'} = \emptyset$ for $j, j' \in \Omega_O$. N is the number of subcarriers in one OFDM symbol. It should be noted that the hybrid pilot distribution in this paper is a generalization of pilot allocation in Fig. 4(a) and Fig. 4(b). When $\alpha = 0$, the hybrid pilot distribution reduces to the case in [32], where all the antennas are equipped with non-orthogonal pilots, while $\alpha = 1$ means all the pilots are orthogonal among all the antennas in the frequency domain [26]. Therefore, for the hybrid pilot distribution, how to perform channel estimation at the receiver is a challenging issue since the hybrid architecture of pilot distribution. At the same time, how to guarantee the optimal channel estimation performance with the hybrid pilot allocation is also a challenging issue, so an optimal pilot allocation scheme is proposed in this paper.

4. Channel estimation

Consider an OFDM based massive MIMO system, which consists one BS with N_T transmit antennas. Each antenna transmits OFDM symbols with N subcarriers, which can be denoted as $\mathbf{X}_i = [x_i(0), x_i(1), \dots, x_i(N-1)]^T$ with $i = 0, 1, 2, \dots, N_T$. Accordingly, the pilot set of each OFDM symbol for the i th transmit antenna is ζ_i , where $\zeta_i \subset \{1, 2, 3, \dots, N\}$. It should be noted that for antennas with orthogonal pilot position, N_{PO} is the number of subcarriers occupied by pilot sequences. $N_{PO} = |\zeta_i|$ for $i \in \Omega_O$, while for antennas equipped with non-orthogonal pilot position, the number of subcarriers is $N_{PN} = |\zeta_i|$ for $i \in \{1, 2, \dots, N_T\} \setminus \Omega_O$. Then, after the inverse fast

Fourier transform (IFFT) operation, the equivalent time-domain signal can be denoted as

$$\mathbf{x}_i = \mathbf{F}^H \mathbf{X}_i \quad (5)$$

where $\mathbf{F} \in \mathbf{C}^{N \times N}$ is the discrete Fourier transform (DFT) matrix with (k, l) element $\frac{1}{\sqrt{N}} \omega^{(k-1)(l-1)}$, $\omega = e^{-j2\pi/N}$, then the cyclic prefix (CP) is added to the time-domain signal to avoid the inter symbol interference (ISI), whose length is no less than channel length L .

At the receiver, after the removal of CP and fast Fourier transform (FFT) operation, the received signal at the user side can be denoted as

$$\mathbf{r} = \sum_{i=1}^{N_T} \text{diag}(\mathbf{X}_i) \mathbf{F} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{0}_{(N-L) \times 1} \end{bmatrix} + \mathbf{w} \quad (6)$$

where $\mathbf{w} \in \mathbf{C}^{N \times 1}$ is the additive white Gaussian noise vector, $\mathbf{w} \sim N(0, \sigma^2 \mathbf{I})$, \mathbf{h}_i is the sparse CIR corresponding to the i th antenna with the user. By extracting the pilot vector according to the index set ζ_i , then the received pilot signal from BS can be expressed by

$$\mathbf{r}_p = \sum_{i=1}^{N_T} \text{diag}(\mathbf{X}_i(\zeta_i)) \mathbf{F}_{|\zeta_i} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{0}_{(N-L) \times 1} \end{bmatrix} + \mathbf{w}_p = \sum_{i=1}^{N_T} \mathbf{D}_{\zeta_i} \mathbf{F}_{L|\zeta_i} \mathbf{h}_i + \mathbf{w}_p \quad (7)$$

where \mathbf{D}_{ζ_i} is a diagonal matrix consisted of the corresponding pilot sequence for antenna i , $\mathbf{F}_{|\zeta_i}$ is the submatrix of \mathbf{F} indexed by ζ_i in rows, $\mathbf{F}_{L|\zeta_i}$ is the partial FFT matrix with first L columns of $\mathbf{F}_{|\zeta_i}$, \mathbf{w}_p is the noise vector of pilot sequences.

The received non-orthogonal pilot signals are mixed together at the user since different BS antennas occupy the completely same subcarrier, which can be denoted as

$$\mathbf{r}_{pN} = \sum_{i=1}^{M_N} \text{diag}(\mathbf{X}_i(\zeta_i)) \mathbf{F}_{L|\zeta_N} \mathbf{h}_i + \mathbf{w}_{pN}, \quad i \in \Omega_N. \quad (8)$$

For the signal form in (8), it can be reformulated as a block signal recovery, where the pilot is uniformly distributed in the subcarriers of OFDM symbols. Then denote $\Phi_i = \text{diag}(\mathbf{X}_i(\zeta_i)) \mathbf{F}_{L|\zeta_N}$, we can have

$$\mathbf{r}_{pN} = \Phi_N \mathbf{h}_N + \mathbf{w}_{pN} \quad (9)$$

where $\Phi_N = [\Phi_1, \Phi_2, \dots, \Phi_{M_N}] \in \mathbf{C}^{N_{PN} \times M_N L}$, $\mathbf{h}_N = [\mathbf{h}_1^T, \mathbf{h}_2^T, \dots, \mathbf{h}_{M_N}^T]^T \in \mathbf{C}^{M_N L \times 1}$ with $i \in \Omega_N$. Extract the element in \mathbf{h}_N with interval L to constitute a new vector $\mathbf{d}_l \in \mathbf{C}^{M_N}$, so an equivalent block vector $\mathbf{d}_N = [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_L^T, \dots, \mathbf{d}_L^T]^T$ is obtained by

$$\mathbf{d}_l = [\mathbf{h}_1^T(l), \mathbf{h}_2^T(l), \dots, \mathbf{h}_{M_N}^T(l)]^T \in \mathbf{C}^{M_N \times 1} \quad (10)$$

which is either all zero or all non-zero with $l = 1, 2, \dots, L$. Accordingly, the sensing matrix can be obtained as

$$\Psi_N = [\mathbf{T}_1^T, \mathbf{T}_2^T, \dots, \mathbf{T}_L^T]^T \quad (11)$$

where

$$\mathbf{T}_l = [\Phi_1(l), \Phi_2(l), \dots, \Phi_{M_N}(l)]. \quad (12)$$

Then (9) can be represented as

$$\mathbf{r}_{pN} = \Psi_N \mathbf{d}_N + \mathbf{w}_{pN}. \quad (13)$$

This means we should recover the equivalent channel \mathbf{d}_N with dimension $M_N L$ from non-orthogonal pilot with dimension N_{PN} . In order to achieve the stable recovery, the required number of pilot should be $N_{PN} \propto M_N K \lg(L/K)$ [23], which is proposition with the number of antennas with non-orthogonal pilots. Furthermore,

$$\mathbf{A}^{(i)} = \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{D}_{\zeta_i}(1) \omega^{\zeta_i(1) \cdot 0} & \dots & \mathbf{D}_{\zeta_i}(1) \omega^{\zeta_i(1) \cdot (L-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{\zeta_i}(N_{PO}) \omega^{\zeta_i(N_{PO}) \cdot 0} & \dots & \mathbf{D}_{\zeta_i}(N_{PO}) \omega^{\zeta_i(N_{PO}) \cdot (L-1)} \end{bmatrix}. \quad (15)$$

A natural approach to estimate channel in (14) is utilizing the CS algorithm, which can be formulated as

$$\min_{\mathbf{h}_i} \|\mathbf{h}_i\|_0, \quad \text{s.t. } \mathbf{r}_p^{(i)} = \mathbf{A}^{(i)} \mathbf{h}_i + \mathbf{w}_p^{(i)}, \quad i \in \Omega_O. \quad (16)$$

For antennas with orthogonal pilots, the downlink channel can be distinguished from different transmit antennas. The trivial CS reconstruction algorithm can be used to address the above problem, i.e. OMP, subspace pursuit (SP), basis pursuit denoising (BPDN) [13]. Also, this problem

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_p^{(1)} \\ \mathbf{r}_p^{(2)} \\ \vdots \\ \mathbf{r}_p^{(M_O)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{0}_{N_{PO} \times L} \\ \mathbf{0}_{N_{PO} \times L} & \mathbf{A}^{(2)} \\ \vdots & \vdots \\ \mathbf{0}_{N_{PO} \times L} & \mathbf{0}_{N_{PO} \times L} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{M_O} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_p^{(1)} \\ \mathbf{w}_p^{(2)} \\ \vdots \\ \mathbf{w}_p^{(M_O)} \end{bmatrix}. \quad (17)$$

In order to acquire the block form of unknown CIRs, it is important to further perform the operation on the sensing matrix, which is the priority for the receiver since the pilot is shared by transmitter and receiver in advance. Furthermore, by rearranging the block diagonal sensing matrix as

$$\mathbf{B} = [\mathbf{B}_1 \quad \dots \quad \mathbf{B}_l \quad \dots \quad \mathbf{B}_L] \in \mathbb{C}^{N_{PO} M_O \times L M_O} \quad (18)$$

where

$$\mathbf{B}_l = \begin{bmatrix} \mathbf{a}_{1,l} & \mathbf{0}_{N_{PO} \times 1} & \dots & \mathbf{0}_{N_{PO} \times 1} \\ \mathbf{0}_{N_{PO} \times 1} & \mathbf{a}_{2,l} & \dots & \mathbf{0}_{N_{PO} \times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{N_{PO} \times 1} & \mathbf{0}_{N_{PO} \times 1} & \dots & \mathbf{a}_{M_O,l} \end{bmatrix} \quad (19)$$

to achieve stable recovery of (13), N_{PN} should be slightly large to maintain the small cross-correlation of columns within Ψ_N [32], which incurs a large pilot overhead N_{PN} for each non-orthogonal antenna. Instead of using traditional CS recovery, the common support motivates us to further exploit the channel estimation using orthogonal pilot.

For the set of antennas allocated with orthogonal pilot, i.e., $\Omega_O \subset \{1, 2, \dots, N_T\}$, the received orthogonal pilot signal can be given by

$$\mathbf{r}_p^{(i)} = \mathbf{D}_{\zeta_i} \mathbf{F}_{L|\zeta_i} \mathbf{h}_i + \mathbf{w}_p^{(i)} = \mathbf{A}^{(i)} \mathbf{h}_i + \mathbf{w}_p^{(i)}, \quad i \in \Omega_O \quad (14)$$

where $\mathbf{A}^{(i)}$ can be expressed as

can be modelled as distributed CS to simultaneously reconstruct a set of joint sparse signals. In this case, (16) can be rearranged as a problem with M_O vectors. Thus the CIR vectors \mathbf{h}_i ($i \in \Omega_O$) can be modeled as JSM-2 problem and be solved using the DCS-SOMP algorithm [31]. For the aim of pilot optimization and exploiting the common sparsity, (16) can be also reformulated into a block signal recovery problem. To exploit the block property among the CIRs of different antennas with orthogonal pilots, the following manipulations are operated:

where $\mathbf{a}_{i,l} \in \mathbb{C}^{N_{PO} \times 1}$ is the l th column of sensing matrix $\mathbf{A}^{(i)}$, which is illustrated clearly in (15). Accordingly, the unknown aggregating channel vectors can be rearranged as the equivalent block CIR vector \mathbf{g} :

$$\mathbf{g} = [\mathbf{g}_1^T \quad \dots \quad \mathbf{g}_l^T \quad \dots \quad \mathbf{g}_L^T]^T \quad (20)$$

where $\mathbf{g}_l = [\mathbf{h}_1(l) \quad \mathbf{h}_2(l) \quad \dots \quad \mathbf{h}_{M_O}(l)]^T$. It should be noted the value of \mathbf{g}_l is either all zero or all non-zero, which results in the equivalent vector \mathbf{g} existing inherent block property due to the common sparsity. Then (17) can be transformed to

$$\mathbf{R} = \mathbf{B} \mathbf{g} + \mathbf{w}. \quad (21)$$

Then for the proposed hybrid pilot distribution, we first recover the CIRs from antennas equipped with orthogonal pilots. The tools to solve this block recovery problem include convex optimization algorithms, like BPDN algorithms with l_1 constraints, and greedy algorithms. Greedy algorithms include block OMP (BOMP) [23], ASSP [32], Block StOMP [33], which are based on the well-known OMP for their fast operation and easy implementation. In terms of the channel estimation, these algorithms have similar recovery performance. However, ASSP, Block StOMP fail to have integrated theoretical analysis. Particularly, for the aim of pilot allocation based on the mutual coherence, the BOMP is employed for its fast operation and easy implementation, especially for a large number of antennas in massive MIMO system. In this case, the Block OMP algorithm for recovering CIRs from antennas with orthogonal pilots is elaborated as below.

Algorithm 1 Block OMP for recovering CIRs from antennas with orthogonal pilots

Input: \mathbf{R} , $\mathbf{A}^{(i)}$, $i = 1, 2, \dots, M_O$

Step 1 (Initialization) Compute equivalent sensing matrix \mathbf{B} according to (18) and (19). Initial residual $\mathbf{R}^{(0)} = \mathbf{R}$, support set $\Lambda^{(0)} = \emptyset$, iteration index $s = 1$.

Step 2 (Correlation operation)

$$\mathbf{C} = \mathbf{B}^H \mathbf{R}^{(s-1)} = [\mathbf{C}_1 \quad \mathbf{C}_2 \quad \dots \quad \mathbf{C}_L].$$

Step 3 (Support estimate) Choose the index number of largest l_2 -norm of correlation block: $\lambda^{(s)} = \arg \max_{1 \leq l \leq L} \|\mathbf{C}_l\|_2^2$.

Step 4 (Support update) $\Lambda^{(s)} = \Lambda^{(s-1)} \cup \lambda^{(s)}$.

Step 5 (Block signal estimate) $\hat{\mathbf{g}}_{\Lambda^{(s)}} = \mathbf{B}_{\Lambda^{(s)}}^\dagger \mathbf{R}$.

Step 6 (Residue update) $\mathbf{R}^{(s)} = \mathbf{R} - \mathbf{B}_{\Lambda^{(s)}} \hat{\mathbf{g}}_{\Lambda^{(s)}}$.

Step 7 If the stop condition is satisfied $\bar{\mathbf{D}} = \Lambda^{(s)}$, go to Step 8;

else

Back to Step 2;

end

Step 8 (Obtain channel)

The equivalent block channel

$$\hat{\mathbf{g}} = [\hat{\mathbf{g}}_1^T \quad \dots \quad \hat{\mathbf{g}}_l^T \quad \dots \quad \hat{\mathbf{g}}_L^T]^T$$

is denoted as

$$\hat{\mathbf{g}}_l^T = \begin{cases} \hat{\mathbf{g}}_l, & l \in \bar{\mathbf{D}} \\ \mathbf{0}_{M_O \times 1}, & l \notin \bar{\mathbf{D}} \end{cases}, 1 \leq l \leq L.$$

Then, the estimated CIRs from different antennas can be obtained according to (20).

Output: estimated channel $\hat{\mathbf{h}}_i$.

Compared with traditional block based OMP in [24], in which all the subcarriers in OFDM symbol are dedicated to

pilot carriers, the Block OMP algorithm here only exploits the part of subcarriers of OFDM symbol. For the above algorithm, some notations should be explained. Correlation vector \mathbf{C} between the residual and the sensing matrix is considered as containing L sub-vectors. In Step 5, $\mathbf{B}_{\Lambda^{(s)}}^\dagger$ is the pseudo inversion of matrix $\mathbf{B}_{\Lambda^{(s)}}$, which only contains sub-block \mathbf{B}_l whose index is $l \in \Lambda^{(s)}$. $\hat{\mathbf{g}}_{\Lambda^{(s)}}$ is the estimate of unknown block vectors consisting of sub-block \mathbf{g}_l where $l \in \Lambda^{(s)}$. That is, block signal estimate in Step 5

$$\text{is equivalent with } \hat{\mathbf{g}}_{\Lambda^{(s)}} = \arg \min_{\hat{\mathbf{g}}_l^s, 1 \leq l \leq L} \left\| \mathbf{R} - \sum_{l \in \Lambda^{(s)}} \mathbf{B}_l \mathbf{g}_l \right\|_2^2.$$

In Step 6, residue update, $\mathbf{B}_{\Lambda^{(s)}}$ is the matrix containing sub-block \mathbf{B}_l whose index is $l \in \Lambda^{(s)}$, which is the same as that in Step 5.

The stop criterion of the above algorithm is similar to the original BOMP. If the receiver contains the priority knowledge of sparsity of wireless channels, the stop condition is the sparsity level, which is a popular stop criterion in the field of sparse channel estimation. On the other hand, the stop criterion can be $\|\mathbf{R}^{(s)}\| \leq \|\mathbf{w}\|$ if the variance of noise is known. After the channel estimation between antennas with orthogonal pilots, the CIRs estimation for antennas equipped with non-orthogonal pilots can be obtained based on this prior support information of wireless channels due to common sparsity illustrated by (4). It should be noted that the CIRs for non-orthogonal pilot antennas remained to be estimated and can be estimated by the algorithm in [32]. Here an alternative scheme is proposed on the channel information obtained by orthogonal pilots. Before the elaboration of the proposed scheme, the Cramer-Rao lower bound (CRLB) [34] is derived for (9).

Firstly, the equivalence between \mathbf{h}_N and \mathbf{d}_N in (9) can be denoted as $\mathbf{d}_N = \mathbf{P}^T \mathbf{h}_N$ with permutation matrix $\mathbf{P} \in \mathbf{C}^{M_N L \times M_N L}$, with the l th $\mathbf{P}_l = [\mathbf{e}_l, \mathbf{e}_{L+l}, \dots, \mathbf{e}_{(M_N-1)L+l}] \in \mathbf{C}^{M_N L \times M_N}$, which holds $\mathbf{I} = \mathbf{P} \mathbf{P}^T \in \mathbf{C}^{M_N L \times M_N L}$, \mathbf{e}_{L+l} means the $(L+l)$ th column of identity matrix $\mathbf{I}_{M_N L \times M_N L}$. Then

$$\begin{aligned} \text{var}\{\hat{\mathbf{d}}_N\} &= \text{E}\{\|\hat{\mathbf{d}}_N - \mathbf{d}_N\|_2^2\} = \text{E}\{\|\mathbf{P}^T \hat{\mathbf{h}}_N - \mathbf{P}^T \mathbf{h}_N\|_2^2\} = \\ &= \text{E}\{\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|_2^2\} = \text{var}\{\hat{\mathbf{h}}_N\}. \end{aligned} \quad (22)$$

For block sparse signal \mathbf{d}_N , once the support $\bar{\mathbf{D}}$ is obtained, then $\mathbf{r}_{pN} = (\boldsymbol{\Psi}_N)_{\bar{\mathbf{D}}}(\mathbf{d}_N)_{\bar{\mathbf{D}}} + \mathbf{w}_{pN}$ with $\mathbf{w}_{pN} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{N_{pN}})$. The probability density function (PDF) of \mathbf{r}_{pN} can be given as

$$\begin{aligned} p_{\mathbf{r}_{pN}|\mathbf{d}_N}(\mathbf{r}_{pN}; (\mathbf{d}_N)_{\bar{\mathbf{D}}}) &= \frac{1}{(2\pi\sigma^2)^{N_{pN}/2}} \cdot \\ &\exp\left(-\frac{1}{2\sigma^2} \|\mathbf{r}_{pN} - (\boldsymbol{\Psi}_N)_{\bar{\mathbf{D}}}(\mathbf{d}_N)_{\bar{\mathbf{D}}}\|_2^2\right). \end{aligned} \quad (23)$$

Note that the estimation of block signal within support set can be obtained by

$$(\hat{\mathbf{d}}_N)_{\bar{D}} = ((\boldsymbol{\Psi}_N)_{\bar{D}})^\dagger \mathbf{r}_{pN} \quad (24)$$

where

$$\begin{aligned} E\{(\hat{\mathbf{d}}_N)_{\bar{D}}\} &= E\{((\boldsymbol{\Psi}_N)_{\bar{D}})^\dagger \mathbf{r}_{pN}\} = \\ E\{((\boldsymbol{\Psi}_N)_{\bar{D}})^\dagger ((\boldsymbol{\Psi}_N)_{\bar{D}}(\mathbf{d}_N)_{\bar{D}} + \mathbf{w}_{pN})\} &= \\ (\mathbf{d}_N)_{\bar{D}} + E\{((\boldsymbol{\Psi}_N)_{\bar{D}})^\dagger \mathbf{w}_{pN}\} &= (\mathbf{d}_N)_{\bar{D}} \end{aligned} \quad (25)$$

which is the unbiased estimation of \mathbf{d}_N . Then the (m_i, m_j) th element of Fisher information matrix for parameters $(\mathbf{d}_N)_{\bar{D}}$ is obtained by

$$\begin{aligned} [\mathbf{J}((\hat{\mathbf{d}}_N)_{\bar{D}})]_{m_i, m_j} &= \\ -E\left\{\frac{\partial^2 \ln(p_{\mathbf{r}_{pN}|\mathbf{d}_N}(\mathbf{r}_{pN}; (\mathbf{d}_N)_{\bar{D}}))}{\partial d_{m_i} \partial d_{m_j}}\right\} &= \\ \frac{1}{\sigma^2} [((\boldsymbol{\Psi}_N)_{\bar{D}})^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}}]_{m_i, m_j} & \end{aligned} \quad (26)$$

where d_{m_i} and d_{m_j} are the m_i th and m_j th element of $(\mathbf{d}_N)_{\bar{D}}$. $1 \leq m_i, m_j \leq |\bar{D}| = M_N K$. Therefore, we have $\text{var}\{\hat{\mathbf{d}}_N\} \geq \text{Tr}(\mathbf{J}((\hat{\mathbf{d}}_N)_{\bar{D}})^{-1})$. Denote the $M_N K$ eigenvalues of $[(\boldsymbol{\Psi}_N)_{\bar{D}}]^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}}$ as $\lambda_1, \lambda_2, \dots, \lambda_{M_N K}$, we can obtain

$$\begin{aligned} \text{Tr}(((\boldsymbol{\Psi}_N)_{\bar{D}})^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}})^{-1}) &= \sum_{k=1}^{M_N K} \lambda_k^{-1} = \\ M_N K \left(\sum_{k=1}^{M_N K} \frac{\lambda_k^{-1}}{M_N K} \right) &\geq M_N K \left(\frac{M_N K}{\sum_{k=1}^{M_N K} \lambda_k} \right) = \\ \left(\frac{(M_N K)^2}{\text{Tr}(((\boldsymbol{\Psi}_N)_{\bar{D}})^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}})} \right) & \end{aligned} \quad (27)$$

where the equality in (27) holds if and only if $\{\lambda_k\}_{k=1}^{M_N K}$ are equal. In this case, to achieve the lower bound, the implication is that the matrix $((\boldsymbol{\Psi}_N)_{\bar{D}})^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}}$ should have identical diagonal elements. This means $\text{Tr}(((\boldsymbol{\Psi}_N)_{\bar{D}})^\text{H} (\boldsymbol{\Psi}_N)_{\bar{D}}) = M_N K N_{PN}$. The Cramer-Rao lower bound (CRLB) of \mathbf{h}_N can be expressed by

$$\begin{aligned} \text{var}\{\hat{\mathbf{h}}_N\} &= E\{\|\hat{\mathbf{h}}_N - \mathbf{h}_N\|_2^2\} = \\ \text{var}\{\hat{\mathbf{d}}_N\} &\geq \frac{M_N K \sigma^2}{N_{PN}} \end{aligned} \quad (28)$$

where N_{PN} is the number of non-orthogonal pilots. CRLB is a theoretical bound to evaluate the algorithms for estimating channel \mathbf{h}_N . It is noteworthy that the CRLB is

unachievable by practical channel estimator because the $(\boldsymbol{\Psi}_N)_{\bar{D}}$ does not contain orthogonal columns. At the same time, it also should be noted that the bound is computationally intractable to be achieved due to the fact the sparse recovery is an NP-hard problem [35]. The important implication behind (28) is that once the support of the aggregate channel vector is obtained, there exists an unbiased estimator to approach the CRLB of channel estimation, where the CRLB is related to the number of significant path and the number of non-orthogonal usage. Fortunately, the CRLB can be proximately approached if the support set of channel can be known as a priority. This can be true based on the support estimation by orthogonal pilots based on Algorithm 1. Particularly, the relationship between support \bar{D} and support \hat{D} can be denoted as

$$\begin{aligned} \bar{D} &= \bigcup_{k=1}^K \bar{D}(k), \\ \bar{D}(k) &= \{(\hat{D}(k) + 1, (M_N - 1) \cdot \hat{D}(k) + 2, \dots, M_N \hat{D}(k))\} \end{aligned} \quad (29)$$

where $\bar{D}(k)$ is the k th block in \bar{D} with M_N elements and $\hat{D}(k)$ is the k th element in support estimation \hat{D} . Upon this prior information of support set, unbiased estimator (24) can be used to acquire the CIRs for antennas equipped with non-orthogonal pilots. Overall, the CIRs estimation is summarized as in Algorithm 2.

Algorithm 2 Recovering CIRs for antennas with non-orthogonal pilots

Input: $\mathbf{r}_{pN}, \boldsymbol{\Psi}_N, \bar{D}$

Step 1 (Support acquisition) Calculate the support set for antennas with non-orthogonal pilots according to (29).

Step 2 (Block signal estimation) Block signal estimation within support set $(\hat{\mathbf{d}}_N)_{\bar{D}}$ is obtained by (24), block signal outside support set $(\hat{\mathbf{d}}_N)_{\{1,2,\dots,M_N L\} \setminus (\bar{D})} = \mathbf{0}$.

Step 3 (Obtain channel) The equivalent CIRs $\hat{\mathbf{h}}_N$ are obtained according to (10).

Output: CIRs for antennas with non-orthogonal pilots.

It should be noted that the channel estimation performance by non-orthogonal pilot is closely related to the support obtained by orthogonal pilots. This means the more exactly the support is obtained, the more accurate channel is characterized, which emphasizes the importance of correct support set. In order to accurately acquire the channel support, the pilot optimization algorithm is proposed for orthogonal pilots.

5. Pilot optimization

5.1 Problem formulation

This section presents the proposed pilot allocation optimization algorithm for antennas with orthogonal pilots. In

CS-based sparse channel estimation, pilot subcarriers are randomly selected and the partial Fourier matrix is constructed as the sensing matrix of channel estimation for each link between the antenna and the user. An alternative and popular evaluation criterion of the sensing matrix is the mutual coherence, which is more strict and feasible than the RIP. Furthermore, due to the hybrid pilot distribution, part of subcarriers have been occupied by the non-orthogonal pilots, it is even challenging to allocate the pilot position with the limited subcarriers to guarantee the stable recovery of channel.

For (15), $\mathbf{A}^{(i)}$ is the sensing matrix of channel estimation formula for antenna $i \in \Omega_O$. The mutual coherence of $\mathbf{A}^{(i)}$ is defined as

$$\mu(\mathbf{A}^{(i)}) = \max_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \frac{|(\mathbf{A}^{(i)}(l))^H \mathbf{A}^{(i)}(l')|}{\|\mathbf{A}^{(i)}(l)\| \|\mathbf{A}^{(i)}(l')\|} = \max_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \frac{|(\mathbf{a}_{i,l})^H \mathbf{a}_{i,l'}|}{\|\mathbf{a}_{i,l}\| \|\mathbf{a}_{i,l'}\|} \quad (30)$$

where $\mathbf{A}^{(i)}(l)$ is the l th column of sensing matrix $\mathbf{A}^{(i)}$. To achieve a lower bound of estimation error for \mathbf{h}_i , the mutual of coherence of sensing matrix $\mathbf{A}^{(i)}$ should be minimized, then the optimization problem is represented as

$$\min \mu(\mathbf{A}^{(i)}) = \min_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \max_{\substack{1 \leq i_p \leq N_{PO}}} \frac{\left| \sum_{i_p=1}^{N_{PO}} |\mathbf{D}_{\zeta_i}(i_p)|^2 \omega^{\zeta_i(i_p) \cdot (l'-l)} \right|}{\sum_{i_p=1}^{N_{PO}} |\mathbf{D}_{\zeta_i}(i_p)|^2}. \quad (31)$$

Due to the constant amplitude of pilot sequences with zero-correlation in OFDM pilot, it can be assumed that all the pilot symbols share the same amplitude, i.e., $|\mathbf{D}_{\zeta_i}(i_p)| = A$, $i \in \Omega_O$, $i_p = 1, 2, \dots, N_{PO}$. In light of this, from (31), we can know that $\mu(\mathbf{A}^{(i)})$ is completely determined by the position of pilot ζ_i . Then (31) can be simplified into

$$\min \mu(\mathbf{A}^{(i)}) = \min_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \max_{\substack{1 \leq i_p \leq N_{PO}}} \left| \sum_{i_p=1}^{N_{PO}} \frac{1}{N_{PO}} \omega^{\zeta_i(i_p) \cdot (l'-l)} \right| \quad (32)$$

which is equivalent with

$$\min_{\zeta_i} \max_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \left| \sum_{\lambda \in \zeta_i} \frac{1}{N_{PO}} \omega^{\lambda \cdot (l'-l)} \right|$$

$$\text{s.t. } |\zeta_i| = N_{PO}, \quad \zeta_i \subset \{1, 2, 3, \dots, N\} \setminus \zeta_N. \quad (33)$$

After the manipulation as (18)–(21), we obtain the block form equation due to spatial common sparsity. Consequently, the block coherence μ_B and sub-coherence v of equivalent sensing matrix in (21) are introduced according to [23]. Specifically, the block coherence μ_B is defined as

$$\mu_B = \frac{1}{M_O} \max_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \rho(\mathbf{B}_l^H \mathbf{B}_{l'}) \quad (34)$$

and sub-coherence v

$$v = \max_{1 \leq l \leq L} \max_{\substack{1 \leq i, j \leq L \\ i \neq j}} |(\mathbf{B}_l(i))^H \mathbf{B}_l(j)|. \quad (35)$$

The physical meaning of μ_B is the largest value of correlation between two different sub-blocks, while v stands for the maximum correlation of two different column atoms in the same sub-block. Upon the above definition, before the elaboration of the proposed pilot optimization algorithm, the error bound of BOMP is presented.

Theorem 1 [23] Given the block signal recovery as in (21), suppose $(1 - (M_O - 1)v)|\mathbf{g}_{\min}| > 2\sigma\sqrt{2\alpha M_O(1 - (M_O - 1)v)\lg(M_O L)} + (2K - 1) \cdot M_O \mu_B |\mathbf{g}_{\min}|$, with $\alpha \geq 1/M_O$ and $|\mathbf{g}_{\min}| = \min_{l \in \text{TrueSupport}} \|\mathbf{g}_l\|_2$, the error bound of the equivalent block vector can be given as

$$\|\hat{\mathbf{g}} - \mathbf{g}\|_2^2 \leq \frac{2\alpha(1 + (M_O - 1)v)}{(1 - (M_O - 1)v - (K - 1)M_O \mu_B)^2} \cdot M_O K \sigma^2 \lg(M_O L) \quad (36)$$

with overwhelming probability.

Remark Theorem 1 provides the theoretical upper bound for block sparse signal recovery with respect to the number of BS antennas, coherence of the sensing matrix consisting of pilot sequence, noise strength and channel equivalent length. Specifically, in this case, we observe that in (18) the column atom of the same sub-matrix of the sensing matrix is orthogonal with each other. Thus we have

$$v = \max_{1 \leq l \leq L} \max_{\substack{1 \leq i, j \leq L \\ i \neq j}} |(\mathbf{B}_l(i))^H \mathbf{B}_l(j)| = 0. \quad (37)$$

Then (36) can be simplified to

$$\|\hat{\mathbf{g}} - \mathbf{g}\|_2^2 \leq \frac{2\alpha}{(1 - (K - 1)M_O \mu_B)^2} M_O K \sigma^2 \lg(M_O L) \quad (38)$$

with $(1 - (2K - 1)M_O \mu_B)|\mathbf{g}_{\min}| > 2\sigma\sqrt{2\alpha M_O \lg(M_O L)}$. To guarantee the smaller error bound of recovery, it is imperative to minimize μ_B , since the lower μ_B provides

lower value of upper bound of (38). According to (34), we have

$$B_l^H B_{l'} = \begin{bmatrix} (\mathbf{a}_{1,l})^H \mathbf{a}_{1,l'} & 0 & \cdots & 0 \\ 0 & (\mathbf{a}_{2,l})^H \mathbf{a}_{2,l'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mathbf{a}_{M_O,l})^H \mathbf{a}_{M_O,l'} \end{bmatrix} \quad (39)$$

and

$$\rho(B_l^H B_{l'}) = \sqrt{\lambda_{\max}(\text{diag}(|(\mathbf{a}_{1,l})^H \mathbf{a}_{1,l'}|^2, \dots, |(\mathbf{a}_{M_O,l})^H \mathbf{a}_{M_O,l'}|^2))} = \max_{1 \leq i \leq M_O} |(\mathbf{a}_{i,l})^H \mathbf{a}_{i,l'}|. \quad (40)$$

Combing (30), (34), (40) and (32), then the μ_B can be represented as

$$\mu_B = \frac{1}{M_O} \max_{1 \leq l, l' \leq L} \max_{\substack{1 \leq i \leq M_O \\ l \neq l'}} |(\mathbf{a}_{i,l})^H \mathbf{a}_{i,l'}| = \frac{1}{M_O} \max_{1 \leq i \leq M_O} \max_{\substack{1 \leq l, l' \leq L \\ l \neq l'}} \left| \sum_{i_p=1}^{N_{PO}} \frac{1}{N_{PO}} \omega^{\zeta_i(i_p) \cdot (l' - l)} \right|. \quad (41)$$

In order to minimize μ_B , it is hard to find an appropriate approach to optimize (41). Furthermore, (41) indicates the average value of the maximum coherence of the sensing matrix in each antenna and fails to take the overall coherence of all the antennas into consideration. Instead, we find another equivalent criterion μ_B^{total} [25,31], which is total coherence and defined as

$$\mu_B^{\text{total}} = \sum_{l=1}^L \sum_{l' \neq l}^L \|B_l^H B_{l'}\|_F^2. \quad (42)$$

The physical meaning of μ_B^{total} is the sum of Frobenius norm of the sub-block product. Considering all the antennas and combining (39) and (41), then we have

$$\mu_B^{\text{total}} = \sum_{i=1}^{M_O} \sum_{l \neq l'}^L \sum_{l=1}^L |(\mathbf{a}_{i,l})^H \mathbf{a}_{i,l'}|^2 = \sum_{i=1}^{M_O} \sum_{l \neq l'}^L \sum_{l=1}^L \left| \sum_{i_p=1}^{N_{PO}} \frac{1}{N_{PO}} \omega^{\zeta_i(i_p) \cdot (l' - l)} \right|^2. \quad (43)$$

The above formula clearly shows that total coherence is the sum of squared absolute of $\sum_{i_p=1}^{N_{PO}} \frac{1}{N_{PO}} \omega^{\zeta_i(i_p) \cdot (l' - l)}$.

According to (43), in order to obtain the optimal pilot locations of all the antennas, the minimization problem of total coherence can be expressed as

$$\min_{\xi_1, \xi_2, \dots, \xi_{M_O}} \sum_{i=1}^{M_O} \sum_{l \neq l'}^L \sum_{l=1}^L \left| \sum_{i_p=1}^{N_{PO}} \frac{1}{N_{PO}} \omega^{\zeta_i(i_p) \cdot (l' - l)} \right|^2$$

$$\text{s.t.} \begin{cases} |\zeta_i| = N_{PO}, \zeta_i \in \{1, 2, 3, \dots, N\} \setminus \zeta_N, \\ 1 \leq i \leq M_O \\ \zeta_i \cap \zeta_j = \emptyset, \quad 1 \leq i, j \leq M_O, i \neq j \end{cases} \quad (44)$$

Still, the above problem is an intractable problem and is a combinational problem. Particularly, to find the optimal solution with exhaustive search, the above problem requires

$$\binom{N - N_{PN}}{N_{PO}} \binom{N - N_{PN} - N_{PO}}{N_{PO}} \cdots \binom{N - N_{PN} - N_{PO}(M_O - 1)}{N_{PO}}$$

combinations to obtain the global minimum value, the computational amount is extremely huge. For instance, if $N = 4096$, $M_O = 4$, $N_{PO} = 12$, $N_{PN} = 192$, then the number of search candidates is $\binom{3904}{12} \binom{3892}{12} \binom{3880}{12} \binom{3868}{12}$, which is nearly impossible to solve this problem with limited time. In order to provide the reliable solution for this pilot optimization problem, the GA based pilot allocation algorithm is proposed, which can converge fast and theoretically provide the global solution with probability one when iterations tend to infinity.

5.2 Proposed GAPA

GA is a computational model inspired by biological evolution process [36,37], which is one of the heuristic methods. The classical GA starts with a population, which are constituted by a certain number of individuals encoded by genes. Since the behavior of individuals is controlled by different genes, the intersection and mutation is manipulated to affect the external appearance of individuals. According to the specific criterion, i.e., fitness function, the next generation is generated with the higher fitness value before crossover and mutation. After several generations, the individuals with the highest fitness value are approximated as the optimal solution. In this case, the set of orthogonal pilot positions for BS antennas can be regarded as the individuals, which is $E = \{e_1, \dots, e_{N_{PO}}, e_{N_{PO}+1}, \dots, e_{M_O N_{PO}-1}, e_{M_O N_{PO}}\}$, where the element of set E stands for the position of each pilot. Particularly, for the

i th antenna, $i \in \Omega_O$, the set of pilot locations is expressed as

$$\zeta_i = \{e_{1+N_{PO}(i-1)} \quad e_{2+N_{PO}(i-1)} \quad \cdots \quad e_{N_{PO}+N_{PO}(i-1)}\}. \quad (45)$$

Note that the aim is to produce the orthogonal pilot position in the frequency domain, due to random procedure of crossover and mutation, it is difficult to guarantee whether the elements of next generation of pilot position are all different from each other. In order to deal with this matter, the illegal individuals are defined to combat with individuals with the superposition elements. The desired individual satisfies $\left| \bigcup_{i \in \Omega_O} \zeta_i \right| = |E| = M_O N_{PO}$. Apart

from that, the elements of individuals should belong to $\{1, 2, \dots, N\}$, which is the range of OFDM subcarriers. The desired individual also satisfies $\min_{1 \leq j \leq N_{PO}} \zeta_i(j) \geq 1$, $\max_{1 \leq j \leq N_{PO}} \zeta_i(j) \leq N, \forall i \in \Omega_O$.

Besides, due to the existence of non-orthogonal pilot, the placement for orthogonal pilot must avoid these positions, which is even harder to address. Specifically, the index for non-orthogonal pilot subcarriers ζ_N [32] satisfies

$$\zeta_N(\kappa) = I_0 + (\kappa - 1) \lfloor N/N_{PN} \rfloor \quad (46)$$

where $\zeta_N \subset \{1, 2, 3, \dots, N\}$, $1 \leq \kappa \leq N_{PN}$, the first index $1 \leq I_0 < \lfloor N/N_{PN} \rfloor$, equal pilot interval $\lfloor N/N_{PN} \rfloor$. In this case, the equivalent sensing matrix composed by non-orthogonal pilot subcarriers can achieve the small cross-correlation and is insensitive to the distribution of pilot subcarriers, which has been demonstrated in [32]. By contrast, the equivalent sensing matrix formed by the orthogonal pilot subcarriers is sensitive to the distribution of pilot subcarriers, which is optimized by the proposed GAPA algorithm.

The individuals outside these three types are defined as illegal individuals. Fortunately, these illegal individuals can be penalized by the fitness function and given smaller fitness value for penalty. The fitness function is defined as

$$\text{fitness function} = \begin{cases} \frac{1}{\mu_B^{\text{total}}}, & \left| \bigcup_{i \in \Omega_O} \zeta_i \right| = M_O N_{PO}; \\ & \zeta_i \subset \{1, 2, 3, \dots, N\} \setminus \zeta_N; \forall i \in \Omega_O \\ P_f, & \text{otherwise} \end{cases} \quad (47)$$

where P_f is the penalty fitness value, which should be smaller than the minimum of $1/\mu_B^{\text{total}}$ and can be chosen empirically. The proposed GAPA algorithm is based on the classical GA, which consists of five main stages. That is,

the initiation of population with a certain number of individuals, calculation of fitness values for each individual, selection operation, crossover operation, and mutation operation among individuals. The proposed GAPA for the optimization problem (44) is elaborated as follows.

Algorithm 3 The GAPA algorithm

Input: $M_O, N_{PO}, N, \zeta_N, P_f$

Step 1 Generate the population size $P_{\text{num}} = 100$, the maximum iteration $M_{\text{max}} = 200$, selection probability $P_s = 0.9$, crossover probability $P_c = 0.7$, mutation probability $P_m = 0.06$.

Step 2 Randomly generate a population with P_{num} individuals and compute the fitness function of each individual according to (47).

Step 3 Carry out the selection operation for the individuals obtained in Step 2 with probability P_s .

Step 4 Carry out the crossover operation between two individuals obtained by Step 3 with probability P_c .

Step 5 Carry out the mutation operation for individuals obtained by Step 4 with probability P_m among $\{1, 2, 3, \dots, N\} \setminus \zeta_N$.

Step 6 Compute the fitness function of each individual obtained by Step 5 according to (47).

Step 7 Combine the individuals obtained by Step 6 and the individuals obtained in Step 2, select P_{num} individuals with the largest fitness value.

Step 8 Reach the maximum iteration, the algorithm terminates; Otherwise, jump to Step 2.

Output: The individual with the largest fitness value is regarded as the set of pilot allocation E . Allocate each transmit antenna with orthogonal pilots $\zeta_i, \forall i \in \Omega_O$ based on (45).

The setting of parameters including $P_{\text{num}}, P_s, P_c, P_m$ according to [37], when the population size and the mutation probability are small, the crossover probability and the selection probability are high, the GA-based algorithm can achieve the global solution with probability one when M_{max} goes to infinity. As shown in Section 5, the proposed GAPA algorithm can converge fast although M_{max} is small. The proposed GAPA for the optimization problem (44) is elaborated as follows. Step 1 above is the initiation of the GAPA algorithm. Step 2 generates a population consisting of a certain number of individuals and the fitness value for each individual is calculated. Then the selection operation and crossover operation for individuals with high probability are conducted in Step 3 and Step 4, respectively. The mutation operation are executed in Step 5 with probability P_m among the legal range $\{1, 2, 3, \dots, N\} \setminus \zeta_N$ to avoid the collision with non-orthogonal pilots. Step 6 re-computes the fitness function

of each individuals obtained by Step 5 according to (33). Step 7 combines the individuals obtained by Step 6 and Step2 and selects P_{num} individuals with the largest fitness value.

To summarize, the whole procedure of pilot optimization for non-orthogonal pilot and orthogonal pilot distribution is depicted in the following Algorithm 4. According to the Algorithm 4, the indices of non-orthogonal pilot carriers are determined firstly. Then, the pilot optimization for orthogonal pilot carriers is carried out based on the remaining subcarriers positions via the proposed GAPA algorithm.

Algorithm 4 The whole procedure of pilot optimization for non-orthogonal pilot and orthogonal pilot distribution

Input: $M_O, N_{PO}, M_N, N_{PN}, N, \zeta_N, P_f$

Step 1 Determine the distribution of non-orthogonal pilots ζ_N according to (46).

Step 2 Determine the distribution of orthogonal pilots ζ_i according to Algorithm 3.

Output: the distribution of non-orthogonal pilots ζ_N , the distribution of orthogonal pilots $\zeta_i, \forall i \in \Omega_O$.

6. Simulation results

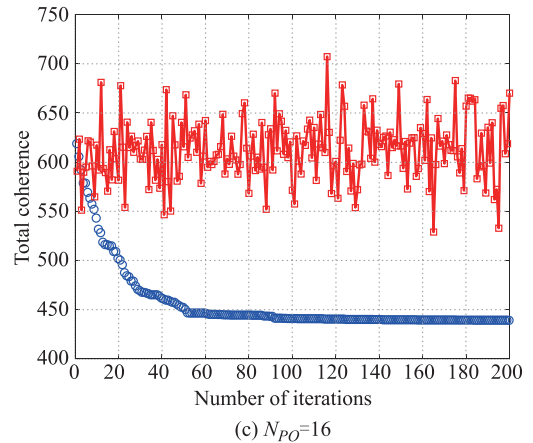
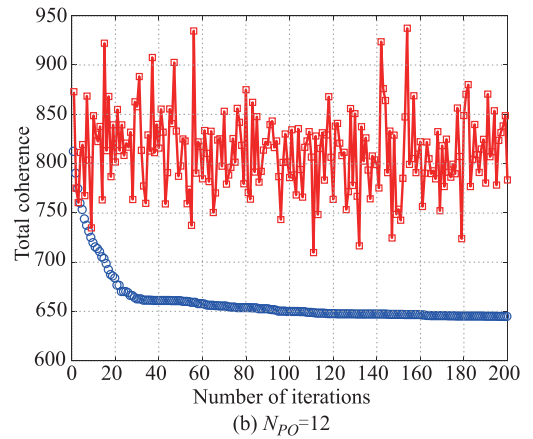
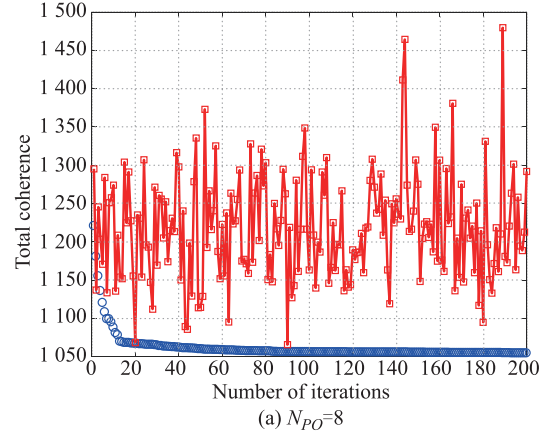
In this section, channel estimation performance of the proposed hybrid pilot distribution and the pilot allocation algorithm for massive MIMO system are verified by a series of experiments. Particularly, $f_c = 2$ GHz, $BW = 30$ MHz, $N_T = 24$, $N = 4096$, 16 quadrature amplitude modulation (QAM) is used for data transmission. The equivalent channel length $L = 50$ with random support, and the channel coefficients are generated according to the International Telecommunications Union Vehicular-a channel model with six significant paths. The support of sparse channels is randomly generated and the average mean squared error (MSE) of the estimated channel is given by

$$\text{MSE} = \frac{1}{N_T} \sum_{i=1}^{N_T} \frac{\|\hat{\mathbf{h}}_i - \mathbf{h}_i\|_2^2}{\|\mathbf{h}_i\|_2^2}. \quad (48)$$

First of all, Fig. 5 provides the convergence of total coherence of the sensing matrix when GAPA and random search are performed for different N_{PO} . The optimal pilot locations are obtained by minimizing total coherence represented by (42). Particularly, the input parameters $M_O = 8$, $P_f = 1/2000$, $I_0 = 1$ and $N_{PN} = 192$. Optimal orthogonal pilots can be selected from the remaining pilot positions $\{1, 2, 3, \dots, N\} \setminus \zeta_N$.

According to the typical process of the proposed GAPA algorithm in Fig. 5, it can be observed that the GAPA algorithm can converge fast with only hundreds iterations,

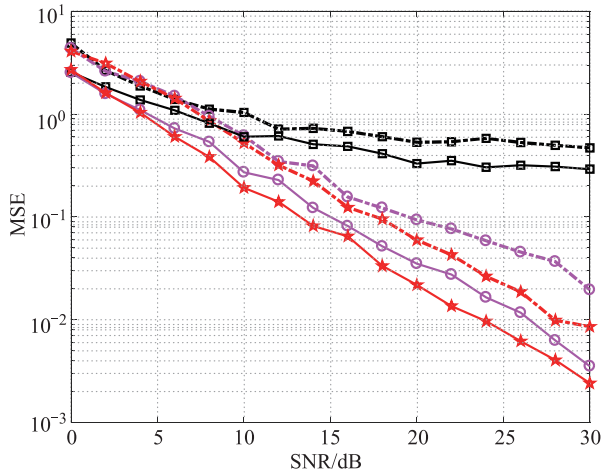
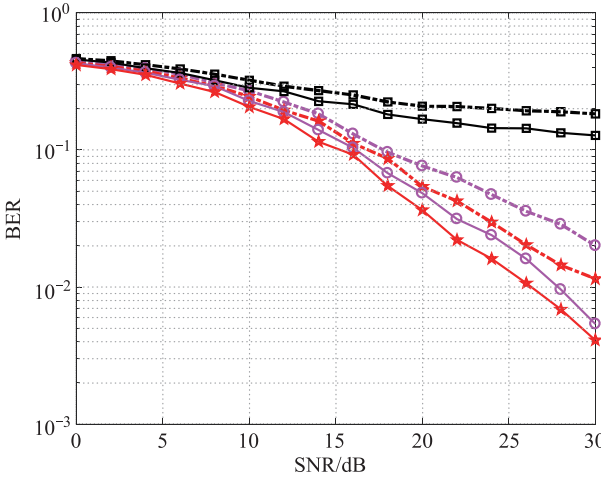
while random search fails to obtain the appropriate solution within few iterations. Specifically, for $N_{PO} = 8$, the total coherence achieves about 1050 using only 80 iterations; $N_{PO} = 12$, the total coherence reaches 650 within 100 iterations and total coherence converges to 440 for $N_{PO} = 16$.



—□— : Random search algorithm;
—○— : Proposed pilot optimization algorithm.

Fig. 5 Total coherence of sensing matrix versus iteration by GAPA and random search

It should be noted that the objective value in (44) decreases with a number of pilots. The gap between GAPA and random search tends to be large with increasing N_{PO} within a few iterations, which demonstrates the convergence of the proposed algorithm. Thus, the optimal location of orthogonal pilots can be quickly obtained using the GAPA algorithm when the minimum total coherence is reached. Based on the optimized pilot allocation, the channel estimation can be further improved and accordingly, pilot overhead is reduced, which is demonstrated in Fig. 6.


 (a) MSE performance when $N_{PO}=8$

 (b) BER performance when $N_{PO}=8$

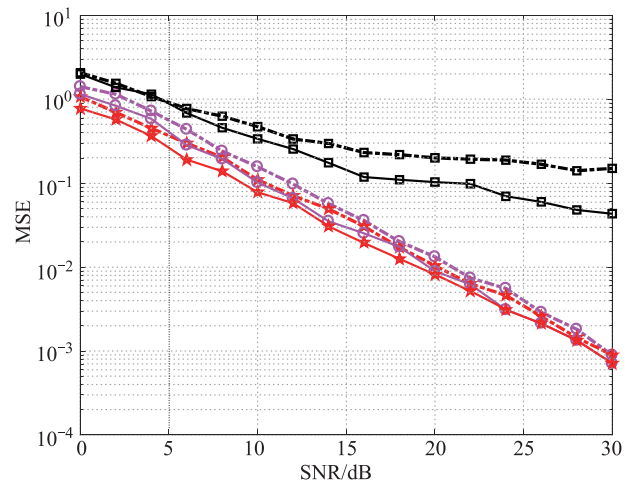
—□— : OMP, unoptimized pilots; —□— : OMP, optimized pilots;
 —○— : BOMP, unoptimized pilots; —○— : BOMP, optimized pilots;
 —★— : Oracle LS, unoptimized pilots;
 —★— : Oracle LS, optimized pilots.

Fig. 6 Comparison between optimized orthogonal pilot and unoptimized orthogonal pilot when $N_{PO} = 8$

To demonstrate the merits of pilot optimization, the MSE and BER performance is compared in difference cases. Particularly, the parameters are set as $M_O = 8$,

$M_N = 16$, $N_{PO} = 8$, $N_{PN} = 192$. Orthogonal pilots are obtained according to Fig. 5(a). The OMP algorithm, BOMP algorithm and Oracle LS algorithm are considered as the typical algorithm for channel estimation of the orthogonal pilot at the BS. According to Fig. 6, although the OMP fails to exploit the block sparsity among massive MIMO channel, the MSE and BER performance could be improved by employing optimized pilots. By using the proposed PAGA algorithm, the MSE gains for the BOMP channel estimation algorithm can be 4.7 dB at most. It should be noticed that the lower bound Oracle LS can even be improved due to the optimized pilots. The reason for this is that the sensing matrix composed by optimized pilots can better capture the channel support than the unoptimized pilots. In terms of BER, the performance is obviously improved by employing the optimized pilots. For the MSE performance, the BOMP algorithm with optimized pilots can be even better than the Oracle LS without the optimized pilot locations. Consequently, the BER performance is further improved due to the improved MSE performance for both BOMP and Oracle LS algorithms.

Fig. 7 depicts the MSE and BER gains achieved by GAPA algorithms for the channel estimation by optimized pilot locations when $N_{PO} = 12$. In this case, due to the increase of the pilot number, the performance of three algorithms all improves and the performance of BOMP with optimized pilots improves 5 dB than $N_{PO} = 8$ in terms of MSE of channel estimation. In particular, the performance gap between BOMP and Oracle LS algorithms gets closer than that of $N_{PO} = 8$. With the optimized pilots, the performance gains for BOMP in terms of MSE and BER are 2 dB and 1 dB, respectively.


 (a) MSE performance when $N_{PO}=12$

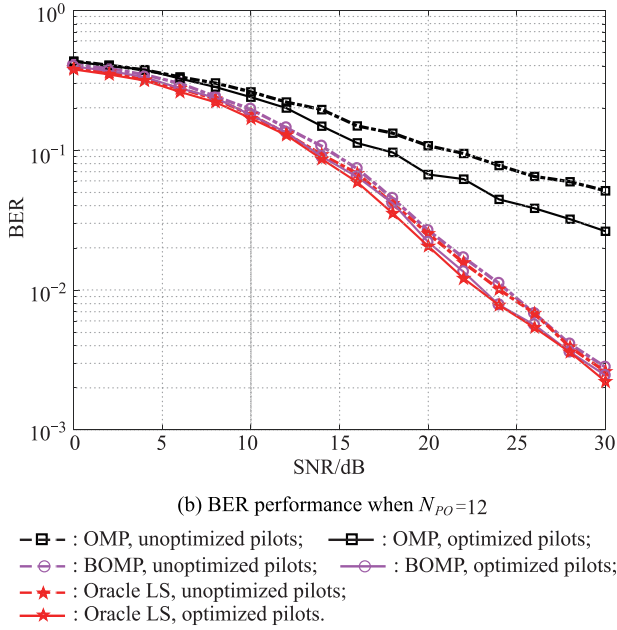


Fig. 7 Comparison between optimized orthogonal pilot and unoptimized orthogonal pilot when $N_{PO} = 12$

Fig. 8 also compares the MSE and BER gains achieved by employing GAPA algorithms to obtain optimized pilot locations when $N_{PO} = 16$. In this case, due to the increase of the pilot number, the performance of three algorithms all improves, whereby the performance of BOMP improves 5 dB than $N_{PO} = 8$ in terms of MSE of channel estimation. Overall, Fig. 6 to Fig. 8 show that the performance gains of the proposed GAPA algorithms can be larger when the number of orthogonal pilots for each antenna is smaller. This indicates that the proposed GAPA algorithm can be even suitable for channel estimation massive MIMO system due to the larger performance gains for a smaller number of pilots.

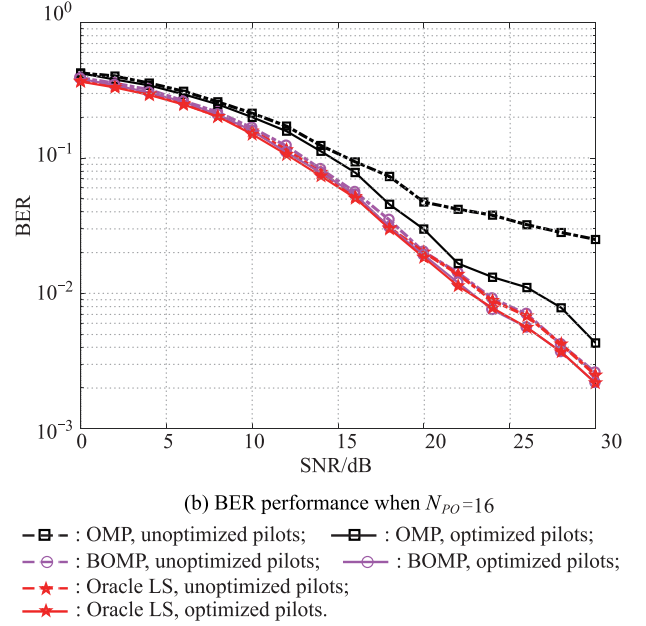
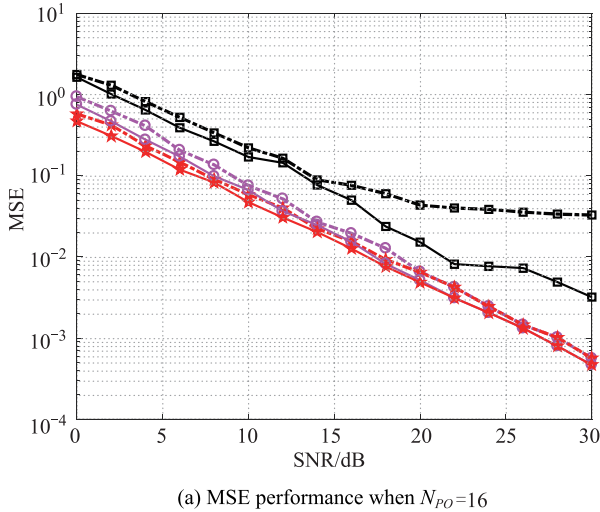


Fig. 8 Comparison between optimized orthogonal pilot and unoptimized orthogonal pilot when $N_{PO} = 16$

The reason behind the improvement of MSE and BER is that the pilot optimization can improve the channel support recovery rate in different cases. Fig. 9 presents the probability of support recovery, which means the ratio of the number of accurate estimated support elements to the total number of true support. It can be clearly observed from Fig. 9 that the successful rate of channel support can be significantly improved by the proposed GAPA algorithm, especially when the number of orthogonal pilots is small. With the increase of the number of pilots, the support recovery rate also rises. When $N_{PO} = 8$, although BOMP with optimized pilots does not reach 100% accurate recovery, it experiences 11% increase of support recovery rate. For $N_{PO} = 12$, the support recovery probability of BOMP with optimized pilots reaches 1 when SNR is more than 24 dB, while this figure reduces to 22 dB when $N_{PO} = 16$, in which the GAPA optimization still has 1.9% performance gain. This performance gain in support recovery rate directly reflects the decrease of MSE and BER in Figs. 6–8.

Fig. 10 presents the MSE performance with hybrid pilot distribution at the BS. Particularly, the OMP algorithm, ASSP algorithm and Oracle LS algorithm are considered as the benchmark for channel estimation of the non-orthogonal pilot at the BS. The proposed Algorithm 2 with optimized pilots is denoted as AOP and Algorithm 2 with unoptimized pilots is denoted as AUP. Channel estimation with OMP indicates the block property fails to be exploited while the AOP, AUP and ASSP are used to exploit the spatial common sparsity. In the case that orthogonal pilots are only $N_{PO} = 8$, the MSE performance of

AOP and AUP outperforms that of ASSP when the number of non-orthogonal pilots is small. This is caused by the fact that both AOP and AUP utilize the prior support obtained by orthogonal pilots, while the ASSP fails to obtain the channel support if the pilot number is small.

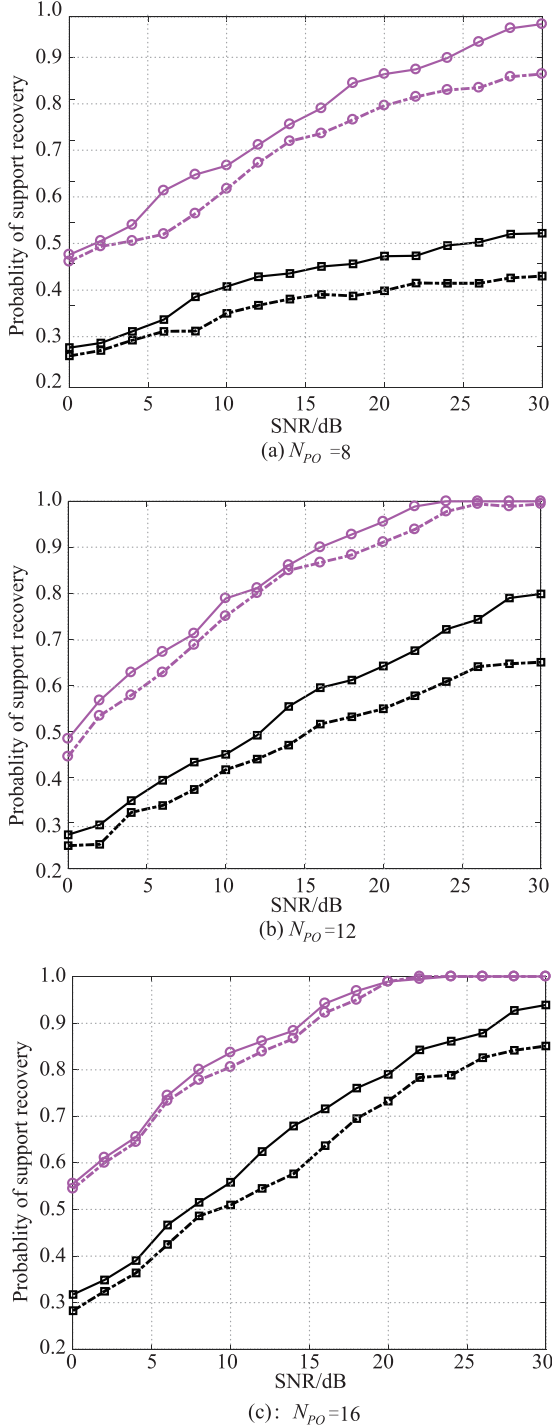


Fig. 9 Comparison of channel support recovery rate between optimized orthogonal pilot and unoptimized orthogonal pilot

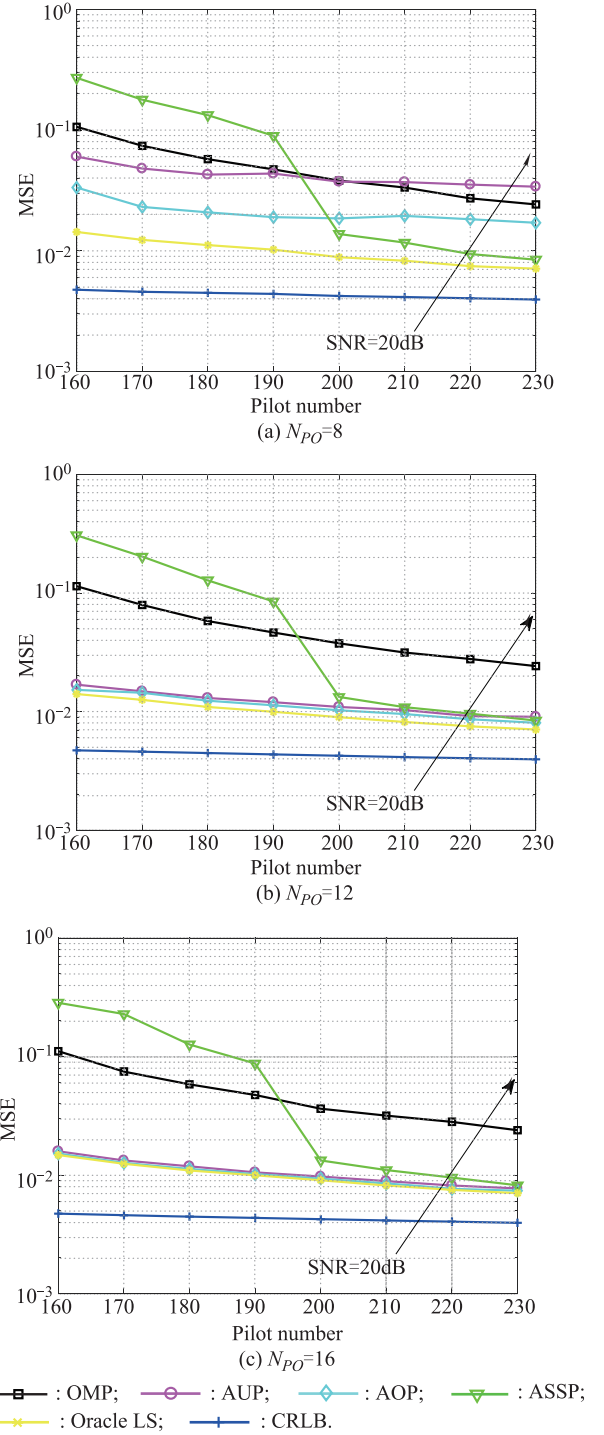


Fig. 10 Comparison of MSE performance for channel estimation with non-orthogonal pilots

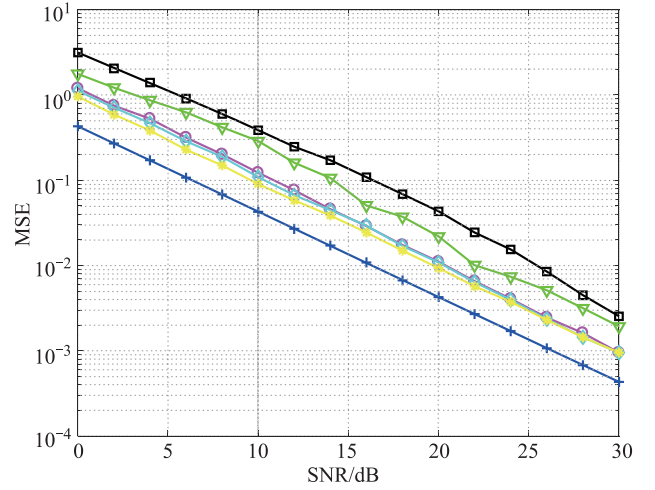
As the number of pilots increases to 200, the ASSP performs better due to the fact that the minimum number of measurements to guarantee the reliable recovery of the support is reached, i.e., $\tilde{N}_{PN} = N_{PN}/M_{PN} = 200/16 > 2 \times 6 = 12$ [12]. With the help of the pilot optimization by the GAPA algorithm, the support recovery rate is improved and the AOP can consistently outperform AUP,

which also can be interpreted by Fig. 9(a). When orthogonal pilots $N_{PO} = 12$, the MSE performances of AOP and AUP both improve and achieve the similar performance as ASSP even with the increase of pilot. Furthermore, it can be observed from Fig. 10(c) that AUP and AOP can obtain almost the same MSE performance, this is because the prior support can be reliably acquired due to the increase of orthogonal pilot, which can be verified in Fig. 9(c).

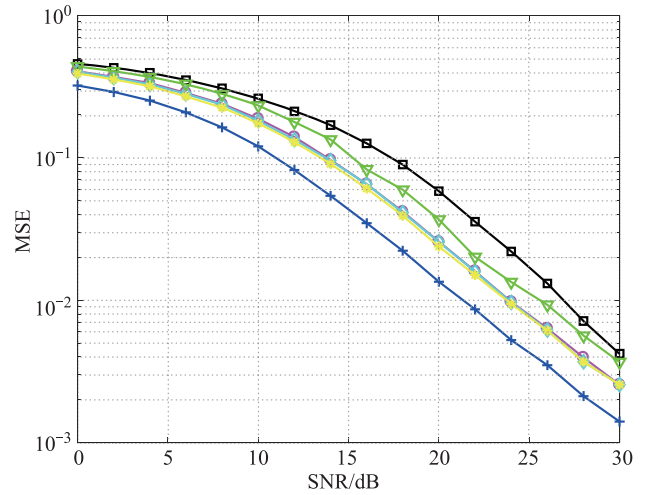
Note that the number of non-orthogonal pilots to guarantee the stable recovery for ASSP is approximately 200 when SNR=20 dB, which is huge in practical system, while AUP and AOP algorithms could consistently approach the lower bound of Oracle LS, especially when the number of orthogonal pilots is slightly large. It should be particularly noted that only $N_{PO} = 12$ pilots are utilized, the MSE of channel estimation can reach the same order of magnitude with the non-orthogonal pilot scheme, in which hundreds of pilots should be deployed for each antenna. Such large pilot overhead results in a substantial loss in data throughput, which motivates us to make use of orthogonal pilot distribution for part of BS antennas to achieve the similar channel estimation performance. In order to alleviate the pilot overhead, a small number of antennas can be equipped with orthogonal pilots during transmission. For example, pilot overhead N_{total} for hybrid pilot distribution can be calculated as $N_{total} = N_{PO}M_O + N_{PN}M_N$. In this case, $N_{total} = 12 \times 8 + 200 \times 16 = 3296$, which is $\frac{24 \times 200 - 3296}{24 \times 200} = 31.33\%$ less than conventional non-orthogonal pilot distribution. In order to further alleviate the pilot burden, the proposed PAGA algorithm is used to optimize the location of orthogonal pilots.

Fig. 11 investigates the MSE and BER performance with pilot number of $N_{PN} = 195$, $N_{PO} = 12$ for FDD system with hybrid pilot at the BS. The proposed AUP and AOP can consistently approach the lower bound of Oracle LS and CRLB and AOP slightly outperforms AUP due to pilot optimization by the GAPA algorithm. The ASSP algorithm does not perform well due to the insufficient number of pilots to track the channel support. In this case, the proposed AUP and AOP only employ $N_{total} = 12 \times 8 + 180 \times 16 = 2976$ pilots and outperform ASSP, which is equipped with $N_{total} = 180 \times 24 = 4320$ pilots, resulting in 31.11% pilot reduction. In Fig. 11(b), the 16QAM and the zero-forcing equilibrium are adopted at the user side to accomplish signal transmission and demodulation. Full CSI indicates that the user has the full CIR as a priority. Since the pilot optimization results in better MSE performance, the proposed AOP algorithm achieves better BER performance than its counterparts. The performance

gain to ASSP can be 1.7 dB in terms of BER.



(a) MSE performance versus SNR



(b) BER performance versus SNR

—■— : OMP; —○— : AUP; —◇— : AOP; —▽— : ASSP;
—★— : Oracle LS; —+— : Full CSI.

Fig. 11 MSE and BER performance versus SNR

7. Conclusions

In this paper, a hybrid orthogonal and non-orthogonal pilot distribution at BS is proposed by exploiting the common sparsity of channel due to the compact antenna arrangement. The proposed pilot distribution generalizes the existing pilot distribution scheme to reduce the overwhelming pilot overhead. Then the block sparsity for antennas with hybrid pilot distribution is derived respectively and can be used to obtain channel impulse response. By utilizing the theoretical analysis of block sparse recovery, the total coherence criterion is proposed to optimize the sensing matrix composed by orthogonal pilots. Due to the huge

complexity of optimal pilot acquisition, a GAPA algorithm is proposed to acquire optimal pilot distribution locations with fast convergence. Furthermore, the CRLB is derived for non-orthogonal pilot-based channel estimation and can be asymptotically approached by the prior support set with optimized pilots. Simulation results show that the proposed hybrid pilot distribution among BS antennas can achieve significant pilot reduction compared with non-orthogonal pilot allocation. Moreover, the proposed GAPA algorithm converges fast and achieves slight improvement in terms of MSE and BER, especially when the number of orthogonal pilots is small.

References

- [1] ANDREWS J G, BUZZI S, WAN C, et al. What will 5G be? *IEEE Journal on Selected Areas in Communications*, 2014, 32(6): 1065–1082.
- [2] LU L, LI G Y, SWINDLEHURST A L, et al. An overview of massive MIMO: benefits and challenges. *IEEE Journal of Selected Topics in Signal Processing*, 2014, 8(5): 742–758.
- [3] WANG C X, HAIDER F, GAO X, et al. Cellular architecture and key technologies for 5G wireless communication networks. *IEEE Communications Magazine*, 2014, 52(2): 122–130.
- [4] EMIL B, LARSSON E G, MARZETTA T L. Massive MIMO: ten myths and one critical question. *IEEE Communications Magazine*, 2015, 54(2): 114–123.
- [5] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 2014, 52(2): 186–195.
- [6] HOYDIS J, TEN BRINK S, DEBBAH M. Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? *IEEE Journal on Selected Areas in Communications*, 2013, 31(2): 160–171.
- [7] CHAN P W C, LO E S, WANG R R, et al. The evolution path of 4G networks: FDD or TDD? *IEEE Communications Magazine*, 2007, 44(12): 42–50.
- [8] NAM Y H, AKIMOTO Y, KIM Y, et al. Evolution of reference signals for LTE advanced systems. *IEEE Communications Magazine*, 2012, 50(2): 132–138.
- [9] BERGER C R, WANG Z, HUANG J, et al. Application of compressive sensing to sparse channel estimation. *IEEE Communications Magazine*, 2010, 48(11): 164–174.
- [10] DAI L, WANG J, WANG Z, et al. Spectrum and energy-efficient OFDM based on simultaneous multichannel reconstruction. *IEEE Trans. on Signal Processing*, 2013, 61(23): 6047–6059.
- [11] BARBOTIN Y, HORMATI A, RANGAN S, et al. Estimation of sparse MIMO channels with common support. *IEEE Trans. on Communications*, 2011, 60(12): 3705–3716.
- [12] DONOHO D L. Compressed sensing. *IEEE Trans. on Information Theory*, 2006, 52(4): 1289–1306.
- [13] DONOHO D L, ELAD M, TEMLYAKOV V N. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. on Information Theory*, 2005, 52(1): 6–18.
- [14] BARON D, WAKIN M B, DUARTE M F, et al. Distributed compressed sensing. Preprint, 2012, 22(10): 2729–2732.
- [15] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on Information Theory*, 2007, 53(12): 4655–4666.
- [16] DAI W, MILENKOVIC O. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. on Information Theory*, 2008, 55(5): 2230–2249.
- [17] TROPP J A, GILBERT A C, STRAUSS M J. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Processing*, 2006, 86(3): 572–588.
- [18] GAO X, EDFORS O, RUSEK F, et al. Linear precoding performance in measured very large MIMO channels. *Proceeding of IEEE Vehicular Technology Conference*, 2011: 1–5.
- [19] RAO X, LAU V K N. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. on Signal Processing*, 2014, 62(12): 3261–3271.
- [20] CHOI J, LOVE D J, BIDIGARE P. Downlink training techniques for FDD massive MIMO systems: open-Loop and closed-Loop training with memory. *IEEE Journal of Selected Topics in Signal Processing*, 2013, 8(5): 802–814.
- [21] SHEN J C, ZHANG J, ALSUSA E, et al. Compressed CSI acquisition in FDD massive MIMO: how much training is needed? *IEEE Trans. on Wireless Communications*, 2016, 15(6): 4145–4156.
- [22] MASOOD M, AFIFY L H, AL-NAFFOURI T Y. Efficient co-ordinated recovery of sparse channels in massive MIMO. *IEEE Trans. on Signal Processing*, 2014, 63(1): 104–118.
- [23] BEN-HAIM Z, ELDAR Y C. Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(5): 1032–1047.
- [24] HOU W, LIM C W. Structured compressive channel estimation for large-scale MISO-OFDM systems. *IEEE Communications Letters*, 2014, 18(5): 765–768.
- [25] ZELNIK-MANOR L, ROSENBLUM K, ELDAR Y C. Sensing matrix optimization for block-sparse decoding. *IEEE Trans. on Signal Processing*, 2011, 59(9): 4300–4312.
- [26] QI C, HUANG Y, JIN S, et al. Sparse channel estimation based on compressed sensing for massive MIMO systems. *IEEE International Conference on Communications*, 2015: 4558–4563.
- [27] CHOI J W, SHIM B, CHANG S. Downlink pilot reduction for massive MIMO systems via compressed sensing. *IEEE Communications Letters*, 2015, 19(11): 1889–1892.
- [28] ZHANG W, XIA X G, CHING P C. Optimal training and pilot pattern design for OFDM systems in Rayleigh fading. *IEEE Trans. on Broadcasting*, 2006, 52(4): 505–514.
- [29] XIA P, ZHOU S, GIANNAKIS G B. Achieving the Welch bound with difference sets. *IEEE Trans. on Information Theory*, 2005, 51(5): 1900–1907.
- [30] CHEN J C, WEN C K, TING P. An efficient pilot design scheme for sparse channel estimation in OFDM systems. *IEEE Communications Letters*, 2013, 17(7): 1352–1355.
- [31] HE X, SONG R, ZHU W P. Pilot Allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO OFDM Systems. *IEEE Trans. on Vehicular Technology*, 2016, 65(5): 2990–3004.
- [32] GAO Z, DAI L, DAI W, et al. Structured compressive sensing based spatio temporal joint channel estimation for FDD massive MIMO. *IEEE Trans. on Communications*, 2016, 64(2): 601–617.
- [33] LEE D. MIMO OFDM channel estimation via block stagewise orthogonal matching pursuit. *IEEE Communications Letters*, 2016, 20(10): 2115–2118.
- [34] KAY S M. Fundamentals of statistical signal processing, volume I: estimation theory. New Jersey, USA: Prentice-Hall, 1993.
- [35] BEN-HAIM Z, ELDAR Y C. The Cramér-Rao bound for estimating a sparse parameter vector. *IEEE Trans. on Signal Processing*, 2010, 58(6): 3384–3389.

- [36] DEB K. Multi-objective optimisation using evolutionary algorithms: an introduction. New York: Wiley, 2001.
- [37] RUDOLPH G. Convergence analysis of canonical genetic algorithms. IEEE Trans. on Neural Networks, 1994, 5(1): 96–101.

Biographies



ZHANG Ruoyu was born in 1992. He received his B.S. degree in communication engineering from Harbin Institute of Technology, Harbin. He is currently pursuing his Ph.D. degree in electrical engineering. His research interests include compressed sensing, channel estimation, massive MIMO system.

E-mail: hitzhangruoyu@163.com



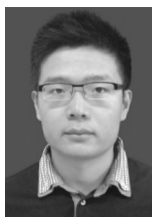
ZHAO Honglin was born in 1969. He received his B.E., M.E., and Ph.D. degrees in communication engineering from Harbin Institute of Technology, Harbin, China, in 1991, 1994, and 1998, respectively. He is currently a professor with the Communication Research Center, School of Electronics and Information Engineering, Harbin Institute of Technology. His research interests include spread spectrum communication, cognitive radio networks and wireless broadband networks.

E-mail: hlzhao@hit.edu.cn



ZHANG Jiayan was born in 1974. He received his Ph.D. degree in communication engineering from Harbin Institute of Technology, Harbin, China, in 2008. He is currently an associate professor with the Communication Research Center, School of Electronics and Information Engineering, Harbin Institute of Technology. His research interests include spread spectrum communication, high speed signal processing and channel coding.

E-mail: jyzhang@hit.edu.cn



JIA Shaobo was born in 1988. He received his B.S. and M.E. degrees from the Harbin Engineering University, Harbin, China, in 2011, 2014, respectively. He is currently pursuing his Ph.D. degree in Harbin Institute of Technology. His research interests lie in the physical layer security, cooperative communications, and cognitive radio networks.

E-mail: jiashaobo2007@126.com