Probability-based pilot allocation for MIMO relay Distributed compressed sensing channel estimation

Abbas Akbarpour-Kasgari[*] [ID] and Mehrdad Ardebilipour

Abstract

Multiple-Input Multiple-Output (MIMO) relay communication systems are used as an efficient system in spectral efficiency and power allocation view point. In these systems, some of the facilities need channel state information (CSI). Besides, new estimation methods based on compressed sensing (CS) are well known for their spectral efficiency and accuracy. In this paper, we have used a Distributed CS-based channel estimation method to improve the accuracy and spectral efficiency of channel estimation for MIMO-Orthogonal Frequency Division Multiplexing relay network. Specifically, using Least Squares (LS) estimation increases the accuracy of well-known Compressive Sampling Matching Pursuit (CoSaMP) algorithm and proposes Block-verified CoSaMP (B-vCoSaMP). To improve the accuracy of estimation, we are encountered with a combinatorial optimization which is dealt with probability-based approaches in this paper. More particularly, three probability-based optimization methods have been proposed to optimize the mutual coherence of measurement matrix called Sequential Cross-Entropy (SCE), Extended Estimating of Distribution Algorithm (EEDA), and Parallel Cross-Entropy (PCE). All these methods are based on sampling from a Probability Density Function (PDF) which is updated in each iteration using elite samples of the population. The simulation results represent the accuracy and speed of the proposed methods, and the comparison is expressed as well.

Keywords: Block-verified compressive sampling matching pursuit, Cross-entropy, Distributed compressed sensing, MIMO-OFDM relay channel estimation, Pilot allocation

1 Introduction

Multiple-Input Multiple-Output (MIMO) relay communication systems make it possible to increase the data rate and coverage area of communication. Different characteristics of MIMO-Orthogonal Frequency Division Multiplexing (OFDM) relaying make it an essential technology to conflict with fading and long-distance impairments [1, 2]. There are some features which need channel state information (CSI) to be available such as coherent demodulation, beamforming, relay selection, and so on [3, 4]. Increasing demand for high data rate communication could be accomplished by increasing the bandwidth efficiency of channel estimation methods which could be met by using compressed sensing (CS)-based procedures. Since there is a small number of significant scatterers

in the wireless environments, and the delay spread is normally large, the wireless channel could be modeled by sparse signal especially when operating in wide bandwidth. Consequently, CS-based channel estimation approaches could be used to increase the bandwidth efficiency and accuracy of channel estimation. Further to sparsity, MIMO communication systems benefit from the joint support of all the channel ensembles, since there are common scatterers in the environment between antennas of a transmit-receive pair [5]. Consequently, the support vector of the different channel groups is distributed identically. Thus, all the channels of MIMO communication between a transceiver pair are jointly sparse and could be represented in a block-sparse form [6]. By taking into account the block sparsity of MIMO channels, we could use the Distributed CS (DCS)-based channel estimation approaches instead of CS-based channel estimation methods. DCS-Simultaneous Orthogonal Matching Pursuit

*Correspondence: aakbarpour@mail.kntu.ac.ir
Department of Electrical and Computer Engineering, K.N. Toosi University of Technology, Shari'ati Street, Tehran, Iran

(SOMP) was one of the first DCS-based approaches which have been used in Single-Input Single-Output (SISO)-OFDM channel estimation [7]. Subsequently, in [8], DCS-SOMP was used to estimate the jointly sparse channel vectors in MIMO transmission. Recently, massive MIMO CSI matrices were estimated using the joint-Orthogonal Matching Pursuit (OMP) algorithm which has been proposed in [9] using the DCS-based approaches.

In DCS-based channel estimation methods for MIMO-OFDM systems, the measurement matrix can be constructed using random pilot patterns. However, it can be optimized to boost the estimation accuracy using Restricted Isometry Property (RIP). Since there is no known polynomial time approach to optimize RIP, alternatively, the mutual coherence is minimized which is more practical [10]. It is shown that the mutual coherence is related to the position of pilots. As a consequence, one may minimize the mutual coherence by exhaustively searching the best positions for pilots which are intractable from the computational point of view. Hence, utilizing computational programs to generate suboptimal pilot patterns are entirely favorable. To this end, some of the papers in the literature perform evolutionary algorithms to create pilot patterns. Specifically, in [11], two different methods are developed for optimizing pilot allocation in CS-based channel estimation including Genetic Algorithm (GA) and minimizing the largest element in mutual coherence set. Moreover, the pilot allocation for DCS-based channel estimation is proposed in [12] using GA-based suboptimal approach. Additionally, Estimating the Distribution Algorithm (EDA) is used as a promising approach to optimize the pilot positions in SISO-OFDM systems [13]. For MIMO-OFDM systems, in [12], the authors try to optimize the pilot allocation using GA-based approach which was the extension of [14]. Other papers including [15–17] focus on the SISO-OFDM systems and try to generate pilot sequences to increase the compressed channel estimation accuracy.

In this paper, we have developed a DCS-based approach called Block-verified Compressive Sampling Matching Pursuit (B-vCoSaMP). Specifically, we have used the joint sparsity behavior of MIMO channels to produce a block sparse measurement matrix. Fortunately, using a permutation matrix, we have generated the measurement matrix and received pilot subcarriers. Using the measurement matrix and received pilot vectors, we have developed a distributed form of Compressive Sampling Matching Pursuit (CoSaMP) algorithm. To utilize CoSaMP successfully, we have verified it by using Least Squares (LS) estimation on the positions of the channel vectors which are suggested by Block-CoSaMP (B-CoSaMP) algorithm. Hence, B-CoSaMP is completed by an LS estimation phase, and B-vCoSaMP is generated. In fact, by taking into account the sparsity order of the channel, CoSaMP is one of the most successful algorithms to estimate the support. By applying block behavior into the recovery algorithm, B-CoSaMP is changed to the more accurate algorithm for estimating the support, as well. Support estimation determines the non-zero paths of channel estimation vector. Furthermore, to increase the accuracy we used LS estimating, as well. Further to estimation algorithm, the pilot allocation approaches are also proposed.

We have proposed three different pilot allocation methods based on probability-based mutual coherence optimization for DCS-based MIMO relay channel estimation, as well. The mutual coherence of the measurement matrix is considered as the fitness function in this paper. It is caused by the impact of mutual coherence on the mean square error of estimation in DCS-based approaches which is addressed in the literature. Moreover, orthogonal pilot allocations are considered in channel estimation, i.e., each antenna transmits neutral subcarrier in pilot subcarriers of the other antennas. At first, we have shown that shifting the measurement matrix rows and columns have not affected the mutual coherence. Furthermore, it is proved that mutual coherence of a block measurement matrix can be calculated using smaller and simpler non-block measurement matrix. Utilizing these phenomena, we have defined the new fitness function for pilot allocation. Modeling the mutual coherence of measurement matrix implies that the fitness function is combinatorial and needs to be optimized using different pilot sequences. Herein, in order to minimize the fitness function, we have used probability-based approaches. Specifically, Sequential Cross-Entropy (SCE)-based pilot allocation, Extended Estimating of Distribution Approximation (EEDA)-based pilot allocation, and Parallel Cross-Entropy (PCE)-based pilot allocation are three different pilot allocation approaches which have been proposed in this paper. SCE-based pilot allocation works based on the sampling from a Probability Density Function (PDF) which is updated in each iteration. Updating the PDF in each iteration is related to the current samples and previous iterations' samples, as well. The relation can be controlled using a control constant τ. On the other hand, considering the marginal value for control constant ($\tau = 1$), the algorithm will update the PDF utilizing current samples. This approach is called EEDA which is an extension of [18]. EEDA approach will increase the convergence speed, but it will increase the risk of local minima trapping. Subsequently, to increase the rate of convergence together with the decrease of local minima trapping, we have proposed PCE-based pilot allocation which considers multiple local extrema in the iterations. It will increase the speed of convergence expectedly, moreover decrease the risk of local minima trapping.

The contributions of the paper can be summarized as:

- The compressed channel estimation using proposed B-vCoSaMP algorithm which uses block representation of the channel vectors to increase the accuracy of support estimation and excess LS estimation stage in the positions determined by estimated support to enhance the channel coefficients accuracy.
- The pilot allocation is optimized using proposed SCE-based sequence determination algorithm and extended to MIMO pilot allocation.
- In order to increase the speed of convergence of SCE-based pilot allocation, a special case is used called EEDA-based pilot allocation.
- To increase the accuracy together with the speed of convergence for pilot allocation, PCE-based pilot allocation algorithm is proposed in which multiple PDFs are used to generate the pilot sequences.

The remainder of the paper is as follows. The system model is represented in Section 2. Section 3 covers channel estimation approach using B-vCoSaMP, and pilot allocation scheme for SISO and MIMO systems and proposed pilot allocation methods are described in Sections 4 and 5, respectively. Numerical results are expressed in Section 6. Eventually, concluding remarks are demonstrated in Section 7.

Notations: Matrices and vectors are denoted by uppercase and lowercase boldfaced letters, respectively. $|.|$ and $(.)^*$ denote the complex modulus and the conjugate of a complex number. For a given matrix $\mathbf{A}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ denote its transpose and conjugate transpose, respectively, and $\mathbf{A}_{i,j}$ denotes the (i,j)th element of $\mathbf{A}$. For a given vector $\mathbf{x}$ with its element denoted by x_l, $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$ represents the Euclidean norm, $\|\mathbf{x}\|_1 = \sum_l |x_l|$ is the l_1-norm, and

diag($\mathbf{x}$) denotes a diagonal matrix with $\mathbf{x}$ on its main diagonal. For two vectors $\mathbf{x}$ and $\mathbf{y}$, $< \mathbf{x}, \mathbf{y} >$ denotes their inner product. For a given set $\mathbf{\Lambda}$, $n(\mathbf{\Lambda})$ is the number of elements in $\mathbf{\Lambda}$. $\mathbb{C}^{m \times n}$ stands for the set of all complex-valued $m \times n$ matrices, and $\emptyset$ denotes the null set.

2 System model

Consider an Amplify-and-Forward (AF) one-way relay network consisting of MIMO terminals (Fig. 1). The network consists of a source node ($\mathbb{S}$), a relay node ($\mathbb{R}$), and a destination node ($\mathbb{D}$), i.e., totally there are three MIMO nodes in a one-way AF network. Besides, source, relay, and destination terminals are equipped by $N_{\mathbb{S}}$, $N_{\mathbb{R}}$, and $N_{\mathbb{D}}$ transceiver antennas, respectively. The channel between mth antenna of $\mathbb{S}$ and qth antenna of $\mathbb{R}$ is denoted by $\mathbf{f}_{mq} = [f_{mq}(0), f_{mq}(1), \ldots, f_{mq}(L-1)]^T$ and between qth antenna of $\mathbb{R}$ and n-antenna of $\mathbb{D}$ is denoted by $\mathbf{g}_{qn} = [g_{qn}(0), g_{qn}(1), \ldots, g_{qn}(L-1)]^T$ for $m = 1, 2, \ldots, N_{\mathbb{S}}$, $q = 1, 2, \ldots, N_{\mathbb{R}}$ and $n = 1, 2, \ldots, N_{\mathbb{D}}$. All the channel ensembles are independent static linear selective channels. Taking into account the distance between transmitting and receiving antenna and the conveying distance of light which is extremely small, the channel delay profile is independent of transmit-receive antenna pairs. Consequently, the antenna spacing can be modeled by a phase rotation into the delay profile. Hence, all the channel pairs in network share common delay profile where there are L scatterers. Among these L scatterers, there are K significant scatterers where the channel coefficients are non-zero. These K significant scatterers are assumed to be $\mathcal{CN}(0, \sigma^2)$ since each path contains multiple zero mean independent identically distributed (i.i.d.) subpaths.

Without loss of generality, we consider OFDM transmission in MIMO terminals. Consider OFDM system with N subcarriers where N_p of them are selected as pilot subcarriers. Besides, to omit the interference of other
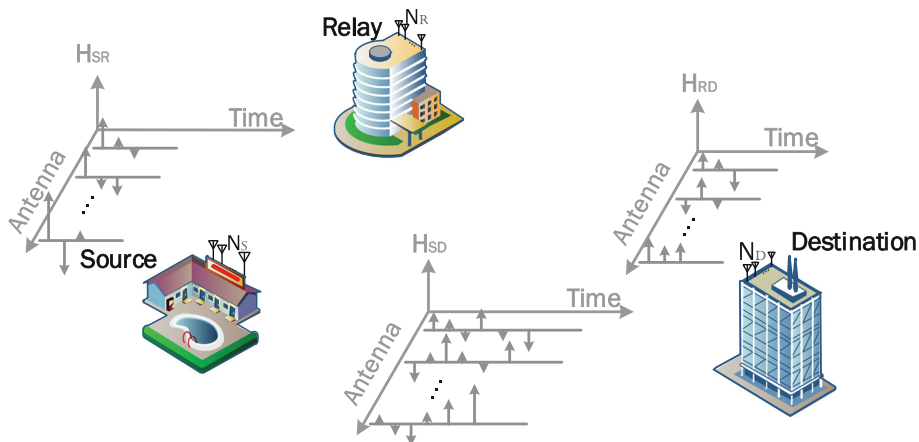
Fig. 1 Conceptual system model for MIMO-OFDM relay network and joint sparsity

antenna's pilots on each other, we consider to utilize the orthogonal pilot allocation, i.e., not only N_p subcarriers are allocated to pilot subcarriers, but also pilot subcarriers assigned to other $N_\mathbb{S} - 1$ transmit antennas are reserved to be zero [19]. Thus, the number of data subcarriers on each of the transmit antenna is equal to $N - N_\mathbb{S}N_p$, assuming $\mathbf{x}_m \in \mathbb{C}^{N \times 1}$ as the OFDM symbol before Cyclic Prefix (CP) adding to be transmitted on the mth transmit antenna of $\mathbb{S}$. In the first time slot of the Time-Division-Duplexing (TDD), the received signal $\mathbf{r}_q$ for $q = 1, 2, \ldots, N_\mathbb{R}$ is formulated as

$$\mathbf{r}_q = \sum_{m=1}^{N_\mathbb{S}} \mathbf{X}_m \mathbf{F}_m^L \mathbf{f}_{mq} + \mathbf{v}_q \tag{1}$$

where $\mathbf{X}_m$ is the diagonal matrix with $\mathbf{x}_m$ as its main diagonal, $\mathbf{F}_m^L$ is the partial Discrete Fourier Transform (DFT) matrix with N_p rows corresponding to the N_p pilot subcarriers of mth transmit antenna and first L columns of $N \times N$ DFT matrix, and $\mathbf{v}_q$ is the zero mean Additive White Gaussian Noise (AWGN) vector with variance σ_v^2 in the qth antenna of relay $\mathbb{R}$. In the second time slot, relay $\mathbb{R}$ amplifies and retransmits the received signal to the destination node $\mathbb{D}$. As a consequence, the received pilots in the nth antenna of destination $\mathbb{D}$ could be formulated as

$$\mathbf{y}_n = \sum_{q=1}^{N_\mathbb{R}} \mathbf{X}_m \mathbf{F}_m^{2L-1} \mathbf{h}_{mn} + \mathbf{z}_n \tag{2}$$

where $\mathbf{h}_{mn} = \sum_{q=1}^{N_\mathbb{R}} \beta \mathbf{f}_{mq} * \mathbf{g}_{qn}$ is the overall channel between $\mathbb{S}$ and $\mathbb{D}$ passing by $\mathbb{R}$ and β is the relay power factor which is selected as $\beta = \sqrt{\frac{P_\mathbb{R}}{K\sigma^2 P_\mathbb{S} + \sigma_v^2}}$ which is selected to keep the long-term observation relay power as $P_\mathbb{R}$ and $\mathbf{F}_m^{2L-1}$ is the partial DFT matrix with N_p rows corresponding to the N_p pilot subcarriers of mth transmit antenna and first $2L - 1$ columns of $N \times N$ DFT matrix. Moreover, $\mathbf{z}_n$ is the zero mean AWGN noise vector with variance σ_z^2. Collecting all the received pilots corresponding to all the antennas as $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_{N_\mathbb{D}}^T\right]^T$, we can represent Eq. (2) using matrix representation as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{h} + \mathbf{z} \tag{3}$$

where $\mathbf{\Phi}$ is the overall measurement matrix defined as Eq. (2), $\mathbf{h} \in \mathbb{C}^{N_\mathbb{S}N_\mathbb{D}(2L-1) \times 1}$ is the overall channel vector defined as $\mathbf{h} = \left[\mathbf{h}_{11}^T, \ldots, \mathbf{h}_{1N_\mathbb{D}}^T, \mathbf{h}_{21}^T, \ldots, \mathbf{h}_{N_\mathbb{S}N_\mathbb{D}}^T\right]^T$, and $\mathbf{z} \in \mathbb{C}^{N_\mathbb{D}N_p \times 1}$ is the overall zero-mean noise vector.

In order to represent the channel vectors in Eq. (3) in a block sparse form, we used a permutation matrix called $\mathbf{W}$ with $\mathbf{w}_j$ as its columns where $\mathbf{w}_j$ is a $N_\mathbb{S}N_\mathbb{D}(2L-1) \times 1$ zero vector except on the jth position which is equal to one. .2Since the different paths of different channels should be positioned beside each other, we define permutation matrix $\mathbf{W} = \{\mathbf{w}_j\}$ with $j = s(2L - 1) + l$ for $s = 0, 1, \ldots, N_\mathbb{S}N_\mathbb{D} - 1$ and $l = 1, 2, \ldots, 2L - 1$ where s is increased first and after $s = N_\mathbb{S}N_\mathbb{D} - 1$ then l is increased one unit and s begins again from the first value. Besides, $\mathbf{W}\mathbf{W}^T = \mathbf{I}$. As an example, for the case of $N_\mathbb{S} = 2$, $N_\mathbb{D} = 2$, and $L = 2$, permutation matrix is defined as $\mathbf{W} = \left[\mathbf{w}_1^T, \mathbf{w}_4^T, \mathbf{w}_7^T, \mathbf{w}_{10}^T, \mathbf{w}_2^T, \mathbf{w}_5^T, \mathbf{w}_8^T, \mathbf{w}_{11}^T, \mathbf{w}_3^T, \mathbf{w}_6^T, \mathbf{w}_9^T, \mathbf{w}_{12}^T\right]^T$. Consequently, using permutation matrix we can reformulate the received pilot as follows;

$$\begin{aligned} \mathbf{y} &= \mathbf{\Phi}\mathbf{W}\mathbf{W}^T\mathbf{h} + \mathbf{z} \\ &= \mathbf{\Psi}\mathbf{e} + \mathbf{z} \end{aligned} \tag{5}$$

where $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{W}$ and $\mathbf{e} = \mathbf{W}^T\mathbf{h}$.

3 Distributed compressed sensing channel estimation

CS-based approaches try to estimate channels individually while DCS-based channel estimation tries to estimate jointly sparse channels altogether. There are different channels in MIMO systems between the various transmitting and receiving antennas. Since all the transmit antennas gathered near each other and receive antennas as well, the significant scatterers which are encountered by the transmitted signals are the same. As a consequence, the sparsity pattern of different transmit-receive pairs would be the same while the channel coefficient would be completely different. Thus, utilizing DCS would be very useful. As a result, we have used the CoSaMP method and define it in a distributed manner to estimate MIMO channels.

3.1 B-vCoSaMP algorithm

In this subsection, we represent the B-vCoSaMP. Apparently, the CoSaMP algorithm utilizes the channel sparsity. Here, we used B-CoSaMP algorithm to estimate the joint support of the channel responses. Finally, after calculating the support of the channels, the channel impulse responses are estimated using LS-based estimation. The algorithm is represented in Algorithm 1.

$$\mathbf{\Phi} = \begin{pmatrix} \beta\mathbf{X}_1\mathbf{F}_1^{2L-1} & \ldots & \beta\mathbf{X}_{N_\mathbb{S}}\mathbf{F}_{N_\mathbb{S}}^{2L-1} & \mathbf{0} & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \beta\mathbf{X}_1\mathbf{F}_1^{2L-1} & \ldots & \beta\mathbf{X}_{N_\mathbb{S}}\mathbf{F}_{N_\mathbb{S}}^{2L-1} & \mathbf{0} & \ldots & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \ldots & \ldots & \ldots & \ldots & \ldots & \mathbf{0} & \beta\mathbf{X}_1\mathbf{F}_1^{2L-1} & \ldots & \beta\mathbf{X}_{N_\mathbb{S}}\mathbf{F}_{N_\mathbb{S}}^{2L-1} \end{pmatrix} \in \mathbb{C}^{N_\mathbb{D}N_p \times N_\mathbb{S}N_\mathbb{D}(2L-1)}$$

$$\tag{4}$$

Algorithm 1 B-vCoSaMP-based channel estimation

   Input: the measurement matrix $\mathbf{\Psi}$, the received pilots $\mathbf{e}$ and sparsity order K

   Output: the estimation of channel impulse responses $\hat{\mathbf{h}}_{mn}$ for $m = 1, 2, \ldots, N_{\mathbb{S}}$ and $n = 1, 2, \ldots, N_{\mathbb{D}}$

   Step 0: $\hat{\mathbf{e}} = \mathbf{0}$, $\mathbf{v} = \mathbf{e}$ and $t = 0$

   Iteration:

   Step 1: Proxy of the signal is formed by $\boldsymbol{\alpha} = \mathbf{\Psi}^H \mathbf{v}$

   Step 2: Reform the proxy to exploit the block sparsity

$$\boldsymbol{\lambda}(l) = \sum_{i=1}^{Q} \boldsymbol{\alpha}((l-1)Q + i)$$

   Step 3: The $2K$ largest components are exploited $\Omega = \text{supp}\{\boldsymbol{\lambda}_{2K}\}$

   Step 4: Merging supports of the pervious iteration and the present one $\hat{T} = \Omega \cup \text{supp}\{\mathbf{a}^{t-1}\}$

   Step 5: Define the indexes corresponding to $\hat{T}$ in other antennas by $T = [1 : Q] + (\hat{T} - 1)Q$

   Step 6: Solve the LS problem to obtain coefficients

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{\Psi}_T \mathbf{b}\|_2^2$$

   Step 7: Reform the LS based coefficients to exploit block sparsity $\hat{\mathbf{b}}(l) = \sum_{i=1}^{Q} |\mathbf{b}((l-1)Q + i)|$

   Step 8: Prune to obtain the next approximation $\mathbf{a}^t = \hat{\mathbf{b}}_K$

   Step 9: Update the samples by $\mathbf{v} = \mathbf{y} - \mathbf{\Psi}_T \mathbf{b}^t$

   Step 10: If stop criteria not met goto **Step 1**, otherwise continue with the last results

   Step 11: Solve the LS problem

$$\min_{\hat{\mathbf{e}}} \|\mathbf{y} - \mathbf{\Psi}_T \hat{\mathbf{e}}\|_2^2$$

   Step 12: $\mathbf{h} = \mathbf{U}^T \hat{\mathbf{e}}$

Obviously, exchanging the formation of the vectors between DCS and CS are repeated in Algorithm 1. Consequently, we try to explain it here. As explained before, $\mathbf{e} = [e_1, e_2, \ldots, e_{Q(2L-1)}]^T$ is the DCS-formed channel model where $Q = N_{\mathbb{S}} N_{\mathbb{D}}$. Actually, e_i for $i = 1, 2, \ldots, Q$ are related to the first path of different channel pairs and $i = Q + 1, Q + 2, \ldots, 2Q$ are related to the second path of different channel pairs. Consequently, we can represent it by e_i for $i = (l-1)Q + [1 : Q]$ and $l = [1 : 2L - 1]$ to clarify the DCS-formed representation. Hence, in Algorithm 1, to sum up all the paths in one vector, we should sum every Q-element of DCS-formed vector and consider it as the corresponding paths. This reformation is represented in **Step 2** of Algorithm 1. Actually, in **Step 2**, we try to accumulate the calculated measurements coherently in order to organize the proxy. Moreover, to represent the DCS form into CS form, we should extend the DCS form

selections by using $(\hat{T} - 1)Q + [1 : Q]$, where $\hat{T}$ is selected path from DCS form representation.

We consider the B-CoSaMP algorithm to represent the support of the vectors. Actually, support estimation determines the non-zero paths of channel estimation vector. To improve the accuracy of support estimation, block representation helps CoSaMP algorithm. Consequently, using B-CoSaMP we determine the significant paths of all the channel pairs in the network. Besides, at the end of the algorithm, we decide to calculate channel coefficients by LS equation to improve the accuracy of the estimation. LS equation is solved based on the support of the algorithm which is extracted from B-CoSaMP. In other words, B-CoSaMP represented that LS estimation should be solved only on the significant paths which are determined. Hence, we combine LS channel estimation and B-CoSaMP together and generate the B-vCoSaMP. We should consider that in the LS part of the channel estimation, the LS equation is only solved for K indices of the measurement matrix which are estimated by B-CoSaMP. In fact, we used two interactive tools called block representation and LS estimation to increase the accuracy of estimation. Block representation enhances the precision of support estimation, and LS estimation improves the accuracy of coefficient estimation. Accordingly, two main features of channel estimation are increased attractively by exploiting the block representation and LS estimation. Consequently, in simulation results, the performance is developed expectedly.

3.2 Complexity analysis

Here, we discuss the complex multiplications which are loaded in each of the states in Algorithm 1 to represent the complexity of the algorithm. Firstly, in **Step 1** of Algorithm 1, the number of complex multiplications to proceed the proxy is $N_{\mathbb{D}} N_p$. Furthermore, in **Step 6** $N_{\mathbb{D}} N_p$ complex multiplications are performed. Besides, $N_{\mathbb{S}} N_{\mathbb{D}} (2L - 1)$ complex numbers are multiplied in **Step 9** and finally in **Step 11** number of complex multiplications is $N_{\mathbb{D}} N_p$. Hence, the overall complexity of the proposed algorithm is of order $3N_{\mathbb{D}} N_p + N_{\mathbb{S}} N_{\mathbb{D}} (2L - 1)$.

4 Pilot allocation for compressed channel estimation

CS-based approaches try to estimate channels individually while DCS-based channel estimation tries to determine jointly sparse channels altogether. As a result of adjacent antenna sitting, the sparsity pattern of different transmit-receive pairs would be the same while the channel coefficient would be completely different. Thus, utilizing DCS-based channel estimation would be very useful. The measurement matrix could be generated using random pilot subcarriers to estimate the channels. Moreover, increasing the accuracy and bandwidth efficiency

of channel estimation could be guaranteed using optimal pilot allocation based on RIP optimization. As stated, since there is no polynomial time approach to evaluating RIP, instead, mutual coherence is used which is computationally tractable. Hence, we introduce the mutual coherence at first and then minimize it using proposed population-based algorithms.

Here, we consider the SISO-OFDM system for simplicity and then extend the results to MIMO case. The mutual coherence of measurement matrix is the orthogonality between the columns of it and can be defined as

$$\mu\{\boldsymbol{\Phi}\} = \max_{i \neq j} \frac{\phi_i^H \phi_j}{\|\phi_i\| \|\phi_j\|} \tag{6}$$

where $\boldsymbol{\Phi}$ is the measurement matrix and ϕ_i is the ith column of it. According to the previous section, the measurement matrix is $\boldsymbol{\Phi} = \mathbf{X}_m \mathbf{F}_m^{2L-1}$. Consequently, by considering equal energy pilot symbols and unit energy partial Fourier transform matrix $\mathbf{F}_m^{2L-1}$, we can define the mutual coherence using

$$\mu\{\boldsymbol{\Phi}\} = \max_{\lambda_i, \lambda_j; \lambda_i \neq \lambda_j} \left| \sum_{l=0}^{2L-2} e^{-j2\pi(\lambda_1 - \lambda_2)l/N} \right| \tag{7}$$

where λ_i and λ_j are pilot subcarriers among N available subcarriers. In other words, mutual coherence can be expressed as the maximum off-diagonal entry of Gram matrix $\boldsymbol{G}\{\boldsymbol{\Phi}\} = \boldsymbol{\Phi}^H \boldsymbol{\Phi}$ if $\boldsymbol{\Phi}$ is orthonormal [20]. Obviously, to minimize the mutual coherence, we have to choose pilot subcarriers which minimize Eq. (7). Moreover, in this equation, the pilot positions λ_i and λ_j are selectable and could be considered as the design criteria. Besides, this problem is combinatorial and should be optimized using iterative search methods. Here we use probability-based methods to minimize Eq. (7). In the following section, the proposed algorithms are explained. To estimate channel using DCS-based approach, we should extend the mutual coherence to MIMO case.

4.1 Extension to MIMO

In order to design optimal pilot sequences for different transmit antennas in MIMO-OFDM relay networks, we should define the mutual coherence for measurement matrix $\boldsymbol{\Psi}$ in Eq. (5). Since $\boldsymbol{\Psi}$ is a column-wise permutation of $\boldsymbol{\Phi}$, at first, we should consider the impact of permutation of columns on the mutual coherence.

Theorem 1. *Assuming $\boldsymbol{\Phi}$ as an orthonormal measurement matrix and its columns as ϕ_i for $i = 1, 2, \ldots, n$. We*

define $\bar{\boldsymbol{\Phi}}$ as a column-wise permutation of $\boldsymbol{\Phi}$. Then, it can be demonstrated that

$$\mu\{\boldsymbol{\Phi}\} = \mu\{\bar{\boldsymbol{\Phi}}\} \tag{8}$$

We consider $\boldsymbol{\Phi} = [\phi_1, \phi_2, \phi_3, \ldots, \phi_n]$ and without loss of generality consider $\bar{\boldsymbol{\Phi}}$ as the substitution of the second and third columns of $\boldsymbol{\Phi}$; consequently, it can be expressed as $\bar{\boldsymbol{\Phi}} = [\phi_1, \phi_3, \phi_2, \ldots, \phi_n]$. According to the definition of mutual coherence, we should calculate $\boldsymbol{G}\{\boldsymbol{\Phi}\} = \boldsymbol{\Phi}^H \boldsymbol{\Phi}$ as below:

$$\boldsymbol{G}\{\boldsymbol{\Phi}\} = \begin{pmatrix} \phi_1^H\phi_1 & \phi_1^H\phi_2 & \phi_1^H\phi_3 & \cdots & \phi_1^H\phi_n \\ \phi_2^H\phi_1 & \phi_2^H\phi_2 & \phi_2^H\phi_3 & \cdots & \phi_2^H\phi_n \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ \phi_n^H\phi_1 & \phi_n^H\phi_2 & \phi_n^H\phi_3 & \cdots & \phi_n^H\phi_n \end{pmatrix}. \tag{9}$$

Thus, $\mu\{\boldsymbol{\Phi}\} = \max_{i \neq j, 1 \leq i,j \leq n} |\phi_i^H\phi_j|$. On the other hand, it can be represented that

$$\boldsymbol{G}\{\bar{\boldsymbol{\Phi}}\} = \begin{pmatrix} \phi_1^H\phi_1 & \phi_1^H\phi_3 & \phi_1^H\phi_2 & \cdots & \phi_1^H\phi_n \\ \phi_3^H\phi_1 & \phi_3^H\phi_3 & \phi_3^H\phi_2 & \cdots & \phi_3^H\phi_n \\ \phi_2^H\phi_1 & \phi_2^H\phi_3 & \phi_2^H\phi_2 & \cdots & \phi_2^H\phi_n \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ \phi_n^H\phi_1 & \phi_n^H\phi_3 & \phi_n^H\phi_2 & \cdots & \phi_n^H\phi_n \end{pmatrix}. \tag{10}$$

Consequently, $\mu\{\bar{\boldsymbol{\Phi}}\} = \max_{i \neq j, 1 \leq i,j \leq n} |\phi_i^H\phi_j| = \mu\{\boldsymbol{\Phi}\}$ which indicates the equality of both mutual coherence. Therefore, the permutation of measurement matrix columns will not effect the mutual coherence value.

Considering Theorem 1, one can deduce that instead of calculating $\mu\{\boldsymbol{\Psi}\}$, it is enough to calculate $\mu\{\boldsymbol{\Phi}\}$ since $\boldsymbol{\Psi}$ is the column-wise permutation of $\boldsymbol{\Phi}$. According to Eq. (2), we can represent $\boldsymbol{\Phi}$ as

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} & \cdots & \mathbf{0} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Gamma} \end{pmatrix} \tag{11}$$

where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \beta\mathbf{X}_1\mathbf{F}_1^{2L-1} & \beta\mathbf{X}_2\mathbf{F}_2^{2L-1} & \cdots & \beta\mathbf{X}_{N_\mathbb{S}}\mathbf{F}_{N_\mathbb{S}}^{2L-1} \end{pmatrix} \tag{12}$$

is $N_P \times N_\mathbb{S}(2L-1)$ complex matrix with $\boldsymbol{\Gamma}_i = \beta\mathbf{X}_i\mathbf{F}_i^{2L-1}$ for $i = 1, 2, \ldots, N_\mathbb{S}$ as the sub-matrices. Moreover, we consider $2L - 1$ columns of $\boldsymbol{\Gamma}_i$ as $\boldsymbol{\gamma}_{i,j}$ for $j = 1, 2, \ldots, 2L-1$. Hence, $\boldsymbol{\Gamma}_i = [\boldsymbol{\gamma}_{i,1}, \boldsymbol{\gamma}_{i,2}, \ldots, \boldsymbol{\gamma}_{i,2L-1}]$. In order to calculate the mutual coherence of $\boldsymbol{\Phi}$, we should consider its Gram matrix $\boldsymbol{G}\{\boldsymbol{\Phi}\} = \boldsymbol{\Phi}^H \boldsymbol{\Phi}$. Accordingly, using Eqs. (11) and (12), we can formulate it as in Eq. (3.2) where $\mathbf{0}$ is $(2L - 1) \times (2L - 1)$ zero matrix.

$$\boldsymbol{G}\{\boldsymbol{\Phi}\} = \begin{pmatrix} \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_{N_\mathbb{S}} & \mathbf{0} & \cdots & & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_{N_\mathbb{S}} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots \cdots & \cdots & & \ddots & \ddots & & \ddots & \cdots \\ \mathbf{0} & \cdots \cdots & \cdots & & \cdots & \mathbf{0} & \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_1^H\boldsymbol{\Gamma}_{N_\mathbb{S}} \end{pmatrix} \in \mathbb{C}^{N_\mathbb{D}N_\mathbb{S}(2L-1) \times N_\mathbb{D}N_\mathbb{S}(2L-1)} \tag{13}$$

Therefore, the mutual coherence of measurement matrix in DCS-based channel estimation for MIMO-OFDM system consists of mutual coherence of different SISO-OFDM systems two by two. Consequently, we can consider the mutual coherence using measurement matrix $\boldsymbol{\Psi}$ or calculating the following:

$$\mu\{\boldsymbol{\Psi}\} = \max_{1 \leq i,j \leq N_{\mathbb{S}}} \mu\{\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_j\} \qquad (14)$$

where $\mu\{\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_i\} = \mu\{\boldsymbol{\Gamma}_i\}$ and $\mu\{\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_j\}$ is the largest off-diagonal absolute value of $\boldsymbol{\Gamma}_i^H \boldsymbol{\Gamma}_j$. Hence, in order to calculate the mutual coherence of the MIMO-OFDM mutual coherence, it is sufficient to calculate all the available SISO-OFDM pairs and consider the largest value as the mutual coherence.

Another important consideration is the orthogonality of the pilot sequences for each of the transmit antennas. In fact, to omit the interference of the different pilot sequences, we generate orthogonal pilot sequences and then transmit zero subcarriers in the other antennas' pilot subcarriers. Consequently, in creating different pilot sequences, we should consider orthogonality of the pilot sequences. The pilot allocation for a $N = 16$ subcarrier MIMO-OFDM system utilizing orthogonal pilot placement and $N_{\mathbb{S}} = 2$ transmit antennas depicted in Fig. 2. In the following, we would consider algorithms to design pilot sequences.

5 Proposed pilot allocation algorithms

In this section, we will propose three algorithms based on the population and probability, to optimize the combinatorial problem in Eq. (6). As mentioned, these algorithms work based on the probability function and try to sample the PDF using different populations and do this until convergence of the PDF to the steady state. Here, $\max_{i,j;i \neq j} \left| \sum_{l=0}^{2L-2} e^{-j2\pi(\lambda_m - \lambda_n)l/N} \right|$ is the fitness function which should be optimized. In order to decrease the computational complexity in fitness function we used Eq. (14). Evidently, there are lots of zero elements in matrix $\boldsymbol{\Psi}$ which are unnecessary to be multiplied. Hence, we used Eq. (14) to optimize the computations in fitness function evaluation. Moreover, this optimization is done over the

search space $\mathcal{S}$ which is composed of $\binom{N}{N_p}$ candidate. Since the exhaustive search over search space $\mathcal{S}$ is intractable, we consider probability-based approaches to optimize it using initial generation $\mathcal{G}^0$ where I different individuals are found in each generation.

5.1 Sequential Cross-Entropy-based pilot allocation approach

As discussed, this method is a probability-based approach which constructs each generation towards better objective function. In fact, SCE samples from a PDF and learns the new population from J best individuals in the current generation. Since sampling from PDF is used as the primary approach in producing different generations, the SCE will be more robust than other population-based algorithms which work by mutation and crossover in trapping into a local minimum. Here, we need a PDF called $\mathcal{P}$ to generate individuals by sampling it. At first, we consider the fitness function, generation, and individuals as before. In this approach, to construct each generation the PDF $\mathcal{P}$ is used. This PDF represents the probability of selection of each element of the individual. Construction of this PDF is based on J elite individuals with best objective function and the PDF of the previous iteration. Mathematically, we assume each generation individuals by a vector of probability

$$p_g(\mathcal{G}) = \left[p_g(\mathcal{G}_1), p_g(\mathcal{G}_2), \ldots, p_g(\mathcal{G}_N) \right]^T \qquad (15)$$

where $p_g(\mathcal{G}_i)$ refers to the probability of obtaining a value of 1 in the ith element of $\mathcal{G}$. At each generation, using the probability vector, $p_g(\mathcal{G})$, I individuals are obtained. In order to do random sampling by the PDF of $p_g(\mathcal{G})$, we first generate the Cumulative Density Function (CDF) $(\mathbb{P})$, and then, using a random number between 0 and 1 and the inverse of CDF $\mathbb{P}$, the sample is randomly generated. Each of these I individuals are evaluated and J best of them $(J \leq I)$ are selected. We denote them by $\mathcal{G}_{1:I}^g, \mathcal{G}_{2:I}^g, \ldots, \mathcal{G}_{J:I}^g$. These selected individuals are used to update the probability vector by

$$p_{g+1}(\mathcal{G}) = (1 - \tau) p_g(\mathcal{G}) + \frac{\tau}{J} \sum_{j=1}^{J} \mathcal{G}_{j:M}^g \qquad (16)$$
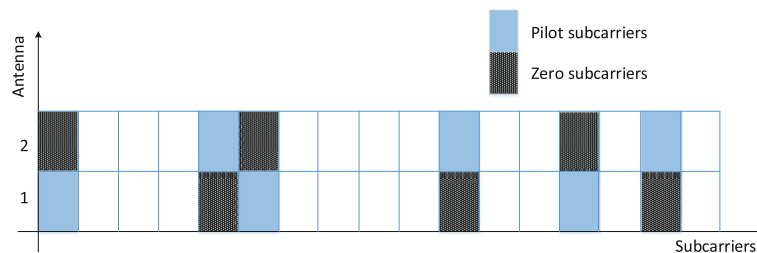
Fig. 2 Orthogonal pilot allocation representation using $N = 16$ subcarriers and $N_p = 3$ pilot

where $\tau \in (0, 1]$ is a parameter of algorithm. Since the probability behavior of populations is evident, possibility of getting stuck in a local minimum is lower. The stage-by-stage algorithm is represented in Algorithm 2. The stop criteria is based on the PDF, and when the probability function contains zero and $1/N_{\mathbb{S}}N_p$, it means that the best possible value for pilot allocation is obtained.

Algorithm 2 Sequential Cross-Entropy-based pilot allocation

1: Initialize the population $\mathcal{G}^0$ using I random individual

2: **while** stop criteria not met **do**
3:      Sort the individuals in $\mathcal{G}^g$ in ascending order according to fitness function
4:      Select J of the individuals in $\mathcal{G}^g$ with the best objective value and establish subset $\mathcal{G}_{j:I}^g$ for $j = 1, 2, \ldots, J$

5:      Construct PDF $p_g(\mathcal{G})$ using $\mathcal{G}_{j:I}^g$ and Eq. (16)
6:      Generate the CDF $\mathbb{P}_g$ using the $p_g(\mathcal{G})$
7:      **for** $i = 1$ **to** I **do**
8:          Generate $N_{\mathbb{S}}N_p$ random numbers in interval $[0, 1]$
9:          Using the inverse of CDF ($\mathbb{P}_g^{-1}$), sample the probability function and generate the new individual without repeat and generate the new generation $\mathcal{G}_{g+1}$
10:     **end for**
11: **end while**

5.2 Special case of SCE ($\tau = 1$)—extension of [18]

Considering PDF update criteria in Eq. (16), if we assume $\tau = 1$, the update process will not be affected by the previous generation, i.e.,

$$p_{g+1}(\mathcal{G}) = \frac{\tau}{J} \sum_{j=1}^{J} \mathcal{G}_{j:M}^g \tag{17}$$

As a result, the memory of PDF will be omitted. Wang et. al. in [18] proposed this approach called as the EDA for the SISO-OFDM system. Here, we extend the EDA-based algorithm for generating $N_{\mathbb{S}}$ different orthogonal sequences for the MIMO-OFDM system. In [18], it is noted that matching the sequences with the best sequence in every stage is mandatory, but in the following theorem, it is proved that matching to the sequence is not mandatory at all, since matching the sequences is obtained using circular shifting of the unmatched sequences. This phenomenon is considered in the following Theorem.

Theorem 2. *Assuming $\mathbf{\Phi}$ as an orthonormal measurement matrix and defining $\bar{\mathbf{\Phi}}$ as a row-wise permutation of $\mathbf{\Phi}$. Then, it can be demonstrated that*

$$\mu\{\mathbf{\Phi}\} = \mu\{\bar{\mathbf{\Phi}}\} \tag{18}$$

Here, we consider $\mathbf{\Phi} = [\phi_1; \phi_2; \phi_3; \ldots; \phi_n]$ where ϕ_i is $1 \times L$ complex vector. On the other hand, we assume $\bar{\mathbf{\Phi}} = [\phi_1; \phi_3; \phi_2; \ldots; \phi_n]$, then mutual coherence for both of the measurement matrices can be calculated as the maximum off-diagonal entry of $\sum_{i=1}^n \phi_i^H \phi_i$. Consequently, the mutual coherence for both of the matrices are the same and row-wise permutation does not change the mutual coherence.

Specifically, in the EDA-based algorithm which is represented in [18], after circular shifting J best sequences, the PDF of the current stage is calculated by counting the number of used subcarriers in J selected placements. Obviously, the counting process is independent of the circular shifting. Another drawback of the algorithm can be considered as trapping in the local minima, since the current PDF is considered without taking into account previous ones. Apparently, the PDF is stuck in local minima more rapidly, and thus, the algorithm would be terminated. Although the speed of convergence is more than SCE algorithm, the accuracy is less since local minima trapping has occurred more rapidly. This algorithm is considered in Algorithm 3.

Algorithm 3 Extended EDA-based pilot allocation

1: Initialize the population $\mathcal{G}^0$ using I random individual

2: **while** stop criteria not met **do**
3:      Sort the individuals in $\mathcal{G}^g$ in ascending order according to fitness function
4:      Select J of the individuals in $\mathcal{G}^g$ with the best objective value and establish subset $\mathcal{G}_{j:I}^g$ for $j = 1, 2, \ldots, J$

5:      Construct PDF $p_g(\mathcal{G})$ using $\mathcal{G}_{j:I}^g$ and Eq. (17)
6:      Generate the CDF $\mathbb{P}_g$ using the $p_g(\mathcal{G})$
7:      **for** $i = 1$ **to** I **do**
8:          Generate $N_{\mathbb{S}}N_p$ random numbers in interval $[0, 1]$
9:          Using the inverse of CDF ($\mathbb{P}_g^{-1}$), sample the probability function and generate the new individual without repeat and generate the new generation $\mathcal{G}_{g+1}$
10:     **end for**
11: **end while**

5.3 Parallel Cross-Entropy-based pilot allocation algorithm

As mentioned, in SCE-based pilot allocation algorithm, the risk of trapping in local minima is lower than the EEDA-based approach. But, the convergence speed is absolutely inferior to the other EEDA-based approach. In order to improve the convergence speed of the SCE-based

algorithm, we have proposed PCE-based pilot allocation plan. PCE considers α number of points in search space $\mathcal{S}$ and tries to converge the PDF of these points to steady state, simultaneously. Hence, since α points are considered simultaneously, trapping in one local minimum is lowered than before. Moreover, working on a different number of PDFs at the same time accelerates the convergence of the proposed method. This algorithm is represented in Algorithm 4.

As demonstrated in Algorithm 4, the difference of the PCE and SCE is considered in pipelining the update PDF and generations. In other words, in each iteration of the algorithm, generation update is performed on α pipelined branches. Moreover, these α pipelined branches are utilized to update α different PDFs. As a consequence, the random sampling is done over α different PDFs, where α different generations are constructed. This approach uses more computational calculations in each iteration which can be handled by parallel-processing supporting hardware. In the other hand, we obtain two significant advantages including conservative behavior in impacting by local minima trapping and lower number of iterations in convergence. Thus, utilizing PCE-based pilot allocation will confirm two critical aspects of pilot allocation algorithms.

Algorithm 4 Parallel Cross-Entropy-based pilot allocation

1: Initialize the population $\mathcal{G}_{1:\alpha}^0$ using I random individual
2: **while** stop criteria not met **do**
3:    **for** $lo = 1$ to α **do**
4:       Sort the individuals in $\mathcal{G}_{lo}^g$ in ascending order
5:       Select J of the individuals in $\mathcal{G}_{lo}^g$ with the best objective value and establish subset $\mathcal{G}_{lo,j:I}^g$ for $j = 1, 2, \ldots, J$
6:       Construct PDF $p_g^{lo}(\mathcal{G})$ using $\mathcal{G}_{lo,j:I}^g$ and Eq. (16)
7:       Generate the CDF $\mathbb{P}_g^{lo}$ using the $p_g^{lo}(\mathcal{G})$
8:       **for** $i = 1$ to I **do**
9:          Generate $N_{\mathbb{S}}N_p$ random numbers in interval $[0, 1]$
10:          Using the inverse of CDF, sample the probability function and generate the new individual without repeat and generate the new generation $\mathcal{G}_{g+1}^{lo}$
11:       **end for**
12:   **end for**
13: **end while**

---

6 Numerical results
In this section, the numerical results are collected to represent the performance of the proposed SCE-, EEDA-, and PCE-based pilot allocation algorithms. In all the simulations, the number of OFDM subcarriers is $N = 512$, and 16-ary quadrature amplitude modulation (16-QAM) with Gray coding is utilized. Among N available subcarriers, there are N_p subcarriers as pilot subcarriers. Various values are assigned to N_p in different simulations. Without loss of generality, in MIMO scenarios the number of transmitting and receiving antennas for all the nodes are 2, i.e., $N_{\mathbb{S}} = N_{\mathbb{R}} = N_{\mathbb{D}} = 2$. Sparse Rayleigh fading channels in the networks are modeled using Finite Impulse Response (FIR) filters with L taps where K number of taps are non-zero and are i.i.d. using zero mean and unit variance complex Gaussian distribution. Furthermore, all the results are averaged over 1000 independent channel ensembles.

6.1 Comparison of the proposed algorithms
In this part, we compare the performance of the proposed algorithms in view of decreasing μ. To evaluate the performance, we consider $N_p = 25$ pilot subcarriers for each antenna which distributed orthogonally using SCE, EEDA, and PCE approaches and $L = 25$ channel taps with $K = 3$ significant non-zero. As a comparison standard, we consider the optimized pilot allocation in [11] where GA was utilized to generate orthogonal pilot sequences. The comparison is evaluated using the minimal mutual coherence μ and distribution of μ. Distribution of μ is illustrated in Fig. 3. Specifically, in EEDA-based approach (Fig. 3b), the convergence is acquired in 52nd iteration, where it is acquired for PCE-based approach (Fig. 3a) and SCE-based approach (Fig. 3c) in 123rd and 186th, respectively. Expectedly, the distribution of μ in each iteration is getting closer in all the three methods. It is caused by the convergence of probability function which is sampled in each iteration. As can be seen, the value of minimum μ in the last iteration is better for the case of PCE-based approach, since the algorithm examines some of the local minimum points together through the global minimum while in the other approaches the algorithm examines just one point. Specifically, the μ values are with a sample pilot sequence which is obtained in each algorithm and are compared with each other in Table 1. GA-based approach algorithm is used to generate the pilot sequence, as well. As represented in Table 1, the minimum achieved mutual coherence is 0.2227 reported by PCE-based approach which was expected for the sake of multiple point examining. Moreover, the supremacy of the proposed SCE-based approach to the EEDA-based approach is represented by the difference between 0.2360 and 0.2439. Finally, the obtained value by GA-based approach in [11] is the worse approach. Hence, three proposed methods can be ordered as PCE-, SCE-, and EEDA-based approaches from minimum mutual coherence approach while all of them are
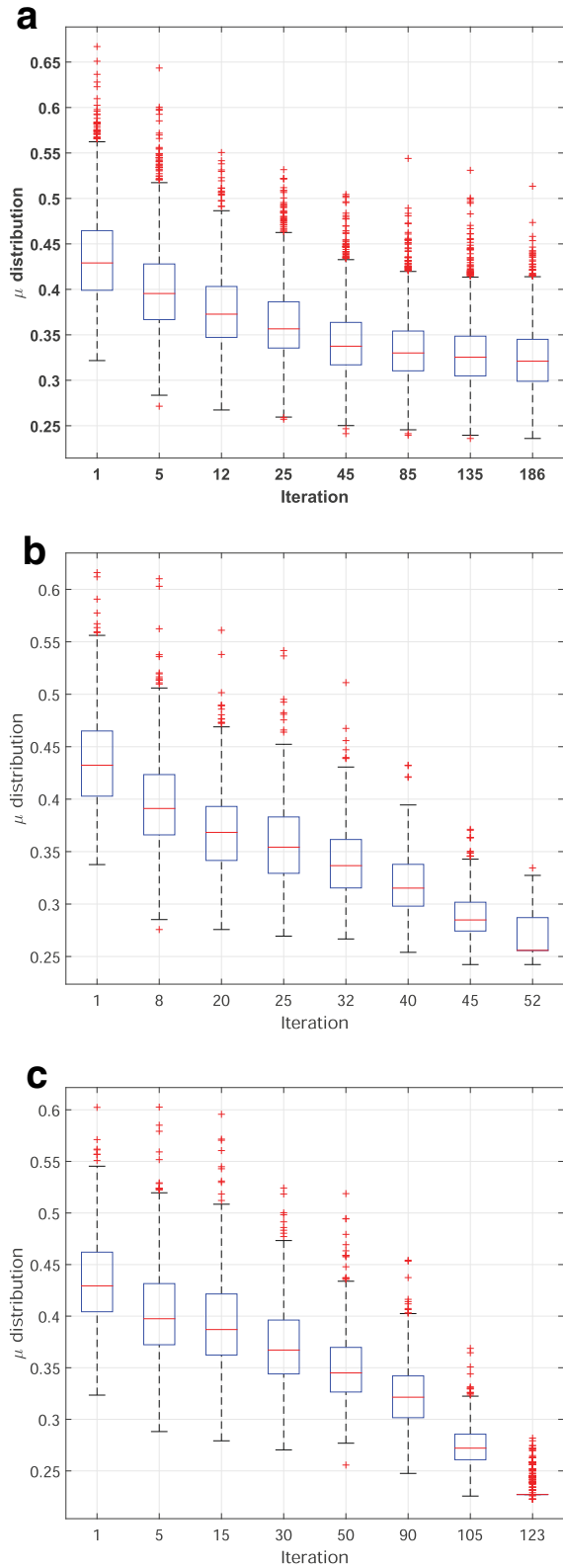
Fig. 3 μ distribution for different proposed algorithms. **a** SCE algorithm, **b** EEDA algorithm, and **c** PCE algorithm

Table 1 Sample pilot sequences and their mutual coherence which have been achieved by different proposed approaches and comparison to the method in [11]

Alg.	SCE		EEDA		PCE		GA	
μ	0.2360		0.2439		0.2227		0.2557	
N_s	1	2	1	2	1	2	1	2
	6	9	31	36	29	31	18	55
	14	22	42	44	47	49	20	61
	33	37	52	71	57	63	36	99
Pilot	40	44	80	87	64	71	44	132
seq.	51	54	92	142	72	110	66	148
	56	59	149	150	115	116	86	160
	62	66	155	186	121	124	116	179
	95	99	187	194	198	200	124	196
	103	105	195	203	207	208	135	203
	112	135	205	210	217	218	145	222
	151	152	215	217	227	228	166	230
	158	212	223	225	236	237	177	238
	219	242	230	233	247	248	185	256
	243	249	237	241	255	257	197	263
	252	257	244	251	278	280	237	277
	276	318	263	291	287	289	252	295
	324	325	292	299	298	308	306	307
	344	349	311	329	317	318	323	331
	354	355	330	335	326	346	348	352
	372	373	349	409	348	355	357	381
	413	420	414	418	357	362	385	419
	430	431	423	426	363	412	421	428
	436	483	429	437	414	434	446	434
	487	493	445	469	439	466	484	447
	502	512	498	500	470	483	490	483

better than GA-based approach. In the following section, we will consider the performance of the proposed pilot sequences utilizing the Normalized Mean Square Error (NMSE) and Bit Error Rate (BER) performance metrics.

6.2 Evaluating MSE and BER performance

Here, we consider the applicability of the proposed algorithm and the obtained pilot sequences utilizing two main characteristics of channel estimator called NMSE and BER. The channel is characterized by $L = 25$ taps where $K = 3$ of them are significant non-zero taps. NMSE is calculated using

$$\mathrm{NMSE} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \frac{\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2}{\|\mathbf{h}_i\|_2^2} \tag{19}$$

where $\mathbf{h}$ is the complete channel vector and $\hat{\mathbf{h}}$ is its estimation. Moreover, N_{MC} is the Monte Carlo iteration which is 1000. BER is evaluated using Monte Carlo simulation using N_{MC} individual simulation.

In Fig. 4 these two performance metrics are compared. Apparently, decreasing the mutual coherence will cause a better NMSE metric and BER performance. As illustrated in Fig. 4a, the MSE of PCE enjoys from the mutual coherence approximately 1 and 2.5 dB rather than SCE- and EEDA-based approaches, respectively. Moreover, from BER point of view, the BER gain caused by mutual coherence supremacy is approximately 3 and 6 dB for $BER = 2 \times 10^{-2}$ to SCE and EEDA-based approaches, respectively. The BER for the case of *Full CSI* is represented as a comparison metric. Evidently, by utilizing $N_p = 25$ optimized pilots on each antenna,

the performance is almost 3 dB away from the optimal performance which could be compensated using other facilities. Actually, using optimized pilots will lead to 90% spectral efficiency which is one of the most critical characteristics of the system.

In order to compare the results with previous approaches in the literature, we have done the simulations using *GA*-based pilot allocation approach together with *PCE*-based approach. Hence, these two methods could be compared from spectral efficiency point of view utilizing NMSE and BER performance metrics (Fig. 5). For the simulations, the number of $N_p = 30$ is selected. The performance comparison represents the superiority of the proposed method from NMSE and BER performance
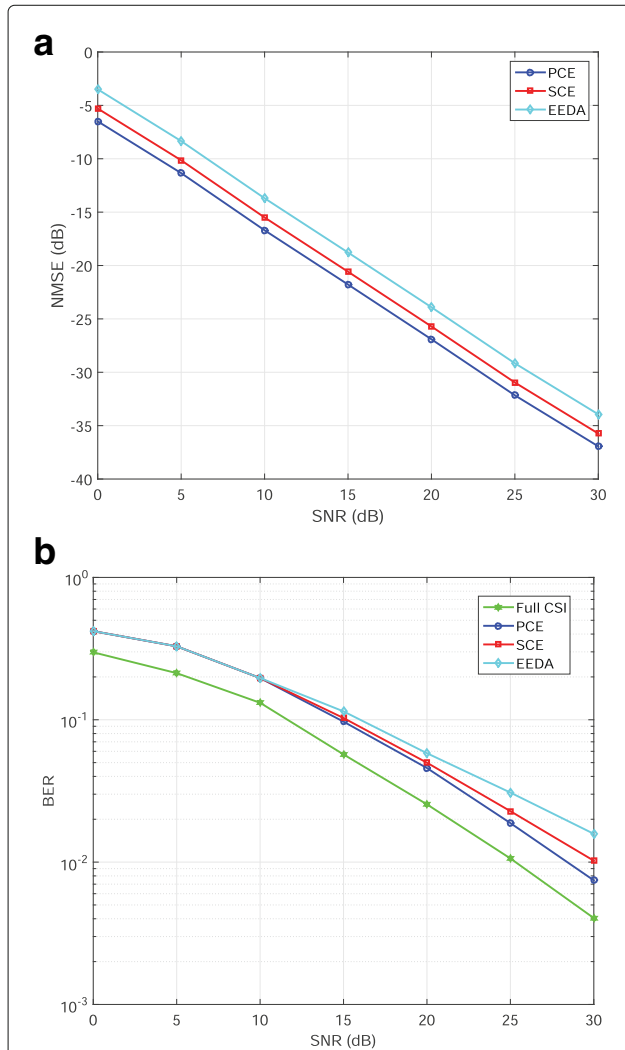
Fig. 4 Performance comparison for different pilot allocation algorithms. **a** NMSE performance metric and **b** BER performance metric
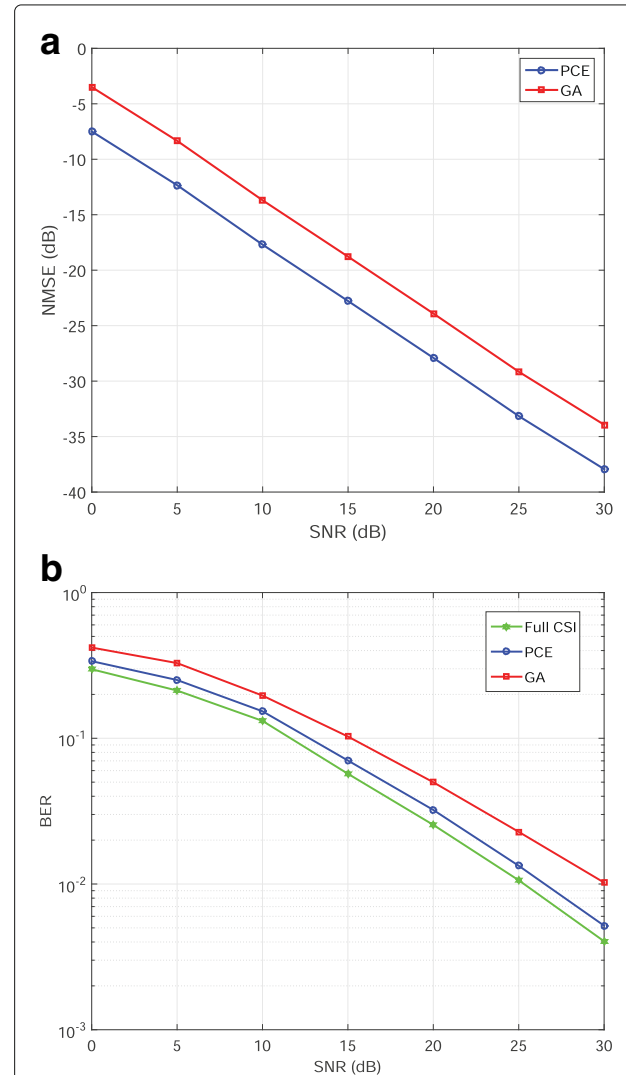
Fig. 5 Performance comparison of proposed method and GA-based pilot allocation algorithm. **a** NMSE performance metric and **b** BER performance metric

metric overview. Apparently, in NMSE and BER the PCE-based approach outperforms the GA-based algorithm almost 4 dB, quantitatively. Furthermore, comparing Figs. 4 and 5, the performance of GA-based algorithm for $N_p = 30$ is identical to the SCE-based approach for $N_p = 25$ pilot numbers. Hence, utilizing the proposed methods, the spectral efficiency will increase more than 1% rather than other spectral efficient DCS-based approaches.

6.3 Performance comparison with the other estimation method

In this part of simulation results, the simulations are performed in comparison to the well-known CoSaMP algorithm in order to evaluate the benefits of the proposed method. In this simulation, the number of channel path is considered to be $L = 30$, where $K = 3$ of them are significant non-zero paths. Moreover, the number of pilot subcarriers for each transmitting antenna is considered to be $N_p = 40$. To evaluate the performance merits of the proposed methods, we considered two different scenarios. In the first scenario the pilot allocation is performed randomly, while in the second scenario the pilot is allocated using the proposed PCE-based algorithm. The BER results of them are represented in Fig. 6. As illustrated in Fig. 6a, using the random pilots together with the CoSaMP-based channel estimation method, results in irreducible BER around 10^{-1} while this is not encountered in case of proposed B-vCoSaMP-based channel estimation method until 10^{-2}. Moreover, utilizing proposed PCE-based pilot allocation which is shown in Fig. 6b, the BER is decreased in both cases but the proposed method is still better than the other method and approximately 2 dB away from the ideal case.

### 6.4 The effect of significant paths of channel

The last simulation results are devoted to the impact of channel significant paths on the channel estimation performance. In order to evaluate the performance, we change the significant paths from $K = 2$ to $K = 4$. The BER performance is compared using random pilot allocation and optimized PCE-based pilot allocation. The channel paths are again $L = 30$ and the pilot number is $N_p = 40$ for each of the transmitting antennas. Expectedly, increasing the number of significant path will decrease the sparsity order of the channel which makes its estimation harder. As a consequence, using $N_p = 40$ random orthogonal pilots, will not be applicable in case of $K = 4$ significant paths (BER is shown in Fig. 7). Moreover, for the case of $K = 3$, utilizing optimized PCE-based pilots will improve the performance while it is comparable with the case of $K = 2$ and approximately 1 dB away from $K = 2$, while this gap is around 3.5 dB in case of random pilot allocation.
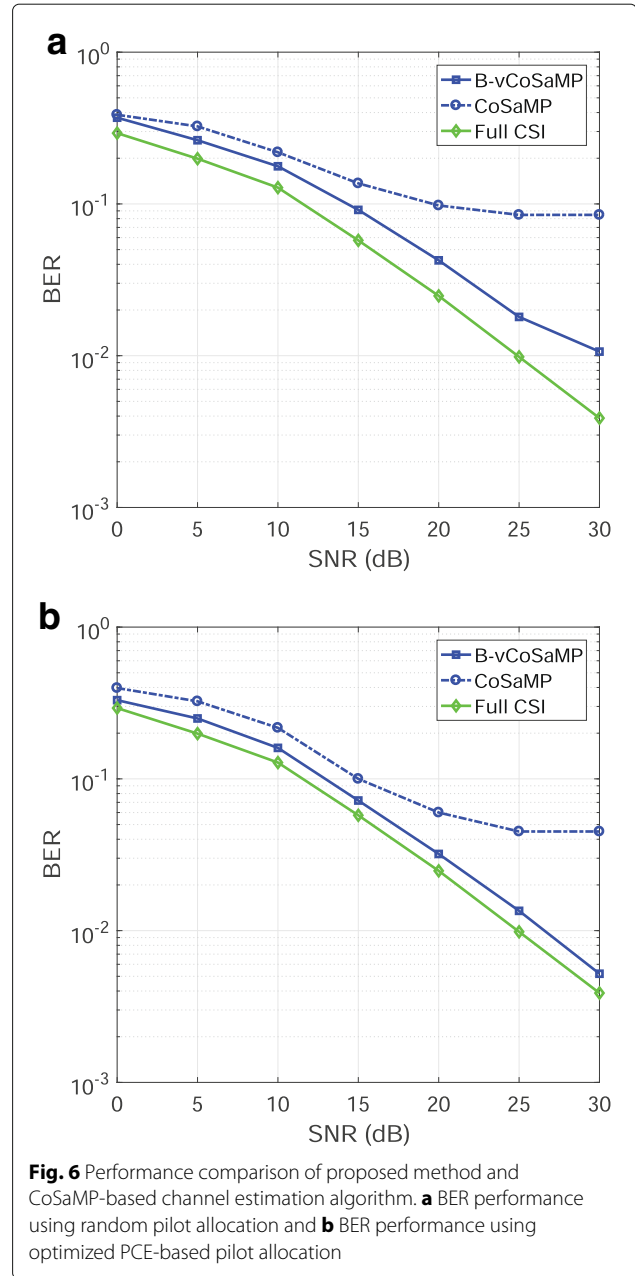
Fig. 6 Performance comparison of proposed method and CoSaMP-based channel estimation algorithm. **a** BER performance using random pilot allocation and **b** BER performance using optimized PCE-based pilot allocation

7 Conclusions

In this paper, we have considered the channel estimation problem in AF MIMO-OFDM relay networks using DCS-based approaches. Firstly, the channel estimation method proposed is called B-vCoSaMP where the channel estimation is improved utilizing block sparsity of MIMO channels and LS estimation. In order to improve the performance of the estimation, we design three different algorithms to minimize the mutual coherence of the resultant channel estimation measurement matrix. The proposed methods are based on the Cross-Entropy optimization and include SCE-, EEDA-, and PCE-based

Fig. 7 Performance comparison of proposed method in case of different significant paths of the channel. **a** BER performance using random pilot allocation and **b** BER performance using optimized PCE-based pilot allocation
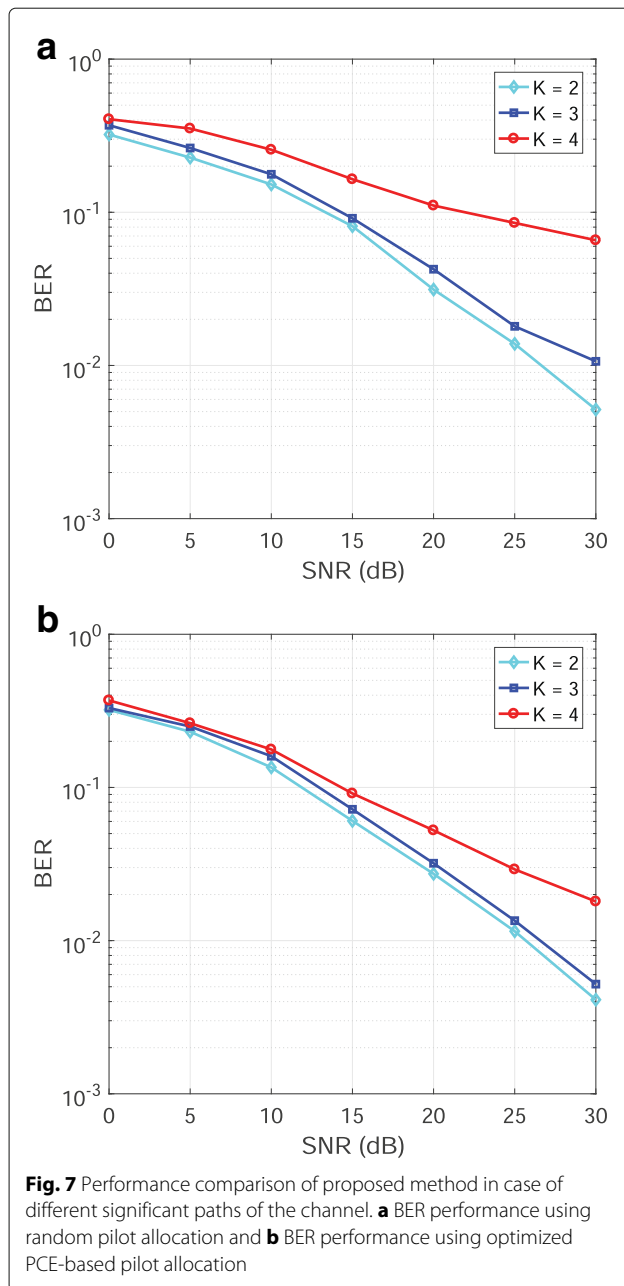
approaches. Utilizing just one local minimum point in SCE and EEDA, the performance is worse than PCE-based approach. Multi-point tracking ability of PCE makes it possible to increase the accuracy and speed of the pilot sequence determination algorithm.

References
1. D Hu, X Wang, L He, A new sparse channel estimation and tracking method for time-varying ofdm systems. IEEE Trans. Veh. Technol. **62**(9), 4648–4653 (2013)
2. E Soleimani-Nasab, A Kalantari, M Ardebilipour, Performance analysis of multi-antenna df relay networks over nakagami-m fading channels. IEEE Commun. Lett. **15**(12), 1372–1374 (2011)
3. C Qi, G Yue, L Wu, N Arumugam, Pilot design for sparse channel estimation in ofdm-based cognitive radio systems. IEEE Trans. Veh. Technol. **63**(2), 982–987 (2014)
4. P Cheng, L Gui, Y Rui, YJ Guo, X Huang, W Zhang, Compressed sensing based channel estimation for two-way relay networks. IEEE Wirel. Commun. Lett. **1**(3), 201–204 (2012)
5. Z Gao, L Dai, B Shim, Z Wang, Structured compressive sensing-based spatio-temporal joint channel estimation for fdd massive mimo. IEEE Trans. Commun. **64**(2), 601–617 (2016)
6. MF Duarte, S Sarvotham, D Baron, MB Wakin, RG Baraniuk, in *Proc. Asilomar Conf. Signals, Syst., Comput.* Distributed compressed sensing of jointly sparse signals (IEEE, Pacific Grove, 2005), pp. 1537–1541
7. P Cheng, Z Chen, Y Rui, YJ Guo, L Gui, M Tao, QT Zhang, Channel estimation for ofdm systems over doubly selective channels: a distributed compressive sensing based approach. IEEE Trans. Commun. **61**(10), 4173–4185 (2013)
8. L Xu, K Niu, Z He, W Xu, Z Zheng, in *Proc. 9th ICICS*. Mimo channel estimation based on distributed compressed sensing for lte-advanced (IEEE, Tainan, 2013), pp. 1–5
9. X Rao, VKN Lau, Distributed compressive csit estimation and feedback for fdd multi-user massive mimo systems. IEEE Trans. Signal Process. **62**(12), 3261–3271 (2014)
10. J Tropp, Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory. **50**(10), 2231–2242 (2004)
11. X He, R Song, WP Zhu, Pilot allocation for sparse channel estimation in mimo-ofdm systems. IEEE Trans. Circuits Syst. II, Exp. Briefs. **60**(9), 612–616 (2013)
12. X He, R Song, WP Zhu, Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in mimo-ofdm systems. IEEE Trans. Veh. Technol. **65**(5), 2990–3004 (2016)
13. C Qi, G Yue, L Wu, Y Huang, A Nallathan, Pilot design schemes for sparse channel estimation in ofdm systems. IEEE Trans. Veh. Technol. **64**(4), 1493–1505 (2015)
14. X He, R Song, WP Zhu, Pilot allocation for sparse channel estimation in mimo-ofdm systems. IEEE Trans. Circuits Syst. II, Exp. Briefs. **60**(9), 612–616 (2013)
15. C Qi, L Wu, A study of deterministic pilot allocation for sparse channel estimation in ofdm systems. IEEE Commun. Lett. **16**(5), 742–744 (2012)
16. C Qi, L Wu, Optimized pilot placement for sparse channel estimation in ofdm systems. IEEE Signal Process. Lett. **18**(12), 749–752 (2011)

17. X He, R Song, in *Proc. IEEE Int. Conf. WCSP*. Pilot pattern optimization for compressed sensing based sparse channel estimation in ofdm systems (IEEE, Suzhou, 2010), pp. 1–5
18. H Wang, Q Guo, G Zhang, G Li, W Xiang, Pilot pattern optimization for sparse channel estimation in ofdm systems. IEEE Commun. Lett. **19**(7), 1233–1236 (2015)
19. A Akbarpour-Kasgari, M Ardebilipour, F Afghah, in *Proc. 49th CISS*. Mimo-ofdm non-regenerative relay channel estimation using compressed sensing (IEEE, Baltimore, 2015), pp. 1–5
20. JM Duarte-Carvajalino, G Sapiro, Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization. IEEE Trans. Image Proc. **18**(7), 1395–1408 (2009)