Decoupling Channel Estimation for FDD Massive MIMO Systems Utilizing Joint Sparsity

XIANGYU YAN*, LI CHEN*, HUARUI YIN*, AND WEIDONG WANG*
Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China
Corresponding author: Li Chen (chenli87@ustc.edu.cn)

This work was supported in part by the National Key Research and Development Project under Grant 2018YFB1801105, and in part by the USTC Research Funds of the Double First-Class Initiative under Grant YD3500002001.

ABSTRACT
Effective channel estimation for massive multiple-input-multiple-output (MIMO) systems based on frequency division duplex (FDD) protocol is a crucial problem. To reduce pilot overhead, we consider the joint sparsity of the multi-path channel in delay-angle domain. Based on the joint sparsity, we propose a decoupling pilot design scheme. In the proposed scheme, the pilot design is decoupled into two parts. In the first part, a global optimal selection greedy iterative algorithm (GOS-GIA) is proposed to obtain the near-optimal pilot subcarrier pattern. In the second part, a random Rademacher distribution pilot matrix is used as the angle-domain pilot matrix. Accordingly, a two-stage channel estimation strategy is also provided and analyzed. The first stage of the strategy is concerned with retrieving the positions of non-zero dominant taps in the delay domain. The second stage focuses on estimating the channel coefficients at these taps. Algorithm complexity analysis shows that the GOS-GIA can reduce the computational complexity at least one order of magnitude, and the proposed channel estimation strategy reduces the computational complexity by more than two orders of magnitude. Simulation results verify that the proposed pilot design scheme achieves better performance with lower pilot overhead.

INDEX TERMS
Angle domain, channel estimation, compressive sensing, delay domain, massive multiple-input-multiple-output, pilot design, sparsity.

I. INTRODUCTION
Massive multiple-input-multiple-output (MIMO) systems have attracted much attention in recent years, which can enhance the spectrum efficiency and energy efficiency by orders of magnitude [1], [2]. Due to these benefits depend on the accuracy of estimating the channel state information (CSI) at the base station, the estimation of CSI has become a great challenge [3].

TDD protocol based on channel reciprocity is usually applied to estimate the CSI of massive MIMO systems. By utilizing the channel reciprocity, downlink channel estimation can be realized through uplink training. The number of orthogonal pilot sequences required is proportional to the number of users instead of the number of BS antennas [1], [2], [4]. While, TDD massive MIMO systems also bring a lot of challenges, such as pilot contamination and reciprocity calibration of hardware chains.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Gao*.

Compared to the TDD protocol, FDD protocol can provide better CSI estimation performance in terms of transmission delay, communication range, and mobility support [5]–[10]. The challenge of CSI estimation for FDD protocol is serious pilot overhead [11]. That is because the number of training pilots required for channel estimation increases rapidly with the number of BS antennas. How to effectively design pilots and estimate the CSI is essential to FDD massive MIMO systems.

A. RELATED WORK
The traditional pilot designs are based on orthogonal schemes, such as time-domain orthogonal pilot design, frequency-domain orthogonal pilot design, code-domain orthogonal pilot design, and time-frequency orthogonal pilot design [12]. All these orthogonal pilot designs require the number of pilots to be proportional to the number of BS antennas and the quantized channel delay spread. The pilot overhead is prohibitively high in massive MIMO systems.
To reduce the pilot overhead, some non-orthogonal pilot designs and channel estimation techniques based on channel statistics have been proposed for massive MIMO systems [9], [10], [13]–[16]. Based on the long-term channel statistics at the user side, [17] proposed an open-loop and closed-loop channel estimation strategy for massive MIMO by exploiting the practical MIMO channels correlation in time and space. An optimal pilot beam pattern design for channel estimation was proposed in [18] by considering the antenna correlation and temporal correlation. By utilizing the structure of the second-order statistics of the channel vectors and the correlation of users, the author [19] proposed a joint spatial division and multiplexing (JSDM) based channel estimation. In [20], [21], blind channel estimation schemes were proposed based on the covariance matrix of the received signals and subspace partitioning. Nevertheless, in these statistics based channel estimation schemes, the acquisition of the perfect long-term channel statistics is challenging in practice. And the pilot overhead is still very high.

To further reduce the pilot overhead, compressive sensing (CS) techniques have also been used to guide the pilot design and channel estimation by exploiting channel sparsity in recent works of literature. In [21]–[26], it was pointed out that the channel impulse response (CIR) in the delay domain tends to be sparse due to the limited local scatterers at the BS. By exploiting the channel sparsity in the delay domain, one superimposed pilot design was employed to reduce pilot overhead in [22]. In [24], the authors presented a novel channel estimation approach that utilizes the sparsity and common support properties in the delay domain. The channel has also been explored to be sparse in the angle domain [12]. Paper [5], [27] considered the hidden joint common and individual sparsity in angle domain among the user channel matrices due to the shared local scatterers in the physical propagation environment. A nonorthogonal downlink pilot design was proposed in [8], [12], [14], [28] by exploiting the spatially common sparsity in angle domain and time correlation of massive MIMO channels. The channel sparsity in angle domain with partial support information was utilized in [29], [30] to acquire compressed CSI. However, all the works including the above-mentioned ones have rarely considered the joint sparsity in the delay-angle domain for massive MIMO systems. To the best of our knowledge, there are no related works of literature about the pilot design and channel estimation for massive MIMO systems with joint sparsity in the delay-angle domain.

**B. CONTRIBUTIONS AND ORGANIZATION**

In this paper, we explore the joint sparsity of multi-path massive MIMO channel in the delay-angle domain. By exploiting the joint sparsity, a decoupling pilot design scheme with low overhead and low complexity is proposed for FDD massive MIMO systems. In the proposed pilot design scheme, the pilot design is decoupled into two parts. In the first part, a global optimal selection greedy iterative algorithm (GOS-GIA) is developed to obtain the near-optimal pilot subcarrier pattern. The proposed GOS-GIA algorithm is based on the mutual incoherence property (MIP) of the pilot matrix. In the second part, a random Rademacher distribution pilot matrix is used as the angle-domain pilot matrix. Based on the proposed pilot design scheme, one corresponding two-stage channel estimation strategy is provided. The strategy is based on CS with exploiting the joint sparsity in the delay-angle domain. The first stage of the strategy is concerned with retrieving the positions of non-zero dominant taps in the delay domain. To retrieve the non-zero taps, one block subspace pursuit (B-SP) algorithm is developed by exploiting the sparse common support (SCS). The second stage focuses on estimating the channel coefficients at these taps. One parallel subspace pursuit (P-SP) algorithm is developed to estimate the coefficients. Additionally, we analyze the algorithm complexity and system performance of the proposed pilot design scheme and the proposed channel estimation. Algorithm complexity analysis shows that the GOS-GIA can reduce the computational complexity at least one order of magnitude, and the proposed channel estimation strategy reduces the computational complexity by more than two orders of magnitude. Simulation results verify that the proposed scheme achieves better performance with lower pilot overhead.

The contributions of this work are summarized as follows:

- **The decoupling pilot design scheme**: By exploiting joint sparsity in the delay-angle domain of massive MIMO systems, we propose a decoupling pilot design scheme with low overhead and low complexity.

- **The novel pilot subcarrier pattern design algorithm**: One novel GOS-GIA algorithm is developed based on the MIP of the pilot matrix. The algorithm can obtain the near-optimal pilot subcarrier pattern with low complexity.

- **The corresponding channel estimation strategy**: Based on the proposed pilot design scheme, one corresponding two-stage channel estimation strategy is provided. The proposed strategy is based on CS with exploiting the joint sparsity in the delay-angle domain.

- **Better performance with algorithm complexity analysis and simulation**: Algorithm complexity analysis shows that the proposed channel estimation strategy reduces the computational complexity by more than two orders of magnitude. Simulation results verify that the proposed scheme achieves better performance with lower pilot overhead.

The rest of this paper is organized as follows. In Section II, we explore the system model and joint sparsity of multi-path massive MIMO channel in the delay domain and angle domain. Section III explore the pilot decoupling scheme. The pilot design and the two-stage channel estimation strategy are proposed in Section IV and V, respectively. Section VI and VII provide algorithm complexity analysis and numerical simulation of the proposed pilot design scheme and the proposed channel estimation strategy. Conclusions are drawn in Section VIII.
Notations: Capital and small bold letters represent matrices and vectors respectively. The transpose, conjugate and Hermitian transpose of the matrix are denoted by $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$, respectively. $\| \cdot \|_F$ and $\| \cdot \|_2$ denote the Frobenius norm and the second-order norm respectively, $\mathbb{C}^{M \times N}$ is the set of complex matrices with $M$ rows and $N$ columns. $\mathcal{C}N(\mu, \sigma^2)$ denotes the circular symmetric complex Gaussian distribution with mean $\mu$ and standard deviation $\sigma$. vec$(X)$ denotes the vectorization of the matrix $X$ formed by stacking the columns of $X$ into a single column vector. $\mathcal{A}\setminus\mathcal{B}$ denotes the difference of sets $\mathcal{A}$ and $\mathcal{B}$.

II. SYSTEM AND CHANNEL MODEL

In this paper, we consider a cell from a massive MIMO cellular system. In the cell, a BS equipped with $M$ antennas simultaneously serves $K$ single-antenna users on $N$ subcarriers. We consider FDD downlink systems, where each frame includes the pilot training phase, feedback phase, and data transmission phase. In the pilot training phase, the BS transmits pilot sequences to users to estimate channel vectors. Then the estimated channel vectors are fed back to the BS in the feedback phase. In the data transmission phase, the BS uses the estimated channel matrix for precoding. In this paper, we focus on the pilot training phase.

To simplify the notations, in the sequel we take an arbitrary user in the cell as an example to introduce the channel model and different representations. In this paper, the block-fading channel is assumed, where the channel gain remains static in each frame but may vary from one to another. Each frame contains $R$ consecutive OFDM symbols.

The multi-path channel is considered. There are $L$ channel paths between the transmit antennas and the user. The CIR of the $l$-th path between the $m$-th transmit antenna and the user is denoted as $h_{l,m}$. Then, the channel matrix between BS antennas and the user is

$$\mathbf{H} = [h_{l,m}]_{L \times M}. \quad (1)$$

Due to the limited scattering in the propagation environment of typical cellular massive MIMO systems with tower-mounted BS, the channels are highly correlated in both frequency and space domains. To explore the correlation of $\mathbf{H}$, we discuss the sparsity of the channel matrix in the delay domain and angle domain respectively.

1) CHANNEL SPARSITY IN DELAY DOMAIN

The CIRs of $L$ paths between the $m$-th transmit antenna and the user are given by

$$\mathbf{h}_m = [h_{0,m}, h_{1,m}, \ldots, h_{L-1,m}]^T \in \mathbb{C}^{L \times 1}. \quad (2)$$

$\mathbf{h}_m$ is the delay channel vector, and $\mathbf{h}_{m_l}$ is the $l$-th column of the channel matrix $\mathbf{H}$.

For multi-path channels, numerous experimental results and theoretical analyses have confirmed that $\mathbf{h}_m$ is sparse in delay domain [22]–[26], [31]–[33].

Meanwhile, the CIRs measured at the user from different antennas share a sparse common support (SCS) [34]. The support of $\mathbf{h}_m$ is written as

$$\mathcal{T}_m = \{ l : |h_{l,m}| > 0 \}. \quad (3)$$

Since different antennas share the SCS channel, the positions of the non-zero elements of $\mathbf{h}_m$ are the same for $m = 1, 2, \ldots, M$, i.e.,

$$\mathcal{T}_1 = \mathcal{T}_2 = \cdots = \mathcal{T}_M \triangleq \mathcal{T}, \quad (4)$$

and

$$|\mathcal{T}_1| = |\mathcal{T}_2| = \cdots = |\mathcal{T}_M| = T \ll L. \quad (5)$$

where $T$ is the number of no-zero entries of all $L$ quantized channel paths, which stands for the number of resolvable paths in the analog domain. As the channel sparsity in the delay domain, $L \ll T$.

2) CHANNEL SPARSITY IN ANGLE DOMAIN

The $l$-th path of the channel between BS antennas and the certain user can be denoted as

$$\mathbf{g}_l = [h_{l,1}, h_{l,2}, \ldots, h_{l,M}] \in \mathbb{C}^{1 \times M}. \quad (6)$$

And $\mathbf{g}_l$ is the $l$-th row of the channel matrix $\mathbf{H}$.

According to the spatial channel model (SCM) in [33], the channel vector $\mathbf{g}_l$ is expressed as

$$\mathbf{g}_l = \sum_{q=1}^{Q} \beta_{l,q} \mathbf{a}(\theta_l + \Delta \theta_{l,q}), \quad (7)$$

where $\beta_{l,q}$ denotes the complex gain; $\theta_l$ is the mean angle of arrival (AoA) of the $l$-th path; $Q$ is the number of subpaths within each path; $\Delta \theta_{l,q}$ is the random angular deviation; and $\mathbf{a}(\theta)$ is the array response vector.

The channel vector $\mathbf{g}_l$ can be modeled using the virtual angular representation [35]

$$\mathbf{g}_l = \mathbf{g}_l^v \mathbf{A}_l^H, \quad (8)$$

where $\mathbf{A}_l^H \in \mathbb{C}^{M \times M}$ is the unitary matrix representing the transformation matrix of the virtual angular domain at the BS side; and $\mathbf{g}_l^v \in \mathbb{C}^{1 \times M}$ is the angle domain channel vector.
The $p$-th entry of $\mathbf{g}_p^T$ is non-zero indicates that there is a spatial path from the $p$-th transmit direction of the BS to the user. As the BS is usually elevated high with few scatterers around, the angle spread at the BS side is small. As indicated by the experimental study [36], since the angle spread is limited at the BS, only a few of the entries in $\mathbf{g}_p^T$ show a significant magnitude. That’s to say, $\mathbf{g}_p^T$ is in general sparse [5], [12], [36]–[38].

Denote $S_l$ as the support of vector $\mathbf{g}_p^T$, i.e.,

$$S_l = \{m : |\mathbf{g}_p^T| > 0\}. \quad (9)$$

As the angle spread for different paths is similar, we assume that

$$|S_0| = |S_1| = \cdots = |S_{L-1}| \triangleq S \ll M. \quad (10)$$

The unitary matrix $\mathbf{A}_l$ is determined by the geometrical structure of the BS antenna array. When the BS antenna array is uniform linear array (ULA) or uniform planar array (UPA), the unitary matrix $\mathbf{A}_l$ becomes a discrete Fourier transform (DFT) matrix $[35], [37]$. To illustrate the basic principle of the proposed pilot design scheme, in the sequel we consider ULA at the BS. Then, the array response vector can be expressed as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \begin{bmatrix} 1, e^{-j2\pi \frac{\lambda}{d} \sin(\theta)}, \ldots, e^{-j2\pi \frac{\lambda}{d} \sin(\theta)(M-1)} \end{bmatrix}, \quad (11)$$

where $\lambda$ is the wavelength and $d$ is the antenna spacing. And

$$\mathbf{A}_0 = \mathbf{A}_1 = \cdots \mathbf{A}_{L-1} \triangleq \mathbf{F}_M, \quad (12)$$

where $\mathbf{F}_M$ is a standard $M$-point DFT matrix.

Combining with Section II-1 and Section II-2, the channel matrix $\mathbf{H}$ is joint sparse in the delay-angle domain. And it can be expressed as

$$\mathbf{H} = \mathbf{H}^a \mathbf{F}_M, \quad (13)$$

where $\mathbf{H}^a = [\mathbf{g}_0^T \cdots \mathbf{g}_{L-1}^T]^T \in \mathbb{C}^{L \times M}$ is the delay-angle domain channel matrix. The matrix $\mathbf{H}^a$ is joint sparse in row and column, and rows are block-sparse.

III. PILOT DECOUPLING

In downlink massive MIMO systems, the amount of pilot overhead, which is proportional to the number of BS antennas, makes channel estimation difficult to implement. However, the sparse structure of $\mathbf{H}^a$ makes low overhead channel training feasible by exploiting CS techniques. By exploiting the joint sparsity of $\mathbf{H}^a$ in the delay-angle domain, we propose a decoupling pilot design scheme to reduce the pilot overhead.

A. CHANNEL TRAINING

As we mentioned before, the block-fading channel is assumed [39]–[42]. Each frame contains $R$ consecutive OFDM symbols, in each frame, $G$ ($G \leq R$) successive OFDM symbols are used to transmit pilots. And pilots are inserted into subcarriers, in each OFDM symbol, $P$ ($P \leq N$) subcarriers are assigned to transmit pilots and the remaining $N - P$ subcarriers are still assigned to transmit data.

As all BS antennas use identical subcarriers to transmit pilot at the same time, the received pilot signal $y(i) \in \mathbb{C}^{P \times 1}$ at the user in $i$-th OFDM symbol is expressed as

$$y(i) = \sum_{m=1}^{M} \text{diag}(\mathbf{x}_m(i))\mathbf{P}_N^L \mathbf{h}_m + \mathbf{z}(i), \quad (14)$$

where $\mathbf{P} = \mathbf{I}_N(\mathbf{P}_\zeta \cdots)$ is a $P \times N$ binary subcarrier selection matrix with only one “1” in each row and $\mathbf{P}^T = \mathbf{I}_P$; and the pilot subcarrier pattern

$$\mathbf{P}_\zeta = [\zeta_1, \zeta_2, \cdots, \zeta_P], \quad 1 \leq \zeta_p \leq N \quad (15)$$

is the index set of pilot subcarriers in $i$-th OFDM symbol. $x_m(i) \in \mathbb{C}^{P \times 1}$ is the pilot signal transmitted by the $m$-th antenna over $\mathbf{P}_\zeta^L$ in the $i$-th OFDM symbol; $\mathbf{F}_N^L \in \mathbb{C}^{N \times L}$ is the first $L$ columns of a $N$-point DFT matrix and $\mathbf{z}(i) \sim CN(0, \sigma_i \mathbf{I}_P)$ is the independent and identically distributed (i.i.d.) additive Gaussian white noise.

B. CHANNEL ESTIMATION PROBLEM FORMULATION

We collect all the $G$ OFDM symbols together to estimate the channel matrix. The collected received signal matrix can be expressed as

$$\text{vec} (\mathbf{Y}) = \mathbf{Y} \text{vec} (\mathbf{H}) + \text{vec} (\mathbf{Z}). \quad (16)$$

where $\mathbf{Y} = [y(1), \cdots, y(G)] \in \mathbb{C}^{P \times G}$ is the collected received signal matrix of all the $G$ OFDM symbols. $\mathbf{Y} = [\mathbf{X}_1^T, \mathbf{X}_2^T, \cdots, \mathbf{X}_G^T]^T$ is the $P \times MLG$ pilot matrix, where

$$\mathbf{X}_i = \{\text{diag}(\mathbf{x}_1(i))\mathbf{P}_N^L, \text{diag}(\mathbf{x}_2(i))\mathbf{P}_N^L, \cdots, \text{diag}(\mathbf{x}_M(i))\mathbf{P}_N^L\}$$

is the $P \times ML$ pilot matrix in $i$-th OFDM symbol. And the matrix $\mathbf{Z} = [\mathbf{z}(1), \mathbf{z}(2), \cdots, \mathbf{z}(G)] \in \mathbb{C}^{M \times G}$ is noise term.

By exploiting the channel sparsity, CS-based channel estimation with non-orthogonal pilot design can provide the efficient compression and reliable recovery of the sparse channel matrix. The CS-based channel estimation problem can be formulated as follows:

$$\hat{\mathbf{H}} = \arg \min_{\mathbf{H}} ||\hat{\mathbf{H}}^a||_0 \quad \text{s.t. vec} (\mathbf{Y}) = \mathbf{Y} \text{vec}(\hat{\mathbf{H}}) + \text{vec} (\mathbf{Z}) \quad \hat{\mathbf{H}} = \hat{\mathbf{H}}^a \mathbf{F}_M \quad (17)$$

The pilot matrix $\mathbf{Y}$ should be designed for channel estimation. The pilot overhead is characterized by $P$ and $G$. To reliably estimate the channel, some traditional orthogonal pilot designs are usually adopted, such as time-domain orthogonal pilot design, frequency-domain orthogonal pilot design, and time-frequency orthogonal pilot design [12]. Time-domain orthogonal pilot design requires $G \geq M$ and $P \geq L$, frequency-domain orthogonal pilot design requires $G = 1$ and $P \geq ML$ and time-frequency orthogonal pilot design.
design requires $PG \geq ML$. In general, all orthogonal pilot designs require $PG \geq ML$. Massive MIMO systems suffer from prohibitively high pilot overhead with orthogonal pilot design. CS-based channel estimation with non-orthogonal pilot design can reduce the pilot overhead. In order to ensure channel estimation accuracy with the lowest pilot overhead, the pilot matrix $Y$ should be optimized.

**Remark 1 (Reduce Algorithm Complexity):** As the size of the pilot matrix $Y$ is very large, the joint optimization of $Y$ is a challenge. It’s difficult to implement with enormous algorithm complexities. To reduce the pilot overhead and algorithm complexity, we propose a decoupling pilot design scheme to design $Y$.

### C. PROPOSED DECOUPLING DOWNLINK PILOT DESIGN SCHEME

In the proposed pilot design scheme, different BS antennas use $P$ completely identical frequency-domain subcarriers to transmit pilots over $G$ successive OFDM symbols. By exploiting the joint sparsity of the channel matrix $H$ in the delay-angle domain, non-orthogonal pilots are considered for CS-based channel estimation.

**Proposition 1 (Decoupling Downlink Pilot Design Scheme):** If we assume that transmitted power vectors of all BS antennas are the same, we can decouple the large scale size pilot matrix $Y \in \mathbb{C}^{P \times ML}$ into two small size matrices $\Phi$ and $\Psi$, where

$$
\Phi = \text{diag}(\phi)PF_N^L \in \mathbb{C}^{P \times L},
$$

$$
\Psi = S^T F_M \in \mathbb{C}^{G \times M}.
$$

And the channel estimation problem can be rewritten as

$$
\hat{H} = \arg \min_{\tilde{H}} \| \tilde{H} \|^2_0
$$

s.t.

$$
Y = \Phi \tilde{H}^a \Psi^H + Z.
$$

(20)

**Proof:** If we assume that transmitted power vectors of all BS antennas are the same, the pilot signal $x_m(i)$ can be divided into two parts:

$$
x_m(i) = s_m(i)\phi,
$$

(21)

where $s_m(i) \in \mathbb{C}^{1 \times 1}$ is normalized angular pilot signal transmitted by the $m$-th antenna in $i$-th OFDM symbol. And $\phi \in \mathbb{C}^{P \times 1}$ is transmitted power vector over the pilot subcarrier pattern $P^L$. According the decoupling expression (21), $s_m(i)$ is independent of the subcarrier index $\tilde{\xi}$, but depends on the antenna index $m$ and the OFDM symbol index $i$. While, $\phi$ and $P$ are independent of $m$ and $i$.

Then $X_i$ can be simplified as

$$
X_i = s(i)^T \otimes (\text{diag}(\phi)PF_N^L),
$$

(22)

where $s(i) = [s_1(i), \ldots, s_m(i), \ldots, s_M(i)]^T$.

The equation (16) can be rewritten as

$$
\text{vec}(Y) = S^T \otimes (\text{diag}(\phi)PF_N^L)\text{vec}(H) + \text{vec}(Z).
$$

(23)

where $S = [s(1), \ldots, s(G), \ldots, s(G)] \in \mathbb{C}^{M \times G}$.

By using the matrix equations property of Kronecker product (43), Lemma 4.3.1, $\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$, the above formula can be changed as

$$
Y = \text{diag}(\phi)PF_N^L H^a F_M S + Z.
$$

(24)

The binary subcarrier selection matrix $P$, the vector $\phi$ and the matrix $S$ should be designed. In formula (24), $\text{diag}(\phi)P$ is on the left side of the channel matrix $H^a$, and $F_M S$ is on the right side of $H^a$.

In summary, we have decoupled the large scale size pilot matrix $Y \in \mathbb{C}^{P \times ML}$ into two small size matrices $\Phi$ and $\Psi$,

$$
\Phi = \text{diag}(\phi)PF_N^L \in \mathbb{C}^{P \times L},
$$

$$
\Psi = S^T F_M \in \mathbb{C}^{G \times M},
$$

and

$$
Y = \Phi H^a \Psi^H + Z.
$$

(27)

By decoupling the large scale size pilot matrix $Y$ into two small size matrices $\Phi$ and $\Psi$, instead of design the pilot matrix $Y$, we can explore the design schemes of pilot matrices $\Phi$ and $\Psi$ respectively. As the sizes of $\Phi$ and $\Psi$ are much less than $Y$, the computational complexity of pilot design can be greatly reduced.

### IV. PILOT DESIGN

In this section, we explore the design schemes of pilot matrices $\Phi$ and $\Psi$ respectively.

#### A. DESIGN OF PILOT MATRIX $\Phi$

First of all, we explore the design scheme of the pilot matrix $\Phi$. The matrix $Y$ is rewritten as

$$
Y = \Phi \tilde{G}^a + Z,
$$

(28)

where $\tilde{G}^a = H^a \Psi^H \in \mathbb{C}^{L \times G}$ is block-sparse with only $T$ rows are non-zero.

The measurement matrix $\Phi$ plays an important role in the sparse signal recovery. Various results have been proposed about what conditions the measurement matrix should satisfy to guarantee a robust signal recovery. Restricted isometry property (RIP) [44] and mutual incoherence property (MIP) [45] are widely used. The MIP condition is stronger than the RIP in that the MIP implies the RIP but the converse is not true [45]. In this paper, we consider the measurement matrix based on MIP.

The MIP of the measurement matrix $\Phi$ is defined as the largest normalized inner product between any two columns of $\Phi$, and it is shown as

$$
\mu(\Phi) \triangleq \max_{0 \leq l_1 < l_2 \leq L-1} \frac{|\langle \Phi(l_1), \Phi(l_2) \rangle|}{\| \Phi(l_1) \|_2 \| \Phi(l_2) \|_2}
$$

$$
= \max_{0 \leq l_1 < l_2 \leq L-1} \left| \sum_{p=1}^{P} \left| \phi_p \right|^2 \exp \left[ -\frac{2\pi}{N} \cos (\beta_p (l_2 - l_1)) \right] \right|
$$

where $l_1$ and $l_2$ are the column indexes of $\Phi$. 

(29)
According to the MIP condition in [45], the recovery performance of sparse signals is better with smaller MIP. \( \Phi \) is determined by \( \mathcal{P}_\zeta \) and \( \phi \). So, we should design \( \mathcal{P}_\zeta \) and \( \phi \) to minimize the MIP, and the problem becomes

\[
\arg\min_{\mathcal{P}_\zeta, \phi} \mu(\Phi).
\]

We assume that the transmit power of the pilot symbols among all subcarriers is equal, i.e.,

\[
|\phi_1|^2 = |\phi_2|^2 = \cdots = |\phi_p|^2 = \frac{1}{P}.
\]

Specifically, we design

\[
\phi_p = e^{-j\psi_p}, \quad p = 1, 2, \cdots, P,
\]

where \( \psi_p \) has the i.i.d. uniform distribution in \((-\pi, \pi]\), namely, the i.i.d \( U(-\pi, \pi] \). Let \( r = l_2 - l_1 \). Then, \( \mu(\Phi) \) is simplified as

\[
\mu(\Phi) = \max_{1 \leq r \leq L-1} |g(r, \mathcal{P}_\zeta)|.
\]

where

\[
g(r, \mathcal{P}_\zeta) = \frac{1}{P} \sum_{p=1}^{P} \exp(-j\frac{2\pi}{N} \zeta_pr) \tag{34}
\]

is the mutual correlation between two columns of \( \Phi \) when the difference between the two column index is \( r \).

From (33), \( \mu(\Phi) \) only depends on \( \mathcal{P}_\zeta \) and it can be denoted as the function \( \mathcal{P}_\zeta \)

\[
\mu(\Phi) \triangleq \eta(\mathcal{P}_\zeta). \tag{35}
\]

As we assume that the transmit power of the pilot symbols among all subcarriers is equal, the problem of minimizing MIP can be translated into the pilot subcarrier pattern optimization problem. The pilot subcarrier pattern \( \mathcal{P}_\zeta \) can be designed to minimize the MIP, i.e.,

\[
\mathcal{P}_\zeta^{*} = \arg\min_{\mathcal{P}_\zeta} \eta(\mathcal{P}_\zeta). \tag{36}
\]

There are some pilot subcarrier patterns have been proposed for the problem (36). Intuitively, by exhaustively searching over all possible pilot patterns, the optimal pilot pattern can be obtained. But it is computationally prohibitive to search from all \( \binom{N}{P} \) candidates when \( N \) and \( P \) are large. For example, if \( N = 1024 \) and \( P = 12 \), there are \( \binom{1024}{12} = 2.6 \times 10^{27} \) different pilot patterns. Random pilot subcarrier pattern and uniform pilot subcarrier pattern are widely adopted [22], [30]. Unfortunately, random pilots and uniform pilots cannot guarantee the performance in practical systems [46]. In [23], it has been shown that the pilot subcarrier pattern generated from the cyclic different set (CDS) is optimal. Nevertheless, the CDS only exists for some specific pair of \((N, P)\). Some iterative algorithms were given in [23], [46]–[49] to obtain near-optimal pilot subcarrier patterns. However, these iterative algorithms still need high algorithm complexities.

In this paper, based on the MIP of \( \Phi \), we propose a greedy-iterative algorithm to obtain near-optimal pilot patterns for any given pair of \((N, P)\) and for any value of \( L \). The algorithm consists of two layers of loops. And it is based on a greedy search, which searches for the near-optimal pilot pattern with two loops of iterations. The main idea of the proposed algorithm is listed as follows:

**Step 1:** In the outer loop, we randomly generate pilot patterns as the initializations of the inner loop.

**Step 2:** In the inner loop, we iteratively update the resulting pilot pattern in a greedy manner. Different from the two stochastic search schemes in [23], in each iteration of the inner loop, we select the worst subcarrier and best subcarrier respectively, and the worst subcarrier is replaced by the best subcarrier.

**Step 3:** The pilot subcarrier pattern with the minimum MIP is selected from the results of inner loops.

Given the maximum numbers of iterations for the outer loop, i.e. \( N_1 \), the global optimal selection greedy iterative algorithm (GOS-GIA) is described as follows.

- In each iteration of the outer loop, we randomly generate a pilot subcarrier pattern \( \mathcal{P}_\zeta^t \) as the initialization of the inner loop. And

\[
\mathcal{P}_\zeta^t = \{\zeta_1, \zeta_2, \cdots, \zeta_P\} \subset \mathcal{N} \tag{37}
\]

with \( \mathcal{N} = \{1, 2, \cdots, N\} \).

- In each iteration of the inner loop, we select the worst subcarrier and best subcarrier respectively, and the worst subcarrier is replaced by the best subcarrier.

First of all, we aim to delete one pilot subcarrier from \( \mathcal{P}_\zeta^t \). According to the greedy idea, we should delete the worst pilot subcarrier index \( \zeta_p^d \) from \( \mathcal{P}_\zeta^t \), which results in the minimum MIP with pilot pattern \( \mathcal{P}_\zeta^t \backslash \zeta_p^d \), i.e., \( \eta(\mathcal{P}_\zeta^t \backslash \zeta_p^d) \).

\[
\zeta_p^d = \arg\min_{\zeta_p \in \mathcal{P}_\zeta^t} \eta(\mathcal{P}_\zeta^t \backslash \zeta_p^d). \tag{38}
\]

- Then, we aim to add one pilot subcarrier from set \( \mathcal{N} \backslash \mathcal{P}_\zeta^t \). The best pilot subcarrier index \( \zeta_* \) in set \( \mathcal{N} \backslash \mathcal{P}_\zeta^t \), which results in the minimal MIP with pilot pattern \( \mathcal{P}_\zeta^t \cup \{\zeta_*\} \), is given as

\[
\zeta_* = \arg\max_{\zeta \in \mathcal{N} \backslash \mathcal{P}_\zeta^t} \eta(\mathcal{P}_\zeta^t \cup \{\zeta\} \backslash \zeta_p^d). \tag{39}
\]

With \( N_1 \) outer loop iterations, we obtain \( N_1 \) optimized inner loop results, from which we select the one with the minimum MIP as the final pilot pattern.

The detailed algorithm is presented in Algorithm 1. As the inner loop is a greedy iteration, when \( \eta(\mathcal{P}_\zeta^t) > \eta(\mathcal{P}_\zeta^{*}) \), it means the inner loop has converged. For the outer loop, initialization has a great impact on the result of the inner loop. So, we repeat the inner loop \( N_1 \) times. For each initial pilot pattern in the outer loop, we obtain a corresponding optimized pilot pattern. With \( N_1 \) outer loop iterations, we then obtain \( N_1 \) optimized results, from which we select the one with the minimum MIP as the final result.
The measurement matrix with its elements following the Rademacher distribution is also commonly adopted due to its simplicity, and it also enjoys a satisfying performance. In this paper, the Rademacher distribution is adopted and $S$ is i.i.d. drawn from $\{-\sqrt{1/M}, \sqrt{1/M}\}$ with equal probability.

V. CORRESPONDING CHANNEL ESTIMATION

With the received compressive measurement $Y$ at the user, one corresponding channel estimation strategy is provided in this section.

The channel estimation can be processed either at the user side or at the BS side. If the channel estimation is processed at the user side, the user needs to have prior knowledge about $P_\xi$, $\phi$ and $S$. To reduce the computational complexity and the storage capacity at the user side, we process the channel estimation at the BS side. The user only needs to feed the compressive measurements $Y$ directly back to the BS. Then the BS should recover the channel matrix $H_a$ from the compressive measurements $Y$ with the prior knowledge of $P_\xi$, $\phi$ and $S$.

The problem of channel estimation is formulated as

$$\hat{H}_a = \arg\min_{H_a} |H_a|_0 \quad \text{s.t.} \quad Y = \Phi H_a \Psi^T + Z. \tag{43}$$

The channel matrix $H_a$ is joint sparse in row and column, and rows are block-sparse. According to the multidimensional compressive sensing based algorithm [51], the two-dimensional problem can degenerate into an one-dimensional problem, which is expressed as

$$\bar{Y} = \Omega \bar{h}_a + \bar{Z}, \tag{44}$$

in which $\bar{Y} = \text{vec}(Y) \in \mathbb{C}^{P\times G}$, $\Omega = \Psi \otimes \Phi \in \mathbb{C}^{PG \times ML}$, $\bar{h}_a = \text{vec}(H_a) \in \mathbb{C}^{ML \times 1}$ and $\bar{Z} = \text{vec}(Z)$.

As the size of $PG \times ML$ is very large, the complexity of the CS-based algorithm will be very enormous. To solve the channel estimation problem (43), we propose one two-stage channel estimation strategy with low algorithm complexity. In the proposed strategy, the large-size problem is decomposed into two small-size problems.

The main idea of the two-stage channel estimation strategy is listed as follows:

Stage 1: The first stage is concerned with retrieving the positions of non-zero dominant taps in the delay domain through exploiting the SCS channel.

Stage 2: The second stage focuses on estimating the channel coefficients at these taps through exploiting the sparsity in the angle domain.

A. FIRST STAGE

In the first stage, we estimate the positions of non-zero dominant taps in delay domain.

The estimated problem can be formulated as

$$\hat{G}_a = \arg\min_{G_a} |G_a|_0 \quad \text{s.t.} \quad Y = \Phi \hat{G}_a + Z. \tag{45}$$

B. DESIGN OF PILOT MATRIX $\Psi$

Then, we explore the design scheme of pilot matrix $\Psi$ in this section. The received signal matrix $Y$ is rewritten as

$$Y^T = \Psi \tilde{H}_a + Z, \tag{40}$$

where $\Psi = S^T F_a \in \mathbb{C}^{G \times M}$ is the angle-domain measurement matrix; and $\tilde{H}_a = H^T(\text{diag}(\phi)P N_0)^T \in \mathbb{C}^{M \times P}$ is sparse with $S$ non-zero entries in each column.

The measurement matrix $\Psi$ is also designed based on MIP. And, we have

$$\mu(\Psi) = \mu(S^T). \tag{41}$$

It implies that the MIP only depends on the $S$. Then, the optimal $S$ is given by

$$S^* = \arg\min_S \mu(S^T). \tag{42}$$

There are also many results for the problem (42). Reference [50] proves the measurement matrix with its elements following the i.i.d. complex Gaussian distribution enjoys a satisfying performance in compressing and recovering sparse signals. The measurement matrix with its elements following

**Algorithm 1 GOS-GIA**

**Input:** $N, L, P$.

**Output:** $P_\xi$.

**Initialization:**

$$P_\xi = [1, 2, \cdots, P] \times \begin{bmatrix} N \end{bmatrix}. \tag{46}$$

**Outer iteration:** Repeat the next steps $N_1$ times.

1) **Outer iteration initialization:**

$$P_\xi^t = \{\xi_1, \xi_2, \cdots, \xi_P\} \subset N; N_2 = 0. \tag{47}$$

2) **Inner iteration:**

- Inner iteration initialization:

  $$P_\xi^t = P_\xi^t; N_2 = N_2 + 1. \tag{48}$$

- Select the worst pilot subcarrier:

  Obtain $\xi_p^d$ according to (38).

- Select the best pilot subcarrier:

  Obtain $\xi^*$ according to (39).

- Update the pilot subcarrier pattern:

  $$P_\xi^t = (P_\xi^t \setminus \xi_p^d) \cup \{\xi^*\}. \tag{49}$$

- Repeat the above steps until

  $$\eta(P_\xi^t) > \eta(P_\xi^t) \tag{50}$$

3) **Update the best pilot subcarrier pattern :**

- if $\eta(P_\xi^t) < \eta(P_\xi^t)$

  $$P_\xi = P_\xi^t. \tag{51}$$

$\eta$ is given by

$$\eta(P_\xi) = \sum_{\xi_i \in P_\xi} \frac{1}{2} |p_\xi^\alpha(\xi_i)|^2 \tag{52}$$

Stage 1: The first stage is concerned with estimating the positions of non-zero dominant taps in the delay domain through exploiting the SCS channel.

Stage 2: The second stage focuses on estimating the channel coefficients at these taps through exploiting the sparsity in the angle domain.
The problem in (45) is a CS problem with the signal matrix \( \tilde{G}^a \) block-sparse. Only \( T \) rows of \( \tilde{G}^a \) are non-zero. Many CS-based algorithms can be adopted to estimate \( \hat{G}^a \), such as orthogonal matching pursuit (OMP), subspace pursuit (BP). In this paper, by exploiting the SCS of the channel, one Block-BP algorithm is applied. The detail of Block-BP algorithm is presented in Algorithm 2. Note that there are three stopping criterions for iteration, \( \| R_{t+1} \|_F > \| R_t \|_F, \| R_{t+1} \|_F < \gamma_t \) and \( t > N_3 \). The first criterion means the iteration has converged. The second criterion means the estimation error is small enough, and the last criterion guarantees the maximum computation time.

**Algorithm 2** Block-SP to Solve Problem (45)

**Input:** \( L, N, G, P, T, Y, \Phi \).

**Output:** Estimated \( \hat{G}^a \) and \( \hat{T} \).

**Initialization:**
- Initialize the chunk support \( \hat{T}_0 = \emptyset \).
- Initialize the residue matrix \( R_0 = Y \).

**Repeat**
- Update the support as follows:
  \[
  \mathcal{T}^* = \arg \max_{|\mathcal{T}|=T} \| (\Phi^H R) (\mathcal{T}, :) \|_F,
  \]
  \[
  \hat{T}_{t+1} = \hat{T}_t \cup \mathcal{T}^*.
  \]
- Estimate the signal with LS, i.e.,
  \[
  \hat{G}_p^a(\hat{T}_t, :) = (\Phi(:, \hat{T}_t)) Y^T, \quad \hat{G}_p^a((1, \ldots, L) \setminus \hat{T}_t, :) = 0.
  \]
- Refine the support as follows:
  \[
  \hat{T}_{t+1} = \arg \max_{|\mathcal{T}|=T} \| \hat{G}_p^a(\mathcal{T}, :) \|_F.
  \]
- Update the signal with LS, i.e.,
  \[
  \hat{G}_{t+1}^a(\hat{T}_{t+1}, :) = (\Phi(:, \hat{T}_{t+1})) Y^T, \quad \hat{G}_{t+1}^a((1, \ldots, L) \setminus \hat{T}_{t+1}, :) = 0.
  \]
- Update the residue, i.e.,
  \[
  R_{t+1} = Y - \Phi \hat{G}_{t+1}^a.
  \]

**Until** \( \| R_{t+1} \|_F > \| R_t \|_F \) or \( \| R_{t+1} \|_F < \gamma_t \) or \( t > N_3 \).

**Output**
- If \( \| R_{t+1} \|_F > \| R_t \|_F \) then
  \[
  \hat{T} = \hat{T}_t, \quad \hat{G}^a = \hat{G}_p^a.
  \]
- Else if \( \| R_{t+1} \|_F < \gamma_t \) or \( t > N_3 \) then
  \[
  \hat{T} = \hat{T}_{t+1}, \quad \hat{G}^a = \hat{G}_{t+1}^a.
  \]

**B. SECOND STAGE**

In the second stage, we would like to recover \( H^o \) from \( \hat{G}^a \).

We assume \( \tilde{G}^a = \hat{G}^a + E \), where \( E \) is the estimation error. The set of the non-zero rows of \( \tilde{G}^a \) is denoted as \( \tilde{T} \), i.e., \( \tilde{T} = \{ l : |\tilde{G}^a(l,:)|_2 > 0, l = 1, 2, \ldots, L \} \). We define a \( T \times L \) binary selection matrix \( T = I_L(\tilde{T}, :) \) with only one “1” in each row and \( T^H = I_T \). Then, we need to estimate \( H_o \) based on \( T \tilde{G}^a \). According to \( \tilde{G}^a = H^o \Psi^T \), we have

\[
  \hat{Y} = \Psi \tilde{X} + \tilde{Z}.
\]

The problem in (47) is also a CS problem with \( \hat{X} \) is sparse. Each column of \( \hat{X} \) has \( S \) non-zero entries. In this paper, one Parallel-BP algorithm is applied to exploit the sparse structure. The detail of the Parallel-BP algorithm is presented in Algorithm 3. The stopping criterions in this algorithm is similar to Algorithm 2.

Then, we can acquire the estimation of the channel \( \hat{H} \)

\[
  \hat{H}(T, :) = \tilde{Y} \tilde{T} \tilde{F}^H, \quad \hat{H}(1, \ldots, L) \setminus \mathcal{T} = 0.
\]

**VI. COMPUTATIONAL COMPLEXITY ANALYSIS**

In this section, we analyze the computational complexities of the proposed decoupling pilot design scheme and the corresponding channel estimation strategy.

**A. ALGORITHM COMPLEXITY OF THE PROPOSED DECOUPLING PILOT DESIGN SCHEME**

The computational complexity of the proposed decoupling pilot design scheme mainly depends on Algorithm 1. In each inner iteration of Algorithm 1, the computational complexity is mainly determined by the searching for the worst index \( \xi^d \) in (38) and the best index \( \xi^e \) in (39). The searching for the worst index \( \xi^d \) has the complexity on the order of \( O(L + P) \). The searching for the best index \( \xi^e \) has the complexity on the order of \( O((N - P)Lp) \) by directly computing (39).

The best pilot subcarrier index \( \xi^* \) in set \( \mathcal{N} \setminus \mathcal{P}_c \), which results in the minimal MIP with pilot pattern \( \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \} \). And \( g(r, \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \}) \) is the mutual correlation between two columns of \( \Phi \) with pilot pattern \( \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \} \) when the difference between the two column index is \( r \). It means \( \min_r g(r, \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \}) \) is the MIP of \( \Phi \) with pilot pattern \( \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \} \). Then, the \( \xi^* \) can be computed by one-step iteration, which is shown in the following

\[
  \xi^* = \arg \max_{\xi \in \mathcal{N} \setminus \mathcal{P}_c} \min_r \left| g(r, \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \}) \right|,
\]

where

\[
  g(r, \{ \mathcal{P}_c \setminus \xi^d \} \cup \{ \xi \}) = g(r, \mathcal{P}_c^o) + \frac{1}{p} (e^{-j2\pi r \xi^d} - e^{-j2\pi r \xi^d}).
\]
The (50) is a one-step iteration with the complexity on the order of $O(1)$. Hence, the computational complexity of searching for the best index $\xi^*$ has been reduced to the order of $O((N-P)L)$. The inner iteration number is denoted as $L_2$, and it’s determined by the speed of convergence of the inner iteration. The outer iteration number is $N_1$.

Then the computational complexity of Algorithm 1 is the order of $O(N_1L_2(N-P)L)$. We discuss the selection of the outer iteration number $N_1$ and the maximum number of inner iteration $N_2$ in Section VII. From the result, it can be seen that the convergence speed of the inner iteration and outer iteration is very fast, and the values of $N_1$ and $N_2$ can be less than 10.

Algorithm 3 Parallel-SP to Solve Problem (47)

**Input:** $M, G, S, T, \hat{\mathbf{F}}, \hat{\mathbf{Y}}, \hat{\mathbf{X}}, \Psi$

**Output:** Estimated $\hat{\mathbf{X}}$ and chunk support $\hat{\mathcal{S}}$.

for $i = 1 : T$
do

- **Initialization:**
  - Set $\mathbf{y}^i = \hat{\mathbf{Y}}(:, i)$ and $\mathbf{x}^i = \tilde{\mathbf{X}}(:, i)$.
  - Initialize the support index $t = 0$, the chunk support $\hat{\mathcal{T}}^i_t = \emptyset$ and the residue matrix $r^i_t = y^i$.

- **repeat**
  - Update the support as follows:
    
    $\hat{\mathcal{T}}^i_{t+1} = \arg \max_{|\mathcal{T}|=3} \|\mathbf{x}^i_{\mathcal{T}}(\mathcal{T})\|_F$.

    - Update the signal with LS, i.e.,
      
      $\mathbf{x}^i_{\hat{\mathcal{T}}^i_{t+1}}(\hat{\mathcal{T}}^i_{t+1}) = (\Psi(:, \hat{\mathcal{T}}^i_{t+1}))^\dagger \mathbf{y}^i$.

    - Update the residue, i.e.,
      
      $r^i_{t+1} = \mathbf{y} - \Psi \hat{\mathbf{x}}^i$.

    - until $\|r^i_{t+1}\|_F > \|r^i_t\|_F$ or $\|r^i_{t+1}\|_F < \gamma \|\mathbf{h}_2\|_2$ or $t > N_4$

- **Output:**
  - if $\|r^i_{t+1}\|_F > \|r^i_t\|_F$ then
    
    $\hat{\mathcal{S}}(:, i) = \hat{\mathcal{T}}^i_t, \hat{\mathbf{x}}(:, i) = \hat{\mathbf{x}}^i$.

  - else if $\|r^i_{t+1}\|_F < \gamma \|\mathbf{h}_2\|_2$ or $t > N_4$ then
    
    $\hat{\mathcal{S}}(:, i) = \hat{\mathcal{T}}^i_{t+1}, \hat{\mathbf{x}}(:, i) = \hat{\mathbf{x}}^i_{t+1}$.

end if

end for

The (50) is a one-step iteration with the complexity on the order of $O(1)$. Hence, the computational complexity of searching for the best index $\xi^*$ has been reduced to the order of $O((N-P)L)$. The inner iteration number is denoted as $L_2$, and it's determined by the speed of convergence of the inner iteration. The outer iteration number is $N_1$.

The outer iteration number is $N_1$, and it's determined by the speed of convergence of the inner iteration. The outer iteration number is $N_1$.

Then the computational complexity of Algorithm 1 is the order of $O(N_1L_2(N-P)L)$. We discuss the selection of the outer iteration number $N_1$ and the maximum number of inner iteration $N_2$ in Section VII. From the result, it can be seen that the convergence speed of the inner iteration and outer iteration is very fast, and the values of $N_1$ and $N_2$ can be less than 10.

**TABLE 1.** Algorithm complexities of pilot subcarrier pattern design algorithms.

| Algorithm | Complexity |
|-----------|------------|
| GOS-GIA   | $(PLP + (N-P)L)N_1N_2$ |
| TLGIA     | $(L + P + (N-P)L)N_1N_2$ |
| SSS       | $P((N-P)L)N_1N_2$ |

Algorithm complexities of different pilot subcarrier pattern design algorithms are indicated in Table 1. The two-layer greedy iteration algorithm (TLGIA) is the result of our previous work in [52], and the SSS algorithm is the result of [23]. Overall, the proposed pilot design scheme has lower computational complexities. It can reduce the computational complexity at least one order of magnitude compared to the SSS algorithm. For example, when $N = 1024$, $P = 12$, and $L = 73$, our proposed GOS-GIA algorithm has almost the same algorithm complexity as the TLGIA algorithm, but at least 10 times less than the SSS algorithm.

**B. ALGORITHM COMPLEXITY OF THE PROPOSED CHANNEL ESTIMATION STRATEGY**

In the proposed corresponding two-stage channel estimation strategy, the B-SP algorithm and P-SP algorithm contribute to most of the computational complexity. The computational complexity of the B-SP algorithm mainly depends on the matrix inversion implemented for the least square (LS) operation. In each iteration of the algorithm, the LS solution has the computational complexity on the order of $O(GT^3)$. And the number of the iteration depends on the size of the received signal $P$. Then the computational complexity of the B-SP algorithm is the order of $O(PGT^3)$. Similarly, the computational complexity of the P-SP algorithm is the order of $O(GTS^3)$.

Overall, our proposed two-stage channel estimation strategy has the computational complexity in the order of $O(PGT^3 + GTS^3)$. If the channel estimation is done by degenerating the two-dimensional problem as a one-dimensional problem, as shown in (44), the computational complexity is the order of $O(PG(TS)^3)$.

The proposed two-stage channel estimation strategy reduces the computational complexity by the order of magnitude $O((S^2, PT^2))$. Obviously, the proposed two-stage channel estimation strategy reduces the computational complexity by more than two orders of magnitude. For example, when $S = 12$, $P = 12$, and $T = 6$, our proposed two-stage channel estimation strategy reduces the computational complexity by more than 400 times.

**VII. SIMULATION RESULTS**

In this section, a simulation study is carried out to investigate the performance of the proposed decoupling pilot design scheme and the corresponding two-stage channel estimation strategy. We consider typical massive MIMO OFDM systems [12], [32], [33], [53] with the basic simulation parameters listed in Table 2. The number of simulation for obtaining
TABLE 2. Massive MIMO OFDM system parameters.

| Parameters                  | Values |
|-----------------------------|--------|
| Number of antennas $M$      | 128    |
| Number of subcarriers $N$   | 1024   |
| Carrier frequency $f_c$ (GHz)| 1.9    |
| Maximal mobile velocity $v$ (km/h) | 35    |
| Coherence time $T_{\text{slot}}$ (ms) | 4.06 |
| Symbols interval $T_s$ (ms)  | 1/14   |
| Usable interval $T_u$ (ms)   | 1/15   |
| Guard interval $T_g$ (ms)    | 1/(14-15) |
| The number of paths $L$     | 73     |
| Coherence OFDM symbols $R$  | 61     |
| The number of no-zero paths $T$ | 6     |
| Subpaths in each resolvable path $Q$ | 20 |
| Angular spread $\sigma_{\text{AS}}$ ($^\circ$) | 15 |
| Sparse level in virtual angle domain $S$ | 12 |

the figures is 1000. We keep these parameters in the following simulation unless otherwise stated.

A. PERFORMANCE OF THE PROPOSED PILOT PATTERN DESIGN

In this section, we analyze the effect on the MIP of the inner iteration and the outer iteration in Algorithm 1. And we discuss the selection of the inner iteration number and the outer iteration number.

In Fig. 2, we compare the average MIP of different pilot subcarrier patterns as the function of the outer iteration number $N_1$. The curve of random generation represents the random generation pilot subcarrier pattern. The curve of one-layer iteration represents the pilot subcarrier pattern result of the inner iteration in Algorithm 1. The curve of GOS-GIA represents our proposed pilot subcarrier pattern result of Algorithm 1. The TLGIA is the result of our previous work in [52], and the SSS algorithm is the result of [23]. In this simulation, the maximum number of inner iterations $N_2 = 10$, the subcarrier overhead $P = 12$ and SNR = 20 dB is adopted. From this figure, it shows that there is a great gap between the MIP of the random generation and the one-layer iteration result. This phenomenon means that the inner iteration can reduce the MIP significantly compared with the random pilot subcarrier pattern. Compared with the SSS algorithm, our proposed GOS-GIA algorithm has a little performance loss. This is because our algorithm greatly reduces the algorithm complexity. The curve of the proposed algorithm is almost converged when $N_1 \geq 6$, and the convergence speed of the outer iteration is very fast.

Fig. 3 describes the average number of inner iteration under various subcarrier overhead $P$. It approximately increases linearly with the subcarrier overhead $P$. Compared with the SSS algorithm in [23], our proposed algorithm greatly reduces the number of inner iteration. Our proposed algorithm has also a fewer number of inner iteration compared with our previous work in [52]. As the average number of iteration is a very small value, the convergence speed of the inner iteration is very fast and the algorithm complexity of the pilot design is very low.

FIGURE 2. The average MIP of different pilot subcarrier patterns as the function of the outer iteration number $N_1$.
reduces the subcarrier overhead by 40%, and our proposed algorithm achieves similar MIP performance with half of the subcarrier overhead $P$. This means the inner iteration and the outer iteration in Algorithm 1 are effective to reduce the pilot overhead. Compared with the SSS algorithm, our proposed algorithm has a slight loss of performance, but greatly reduces the number of inner iteration.

B. PERFORMANCE OF THE PROPOSED CHANNEL ESTIMATION STRATEGY

In this section, we investigate and analyze the performance of the proposed channel estimation strategy.

We compare the mean square error (MSE) of the proposed channel estimation strategy with different pilot subcarrier patterns. Random pilot subcarrier patterns, the proposed GOS-GIA pilot patterns, and pilot patterns with the SSS algorithm are considered. We also take LS channel estimation as a reference. In the LS channel estimation, the sparse support in the delay-angle domain is assumed fully known in advance at the receiver side.

In Fig. 5, we show the channel estimation MSE performance under various subcarrier overhead $P$. In this simulation, $G = 48$ and SNR = 20dB are adopted. Compared with the random pilot subcarrier pattern, our proposed GOS-GIA pilot subcarrier pattern design achieves significant MSE performance gain when $P < 16$. This is because the proposed pilot subcarrier pattern design has better MIP performance. When $P \geq 16$, the MSE curves tend to unchanged. The reason is the detection probability in the delay domain approximate to zero, and the MSE performance is mainly determined by the estimation error in the angle domain which is independent with $P$. Compared with the pilot patterns with the SSS algorithm and the referenced LS channel estimation, our proposed channel estimation with the proposed pilot pattern has almost the same performance. This means the proposed channel estimation and the proposed pilot pattern are near-optimal for various $P$.

Fig. 6 shows the channel estimation MSE performance under various signal to noise ratio (SNR). In this simulation, $P = 12$ and SNR = 20dB are adopted. Compared with the random pilot subcarrier pattern, our proposed GOS-GIA pilot subcarrier pattern design achieves significant MSE performance gain. The gain increases with the SNR. This is because when SNR is high, the influence of noise weakens and the pilot pattern becomes the dominant factor. Compared with the pilot patterns with the SSS algorithm and the referenced LS channel estimation, our proposed channel estimation with the proposed pilot pattern has almost the same performance. This means the proposed channel estimation and the proposed pilot pattern are near-optimal for various SNR.

In Fig. 7, we show the channel estimation performance under various OFDM symbols overhead $G$. In this simulation, $P = 12$ and SNR = 20dB are adopted. Compared with the random pilot subcarrier pattern, our proposed GOS-GIA pilot subcarrier pattern design can achieve significant MSE performance gain. The performance gain increases with $G$. This is because when $G$ is small, the MSE is mainly influenced by the estimation error in the angle domain, and when $G$ is large, the MSE is mainly influenced by the estimation error in the delay domain. And the proposed GOS-GIA pilot pattern is in the delay domain. Compared with the pilot patterns with the SSS algorithm, our proposed channel estimation with the proposed pilot pattern has almost the same performance. This means the proposed channel estimation and the proposed
pilot pattern are near-optimal for various $G$. Compared with the referenced LS channel estimation, our proposed channel estimation with the proposed GOS-GIA pilot pattern has almost the same performance when $G \geq 40$, while there exists a large performance gap when $G$ is small. The reason is the sparse support detection error in the angle domain has a great effect on performance.

VIII. CONCLUSION

This paper has addressed the issue of pilot design for FDD massive MIMO systems by utilizing the joint sparsity in the delay-angle domain. Based on the joint sparsity, a decoupling pilot design scheme has been proposed, in which the pilot design has been decoupled into two domains. In the delay domain, a GOS-GIA algorithm has been proposed to obtain the near-optimal pilot subcarrier pattern. In the angle domain, a random Rademacher distribution pilot matrix has been adopted. At the receiver, the corresponding channel estimation strategy with two-stage has been provided. Algorithm complexity analysis has shown that the GOS-GIA can reduce the computational complexity at least one order of magnitude, and the proposed two-stage channel estimation strategy reduces the computational complexity by more than two orders of magnitude. Simulation results have verified the proposed decoupling FDD pilot design scheme with the corresponding channel estimation strategy achieves better performance with lower pilot overhead compared with the random pilot pattern.

REFERENCES

[1] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Process. Mag., vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[3] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, “Channel capacity and state estimation for state-dependent Gaussian channels,” IEEE Trans. Inf. Theory, vol. 51, no. 4, pp. 1486–1495, Apr. 2005.

[4] L. Gu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] X. Rao and V. K. N. Lau, “Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems,” IEEE Trans. Signal Process., vol. 62, no. 12, pp. 3261–3271, 2014.

[6] J. Dai, A. Liu, and V. K. N. Lau, “FDD massive MIMO channel estimation with arbitrary 2D-array geometry,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2584–2599, May 2018.

[7] W. Wang, Y. Xiu, B. Li, and Z. Zhang, “FDD downlink channel estimation solution with common sparsity learning algorithm and zero-partition enhanced GAMP algorithm,” IEEE Access, vol. 6, pp. 11123–11145, 2018.

[8] L. Lian, A. Liu, and V. K. N. Lau, “Exploiting dynamic sparsity for downlink FDD-massive MIMO channel tracking,” IEEE Trans. Signal Process., vol. 67, no. 8, pp. 2007–2021, Apr. 2019.

[9] V. Kafedziski, “Compressed sensing based channel estimation in FDD multi-user massive MIMO using angle domain sparsity and transmit antenna correlation,” in Proc. Eur. Conf. Netw. Commun. (EuCNC), Jun. 2018, pp. 1–9.

[10] A. Liu, L. Lian, V. K. N. Lau, and X. Yuan, “Downlink channel estimation in multiuser massive MIMO with hidden Markovian sparsity,” IEEE Trans. Signal Process., vol. 66, no. 18, pp. 4796–4810, Sep. 2018.

[11] X. Yang, D. Fan, A. Ren, N. Zhao, and M. Alam, “5G-based user-centric sensing at C-band,” IEEE Trans. Ind. Informat., vol. 15, no. 5, pp. 3040–3047, May 2019.

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

[13] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, “Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO,” IEEE Trans. Signal Process., vol. 68, pp. 764–779, 2020.

[14] G. Liu, S. Ma, G. Liu, H. He, H. Deng, and X. Qian, “An novel sparse adaptive blocked matching pursuit algorithm for spatio-temporal joint channel estimation in FDD massive MIMO system,” in Proc. IEEE Int. Conf. Comput. Technol. (CCET), Aug. 2018, pp. 1–4.

[15] I. Dai, A. Liu, and H. C. So, “Non-uniform burst-sparse learning for massive MIMO channel estimation,” IEEE Trans. Signal Process., vol. 67, no. 4, pp. 1075–1087, Feb. 2019.

[16] L. Ge, Y. Zhang, G. Chen, and J. Tong, “Compression-based LMMSE channel estimation with adaptive sparsity for massive MIMO in 5G systems,” IEEE Syst. J., vol. 13, no. 4, pp. 3847–3857, Dec. 2019.

[17] J. Choi, D. J. Love, and P. Bidigare, “Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 802–814, Oct. 2014.

[18] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, “Pilot beam pattern design for channel estimation in massive MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 787–801, Oct. 2014.

[19] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, “Joint spatial division and multiplexing for mm-wave channels,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1239–1255, Jun. 2014.

[20] R. R. Muller, L. Cottatellucci, and M. Vehkapera, “Blind pilot decontamination,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 773–786, Oct. 2014.

[21] L. Chen and X. Yuan, “Blind mutliuser detection in massive MIMO channels with clustered sparsity,” IEEE Wireless Commun. Lett., vol. 8, no. 4, pp. 1052–1055, Aug. 2019.

[22] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, “Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO,” IEEE Trans. Commun., vol. 64, no. 2, pp. 601–617, Feb. 2016.

[23] C. Qi, Y. Huang, S. Jin, and L. Wu, “Sparse channel estimation based on compressed sensing for massive MIMO systems,” in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp. 4558–4563.

[24] M. Masood, L. H. Affify, and T. Y. Al-Naffouri, “Efficient coordinated recovery of sparse channels in massive MIMO,” IEEE Trans. Signal Process., vol. 63, no. 1, pp. 104–118, Jan. 2015.

[25] A. Nasser and M. Elsabrouty, “Frequency-selective massive MIMO channel estimation and feedback in angle-time domain,” in Proc. IEEE Symp. Comput. Commun. (ISCC), Jun. 2016, pp. 1018–1023.

[26] L. Dai, Z. Gao, and Z. Wang, “Joint channel estimation and feedback with low overhead for FDD massive MIMO systems,” in Proc. IEEE/CIC Int. Conf. Commun. China (ICC), Nov. 2015, pp. 1–6.

[27] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, “A block sparsity based estimator for mmWave massive MIMO channels with beam squint,” IEEE Trans. Signal Process., vol. 68, pp. 49–64, 2020.

[28] Q. Qin, L. Gui, B. Gong, and S. Luo, “Sparse channel estimation for massive MIMO-OFDM systems over time-varying channels,” IEEE Access, vol. 6, pp. 33740–33751, 2018.

[29] S. Bazzi, S. Stefanatos, L. De Magaorou, S. E. Hajiri, M. Assaad, S. Paquelter, G. Wunder, and W. Xu, “Exploiting the massive MIMO channel structural properties for minimization of channel estimation error and training overhead,” IEEE Access, vol. 7, pp. 32434–32452, 2019.

[30] J.-C. Shen, Z. Zhang, E. Alusa, and K. B. Letaief, “Compressed CSI acquisition in FDD massive MIMO: How much CS?” IEEE Trans. Wireless Commun., vol. 15, no. 6, pp. 4145–4156, Jun. 2016.

[31] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressive channel sensing: A new approach to estimating sparse multipath channels,” Proc. IEEE, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[32] D. S. Baum, J. Hansen, G. D. Galdo, M. Milojic, J. Salo, and P. Kyosti, “An interlin channel model for beyond-3G systems: Extending the 3GPP spatial channel model (SCM),” in Proc. IEEE 61st Veh. Technol. Conf. (VTC Spring), vol. 5, May 2005, pp. 3132–3136.
X. Yan et al.: Decoupling Channel Estimation for FDD Massive MIMO Systems Utilizing Joint Sparsity

XIANGYU YAN received the B.E. degree in communication engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2012, where he is currently pursuing the Ph.D. degree in wireless communications. His research interests include sparse signal processing, pilot decontamination, and channel estimation for massive MIMO systems.

LI CHEN received the bachelor’s degree in electrical and information engineering from the Harbin Institute of Technology (HIT), in 2009, and the Ph.D. degree in electrical engineering from the University of Science and Technology of China (USTC), in 2014. He joined as a Postdoctoral Researcher with USTC, from 2014 to 2016, where he is currently a faculty. In 2014, he won the Chinese Academy of Sciences President Award.

HUARUI YIN received the bachelor’s and Ph.D. degrees in electronic engineering and information science from the University of Science and Technology of China (USTC), Hefei, Anhui, in 1996 and 2006, respectively. Since 2010, he has been with the Department of Electronic Engineering and Information Science, USTC, as an Associate Professor. His research interests include digital signal processing, low-complexity receiver design, multiple access for massive connections, and throughput analysis of wireless networks.

WEIDONG WANG received the B.S. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989, and the M.S. degree from the University of Science and Technology of China (USTC), Hefei, China, in 1993. He is currently a Full Professor with the Department of Electronic Engineering and Information Systems, USTC. His research interests include wireless communication, microwave and millimeter, and radar technology. He is a member of the Committee of Optoelectronic Technology, Chinese Society of Astronautics.