IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 January 2023, accepted 11 February 2023, date of publication 22 February 2023, date of current version 8 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247662

RESEARCH ARTICLE

Second-Order Statistics-Aided Channel Estimation for Multipath Massive MIMO-OFDM Systems

MOHAMMAD REZA MEHRABANI[1], BAHMAN ABOLHASSANI[1], FARZAN HADDADI[1], AND CHINTHA TELLAMBURA[2], (Fellow, IEEE)

[1]School of Electrical Engineering, Iran University of Science & Technology, Tehran 16846-13114, Iran
[2]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada

Corresponding author: Farzan Haddadi (farzanhaddadi@iust.ac.ir)

ABSTRACT This paper develops an efficient channel estimation algorithm based on second-order statistics of time division duplex (TDD) multiuser massive multiple-input multiple-output (MIMO) systems. The algorithm uses the received signal correlation to determine the most significant lags (MSLs) of the received signal. We first employ these MSLs to propose a novel set containing the channel's four most significant taps (MSTs). Then, by using them, we propose an efficient semi-blind iterative algorithm called enhanced modified-subspace pursuit (EM-SP). It uses the set mentioned above and two theoretical results (Lemma 1 and Theorem 1) to estimate an arbitrary number of MSTs efficiently. Simulation results show that the normalized mean square error (NMSE) of the proposed EM-SP algorithm is much smaller than that of the subspace pursuit (SP) and orthogonal matching pursuit (OMP) algorithms at the cost of 0.3 % and 2 % more computational complexity for the channels with three and six nonzero paths, respectively. Moreover, the NMSE of it is very close to that of the optimal genie-aided least square algorithm.

INDEX TERMS Compressive sensing, massive MIMO-OFDM, frequency-selective channel estimation, most significant tap (MST) detection, second-order statistics.

I. INTRODUCTION

Massive MIMO systems employ hundreds of base station (BS) antennas to serve multiple users in the same time-frequency resource [1]. The spatial diversity gains of a massive number of antennas tremendously boost both energy efficiency and spectrum efficiency [2]. Besides those gains, increasing spatial resolution and degrees of freedom make massive MIMO systems very robust [3], [4]. Due to these advantages, the massive MIMO concept has been widely researched. This paper focuses on multiuser (MU) massive MIMO, where a BS with a massive number of antennas serves multiple single-antenna users.

The accuracy of channel estimation has a decisive effect on the receiver's performance because the receiver's data detection and other operations depend on channel state information (CSI) availability. Any error in channel estimation thus degrades the overall performance [5]. Therefore, the trade-off between computational complexity and the affordable training sequence (pilot) overhead in multipath massive MIMO channel estimation becomes a challenging problem. In massive MIMO systems, channels often have a sparse structure due to a limited number of local scatterers, exhibiting a few resolvable paths [6]. The sparsity of the channel makes compressed sensing (CS) a convenient tool for channel estimation [7]. CS requires fewer pilots than those of conventional schemes such as least square (LS) and minimum mean square error (MMSE) [8].

Frequency division duplexing (FDD) and time-division duplexing (TDD) are two transmission protocols suitable for MU massive MIMO systems [9]. In TDD, the uplink is separated from the downlink by allocating different time slots in the same frequency band. Whereas in FDD, they are allocated separate frequency bands. And this channel separation is much bigger than the coherence bandwidth; consequently, these two channels experience uncorrelated fading. Thus, the BS cannot estimate the uplink CSI and send it to the mobile station (MS) to be used as the downlink CSI. Furthermore,

The associate editor coordinating the review of this manuscript and approving it for publication was Parul Garg.

because the BS has a massive number of antennas, FDD imposes a high computation and pilot overhead at the MS for CSI estimation [10], decreasing the spectrum efficiency of MU massive MIMO systems [11].

The TDD protocol can overcome these problems. In TDD, channel reciprocity holds (e.g., downlink and uplink channel responses are nearly identical) if a transmission burst in each link is much shorter than the channel coherence time [12]. In many MU networks with TDD, it is typical to have single-antenna users. The BS can acquire the uplink CSI and use it to precode the downlink signals to compensate for channel effects [13]. The uplink CSI estimate thus works for the downlink [14]. This critical advantage ensures the prevalency of TDD [15], and it is also the choice in our paper. In some situations, including transmitter-receiver imbalance (such as more receiving antennas on the user side), significant Doppler effects, and interference of mobile systems [16], channel reciprocity holds partially, and precoding alone is insufficient. Some feedback is still necessary [17]. Some works use precoding and a few pilot sequences as feedback [18], and some consider precoding along with blind downlink channel estimation [13].

Channel estimation methods can be training-based, blind, or semi-blind (see [19], [20], [21] and references therein for a more detailed discussion of these topics). Training-based methods transmit a known sequence of symbols (pilots) to the receiver. The receiver uses the pilot sequence to estimate the CSI. While this approach is simple, the use of pilots amounts to bandwidth or spectral efficiency loss. Blind methods take a diametrically opposite approach. They perform channel estimation and data detection jointly using no pilot at the expense of much computational complexity, making them impractical for delay-sensitive applications.

Semi-blind methods are a trade-off between the above two methods. They use fewer pilot subcarriers. Instead, they employ unknown data symbols for channel estimation with an acceptable computational complexity [22].

In blind and semi-blind methods, channel estimation exploits statistical features. Among all such methods, the second-order statistics-based algorithms dominate blind and semi-blind methods due to their special properties. In those algorithms, the covariance [23], [24], [25], [26], [27], [28] or the correlation [29], [30], [31] of the incoming signal is used to obtain the CSI. One of the most commonly used algorithms is the subspace method [23]. It uses the covariance matrix of the received signal to estimate the signal subspace.

These methods include the eigenvalue decomposition (EVD) and singular value decomposition (SVD) channel estimation [23]. EVD methods analyze the covariance matrix of the received signal to estimate the CSI [24]. Reference [25] shows that in MU massive MIMO systems, the small-scale fading matrix is a product of the received signal eigenvectors matrix and an ambiguity matrix. Using this, [26] shows that the channel vectors lie in a subspace spanned by the eigenvectors of the received signal covariance matrix. They express each channel vector as a linear combination of the eigenvectors. Conversely, SVD methods use the singular vector of the received signal to separate desired signals from interference and eliminate the pilot contamination [27].

Subspace methods always have an error because of a finite number of BS antennas (which are non-orthogonal to the channel vectors) and a limited length of data symbols [28]. The error may be mitigated with numerous antennas.

The other group of the second-order-based algorithms is those which use the correlation of the incoming signal. References [29] and [30] have proposed semi-blind algorithms based on second-order statistics and a linear prediction method. These two algorithms estimate the nonzero channel taps named the most significant taps (MSTs) using the autocorrelation of the received signal. These methods combine the analysis of the second-order constraint with a training-based criterion to arrive at a new linear prediction-based algorithm, which has higher efficiency in computation and convergence speed.

In [29], authors use nonzero values of the autocorrelation function whose corresponding indexes are called the most significant lags (MSLs). Pilot subcarriers are thus employed to estimate the most significant taps' actual locations and amplitudes. Although for channels with long multipath delays, this algorithm becomes complicated. Algorithms [29] and [30] are suitable only for a specific and limited number of channel coefficients. However, our proposed algorithm estimates MST locations from MSLs in a general mode. Reference [31] develops some semi-blind methods using a reduced data matrix, and the channel estimation utilizes the estimated MSLs.

Reducing the computational complexity is a crucial issue for a practical receiver [32]. The complexity of a receiver highly depends on the computational complexity of the channel estimator. Reducing the latter while maintaining receiver performance targets are mutually contradictory goals because the performance critically depends on the quality of channel estimates. Therefore, high-quality channel estimates come at the cost of higher computational complexity [33]. To reduce this complexity, [34] generates a training set of reconfigurable intelligent surface (RIS) reflection coefficient vectors firstly offline. For each of these vectors, this work performs the composite channel estimation and transmit power allocation similarly to the conventional OFDM systems without RIS. This simplifies power allocation and channel estimation to maximize achievable data rates. In contrast, our proposed channel estimation scheme balances the complexity and performance by exploiting the received signal's correlation.

Suppose the largest dimension of the BS antenna array is minimal compared to the distance between the BS and an MS. In that case, the signal transmitted by the BS antennas experiences similar or common sparse multipath channels. This similarity arises from the similar scatterers between the BS co-located antennas and a mobile station [35].

Based on this joint sparsity, we propose a novel second-order statistics-based channel estimation using structured compressive sensing (SCS) for TDD MU massive MIMO systems. First, we use the relation between estimated MSLs and MSTs to prove Lemma 1 and Theorem 1. Using those, we propose a novel set of four channel MST positions. Second, we propose a novel, efficient iterative semi-blind channel estimation algorithm called enhanced modified-subspace pursuit (EM-SP), a combination of the SP [36] and Theorem 1. This algorithm decreases the number of pilot subcarriers required for channel estimation. Moreover, it has low computational complexity, but much more accurate channel estimation than existing algorithms such as SP [36] and OMP [37]. The contributions of this paper may be summarized as follows.

- We propose the EM-SP algorithm, an upgrade of the SP algorithm, by utilizing second-order statistics of the received signal. As evidenced by the simulation results, it requires fewer pilot subcarriers to estimate the channel compared to other channel estimation algorithms.
- Unlike subspace methods [23], [24], [25], [26], [27], which are most effective when the number of BS antennas is vast, the proposed EM-SP algorithm can reduce pilot overheads and achieve good performance for a BS with a few antennas.
- In [29] and [30], MSTs are detected based on the training-based least square criterion and MSL's positions, where relations are not suitable for the general case. In contrast, our algorithm estimates the channel with an arbitrary number of taps.

The rest of the paper is organized as follows. Section II introduces signal and sparse channel models for a single-cell MU massive MIMO-OFDM and analyzes the received signal's second-order statistics. Section III contains a CS-based channel scheme, analyzes the relation of MSL positions and MSTs, and proposes a set that contains four MST positions. Then, it develops a new algorithm to estimate the uplink MU massive MIMO-OFDM channels. Section IV evaluates our proposed algorithm using simulation results, which show that it outperforms existing algorithms such as OMP and SP, and its performance is very close to that of the benchmark Genie-aided LS algorithm. However, its complexity is slightly more than that of the SP algorithm. Section V concludes the paper by emphasizing the essential points of the work.

Notations: Uppercase and lowercase boldface letters denote matrices and column vectors, respectively. The operators $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^\dagger$ and $|\cdot|$ are stand for transpose, conjugate, conjugate transpose, Moore-Penrose pseudoinverse and cardinality operator, respectively. Finally set($\mathbf{b}$), $\|\mathbf{\Psi}\|_F$ and supp($\mathbf{b}$) denote the set formed by collecting the entries of $\mathbf{b}$, Frobenius norm of $\mathbf{\Psi}$ and the index set of nonzero entries of vector $\mathbf{b}$ respectively.

II. SYSTEM MODEL

This paper aims to understand the effects of semi-blind estimation in the uplink of single-cell scenarios with massive antenna arrays. We thus consider a single-cell multipath MU massive MIMO system with large-scale and small-scale fading components. For simplicity's sake, we omit the symbol's time stamp and assume the channel remains fixed during one OFDM symbol.

A. MIMO CHANNEL MODEL

Suppose the BS is equipped with M antennas, serving K single-antenna users and using N subcarriers. At the BS, after the removal of the CP, the received time domain OFDM symbol of the m-th antenna $\tilde{\mathbf{y}}_m = [\tilde{y}_m(0), \tilde{y}_m(1), \ldots, \tilde{y}_m(N-1)]^T$ can be expressed as

$$\tilde{\mathbf{y}}_m = \sum_{k=1}^{K} \tilde{\mathbf{H}}_{m,k}\tilde{\mathbf{x}}_k + \tilde{\mathbf{z}}_m, \tag{1}$$

where $\tilde{\mathbf{x}}_k = [\tilde{x}_k(0), \tilde{x}_k(1), \ldots, \tilde{x}_k(N-1)]^T$ is time-domain OFDM symbol which is transmitted by the k-th user and $\tilde{\mathbf{z}}_m$ is corresponding zero-mean complex Gaussian noise vector with variance $\sigma^2_{\tilde{\mathbf{z}}}$. Furthermore, $\tilde{\mathbf{H}}_{m,k}$ is the N by N circulant time-domain channel matrix and $\tilde{\mathbf{H}}_{m,k} = \mathbf{F}^H \text{diag}\{\sqrt{N}\mathbf{F}[[\mathbf{h}_{m,k}]^T, [0_{1,N-L}]^T]\}\mathbf{F}$, where $\mathbf{h}_{m,k} = [h_{m,k}(0), h_{m,k}(1), \ldots, h_{m,k}(L-1)]^T$ is the channel impulse response (CIR) with length L. And $h_{m,k}(\ell)$ denotes the ℓ-th channel tap from the k-th user to the m-th receive antenna of the BS and $\mathbf{F}$ is a N-point unitary DFT matrix [38].

Here $h_{m,k}(\ell) \triangleq g_{m,k}(\ell)\sqrt{\beta_{k,\ell}}$, where $g_{m,k}(\ell)$ denotes the small-scale channel coefficient and $\beta_{k,\ell}$ denotes the large-scale fading of the ℓ-th channel tap. The small-scale fading coefficients are complex Gaussian random variables, independent and identically distributed (i.i.d.), with zero mean and unit variance. The large-scale fading $\beta_{k,\ell}$ is modeled as [39]

$$\beta_{k,\ell}(dB) = \nu_k(dB) - 10\frac{\ell}{L}\log_{10}(e), \tag{2}$$

where $\nu_k(dB) = 20\log_{10}(\frac{4\pi r_{ref}}{\lambda}) + 10\alpha\log_{10}(\frac{r_k}{r_{ref}}) - \psi_k$. And r_{ref} is a reference distance, r_k is the k'th user's distance to the BS, λ is the wavelength, α is the path loss exponent and ψ_k is the shadowing with standard deviation σ_ψ. The second term in (2) comes from the power delay profile, which has been modeled as an exponentially decaying function with $\ell = 0, 1, \ldots, L-1$.

Taking the DFT of $\tilde{\mathbf{y}}_m$, we obtain

$$\mathbf{y}_m = \sum_{k=1}^{K} \text{diag}\{\mathbf{F}_L\mathbf{h}_{m,k}\}\mathbf{x}_k + \mathbf{z}_m, \tag{3}$$

where $\mathbf{y}_m = [y_m(0), y_m(1), \ldots, y_m(N-1)]^T$ is the received frequency domain symbol of the m-th antenna, $\mathbf{x}_k = [x_k(0), x_k(1), \ldots, x_k(N-1)]^T$ is the frequency domain symbol of the k-th user, $\mathbf{F}_L$ is $\sqrt{N}$ times of the first L columns of matrix $\mathbf{F}$ and $\mathbf{z}_m = \mathbf{F}\tilde{\mathbf{z}}_m$ is the DFT of the noise samples.

Now, set $\mathcal{P}_k$ is the index set of pilot subcarriers for the k-th transmitter with the length $N_{\mathcal{P}} = |\mathcal{P}_k|$, which is randomly selected from $\{0, 2, \ldots, N-1\}$. We employ frequency-orthogonal pilot placement for different transmit antennas. For better performance, when the k-th antenna transmit a pilot, all the other antennas remain silent. Using the properties of diagonal matrices, we have $\text{diag}\{\mathbf{F}_L\mathbf{h}_{m,k}\}\mathbf{x}_k = \text{diag}\{\mathbf{x}_k\}\mathbf{F}_L\mathbf{h}_{m,k}$. Consequently define k-th user's pilot sequence as $\mathbf{P}_k \triangleq \text{diag}(\mathbf{x}_k^{\mathcal{P}_k}) \in \mathbb{C}^{N_{\mathcal{P}}\times 1}$, then the received pilot sequence $\mathbf{y}_m^{\mathcal{P}} \in \mathbb{C}^{N_{\mathcal{P}}K\times 1}$ is

$$\begin{aligned}\mathbf{y}_m^{\mathcal{P}} &= \begin{bmatrix}\mathbf{y}_{m,1}^{\mathcal{P}_1}\\ \mathbf{y}_{m,2}^{\mathcal{P}_2}\\ \vdots\\ \mathbf{y}_{m,K}^{\mathcal{P}_K}\end{bmatrix}\\ &= \underbrace{\begin{bmatrix}\mathbf{P}_1\mathbf{F}_L^{\mathcal{P}_1} & 0 & \cdots & 0\\ 0 & \mathbf{P}_2\mathbf{F}_L^{\mathcal{P}_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \mathbf{P}_K\mathbf{F}_L^{\mathcal{P}_K}\end{bmatrix}}_{\mathbf{\Phi}} \underbrace{\begin{bmatrix}\mathbf{h}_{m,1}\\ \mathbf{h}_{m,2}\\ \vdots\\ \mathbf{h}_{m,K}\end{bmatrix}}_{\mathbf{h}_m} + \mathbf{z}_m^{\mathcal{P}}\\ &= \mathbf{\Phi}\mathbf{h}_m + \mathbf{z}_m^{\mathcal{P}}, \end{aligned} \tag{4}$$

where $\mathbf{F}_L^{\mathcal{P}_k} \in \mathbb{C}^{N_{\mathcal{P}}\times L}$ is the sub-matrix by selecting rows of $\mathbf{F}_L$ according to $\mathcal{P}_k$ and $\mathbf{z}_m^{\mathcal{P}}$ is the noise vector, respectively.

By stacking all the channel vectors, we define the channel matrix $\bar{\mathbf{H}} \triangleq [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_M] \in \mathbb{C}^{KL\times M}$. Therefore, the linear model of the MU massive MIMO-OFDM can be described as

$$\mathbf{Y}^{\mathcal{P}} = [\mathbf{y}_1^{\mathcal{P}}, \mathbf{y}_2^{\mathcal{P}}, \ldots, \mathbf{y}_M^{\mathcal{P}}] = \mathbf{\Phi}\bar{\mathbf{H}} + \mathbf{Z}^{\mathcal{P}}, \tag{5}$$

where $\mathbf{Y}^{\mathcal{P}} \in \mathbb{C}^{KN_{\mathcal{P}}\times M}$, $\mathbf{Z}^{\mathcal{P}} = [\mathbf{z}_1^{\mathcal{P}}, \mathbf{z}_2^{\mathcal{P}}, \ldots, \mathbf{z}_M^{\mathcal{P}}] \in \mathbb{C}^{KN_{\mathcal{P}}\times M}$.

Now, if the MU massive MIMO channel is sparse and also the positions of the significant or nonzero taps are the same for all transmit-receive antenna pairs [35], then (5) must be rewritten in such a way that the channel matrix $\bar{\mathbf{H}}$, has $\mathcal{D}$-th chunk sparsity level. For this, define matrix $\mathbf{B}$ as

$$\mathbf{B} = [\mathbf{H}^T(0), \mathbf{H}^T(1), \ldots, \mathbf{H}^T(L-1)] \in \mathbb{C}^{KL\times M}, \tag{6}$$

where $\mathbf{H}(\ell)$ is the ℓ-th tap channel matrix and is defined as

$$\mathbf{H}(\ell) \triangleq \begin{bmatrix} h_{1,1}(\ell) & h_{1,2}(\ell) & \cdots & h_{1,K}(\ell)\\ h_{2,1}(\ell) & h_{2,2}(\ell) & \cdots & h_{2,K}(\ell)\\ \vdots & \vdots & \ddots & \vdots\\ h_{M,1}(\ell) & h_{M,2}(\ell) & \cdots & h_{M,K}(\ell)\end{bmatrix}. \tag{7}$$

Now, we rewrite system model of (5) as

$$\mathbf{Y}^{\mathcal{P}} = \mathbf{\Psi}\mathbf{B} + \mathbf{Z}^{\mathcal{P}}, \tag{8}$$

where $\mathbf{\Psi} = [\mathbf{\Psi}_0, \mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_{L-1}] \in \mathbb{C}^{KN_{\mathcal{P}}\times KL}$ and $\mathbf{\Psi}_{L-1} = [\mathbf{\Phi}_\ell, \mathbf{\Phi}_{L+\ell}, \ldots, \mathbf{\Phi}_{(K-1)L+\ell}]$ is extracted from the corresponding columns of matrix $\mathbf{\Phi}$. In (8), $\mathbf{H}(\ell)$ is the ℓ-th chunk of the channel matrix $\mathbf{B}$, and due to the spatial common sparsity, $\mathbf{B}$ exhibits structured sparsity with chunk size $K \times M$ [40].

B. SECOND-ORDER STATISTICS

In blind and semi-blind channel estimation, the received signal correlation function plays a vital role [29], which may be defined as

$$\mathfrak{R}(\ell) \triangleq \text{E}\{\tilde{\mathbf{y}}(n)\tilde{\mathbf{y}}^H(n)\}, \tag{9}$$

where $\tilde{\mathbf{y}}(n) = [\tilde{\mathbf{y}}_1(n), \tilde{\mathbf{y}}_2(n), \ldots, \tilde{\mathbf{y}}_m(n)]^T$. Now, suppose that the channel is sparse with $\mathcal{D}$ nonzero paths, and the significant taps positions are the same for all transmit-receive antenna pairs [35]. We denote $\mathfrak{Z}_A$ as the MST of nonzero tap matrix

$$\mathfrak{Z}_A \triangleq [\mathbf{H}(\ell_0), \mathbf{H}(\ell_1), \ldots \mathbf{H}(\ell_{\mathcal{D}-1})], \tag{10}$$

where ℓ_d is an integer such that $0 = \ell_0 < \ell_1 < \ldots < \ell_{\mathcal{D}-1}$. In the absence of noise, using (9) and (10), we obtain

$$\mathfrak{R}(\ell) = \mathfrak{Z}_A\mathfrak{R}_{\tilde{x},\mathcal{D}}(\ell)\mathfrak{Z}_A^H, \tag{11}$$

where

$$\begin{aligned}&\mathfrak{R}_{\tilde{x},\mathcal{D}}(\ell)\\ &= \text{E}\left\{\begin{bmatrix}\tilde{\mathbf{x}}_K(n-\ell_0)\\ \tilde{\mathbf{x}}_K(n-\ell_1)\\ \vdots\\ \tilde{\mathbf{x}}_K(n-\ell_{\mathcal{D}-1})\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}_K(n-\ell-\ell_0)\\ \tilde{\mathbf{x}}_K(n-\ell-\ell_1)\\ \vdots\\ \tilde{\mathbf{x}}_K(n-\ell-\ell_{\mathcal{D}-1})\end{bmatrix}^H\right\},\end{aligned} \tag{12}$$

with $\tilde{\mathbf{x}}_K(n) = [\tilde{x}_1(n), \tilde{x}_2(n), \ldots, \tilde{x}_K(n)]^T$. Suppose that the source signal is a sequence with zero mean and unit variance, i.e., $\sigma_x^2 = 1$. Therefore, we have [29]

$$\mathfrak{R}(\ell) = \sum_{d=0}^{\mathcal{D}-1}\sum_{d'=0}^{\mathcal{D}-1}\delta(\ell-\ell_d-\ell_{d'})\mathbf{H}(\ell_d)\mathbf{H}^H(\ell_{d'}). \tag{13}$$

Equation (13) states that $\mathfrak{R}(\ell_j)$ is nonzero when, $\ell_j = \ell_d - \ell_{d'}$ for $d' < d$, $d = 1, 2, \ldots, \mathcal{D}-1$ and $0 \le \ell_j \le L-1$.

Note that (13) is for the ideal situation (no noise or interference). So, for real situations, rather than using (13) for estimation of $\mathfrak{R}(\ell)$, we define the estimated version as

$$\hat{\mathfrak{R}}^{\zeta}(\ell) \triangleq \frac{1}{N}\sum_{n=0}^{N-1}\tilde{\mathbf{y}}^{(\zeta)}(n)(\tilde{\mathbf{y}}^{(\zeta)})^H(n-\ell), \tag{14}$$

where $\tilde{\mathbf{y}}^{(\zeta)}(n) \triangleq [\tilde{y}_1(n), \tilde{y}_2(n), \ldots, \tilde{y}_\zeta(n)]^T \in \mathbb{C}^{\zeta\times 1}$ in which $\tilde{\mathbf{y}}^{(\zeta)}(n) = \tilde{\mathbf{y}}^{(\zeta)}(N+n)$ for $n < 0$. It should be pointed out that unlike conventional estimated version of the received signal's correlation, which is calculated over all the receiver antennas signal, due to the common sparsity of the channels, relation (14) is calculated over $\zeta \ll M$ signals. This leads to a significant reduction of the computational complexity especially when the number of receiver antennas increases in MU massive MIMO systems.

Now we can rewrite (14) as

$$\hat{\mathfrak{R}}^{\zeta}(\ell) = \mathfrak{R}(\ell) + \mathcal{E}_{\tilde{x}} + \mathcal{E}_{\tilde{z}}, \tag{15}$$

where $\mathcal{E}_{\tilde{z}}$ and $\mathcal{E}_{\tilde{x}}$ are noise and signal error perturbations. Therefore in general, even if $\mathfrak{R}(\ell) = 0$, $\hat{\mathfrak{R}}^{\zeta}(\ell)$ is a nonzero value.

Definition 1 (MSL Set): Define an ascending set containing ζ significant lags of the received signal correlation as

$$\Gamma_{corr} \triangleq \{\gamma_{\varsigma} : \|\hat{\mathfrak{R}}^{\zeta}(\gamma_{\varsigma})\|_F \geq P_{th}, \quad 0 \leq \varsigma \leq \Sigma - 1\}, \tag{16}$$

where $\gamma_0 < \gamma_1 < \ldots < \gamma_{\Sigma-1}$, $\gamma_0 = 0$, and P_{th} is a threshold where is defined as

$$P_{th} = \frac{K_e}{L} \sum_{\ell=0}^{L-1} \|\hat{\mathfrak{R}}^{\zeta}(\ell)\|_F, \tag{17}$$

where K_e is a constant to adjust the average of the $\|\hat{\mathfrak{R}}^{\zeta}(\ell)\|_F$.

C. COMPRESSIVE SENSING MODEL

Suppose we have compressive measurements $\mathbf{Y} \in \mathbb{C}^{N_{\mathcal{P}} \times M}$ and an unknown sparse matrix $\mathbf{B} \in \mathbb{C}^{KL \times M}$, with the relation as

$$\mathbf{Y} = \mathbf{\Psi}\mathbf{B} + \mathbf{Z}, \tag{18}$$

where $\mathbf{\Psi} \in \mathbb{C}^{N_{\mathcal{P}} \times KL}$, $(N_{\mathcal{P}} \ll KL)$ is a known measurement matrix and $\mathbf{Z} \in \mathbb{C}^{N_{\mathcal{P}} \times M}$ is the noise. Our objective is to reconstruct $\mathbf{B}$ based on measurements $\mathbf{Y}$. (18) is an underdetermined set of equations and accepts infinite solutions, but it is proved in [41] that if $\mathbf{B}$ is $\mathcal{D}$-sparse and $\mathbf{\Psi}$ satisfies the Restricted Isometry Property (RIP), then $\mathbf{B}$ can be reconstructed correctly.

Definition 2 (Signal Sparsity Model): Suppose the sparse matrix $\mathbf{B} \in \mathbb{C}^{KL \times M}$ consists of L matrixes of size $K \times M$ named chunk. Define the chunk support set of $\mathbf{B}$ as [40]

$$\Gamma = \text{Supp}_{\text{chunk}}(\mathbf{B}) \triangleq \{\ell_d : \|\boldsymbol{B}^{[\ell_d]}\|_F \geq P_{th}, \quad 0 \leq d \leq L. \tag{19}$$

So, $\mathbf{B}$ is said to be $\mathcal{D}$-sparse if the chunk support of $\mathbf{B}$ satisfies $|\Gamma| = \mathcal{D} \ll K$.

The other conventional notion which is used to facilitate the performance analysis is define as [42]

Definition 3 (Block Restricted Isometry Property): A matrix $\mathbf{\Psi} \in \mathbb{C}^{N_{\mathcal{P}} \times KL}$ is said to satisfy the Block Restricted Isometry Property (Block-RIP) with $\mathcal{D}$-th order Block-RIP constant $\delta_{\mathcal{D}|K}$ for block size of K as

$$\delta_{\mathcal{D}|K} \triangleq \{\delta : (1-\delta)\|\mathbf{b}\|_2^2 \leq \|\mathbf{\Psi}\mathbf{b}\|_2^2 \leq (1+\delta)\|\mathbf{b}\|_2^2, \tag{20}$$

where $0 \leq \delta_{\mathcal{D}|K} < 1$ and $\text{Supp}_{\text{K}}(\mathbf{b}) \leq \mathcal{D}$, in which b is containing L blocks with size $K \times 1$.

III. SPARSE CHANNEL ESTIMATION

In this section, we show the properties of MSL positions and prove the relations between MSLs and MSTs. We present and prove Lemma 1 and Theorem 1. And based on them, we define a new set containing four MST positions. Furthermore, we propose an algorithm to estimate the sparse channel based on the proposed set.

TABLE 1. List of symbols used in the paper.

Symbol	Description
$\mathfrak{R}(\ell)$	correlation of the received signal over ℓ-th lag
$\hat{\mathfrak{R}}^{\zeta}(\ell)$	estimated version of $\mathfrak{R}(\ell)$ over ζ received signal
$\Gamma = \{\ell_0, \ell_1, ..., \ell_{\mathcal{D}-1}\}$	set of MST positions
$\Gamma_{diff} = \{\varepsilon_0, ..., \varepsilon_{\Sigma'-1}\}$	set of the MST positions difference $(\varepsilon_{\varsigma'} = \ell_i - \ell_j; i > j)$
$\Gamma_{corr} = \{\gamma_0, ..., \gamma_{\Sigma-1}\}$	set of MSL positions
$\Gamma_{diff}^{tap} = [\ell_0, ..., \ell_{\mathcal{D}-1}]$	subset of Γ_{diff} contains $\ell_i - \ell_j; \ell_j = 0$
$\Gamma_{diff}^{lag} = [\ell_2 - \ell_1, ..., \ell_{\mathcal{D}-1} - \ell_1]$	subset of Γ_{diff} contains $\ell_i - \ell_j; \ell_i, \ell_j \neq 0$
$\mathcal{T}$	the set made from MSLs, which contains four MSTs
P_{th}	the threshold to detect MSLs defined in (17)
$\delta_{\mathcal{D}|K}$	Block-RIP constant

A. POSITION MATRIX

In this sub-section, we define a new ascending set which includes the difference of all the MST positions.

Definition 4: Suppose that the channel has L taps in which $\mathcal{D}$ of them are nonzero (for the noiseless case) and the MST positions set is $\Gamma = \{\ell_0, \ell_1, \ldots, \ell_{\mathcal{D}-1}\}$ where $\ell_0 = 0$. Define a new ascending set as

$$\Gamma_{diff} \triangleq \{\varepsilon_{\varsigma'} : \varepsilon_{\varsigma'} = \ell_i - \ell_j;\ i > j \,\&\, 0 \leq \varepsilon_{\varsigma'} \leq \Sigma' - 1\}, \tag{21}$$

where $\ell_i \in \Gamma$ is the i'th position of the MSTs.

The set Γ_{diff}, contains all of potential MSLs positions and would has some identical elements. For more explanation, some of $\varepsilon_{\varsigma'} = \ell_i - \ell_j$, may have the same values, and the corresponding MSLs at that positions will be overlapped. In this situation, duplicate members in Γ_{diff} will have one representative in Γ_{corr}, therefore, we have $|\Gamma_{corr}| \leq \|\Gamma_{diff}|$. Now suppose the channel has $\Gamma_{corr} = \{\gamma_0, \gamma_1, \ldots, \gamma_{\Sigma-1}\}$. It is clear that the first and the last positions of MSTs can be obtained as

$$\begin{aligned} \ell_0 &= \gamma_0 \\ \ell_{\mathcal{D}-1} &= \gamma_{\Sigma-1}. \end{aligned} \tag{22}$$

We assume that $\mathcal{D}$ is known for obtaining the other MST positions. So, we define a new set $\mathcal{T}$ from estimated MSLs, which contains some information about the MST positions. As $\mathcal{T}$ contains at least four MST positions, we divide the problem into two parts. a) The channel has three or less than three nonzero taps ($\mathcal{D} \leq 3$), which $\mathcal{T}$ can contain two or three MSTs, and b) the channel has more than three nonzero taps ($\mathcal{D} > 3$), which $\mathcal{T}$ contains four MST positions. Therefore, we define $\mathcal{T}$ for two cases $\mathcal{D} \leq 3$ and $\mathcal{D} > 3$ as follows.

To facilitate our explanation, we first define a set of symbols in Table 1.

1) CHANNEL HAS $\mathcal{D} \leq 3$ NONZERO TAPS

If $\mathcal{D} = 2$, MSLs, and MSTs positions will be equal and channel estimation will be straightforward. Now suppose that

$\mathcal{D}=3$ with support set $\Gamma=\{\ell_0,\ell_1,\ell_2\}$. Using (22), ℓ_0 and ℓ_2 can be found, so we are trying to find ℓ_1 from estimated Γ_{corr}. if $\Gamma_{corr}=\{\gamma_0,\gamma_1,\gamma_2\}$, it is clear that ℓ_1 and $\ell_2-\ell_1$ have been overlapped in MSLs, so $\ell_1=\ell_2-\ell_1=\gamma_1$. On the other hand, if $\Gamma_{corr}=\{\gamma_0,\gamma_1,\gamma_2,\gamma_3\}$, then

$$\begin{cases}\gamma_2=\ell_1 \text{ and } \gamma_1=\ell_2-\ell_1 & \text{if } \ell_1>\ell_2-\ell_1\\ \gamma_1=\ell_1 \text{ and } \gamma_2=\ell_2-\ell_1 & \text{if } \ell_2-\ell_1>\ell_1,\end{cases} \tag{23}$$

so we are sure that ℓ_1 is equal to γ_2 or γ_1. Therefore, for $\mathcal{D}=3$ we can define the set $\mathcal{T}$ as

$$\mathcal{T}=\{\gamma_0,\gamma_1,\gamma_2,\gamma_3\}, \tag{24}$$

in which contains ℓ_0, ℓ_1 and ℓ_2.

2) CHANNEL HAS $\mathcal{D}>3$ NONZERO TAPS

For this situation, we present Lemma 1 and Theorem 1 as follows.

Lemma 1: Suppose a channel has $\mathcal{D}>3$ nonzero taps in positions $\Gamma=\{\ell_0,\ell_1,\ldots,\ell_{\mathcal{D}-1}\}$ in which $\ell_0=0$. If $\Gamma_{diff}=\{\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{\Sigma'-1}\}$ satisfies $\varepsilon_{\Sigma'-1}\neq\varepsilon_{\Sigma'-2}$, and $\Gamma_{corr}=\{\gamma_0,\gamma_1,\ldots,\gamma_{\Sigma-1}\}$, be the estimated MSL positions from (16) then

$$[\gamma_{\Sigma-3},\gamma_{\Sigma-2}]\in\big\{[\ell_{\mathcal{D}-1}-\ell_1,\ell_{\mathcal{D}-2}],[\ell_{\mathcal{D}-3},\ell_{\mathcal{D}-2}],\\ [\ell_{\mathcal{D}-2},\ell_{\mathcal{D}-1}-\ell_1],[\ell_{\mathcal{D}-1}-\ell_2,\ell_{\mathcal{D}-1}-\ell_1]\big\}. \tag{25}$$

Proof: To prove Lemma 1, at first we derive relation (25) for assumption $\varepsilon_i\neq\varepsilon_j;\forall\varepsilon_i,\varepsilon_j\in\Gamma_{diff}$, and then prove that it is correct for $\varepsilon_{\Sigma'-1}\neq\varepsilon_{\Sigma'-2}$. Now Suppose that $\varepsilon_i\neq\varepsilon_j;\forall\varepsilon_i,\varepsilon_j\in\Gamma_{diff}$. According to (16) and (21), the members of Γ_{corr} and Γ_{diff} are the same, i.e., $\varepsilon_{\varsigma'}=\gamma_\varsigma$ for $\varsigma=\varsigma'\in\{0,1,\ldots,\Sigma\}$, where $\Sigma=|\Gamma_{corr}|=|\Gamma_{diff}|$.

First we prove that $\varepsilon_{\Sigma-2}\in\{\ell_{\mathcal{D}-2},\ell_{\mathcal{D}-1}-\ell_1\}$. Define a new set $\bar{\Gamma}_{diff}\triangleq\Gamma_{diff}\setminus\varepsilon_{\Sigma'-1}=\{\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{\Sigma'-2}\}$, so $\gamma_{\Sigma-2}=\max(\bar{\Gamma}_{diff})$. Now divide $\bar{\Gamma}_{diff}$ into two subset as

$$\begin{aligned}\bar{\Gamma}_{diff}^{tap}&\triangleq[\ell_0,\ell_1,\ldots,\ell_{\mathcal{D}-2}]\\ \bar{\Gamma}_{diff}^{lag}&\triangleq\&[\ell_2-\ell_1,\ldots,\ell_{\mathcal{D}-1}-\ell_2,\ell_{\mathcal{D}-1}-\ell_1],\end{aligned} \tag{26}$$

where $\bar{\Gamma}_{diff}^{tap}\cup\bar{\Gamma}_{diff}^{lag}=\bar{\Gamma}_{diff}$ and $\bar{\Gamma}_{diff}^{tap}\cap\bar{\Gamma}_{diff}^{lag}=\emptyset$. We have $\max(\bar{\Gamma}_{diff})\in\{\max(\bar{\Gamma}_{diff}^{tap}),\max(\bar{\Gamma}_{diff}^{lag})\}$, while $\max(\bar{\Gamma}_{diff}^{tap})=\ell_{\mathcal{D}-2}$ and $\max(\bar{\Gamma}_{diff}^{lag})=\ell_{\mathcal{D}-1}-\ell_1$, therefore $\gamma_{\Sigma-2}=\varepsilon_{\Sigma'-2}=\max(\bar{\Gamma}_{diff})\in\{\ell_{\mathcal{D}-2},\ell_{\mathcal{D}-1}-\ell_1\}$. Now consider two case

Case 1: Let $\gamma_{\Sigma-2}=\ell_{\mathcal{D}-2}$ (we notice that $\gamma_\varsigma=\varepsilon_{\varsigma'}$ for $\varsigma=\varsigma'\in\{0,1,\ldots,\Sigma\}$), so to find $\gamma_{\Sigma-3}$, we define a new subset by dropping $\ell_{\mathcal{D}-2}$ from $\bar{\Gamma}_{diff}^{tap}$ as

$$\bar{\Gamma}_{diff}^{tap\setminus\ell_{\mathcal{D}-2}}\triangleq\bar{\Gamma}_{diff}^{tap}\setminus\ell_{\mathcal{D}-2}=\{\ell_0,\ell_1,\ldots,\ell_{\mathcal{D}-3}\}. \tag{27}$$

Therefore we have $\gamma_{\Sigma-3}\in\{\bar{\Gamma}_{diff}^{tap\setminus\ell_{\mathcal{D}-2}},\max(\bar{\Gamma}_{diff}^{lag})\}=\{\ell_{\mathcal{D}-3},\ell_{\mathcal{D}-1}-\ell_1\}$.

Case 2: If $\gamma_{\Sigma-2}=\ell_{\mathcal{D}-1}-\ell_1$, we will have $\gamma_{\Sigma-3}\in\{\ell_{\mathcal{D}-2},\max(\bar{\Gamma}_{diff}^{lag}\setminus(\ell_{\mathcal{D}-1}-\ell_1)\}$. Now if $\gamma_{\Sigma-3}=\max(\bar{\Gamma}_{diff}^{lag}\setminus(\ell_{\mathcal{D}-1}-\ell_1))>\ell_{\mathcal{D}-2}$, divide $\bar{\Gamma}_{diff}^{lag}$ to some new subsets as

$$\begin{aligned}\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)}&\triangleq\{\ell_{\mathcal{D}-1}-\ell_{\mathcal{D}-2},\ldots,\ell_{\mathcal{D}-1}-\ell_2,\ell_{\mathcal{D}-1}-\ell_1\}\\ \bar{\Gamma}_{diff}^{lag(\mathcal{D}-2)}&\triangleq\{\ell_{\mathcal{D}-2}-\ell_{\mathcal{D}-3},\ldots,\ell_{\mathcal{D}-2}-\ell_2,\ell_{\mathcal{D}-2}-\ell_1\}\\ &\vdots\\ \bar{\Gamma}_{diff}^{lag(2)}&\triangleq\{\ell_2-\ell_1\}.\end{aligned} \tag{28}$$

It is clear that $\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)})>\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-2)})>\ldots>\max(\bar{\Gamma}_{diff}^{lag(2)})$. If we remove $\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)})=\ell_{\mathcal{D}-1}-\ell_1$ from $\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)}$ and form the new set $\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)}\setminus\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)})=\{\ell_{\mathcal{D}-1}-\ell_{\mathcal{D}-2},\ldots,\ell_{\mathcal{D}-1}-\ell_2\}$, we can not decide $\max(\bar{\Gamma}_{diff}^{lag}\setminus\ell_{\mathcal{D}-1}-\ell_1)=\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)}\setminus\ell_{\mathcal{D}-1}-\ell_1)=\ell_{\mathcal{D}-1}-\ell_2$ or $\max(\bar{\Gamma}_{diff}^{lag}\setminus\ell_{\mathcal{D}-1}-\ell_1)=\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-2)})=\ell_{\mathcal{D}-2}-\ell_1$, but it is clear that $\ell_{\mathcal{D}-2}>\ell_{\mathcal{D}-2}-\ell_1$. Therefore if $\ell_{\mathcal{D}-1}-\ell_2>\ell_{\mathcal{D}-2}$ we can conclude that $\ell_{\mathcal{D}-1}-\ell_2>\ell_{\mathcal{D}-2}-\ell_1$ and so $\max(\bar{\Gamma}_{diff}^{lag}\setminus\ell_{\mathcal{D}-1}-\ell_1)=\ell_{\mathcal{D}-1}-\ell_2$. Thus, if $\gamma_{\Sigma-2}=\ell_{\mathcal{D}-1}-\ell_1$, then $\gamma_{\Sigma-3}\in\{\max(\bar{\Gamma}_{diff}^{tap}),\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-1)}\setminus\ell_{\mathcal{D}-1}-\ell_1)\}=\{\ell_{\mathcal{D}-2},\ell_{\mathcal{D}-1}-\ell_2\}$.

From the above discussions, relation (25) is derived for assumption $\varepsilon_i\neq\varepsilon_j;\forall\varepsilon_i,\varepsilon_j\in\bar{\Gamma}_{diff}$. Now we prove that (25) is correct for $\varepsilon_{\Sigma'-2}\neq\varepsilon_{\Sigma'-3}$ and the assumption $\varepsilon_i\neq\varepsilon_j;\forall\varepsilon_i,\varepsilon_j\in\bar{\Gamma}_{diff}$ is not necessary. When $\varepsilon_{\Sigma'-2}\neq\varepsilon_{\Sigma'-3}$, $\gamma_{\Sigma-2}$ will be equal to $\varepsilon_{\Sigma'-2}$ and $\gamma_{\Sigma-2}\neq\varepsilon_i$ for $i=\Sigma'-3,\Sigma'-4,\ldots,0$; therefore to determine $\gamma_{\Sigma-2}$, there is no difference between $\varepsilon_i\neq\varepsilon_j;\forall\varepsilon_i,\varepsilon_j\in\bar{\Gamma}_{diff}$ and $\varepsilon_{\Sigma'-2}\neq\varepsilon_{\Sigma'-3}$.

For $\gamma_{\Sigma-3}$, we have $\gamma_{\Sigma-3}=\max(\bar{\Gamma}_{diff}\setminus\varepsilon_{\Sigma'-2})=\varepsilon_{\Sigma'-3}$. Now suppose that $\varepsilon_{\Sigma'-3}=\varepsilon_{\Sigma'-4}=\varepsilon_{\Sigma'-3-n}$ for $1\leq n\leq\Sigma'-3$, in this situation $\max(\bar{\Gamma}_{diff}\setminus\varepsilon_{\Sigma'-2})$ will not change and $\gamma_{\Sigma-3}=\varepsilon_{\Sigma'-3}=\varepsilon_{\Sigma'-4}=\ldots=\varepsilon_{\Sigma'-3-n}$. So Lemma 1 is proved. ■

In fact, (25) stats that, $[\gamma_{\Sigma-3},\gamma_{\Sigma-2}]$ is definitely a member of the set $\big\{[\ell_{\mathcal{D}-1}-\ell_1,\ell_{\mathcal{D}-2}],[\ell_{\mathcal{D}-3},\ell_{\mathcal{D}-2}],[\ell_{\mathcal{D}-2},\ell_{\mathcal{D}-1}-\ell_1],[\ell_{\mathcal{D}-1}-\ell_2,\ell_{\mathcal{D}-1}-\ell_1]\big\}$. Now based on Lemma 1, we are going to define a set containing two positions of channel's MST.

Theorem 1: Suppose a channel with $\mathcal{D}>3$ nonzero taps in positions $\Gamma=\{\ell_0,\ell_1,\ldots,\ell_{\mathcal{D}-1}\}$ where $\ell_0=0$ and the channel's MSL positions be $\Gamma_{corr}=\{\gamma_0,\gamma_1,\ldots,\gamma_{\Sigma-1}\}$. Let define the set $\bar{\mathcal{T}}$ as

$$\begin{aligned}\bar{\mathcal{T}}\triangleq\{&[\gamma_{\Sigma-3},\gamma_{\Sigma-1}-\gamma_{\Sigma-2}]^T,[\gamma_{\Sigma-1}-\gamma_{\Sigma-3},\gamma_{\Sigma-1}-\gamma_{\Sigma-2}]^T,\\ &[\gamma_{\Sigma-1}-\gamma_{\Sigma-3},\gamma_{\Sigma-2}]^T,\\ &[\gamma_{\Sigma-2}-\gamma_{\Sigma-3},\gamma_{\Sigma-2}]^T\}.\end{aligned} \tag{29}$$

Then $\bar{\mathcal{T}}$ surely includes a vector with two channel tap positions.

Proof: See Appendix ■

In Theorem 1, we defined a set of five elements, one of which contains two positions of MSTs. Now suppose that Σ taps of the MSLs have been detected and the channel has $\mathcal{D}$ nonzero

TABLE 2. Notations used in algorithm 1.

$\mathbf{b}^{\Gamma}$	sub-vector formed by collecting the entries of $\mathbf{b}$ indexed by Γ
$\mathbf{B}^{[\Gamma]}$	sub-matrixes formed by collecting the chunks of $\mathbf{B}$ indexed by Γ
$\mathbf{\Psi}^{\Gamma}$	sub-matrixes formed by collecting the columns of $\mathbf{\Psi}$ indexed by Γ
$\mathbf{\Psi}_{[\Gamma]}$	sub-matrix formed by collecting columns of $\mathbf{\Psi}$ indexed by $\{(k-1)L+1, ..., kL : \forall k \in \Gamma\}$

taps, so we can define set $\mathcal{T}$ from (23) and (29) as

$$\mathcal{T} \triangleq \{\gamma_0, \gamma_{\Sigma-1}-\gamma_{\Sigma-2}, \gamma_{\Sigma-1}-\gamma_{\Sigma-3}, \gamma_{\Sigma-2}-\gamma_{\Sigma-3}, \gamma_{\Sigma-3}, \gamma_{\Sigma-2}, \gamma_{\Sigma-1}\}. \quad (30)$$

According to Theorem 1, the set $\mathcal{T}$ includes four MST positions of the channel.

B. ALGORITHM DESIGN

If the aperture of the BS antenna array is not very large, then there is a common sparsity in channels between a user and different receive antennas [40]. Therefore, the equivalent channel matrix $\mathbf{B}$ in (8) exhibits structured sparsity. Using this property and exploiting information about the nonzero tap's positions, we propose the Enhanced Modified-SP (EM-SP) algorithm to estimate the MU massive MIMO channel.

Our proposed algorithm has been developed from the subspace pursuit (SP) algorithm [36], a greedy CS recovery algorithm with a good restricted RIP constant, uniform recovery guarantee, and resistance against noise [36], making it suitable for our work. Our algorithm combines the SP algorithm and Theorem 1. Here $\hat{\mathcal{P}}_i$ and $\hat{\mathbf{B}}_{(i)}$ are the estimated support and channel outputs in the i'th iteration respectively. $\mathcal{D}$ is the sparsity of the signal which is assumed to be known, $\tilde{\mathbf{Y}} = [\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \ldots, \tilde{\mathbf{y}}_M] \in \mathbb{C}^{N\times M}$ is the time-domain received signal and $\mathbf{Y}^{\mathcal{P}} \in \mathbb{C}^{N_{\mathcal{P}}\times M}$ is the frequency-domain received signal at pilot subcarriers. As explained in (8), the sparse matrix $\mathbf{B}$ is a chain of L chunks with size $K \times M$.

In **Step 1** A, using 16, the MSL's positions of the channel is calculated. If $\mathcal{D} = 2$, it is clear from 22 that $\gamma_0 = \ell_0$ and $\gamma_1 = \ell_1$ and so MST positions are obtained and $\hat{\mathbf{B}}$ is calculated. On the other hand, if $|\Gamma_{corr}| = \mathcal{D}$, it means that all positions of MSLs fall on the positions of the MSTs. So, $\gamma_{\varsigma} = \ell_d$ for $\varsigma = d = 0, 1, \ldots, \mathcal{D}-1$ and $\hat{\mathbf{B}}$ can be calculated. Otherwise, if $\mathcal{D} \neq 2$ and $|\Gamma_{corr}| \neq \mathcal{D}$, form set $\mathcal{T}$ according to (24) or (30), using members of calculated Γ_{corr}. As can be seen in **Step 2** A (31), Γ_b is selected from $\mathcal{T}$.

The proposed (EM-SP) algorithm is detailed in Algorithm 1. For brevity, the associated notations are given in Table 2.

In **Step 2** A (31), Γ_c is defined as $\Gamma_c = \{\gamma_1, \gamma_2, \ldots, \gamma_{\Sigma-2}-1\}\setminus\Gamma_b$ and then is used in (32) to form Γ_d. The reason we used $\{\gamma_1, \gamma_2, \ldots, \gamma_{\Sigma-2}-1\}\setminus\Gamma_b$ instead of $\{0, 1, \ldots, L-1\}\setminus\Gamma_b$ to form Γ_c, where L is the channel length, is that as shown in Theorem 1, $\gamma_{\Sigma-2}$ in the set $\mathcal{T}$ is $\ell_{\mathcal{D}-2}$ or $\ell_{\mathcal{D}-1} = \ell_1$. Now if $\gamma_{\Sigma-2} = \ell_{\mathcal{D}-2}$, then $\ell_{\mathcal{D}-2}$ will be a member of Γ_b. So we are sure that the largest MST position in Γ_d is less than $\ell_{\mathcal{D}-2}$, and therefore the search range of Γ_d is up to $\ell_{\mathcal{D}-2}-1 = \gamma_{\Sigma-2}-1$. On the other hand, if $\gamma_{\Sigma-2} = \ell_{\mathcal{D}-1} - \ell_1$, since $\ell_{\mathcal{D}-2} \leq \ell_{\mathcal{D}-1} - \ell_1$, searching in range $\{\gamma_1, \gamma_2, \ldots, \gamma_{\Sigma-2}-1\}\setminus\Gamma_b$ will cover all modes of Γ_d. Also in **Step 2** C (33), with a similar argument to **Step 2** A (31)-(33), the signal support $\hat{\Gamma}_{i+1}$ from $\mathcal{T}$ and $\mathcal{D}-|\Gamma_1|$ are formed. Finally, the stopping criteria **Step 2** F is similar to the SP algorithm.

Algorithm 1 Enhanced Modified-SP (EM-SP) Algorithm

Input: $\tilde{\mathbf{Y}}$, $\mathbf{Y}^{\mathcal{P}}$, $\mathbf{\Psi}$, $\mathcal{D}$

Output: Output: Estimated chunk support $\hat{\Gamma}$ and channel's matrix $\hat{\mathbf{B}}$.

Step 1 (Initialization):

- A: Calculate $\Gamma_{corr} \triangleq \{\gamma_0, \gamma_1, \ldots, \gamma_{\Sigma-1}\}$ using (16). If $\mathcal{D} = 2$ or $|\Gamma_{corr}| = \mathcal{D}$, stop and output $\hat{\Gamma} = \Gamma_{corr}$, $\hat{\mathbf{B}} = \mathbf{\Psi}^{\dagger}_{[\hat{\Gamma}]}\mathbf{Y}^{\mathcal{P}}$ and $\hat{\mathbf{B}}^{[\{1,2,\ldots,L\}\setminus\hat{\Gamma}]} = 0$; Otherwise, define the set $\mathcal{T}$ for $\mathcal{D} = 3$ using Equation (24) and for $\mathcal{D} > 3$ using (30).
- B: Initialize the iteration index $i = 0$, channel support $\hat{\Gamma}_i = \emptyset$, and the residue matrix $\mathbf{r}_{res}(0) = \mathbf{Y}^{\mathcal{P}}$.

Step 2 (Iteration)):Repeat the following steps.

- A (Support Merge): Set $\Gamma_a = \hat{\Gamma}_i \cup (\Gamma_b \cup \Gamma_d)$ where

$$\Gamma_b = \arg\max_{|\Gamma_1|=4, \Gamma_1\in\mathcal{T}} \|(\mathbf{\Psi}^H\mathbf{r}_{res}(i))^{[\Gamma_1]}\|_F, \quad (31)$$

$$\Gamma_c = \{\gamma_1, \ldots, \gamma_{\Sigma-2}\}\setminus\Gamma_b, \quad (32)$$

$$\Gamma_d = \arg\max_{|\Gamma_2|=\mathcal{D}-|\Gamma_1|, \Gamma_2\subseteq\Gamma_c} \|(\mathbf{\Psi}^H\mathbf{r}_{res}(i))^{[\Gamma_2]}\|_F. \quad (33)$$

- B (LS Estimation): Set $\mathbf{Z}^{[\Gamma_a]} = \mathbf{\Psi}^{\dagger}_{[\Gamma_a]}\mathbf{Y}^{\mathcal{P}}$ and $\hat{\mathbf{Z}}^{[\{1,2,\ldots,L\}\setminus\Gamma_a]} = 0$.
- C (Support Refinement): Select $\hat{\Gamma}_{i+1}$ as follows

$$\hat{\Gamma}_{i+1} = \left\{\arg\max_{|\Gamma_1|=4, \Gamma_1\in\mathcal{T}} \|\mathbf{Z}^{[\Gamma_1]}\|_F\right\} \cup \left\{\arg\max_{|\Gamma_2|=\mathcal{D}-|\Gamma_1|, \Gamma_2\subseteq\Gamma_c} \|\mathbf{Z}^{[\Gamma_2]}\|_F\right\}. \quad (34)$$

- D (Signal Estimation): Set $\hat{\mathbf{B}}^{[\hat{\Gamma}_{i+1}]}_{i+1} = \mathbf{\Psi}^{\dagger}_{[\hat{\Gamma}_{i+1}]}\mathbf{Y}^{\mathcal{P}}$ and $\hat{\mathbf{B}}^{[\{1,2,\ldots,L\}\setminus\hat{\Gamma}_{i+1}]} = 0$.
- E (Residue): Compute $\mathbf{r}_{res}(i+1) = \mathbf{Y}^{\mathcal{P}} - \mathbf{\Psi}_{[\hat{\Gamma}_{i+1}]}\hat{\mathbf{B}}^{[\hat{\Gamma}_{i+1}]}_{i+1}$.
- F (Stopping Condition and Output): if $\|\mathbf{r}_{res}(i+1)\|_F \geq \|\mathbf{r}_{res}(i)\|_F$ stop and output $\hat{\Gamma} = \hat{\Gamma}_i$ and $\hat{\mathbf{B}} = \hat{\mathbf{B}}_i$; Otherwise $i = i+1$ and go to **Step2** A.

C. RECONSTRUCTION ERROR BOUND

The reconstruction error bound is described in terms of the restricted isometry property (RIP), which is given in [36, Definition 3]. Suppose that the measurement matrix Ψ has block-RIP properties with block- restricted isometry constant (block-RIC) given by $\delta_{\mathcal{D}|K}$, and define $\mathcal{D}_1$ based on Γ_a in

Step 2 A of the proposed EM-SP Algorithm as

$$|\Gamma \cup \Gamma_a| \leq \mathcal{D}_1 \triangleq 3\mathcal{D} + (|\mathcal{T}| - 12) + \min(0, |\Gamma_{corr}| - 3(\mathcal{D} - 4)). \quad (35)$$

Suppose that $\mathcal{D}_1$'th block-RIC satisfies $\delta_{\mathcal{D}_1|K} < 0.2412$, then $\mathcal{D}$-sparse matrix $\mathbf{B}$ will be reconstructed with the error bounded as [43]

$$\|\mathbf{B} - \hat{\mathbf{B}}\|_F \leq C_1 \|\mathbf{Z}\|_F, \quad (36)$$

where $\mathbf{Z}$ is the noise matrix and

$$C_1 = \left(1 + \frac{\delta_{\mathcal{D}_1|K}\sqrt{1+\delta_{\mathcal{D}_1|K}}}{\sqrt{1-\delta_{\mathcal{D}_1|K}}}\right) \times \frac{2 + \sqrt{1+\delta_{\mathcal{D}_1|K}} \times C_2}{\sqrt{1-\delta_{\mathcal{D}_1|K}} - \sqrt{1+\delta_{\mathcal{D}_1|K}} \times C_3} + \frac{1}{\sqrt{1-\delta_{\mathcal{D}_1|K}}}, \quad (37)$$

in which

$$C_2 = \frac{2\delta_{\mathcal{D}_1|K}}{1-\delta_{\mathcal{D}_1|K}} \sqrt{1 + \delta^2_{\mathcal{D}_1|K} \frac{1+\delta_{\mathcal{D}_1|K}}{1-\delta_{\mathcal{D}_1|K}}} \times \sqrt{1 + 4\delta^2_{\mathcal{D}_1|K} \frac{1+\delta_{\mathcal{D}_1|K}}{1-\delta_{\mathcal{D}_1|K}}}, \quad (38)$$

and

$$C_3 = \frac{2\sqrt{1+\delta_{\mathcal{D}_1|K}}}{1-\delta_{\mathcal{D}_1|K}} \sqrt{1 + 4\delta^2_{\mathcal{D}_1|K} \frac{1+\delta_{\mathcal{D}_1|K}}{1-\delta_{\mathcal{D}_1|K}}} + \frac{2}{\sqrt{1-\delta_{\mathcal{D}_1|K}}}. \quad (39)$$

D. COMPUTATIONAL COMPLEXITY

Each iteration of the proposed EM-SP algorithm has the following computational complexity levels. **Step 1** implement the received signal correlation using FFT with complexity of $\mathcal{O}(\zeta^2(N\log_2 N)/2)$. For **Step 2** A the complexity of the support merger is $\mathcal{O}(L)$ [35] and for **Step 2** A (31), the complexity of correlation operation is $\mathcal{O}(MLN_{\mathcal{P}}K^2)$, norm operation is $\mathcal{O}(|\mathcal{T}| \times MK)$ and the max operation is $\mathcal{O}(|\mathcal{T}|)$. For **Step 2** A (32), the complexity of correlation operation is $\mathcal{O}(MLN_{\mathcal{P}}K)$, norm operation is $\mathcal{O}((\Sigma - 4) \times MK)$, and the max operation complexity is $\mathcal{O}(\max(0, \mathcal{D} - 4))$. For **Step 2** B, the complexity of Moore-penrose matrix inversion operation is $\mathcal{O}(2KN_{\mathcal{P}}(K\mathcal{D})^2 + (K\mathcal{D})^3)$ [44]. For **Step 2** C (34), the complexity of norm operations are $\mathcal{O}(|\mathcal{T}| \times MK)$ and $\mathcal{O}((\Sigma - 4) \times MK)$, and the complexity of the max operations are $\mathcal{O}(5)$ and $\mathcal{O}(\max(0, \mathcal{D} - 4))$. For **Step 2** D and E, the Moore-penrose matrix inversion has complexity of $\mathcal{O}(2KN_{\mathcal{P}}(K\mathcal{D})^2 + (K\mathcal{D})^3)$ and the residue update has complexity of $\mathcal{O}(MLN_{\mathcal{P}}K^2)$. So the total computational complexity of EM-SP algorithm is $C_{EM-SP} = \mathcal{O}(\zeta^2(N\log_2 N)/2) + \mathcal{O}(L) + 2 \times \big(\mathcal{O}(2KN_{\mathcal{P}}(K\mathcal{D})^2 + (K\mathcal{D})^3) + \mathcal{O}(MLN_{\mathcal{P}}K^2) + \mathcal{O}(|\mathcal{T}| \times MK) + \mathcal{O}(\max(0, \mathcal{D} - 4)) + \mathcal{O}(|\mathcal{T}|) + \mathcal{O}((\Sigma - 4) \times MK)\big)$.

TABLE 3. Simulation parameters.

Carrier Frequency (f_c)	2 GHz
Bandwidth	10 MHz
Doppler frequency (f_d)	110 Hz
Reference distance (r_{ref})	100 m
Number of carriers (N)	1024
Modulation	QPSK
Number of BS antennas (M)	128
Number of users (K)	8
Channel length (L)	64
Path loss exponent (α)	3.5
Standard deviation of shadowing (σ_ψ)	4 dB
Threshold constant (K_e)	1

We notice that the EM-SP algorithm is an upgraded version of the SP [36], by implementation of the received signals correlation. The computational complexity of the SP is just as EM-SP except that EM-SP has additional $\mathcal{O}(\zeta^2(N\log_2 N)/2)$ computation for received signals correlation, and the complexity of the norm and max operations in (31)-(33) of the SP are $\mathcal{O}(MKL)$ and $\mathcal{O}(L)$, respectively. Therefore, the total complexity of the SP algorithm is $C_{SP} = \mathcal{O}(L) + 2 \times \big(\mathcal{O}(2KN_{\mathcal{P}}(K\mathcal{D})^2 + (K\mathcal{D})^3) + \mathcal{O}(MLN_{\mathcal{P}}K^2) + \mathcal{O}(MLK) + \mathcal{O}(L)\big)$.

IV. SIMULATION RESULTS

This section considers uplink channel estimation in MU massive MIMO-OFDM systems in a single-cell scenario. We assume $N = 1024$ subcarriers, and $N_{\mathcal{P}}$ of them are pilot subcarriers. The pilot overhead ratio is $\eta_{\mathcal{P}} = (KN_{\mathcal{P}})/N$. The simulations are done over the channel with common spatial sparsity and the orthogonal pilot design with random pilot subcarrier locations in each transmitter. The sparse multi-path Rayleigh fading channel is modeled by $L = 64$ taps in which the locations of $\mathcal{D}$ nonzero taps are determined randomly in 1000 iterations. The small-scale fading coefficients are i.i.d. complex Gaussian random variables with zero means and unit variance, while the large-scale coefficient is modeled per (2). Table 3 gives the simulation parameters and their corresponding values.

To illustrate the efficacy of our proposed algorithm, we consider the sparsity $\mathcal{D} \leq 3$ and $\mathcal{D} > 3$ to examine the correctness of (24) and (30). For this, the simulation study is done for $\mathcal{D} = 3$ and $\mathcal{D} = 6$. It should be mentioned that equations (24) and (30) are correct for any values of D. For performance comparison, we consider the normalized mean squared error (NMSE) of the Genie-aided LS (GaLS) as the benchmark, (which is $\frac{1}{Mc}\sum_{i=1}^{Mc}\frac{\|h_i - \hat{h}_i\|_2^2}{\|h_i\|_2^2}$, where h_i and $\hat{h}_i$ are i-th true and estimated channel and Mc is the number of Monte Carlo iterations). In GaLS, we assume that the least square algorithm knows the actual channel support t (hence, Genie-aided least square), and the least square recovers the channel coefficients. Moreover, we compare the proposed EM-SP with two other algorithms, SP [36] and OMP [37], to quantify its better NMSE and the extra cost of computational complexity.

A. MSE PERFORMANCES

In the first step, we determine the value for ζ in equation (14). For this, we compare the NMSE of the EM-SP algorithm with Genie-aided LS versus ζ. For a better numerical comparison, we define Δ_{NMSE} as the difference between the NMSE of the EM-SP ($NMSE_{EM-SP}$) and the NMSE of the GaLS ($NMSE_{GaLS}$) as $\Delta_{NMSE} = NMSE_{EM-SP} - NMSE_{GaLS}$.

Fig. 1 compares Δ_{NMSE} of the channels with $\mathcal{D} = 3$ and 6 for SNRs equal to 10 dB, 20 dB, and 30 dB versus the value of ζ. As per Fig. 1, Δ_{NMSE} is almost the same for $\zeta \geq 5$ and $\zeta \geq 8$ for the channels with the sparsity of $\mathcal{D} = 3$ and 6. Therefore, setting $\zeta = 5$ for simulations of the channel with $\mathcal{D} = 3$ and $\zeta = 9$ for $\mathcal{D} = 6$, can provide accurate channel estimation.

In the last section, we show that the computational complexity of the received signal correlation is proportional to ζ^2, so a smaller ζ can significantly reduce computational complexity. To quantitatively compare the computational complexity of the EM-SP algorithm with the SP algorithm, for the channels with $\mathcal{D} = 3$ and $\mathcal{D} = 6$, the ratio of the computational complexity of EM-SP and SP algorithms, C_{EM-SP}/C_{SP}, is about 1.003 and 1.02. Thus, the EM-SP algorithm's computational complexity is about 0.3% and 2% more than that of the SP algorithm for $\mathcal{D} = 3$ and $\mathcal{D} = 6$, respectively.

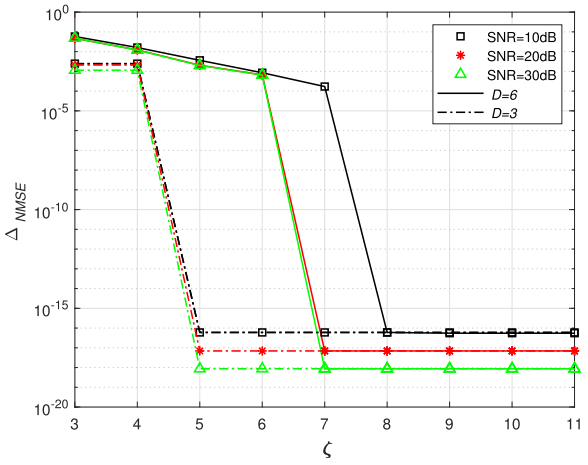

FIGURE 1. The difference between NMSE of EM-SP and GaLS versus ζ.

Figs. 2 and 3, provide the NMSE comparison of the proposed EM-SP algorithm with SP and GaLS algorithms, considering spars channel with $\mathcal{D} = 3$ and $\mathcal{D} = 6$ nonzero taps, for one OFDM symbol, $K = 8$ users (with random locations in the cell) and $M = 128$ antennas. We observe that the NMSE of the EM-SP is close to that of the GaLS across different pilot overhead ratios at $SNR = 10$ dB, 20 dB, and 30 dB, while the SP algorithm is far from the ideal case for small numbers of pilot subcarriers. In other words, it is inefficient if the number of pilot subcarriers is insufficient. Thus, the EM-SP algorithm requires fewer pilot subcarriers to reach the desired error, representing a solid improvement in spectral efficiency. This feature is a significant advantage

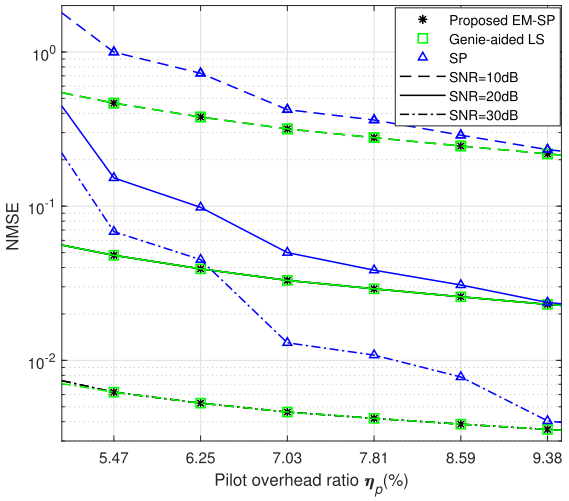

FIGURE 2. NMSE versus pilot overhead ratio, channel has $\mathcal{D} = 3$ nonzero taps, for $K = 8$ and $M = 128$.

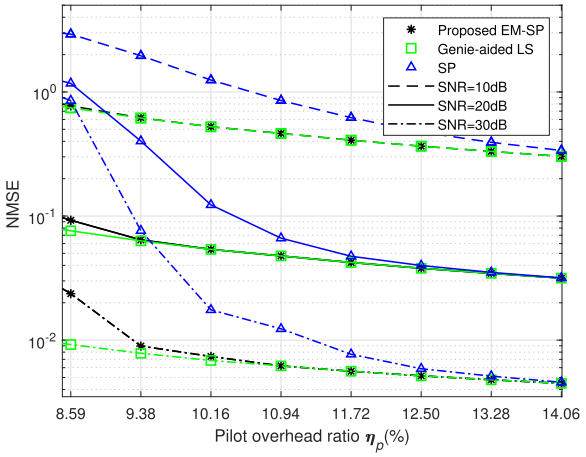

FIGURE 3. NMSE versus pilot overhead ratio, channel has $\mathcal{D} = 6$ nonzero taps, for $K = 8$ and $M = 128$.

of our proposed method since, in MU massive MIMO systems, the small number of pilot subcarriers retains spectral efficiency because of the large number of antennas.

This reduction of pilot subcarriers was predictable based on theory. For measurement matrix Φ, the number of pilots to achieve $\mathcal{D}_1$'th order RIP with block-RIC of $\delta_{\mathcal{D}_1|K}$ is $N_{\mathcal{P}} = \mathcal{O}(\delta_{\mathcal{D}_1|K} \ln \delta^{-1}_{\mathcal{D}_1|K} + \delta^{-1}_{\mathcal{D}_1|K} \mathcal{D}_1 \log L)$ [42]. In other words, the pilot overhead is proportional to the inverse of parameter $\delta_{\mathcal{D}_1|K}$, and a smaller block-RIC leads to a reduction in the number of required pilots. Furthermore, for above conditions, from (35), we have $\mathcal{D}_1 < 3\mathcal{D}$ and therefore, $\delta_{\mathcal{D}_1|K} < \delta_{3\mathcal{D}|K}$ based on monotonicity of RIC [36]. From the above analysis, we conclude that by using the correlation of the received signal, our proposed EM-SP can reduce the number of required training pilots in massive MIMO systems. This better performance costs about 0.3% and 2% more computational complexity than the SP algorithm.

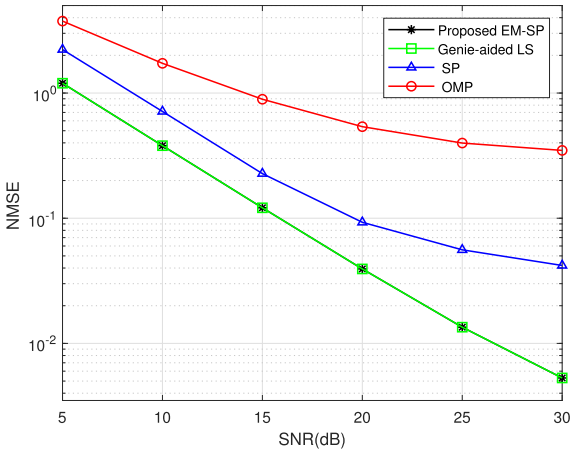

FIGURE 4. NMSE versus SNR, channel has $\mathcal{D} = 3$ nonzero taps, for $K = 8$, $M = 128$, and $\eta_{\mathcal{P}} = 6.25\%$.

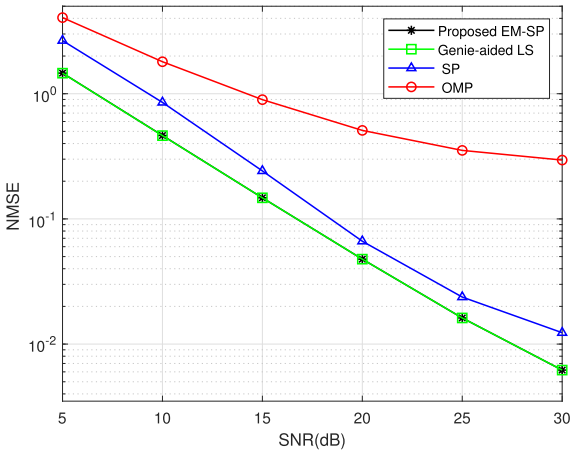

FIGURE 5. NMSE versus SNR, channel has $\mathcal{D} = 6$ nonzero taps, for $K = 8$, $M = 128$, and $\eta_{\mathcal{P}} = 10.93\%$.

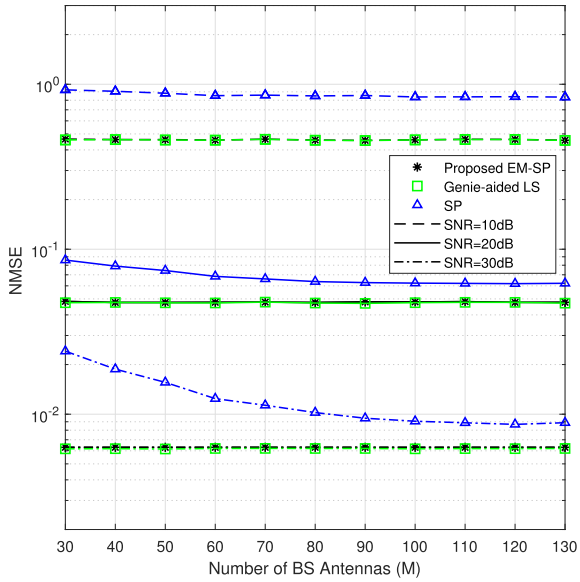

FIGURE 6. NMSE versus number of BS antennas, channel has $\mathcal{D} = 6$ nonzero taps, for $K = 8$, $M = 128$, and $\eta_{\mathcal{P}} = 10.93\%$.

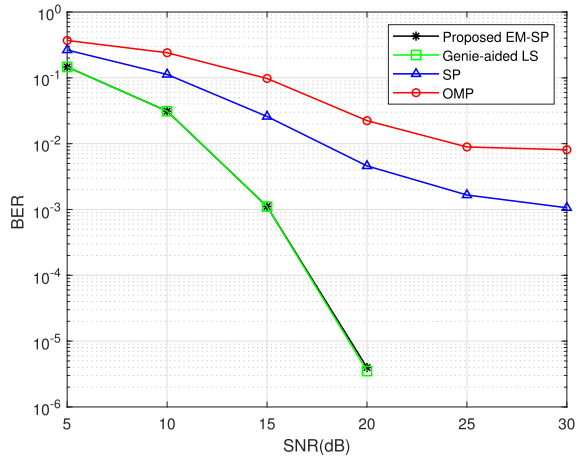

FIGURE 7. BER versus SNR, channel has $\mathcal{D} = 3$ nonzero taps, for $K = 8$, $M = 128$, and $\eta_{\mathcal{P}} = 6.25\%$.

Figs. 4 and 5 compare the NMSE performance of the proposed EM-SP algorithm with those of OMP, SP, and GaLS in different SNRs with $\eta_{\mathcal{P}} = 5.46$ and $\eta_{\mathcal{P}} = 11.71$ pilot overhead ratio in an MU massive MIMO system with $\mathcal{D}$ and six nonzero taps. Notably, we calculate the set $\mathcal{T}$ in our proposed algorithm using (24) for $\mathcal{D} = 3$ and (30) for $\mathcal{D} = 6$. We can observe that the NMSE of the EM-SP algorithm is close to that of the GaLS. With a lower pilot overhead ratio, typical estimators' performances are far from that of the GaLS, while the EM-SP algorithm is close. From Fig. 4, we observe that our proposed EM-SP algorithm has about 10.3 dB margin over the SP and 19 dB over the OMP at SNR=30 dB, as in Fig. 5, the EM-SP algorithm has 3 dB margin over the SP and 17 dB over the OMP. The proposed EM-SP algorithm leverages the correlation of the received signal for better NMSE performance.

We present the performance of different estimators versus the number of BS antennas in Fig. 6. As it shows, unlike subspace methods [23], [24], [25], [26], [27], decreasing the number of BS antennas does not affect the performance of the proposed EM-SP, which is close to the Genie-aided estimator.

B. BER PERFORMANCES

We have illustrated the superiority of our proposed algorithm in Figs. 2-6. In our evaluations, we use a more convincing and straightforward metric, Bit Error Rate (BER). In Figs. 7-8, we compare the BER performance versus the SNR for the EM-SP algorithm and the conventional channel estimators. Fig. 7, compares the BER versus the SNR when the channel has $\mathcal{D} = 3$ nonzero taps, as well as Fig. 8, which compares the BER versus the SNR when the channel has $\mathcal{D} = 6$ nonzero taps. As shown in both Figs 7 and 8, the proposed algorithm significantly outperforms the conventional estimators, especially for higher SNRs, and is slightly off from the ideal case with perfect channel support knowledge. This behavior shows

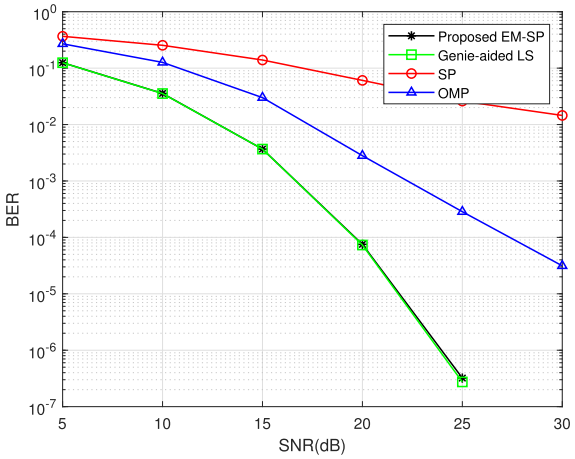

FIGURE 8. BER versus SNR, channel has $\mathcal{D} = 6$ nonzero taps, for $K = 8$, $M = 128$, and $\eta_{\mathcal{P}} = 10.93\%$.

that our algorithm achieves a favorable tradeoff between high spectral efficiency and complexity for MU massive MIMO systems.

V. CONCLUSION

In this paper, we proposed an MST detection approach for sparse channel estimation by getting information about channel support from the second-order statistics of the received signal. We first proved relations between the MSTs and MSLs of the correlation function of the received OFDM signal. Then, we proved Lemma 1 and Theorem 1, which have been used to propose a novel set with four MST positions. Moreover, for any number of taps, we exploited this novel set in our proposed EM-SP algorithm to enhance the channel estimation for MU massive MIMO-OFDM systems. The proposed algorithm significantly outperforms the existing channel estimators.

APPENDIX
PROOF FOR THE THEOREM 1

Define the ascending set $\Gamma_{diff} = \{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{\Sigma'-1}\}$ from (21), where contains all possible positions for MSLs, in which some $\varepsilon_i \in \Gamma_{diff}$ may be duplicates. Now we consider three cases as follow

Case 1: Suppose that $\varepsilon_{\Sigma'-2} \neq \varepsilon_{\Sigma'-3}$, then

$$\begin{aligned} \gamma_{\Sigma-1} &= \varepsilon_{\Sigma'-1}, \\ \gamma_{\Sigma-2} &= \varepsilon_{\Sigma'-2}, \\ \gamma_{\Sigma-3} &= \varepsilon_{\Sigma'-3}. \end{aligned} \tag{40}$$

From Lemma 1, if $\gamma_{\Sigma-2} = \varepsilon_{\Sigma'-2} = \ell_{\mathcal{D}-2}$ then from (22) and (40), we have $[\gamma_{\Sigma-3}, \gamma_{\Sigma-2}] = [\ell_{\mathcal{D}-3}, \ell_{\mathcal{D}-2}]$ when $\varepsilon_{\Sigma'-2} = \ell_{\mathcal{D}-2}$, and $[\gamma_{\Sigma-1} - \gamma_{\Sigma-3}, \gamma_{\Sigma-2}] = [\ell_1, \ell_{\mathcal{D}-2}]$ when $\varepsilon_{\Sigma'-2} = \ell_{\mathcal{D}-1} - \ell_1$.

Now suppose $\gamma_{\Sigma-2} = \varepsilon_{\Sigma'-2} = \ell_{\mathcal{D}-1} - \ell_1$, then we will have $\gamma_{\Sigma-1} - \gamma_{\Sigma-2} = \ell_1$, and from Lemma 1, $[\gamma_{\Sigma-1}, \gamma_{\Sigma-3}] = [\ell_1, \ell_{\mathcal{D}-2}]$ when $\varepsilon_{\Sigma'-3} = \ell_{\mathcal{D}-3}$, and $[\gamma_{\Sigma-1} - \gamma_{\Sigma-2}, \gamma_{\Sigma-1} - \gamma_{\Sigma-3}]$ when $\varepsilon_{\Sigma'-3} = \ell_{\mathcal{D}-1} - \ell_1$.

Case 2: Suppose that $\gamma_{\Sigma-2} = \varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-3} \neq \varepsilon_{\Sigma'-4} = \gamma_{\Sigma-3}$. From Lemma 1, $[\varepsilon_{\Sigma'-3}, \varepsilon_{\Sigma'-2}] \in \mathcal{T}_1$, where $\mathcal{T}_1$ is define as

$$\begin{aligned} \bar{\mathcal{T}}_1 = \{&[\ell_{\mathcal{D}-1} - \ell_1, \ell_{\mathcal{D}-2}], [\ell_{\mathcal{D}-3}, \ell_{\mathcal{D}-2}], \\ &[\ell_{\mathcal{D}-2}, \ell_{\mathcal{D}-1} - \ell_1], [\ell_{\mathcal{D}-1} - \ell_2, \ell_{\mathcal{D}-1} - \ell_1]\}. \end{aligned} \tag{41}$$

Now we have two situations

$$\begin{cases} [\varepsilon_{\Sigma'-3}, \varepsilon_{\Sigma'-2}] = [\ell_{\mathcal{D}-3}, \ell_{\mathcal{D}-2}] \\ \text{or} \\ [\varepsilon_{\Sigma'-3}, \varepsilon_{\Sigma'-2}] = [\ell_{\mathcal{D}-1} - \ell_2, \ell_{\mathcal{D}-1} - \ell_1]. \end{cases} \tag{42}$$

From (42) and the assumption $\ell_d \neq \ell_{d'}$ for $d \neq d'$, we will have $\varepsilon_{\Sigma'-2} \neq \varepsilon_{\Sigma'-3}$, and

$$\begin{cases} [\varepsilon_{\Sigma'-3}, \varepsilon_{\Sigma'-2}] = [\ell_{\mathcal{D}-1}, \ell_{\mathcal{D}-2}] \\ \text{or} \\ [\varepsilon_{\Sigma'-3}, \varepsilon_{\Sigma'-2}] = [\ell_{\mathcal{D}-2}, \ell_{\mathcal{D}-1} - \ell_1]. \end{cases} \tag{43}$$

We can see that the situation in (43) is contradictory to the assumption $\varepsilon_{\Sigma'-2} \neq \varepsilon_{\Sigma'-3}$, so Case 2, can be happen only for situation in (43). Now if $\varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-3}$ then $\ell_{\mathcal{D}-1} - \ell_1 = \ell_{\mathcal{D}-2}$. Define a new subset from (26) as

$$\bar{\Gamma}_{diff}^{lag \setminus \ell_{\mathcal{D}-2} - \ell_1} \triangleq \bar{\Gamma}_{diff}^{lag} \setminus \ell_{\mathcal{D}-2} - \ell_1. \tag{44}$$

From (28), and the fact that $\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-2)}) > \max(\bar{\Gamma}_{diff}^{laglag(\mathcal{D}-3)}) > \ldots > \max(\bar{\Gamma}_{diff}^{lag(2)})$, three possibilities will left for $\varepsilon_{\Sigma'-4}$ as

$$\begin{aligned} \varepsilon_{\Sigma'-4} \in \Big\{&\max(\bar{\Gamma}_{diff}^{tap \setminus \ell_{\mathcal{D}-2}}), \max(\bar{\Gamma}_{diff}^{lag \setminus \ell_{\mathcal{D}-2} - \ell_1}), \\ &\max(\bar{\Gamma}_{diff}^{lag(\mathcal{D}-2)})\Big\}, \end{aligned} \tag{45}$$

so we have $\varepsilon_{\Sigma'-4} \in \{\ell_{\mathcal{D}-3}, \ell_{\mathcal{D}-1} - \ell_2, \ell_{\mathcal{D}-2} - \ell_1\}$.

We note that $\gamma_{\Sigma-2} = \varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-3}$ and $\gamma_{\Sigma-3} = \varepsilon_{\Sigma'-4}$. Therefore, based on (45), one of the following possibilities will occur $\gamma_{\Sigma-3} = \ell_{\mathcal{D}-3}$, $\gamma_{\Sigma-1} - \gamma_{\Sigma-3} = \ell_2$, or $\gamma_{\Sigma-2} - \gamma_{\Sigma-3} = \ell_1$. So, if $\gamma_{\Sigma-2} = \ell_{\mathcal{D}-1} - \ell_1 = \ell_{\mathcal{D}-2}$, a member of the set $\mathcal{T}_2$ in (46), certainly is a vector of channel tap positions.

$$\begin{aligned} \bar{\mathcal{T}}_2 = \{&[\gamma_{\Sigma-3}, \gamma_{\Sigma-2}], [\gamma_{\Sigma-1} - \gamma_{\Sigma-3}, \gamma_{\Sigma-2}], \\ &[\gamma_{\Sigma-2} - \gamma_{\Sigma-3}, \gamma_{\Sigma-2}]\}. \end{aligned} \tag{46}$$

Case 3: Suppose that $\varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-3} = \varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-4}$. From (46), to have this assumption, three different situations will happen as

$$\begin{cases} \ell_{\mathcal{D}-3} = \ell_{\mathcal{D}-2} \\ \ell_{\mathcal{D}-1} - \ell_2 = \ell_{\mathcal{D}-1} - \ell_1 \\ \ell_{\mathcal{D}-2} - \ell_1 = \ell_{\mathcal{D}-1} - \ell_1 \end{cases} \Longrightarrow \begin{cases} \ell_{\mathcal{D}-3} = \ell_{\mathcal{D}-2} \\ \ell_1 = \ell_2 \\ \ell_{\mathcal{D}-1} = \ell_{\mathcal{D}-2}. \end{cases} \tag{47}$$

All statements in (47) are impossible because $\ell_d \neq \ell_{d'}$ for $d \neq d'$, therefore, Case 3 will be impossible to happen. According to the above discussion, the general case is obtained by considering two cases $\varepsilon_{\Sigma'-2} \neq \varepsilon_{\Sigma'-3}$ and

$\varepsilon_{\Sigma'-2} = \varepsilon_{\Sigma'-3} \neq \varepsilon_{\Sigma'-4}$. So, $\bar{\mathcal{T}}$ is obtained as $\bar{\mathcal{T}} = \bar{\mathcal{T}}_1 \cup \bar{\mathcal{T}}_2 = \{[\gamma_{\Sigma-3}, \gamma_{\Sigma-1} - \gamma_{\Sigma-2}], [\gamma_{\Sigma-1} - \gamma_{\Sigma-3}, \gamma_{\Sigma-1} - \gamma_{\Sigma-2}], [\gamma_{\Sigma-3}, \gamma_{\Sigma-2}], [\gamma_{\Sigma-2} - \gamma_{\Sigma-3}, \gamma_{\Sigma-2}]\}$, and Theorem 1, is proved.

REFERENCES

[1] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 121–135, Jan. 2019.

[2] D. Zhao and T. Han, "Low-complexity compressed sensing downlink channel estimation for multi-antenna terminals in FDD massive MIMO systems," *IEEE Access*, vol. 8, pp. 130183–130193, 2020.

[3] L. Lu, 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.

[4] W. Ji, F. Zhang, and L. Qiu, "Multipath extraction based UL/DL channel estimation for FDD massive MIMO-OFDM systems," *IEEE Access*, vol. 9, pp. 75349–75361, 2021.

[5] S. Silva, G. A. A. Baduge, M. Ardakani, and C. Tellambura, "Performance analysis of massive MIMO two-way relay networks with pilot contamination, imperfect CSI, and antenna correlation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4831–4842, Jun. 2018.

[6] P. Liang, J. Fan, W. Shen, Z. Qin, and G. Y. Li, "Deep learning and compressive sensing-based CSI feedback in FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9217–9222, Aug. 2020.

[7] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195–208, 2015.

[8] Y. Nan, L. Zhang, and X. Sun, "Weighted compressive sensing based uplink channel estimation for time division duplex massive multi-input multi-output systems," *IET Commun.*, vol. 11, no. 3, pp. 355–361, Feb. 2017.

[9] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: New design methods and analysis," *IEEE Access J.*, vol. 2, pp. 947–959, 2014.

[10] W. Lu, Y. Wang, X. Wen, X. Hua, S. Peng, and L. Zhong, "Compressive downlink channel estimation for FDD massive MIMO using weighted l_p minimization," *IEEE Access*, vol. 7, pp. 86964–86978, 2019.

[11] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, Apr. 2020.

[12] S. R. Kenarsari and M. F. Naeiny, "Mobility-aware reconstruction algorithm for correlated nonzero neighborhood structured downlink channel in massive MIMO," *IEEE Access*, vol. 6, pp. 17080–17092, 2018.

[13] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, May 2017.

[14] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[15] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 525–536, Jan. 2016.

[16] Y. Shi, M. Badi, D. Rajan, and J. Camp, "Channel reciprocity analysis and feedback mechanism design for mobile beamforming systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6029–6043, Jun. 2021.

[17] J. Tan and L. Dai, "Channel feedback in TDD massive MIMO systems with partial reciprocity," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 12960–12974, Dec. 2021.

[18] A. Khansefid and H. Minn, "Achievable downlink rates of MRC and ZF precoders in massive MIMO with uplink and downlink pilot contamination," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4849–4864, Dec. 2015.

[19] T. Cui, C. Tellambura, and Y. Wu, "Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2005, pp. 1980–1984.

[20] T. Cui and C. Tellambura, "Semiblind channel estimation and data detection for OFDM systems with optimal pilot design," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 1053–1062, May 2007.

[21] S. Ma, G. Wang, R. Fan, and C. Tellambura, "Blind channel estimation for ambient backscatter communication systems," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1296–1299, Jun. 2018.

[22] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[23] C. Hu, H. Wang, and R. Song, "Pilot decontamination in multi-cell massive MIMO systems via combining semi-blind channel estimation with pilot assignment," *IEEE Access*, vol. 8, pp. 152952–152962, 2020.

[24] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[25] G. Wang, Y. Ma, N. Yi, and R. Tafazolli, "Semi-blind MAP channel estimator for pilot decontamination," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–5.

[26] J. Mirzaei, R. S. Adve, and S. Shahbazpanahi, "Semi-blind time-domain channel estimation for frequency-selective multiuser massive MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1045–1058, Feb. 2019.

[27] C.-Y. Wu, W.-J. Huang, and W.-H. Chung, "Low-complexity semiblind channel estimation in massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6279–6290, Sep. 2017.

[28] W. Xu, W. Xiang, Y. Jia, Y. Li, and Y. Yang, "Downlink performance of massive-MIMO systems using EVD-based channel estimation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3045–3058, Apr. 2017.

[29] F. Wan, W. P. Zhu, and M. N. S. Swamy, "Semi-blind most significant tap detection for sparse channel estimation of OFDM systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 3, pp. 703–713, Mar. 2010.

[30] F. Wan, W.-P. Zhu, and M. Swamy, "Semiblind sparse channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2569–2582, Jul. 2011.

[31] P. De, "Semi-blind sparse channel estimation using reduced rank filtering," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1418–1431, Mar. 2018.

[32] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, May 2014.

[33] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2nd Quart., 2022.

[34] J. An, Q. Wu, and C. Yuen, "Scalable channel estimation and reflection optimization for reconfigurable intelligent surface-enhanced OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 796–800, Apr. 2022.

[35] 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. 2015.

[36] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[37] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[38] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[39] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[40] X. Rao and V. K. N. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

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

[42] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[43] L.-H. Chang and J.-Y. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5702–5715, Sep. 2014.
[44] Å. Björck, *Numerical Methods in Matrix Computations*, vol. 59. Cham, Switzerland: Springer, 2015.

MOHAMMAD REZA MEHRABANI received the B.Sc. and M.Sc. degrees from the Iran University of Science & Technology, Tehran, Iran, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His research interests include channel estimation and compressive sensing in wireless communications.

BAHMAN ABOLHASSANI was born in Tehran, Iran. He received the B.Sc. degree from the Iran University of Science & Technology (IUST), Tehran, and the M.Sc. and Ph.D. degrees from the University of Saskatchewan, Saskatoon, SK, Canada, all in electrical engineering. He was an Instrumentation Engineer with the College of Water and Power Technology, Iranian Ministry of Energy, for three years. Then, he worked as a communication system engineer in a number of private and government companies. He joined the School of Electrical Engineering, IUST, where he is currently an Associate Professor. He worked as the Dean of the School of Electrical Engineering and the Associate Dean of Research. He also worked as a Sessional Lecturer with the University of Saskatchewan. His research interests include the fields of wireless communication systems, network planning, spread spectrum, cognitive radio networks, resource allocation, NOMA, VANETs, and optimization of large systems.

FARZAN HADDADI was born in 1979. He received the B.Sc., M.Sc., and Ph.D. degrees in communication systems from the Sharif University of Technology, Tehran, Iran, in 2001, 2003, and 2010, respectively. He joined as a Faculty Member of the Iran University of Science & Technology, in 2011. His research interests include array signal processing, statistical signal processing, subspace tracking, and compressed sensing.

CHINTHA TELLAMBURA (Fellow, IEEE) received the B.Sc. degree (Hons.) from the University of Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the King's College London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Canada, in 1993. He was with Monash University, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. He has authored or coauthored over 500 journals and conference papers with total citations over 14 000 and a H-index of 62 on Google Scholar. His current research interests include the design, modeling, analysis of cognitive radio, heterogeneous cellular networks, and 5G wireless networks. In 2011, he was elected as a fellow of IEEE for his contributions to physical layer wireless communication theory, and in 2017, elected as a fellow of the Canadian Academy of Engineering. He received the best paper awards from the Communication Theory Symposium, in 2012, the IEEE International Conference on Communications (ICC) in Canada, and the 2017 ICC in France. He was a recipient of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (1999–2011). From 2001 to 2007, he was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and continued as an Area Editor for *Wireless Communications Systems and Theory*, from 2007 to 2012.