Received April 19, 2019, accepted May 14, 2019, date of publication June 5, 2019, date of current version June 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920861

FDD Cooperative Channel Estimation and Feedback for 3D Massive MIMO System

AHMED NASSER [1,3], MAHA ELSABROUTY [1], AND OSAMU MUTA [2]

[1]Department of Electronics and Communication, Egypt-Japan University of Science and Technology, Alexandria 21934, Egypt
[2]Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University, Fukuoka 819-0395, Japan
[3]Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan

Corresponding author: Ahmed Nasser (ahmed.nasser@ejust.edu.eg)

This work was financially supported and was done as part of the research project "Super-HETs, Empowering 5G HetNets for better Performance" funded by the National Telecommunication Regulation Authority (NTRA Egypt). Also, This project was supported in part by the Center for Japan-Egypt Cooperation in Science and Technology, JSPS KAKENHI, under Grant JP17K06427, and in part by the Egyptian Ministry of Higher Education (MOHE).

ABSTRACT In this paper, a downlink cooperative channel estimation scheme is proposed for the three-dimensional massive multiple inputs multiple outputs (3D-mMIMO) system operating in the frequency division duplexing (FDD) mode. In the proposed cooperative scheme, users have to cooperate with each other via device-to-device (D2D) communication protocol to jointly exploit the sparsity structure property of the channel. Motivated by the sparsity property of the mMIMO channel in the angle-time domain, a parametric feedback scheme is proposed, where the feedback overhead is decreased by sending a limited version of the estimated coefficients rather than all the coefficients back to the BS. Then, a compressive sensing (CS) algorithm is proposed, which we named weighted fast iterative shrinkage thresholding (WFISTA). In the WFISTA, we first introduce new weights and threshold function to the original FISTA to enhance its sparsity-undersampling trade-off in the single measurement vector (SMV) case, then we extend the proposed WFISTA to the case of multiple measurement vector (MMV) problem by adopting ReMBo (reduce MMV and boost) strategy. The proposed WFISTA has the ability to estimate the mMIMO system's channel coefficients by exploiting the joint sparsity structure through cooperation among users' equipment (UEs). Complexity analysis and the probability of decreasing the feedback overhead are provided for the proposed cooperative estimation scheme. The simulation results verify the efficiency of the proposed cooperative algorithm scheme compared to several joint channel estimations.

INDEX TERMS Channel estimation, device to device (D2D) communication, compressive sensing, fast iterative shrinkage thresholding algorithm (FISTA), massive multiple input multiple output (mMIMO), multiple measurement vectors (MMV), frequency division duplexing (FDD).

I. INTRODUCTION

Future wireless communication network must overcome the challenges of current cellular networks such as supporting very high data rates with low latency and improving the energy efficiency. To meet these towering requirements, employing a huge number of antennas at the base station (BS) side that serves a large number of multiple antenna users contributing to massive multiple inputs multiple outputs (mMIMO) system is considered as a decisive solution [1], [2]. However, the required minimum spacing between antennas and the confined space at the BS are considered as a stumbling block towards increasing the number of antennas at the BS. For the BS, to support these huge number of antennas, various antenna arrays structures were proposed in the literature [1]. Among these different structures, a uniform planar array (UPA) is considered to be convenient configuration to accommodate a large number of antennas in a small bounded area. Massive MIMO (mMIMO) systems with UPAs have the attribute of controlling the transmitted beam in both vertical and horizontal directions, so it is known as 3D-mMIMO [3].

On the other hand, realizing the gains of the mMIMO is restricted by the availability of the channel state information (CSI) at both sides [2]. However, over the available frequency or time resources, acquiring precise CSI with a plausible number of pilot symbols and feedback overhead

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

is turned into a challenging problem with the deployment of mMIMO [1]. Although the channel reciprocity property of the time division duplex (TDD) mode is considered as a suitable solution for the CSI feedback overhead, challenges such as pilot contamination affect the practical application of TDD [4]. In contrast, symmetric traffic with low-latency communication can be achieved by deploying the frequency division duplexing (FDD) mode. However, the tremendous CSI feedback overhead must be carefully considered.

A. EXISTING RESEARCH WORKS

Benefiting from the sparsity property of the mMIMO channels in some spatial domains, compressive sensing (CS) [5] emerges as a valuable tool that can provide high channel estimation performance with a reasonable pilot and CSI feedback overhead in FDD mode. Based on this idea, two main schemes are proposed in the literature to reduce the feedback overhead in the FDD transmission mode. In the first scheme, each user's equipment (UE) receives the downlink pilot and estimates its channel separately. Then, they feed a compressed version from the estimated channel back to the BS. This scheme is studied in different literature works, e.g. [6], [7], where different CS recovery algorithms are proposed as structured subspace pursuit in [6], and split Bregman in [7]. These CS algorithms are proposed to recover single measurement vector (SMV) channels, so they are called SMV CS algorithms. On the other hand, the second scheme is called distributed CS estimation [8]–[10]. The distributed scheme proposes to estimate the channel at BS instead of estimating it individually at each UE. In this case, each UE sends its received pilots back to the BS, which tackles the role of recovering the channels. This scheme can benefit from the inter-channel correlation between the adjacent UEs to efficiently estimate the joint coefficients between them and to decrease the feedback. In that case, The SMV CS algorithms must be adopted to recover multiple measurement vectors (MMV) channels, where multiple vectors are jointly recovered taking into account the existing of joint coefficient due to inter-channel correlation. Different joint MMV recovery algorithms have been proposed for the distributed schemes as alternative direction of multiplier (ADM) [8], joint orthogonal matching pursuit (J-OMP) [9], and sparsity adaptive matching pursuit (SAMP) [10]. However, the distributed scheme is not suitable for the applications that need CSI at the UE side such as interference management techniques, also the feedback overhead is still quite large. In the last few years, device to device communication (D2D) technology became a reality [11]. Now, billions of devices are connected with different protocols, and this trend will sustain its exponential increase. At the same time, UEs are now embedded with powerful processors and huge storage capability. As a result, the decentralization is considered as an appropriate way to deal with such tremendous communication traffic. In line with this trend, we incorporate D2D communication as an enabling tool for exchanging the CSI information.

B. MOTIVATION AND CONTRIBUTION

Motivated by the critical importance of channel estimation and feedback challenges, and since the decrease in the CSI feedback overhead proposed in the previous literature is still not enough, in addition to the unavailability of CSI at the UE sides in their proposal, we present the work in this paper aiming at improving the solution for these challenges.

Particularly, and in contrast to the existing FDD estimation schemes, this paper proposes[1] a cooperative channel estimation and feedback scheme for the 3D-OFDM mMIMO operating in FDD mode based on D2D communication protocols. The main contribution of this paper can be summarized as follows:

- We propose a cooperative channel estimation scheme to utilize the sparsity property of the mMIMO channel. In the proposed cooperative scheme, UEs within the cell are clustered according to their relative distance and a cluster head (CH) is chosen for each cluster. Then, the channel estimation is done cooperatively among UEs in the same cluster using D2D protocol before feeding the estimated channels' coefficients back to the BS.
- We propose a parametric feedback scheme that benefits from the joint sparsity property of the proposed cooperative channel estimation. In this scheme, the feedback overhead is decreased by sending a limited version of the estimated coefficients rather than all the coefficients back to the BS.
- To exploit the sparsity property of the estimated mMIMO channel, we propose a weighted fast iterative shrinkage-thresholding algorithm (WFISTA) for the cooperative channel estimation strategy, where weighted terms are added to a recursive equation of the estimated vector. Thus, unlike the conventional FISTA, the WFISTA is able to enhance the sparsity-undersampling trade-off by expanding the search space based on mutation strategy of heuristic algorithms.
- We validate the channel estimation capability of the proposed cooperative scheme in utilizing the sparsity property of the mMIMO while reducing the required complexity compared with the conventional schemes. In addition, the complexity analysis and the probability of decreasing the feedback overhead are provided for the proposed cooperative estimation scheme.

C. PAPER ORGANIZATION AND NOTATION

The rest of the paper is organized as follows: Section II introduces the system transmission model, assumptions, and the channel sparsity representation of our proposed schemes. Section III discusses the proposed cooperative channel estimation and parametric feedback schemes. Section IV presents the problem formulation of the CS model, the proposed WFISTA algorithm for joint channel estimation, and the complexity analysis of the proposed algorithm. The prob-

[1]The material in this paper was presented in part at the 2017 Personal, Indoor, and Mobile Radio Communications (PIMRC) [12].

ability that the proposed cooperative scheme can decrease the feedback overhead is discussed in Section V, while Section VI simulates the performance of the proposed technique. Section VII concludes the paper.

Notation: Variable, i.e. matrices and vectors are, respectively, indicated with upper case and lower case bold letters. The normal math font '$\boldsymbol{X}, \mathbf{x}$' expresses matrices and vectors in the frequency domain, respectively, while the calligraphic font '$\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{x}}$' indicates the time domain matrices and vectors. The Frobenius norm is represented by $\| . \|_f$, while the notation $\|\boldsymbol{X}\|_{p,q} = (\sum_{j=1}^{N} \|\mathbf{x}_j\|_p^q)^{1/q}$ refers to $l_{p,q}$ mixed norm. Transpose, conjugate transpose, cardinality, and pseudo-inverse operators are represented, respectively, by the notations $(\,.\,)^T$, $(\,.\,)^H$, $|\;.\;|$, and $(\,.\,)^\dagger$. The operator $supp(\mathbf{x})$ estimates the non-zero indices of vector $\mathbf{x}$. The notation $Blkdiag[\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}]$ refers to a matrix assembled using the matrices $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ stacked along its diagonal, while the transpose of multiple matrices $[\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}]^T = [\boldsymbol{X}; \boldsymbol{Y}; \boldsymbol{Z}]$ convert the matrices, $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, from the concatenation over a row vector to concatenation over a column vector. The sub-matrices $\boldsymbol{X}\mid_\Omega$, and $\boldsymbol{X}\mid^\Omega$ are assembled, respectively, from the Ω indices set of rows and columns. The notation $\boldsymbol{A}(j, :)$ denotes the j^{th} row of the matrix $\boldsymbol{A}$. The Kronecker product, real number field, and the complex number field are denoted by $\otimes$, $\mathbb{R}$, and $\mathbb{C}$ respectively. $\boldsymbol{I}_x$ denotes the x $\times$ x identity matrix.

II. SYSTEM MODEL AND CHANNEL SPARSITY REPRESENTATION

As shown in Fig.1, the proposed system model considers a single cell of a 3D multi-user mMIMO system. The BS is equipped with a UPA of N_{BS} antennas distributed across N_{ro} rows and N_{co} columns, where $N_{BS} = N_{co} \times N_{ro}$. According to the proposed cooperative scheme, the UEs within the cell are clustered into N_c cluster depending on their relative distance; each cluster has M UEs of K antennas distributed

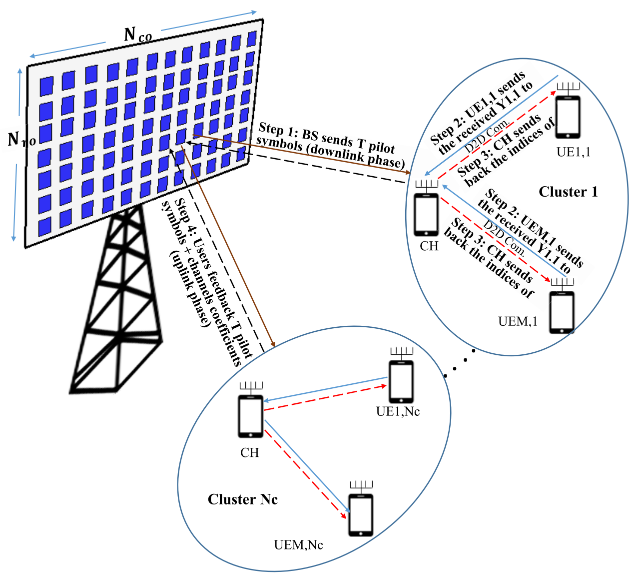

FIGURE 1. The system transmission model for the proposed cooperative channel estimation scheme.

in a ULA configuration, where $N_c \times M \times K = N_u < N_{BS}$. This paper proposes a cooperative channel estimation scheme for the systems operate in FDD mode, where the BS sends pilot symbols to UEs in the downlink phase, then the UEs within the cluster cooperate to jointly estimate the channel and feedback the estimated coefficients to the BS.

A. 3D MASSIVE MIMO SYSTEM TRANSMISSION AND CHANNEL MODEL

An OFDM scheme with N_s orthogonal subcarriers is considered as a signaling mechanism for the frequency-selective multi-user mMIMO system. Each orthogonal subcarrier is treated as a flat block fading channel. Within the coherence time, the received N_p sequence of pilot symbols transmitted from the BS over the j^{th} subcarriers to the i^{th} UE in the n^{th} cluster can be expressed as:

$$\boldsymbol{Y}_{i,n}(j) = \boldsymbol{H}_{i,n}(j)\boldsymbol{X}(j) + \boldsymbol{W}_{i,n}(j), \tag{1}$$

where $\boldsymbol{X}(j) \in \mathbb{C}^{N_{BS} \times N_p}$ is the N_p transmitted pilot symbols at j^{th} sub-carrier. $\boldsymbol{H}_{i,n}(j) \in \mathbb{C}^{K \times N_{BS}}$ is the channel matrix at the j^{th} subcarrier from the BS to the i^{th} UE in the n^{th} cluster, while $\boldsymbol{w}_{i,n}(j) \in \mathbb{C}^{K \times N_p}$ is an additive white Gaussian noise (AWGN) with zero mean and variance σ_n^2.

The $K \times N_{BS}$ channel frequency response of i^{th} UE in the n^{th} cluster, $\boldsymbol{H}_{i,n}(j)$, can be modeled as in [13]:

$$\boldsymbol{H}_{i,n}(j) = \sum_{m=1}^{N_h} \beta_{i,n}^m \boldsymbol{\mathcal{V}}_r(\psi_{r_{i,n}}^m) \boldsymbol{\mathcal{V}}_t^T(\psi_{t_{i,n}}^m) e^{-j2\pi \tau_{i,n}^m f_j}, \tag{2}$$

where $\beta_{i,n}^m$ and $\tau_{i,n}^m$ are the m^{th} path loss and the m^{th} path delay within the delay spread for the i^{th} UE in the n^{th} cluster, respectively, while f_j is the frequency of the j^{th} subcarrier. The $\boldsymbol{\mathcal{V}}_t$ and $\boldsymbol{\mathcal{V}}_r$ are, consecutively, the receive and the transmit steering vectors. The receive steering vector $\boldsymbol{\mathcal{V}}_r$ can be written as in [13], [14]:

$$\boldsymbol{\mathcal{V}}_r(\psi_r) = \frac{1}{\sqrt{K}}[1, e^{-j2\pi \psi_r}, \ldots, e^{-j2\pi(K-1)\psi_r}]^T, \tag{3}$$

where $\psi_r = \frac{d_r}{\lambda} sin(\alpha)$, while α is the angle of arrival (AoA) at UEs within the angle spread $[-\pi/2, \pi/2]$, the symbol λ is the wavelength of the carrier, and d_r is the relative displacement between any two neighboring antennas at UE. As in [15], the transmit steering vector $\boldsymbol{\mathcal{V}}_t$ is related to the UPA configuration and can explicit as:

$$\boldsymbol{\mathcal{V}}_t(\psi_t) = \boldsymbol{\mathcal{V}}_h(\psi_h) \otimes \boldsymbol{\mathcal{V}}_v(\psi_v) \tag{4}$$

where $\boldsymbol{\mathcal{V}}_h$ and $\boldsymbol{\mathcal{V}}_v$ are the horizontal and the vertical steering vectors, respectively. The vectors $\boldsymbol{\mathcal{V}}_h$ and $\boldsymbol{\mathcal{V}}_v$ can be expressed from [3], [16] as:

$$\boldsymbol{\mathcal{V}}_h(\psi_h) = \frac{1}{\sqrt{N_{ro}}}[1, e^{-j2\pi \psi_h}, \ldots, e^{-j2\pi(N_{ro}-1)\psi_h}]^T \tag{5}$$

$$\boldsymbol{\mathcal{V}}_v(\psi_v) = \frac{1}{\sqrt{N_{co}}}[1, e^{-j2\pi \psi_v}, \ldots, e^{-j2\pi(N_{co}-1)\psi_v}]^T \tag{6}$$

where $\psi_v = \frac{d_{co}}{\lambda} cos(\theta)$, $\psi_h = \frac{d_{ro}}{\lambda} sin(\theta)cos(\phi)$, while d_{co} and d_{ro} are the relative displacement between two neighboring antennas in a column and a row at the BS side, respectively. The angles $\phi \in [0, \pi]$ and $\theta \in [0, \pi/2]$ are, consecutively, the azimuth and the elevation angles of departure (AoD) from the BS to UEs. By stacking all subcarriers, the concatenated 3D-mMIMO channel for the i^{th} UE in the n^{th} cluster can be rewritten as:

$$\boldsymbol{H}_{i,n} = [\boldsymbol{H}_{i,n}(1), \boldsymbol{H}_{i,n}(2), \ldots, \boldsymbol{H}_{i,n}(N_s)]^T \in \mathbb{C}^{KN_s \times N_{BS}}. \quad (7)$$

B. SPARSITY REPRESENTATION OF THE CHANNEL

In the environment of poor scatterers, the angular domain is considered a proper sparse representation for mMIMO channels [14], [17]. The angular domain decomposes the transmitted signals into different beams across dedicated directions, while the remaining directions turn into sparse. By deploying the angular domain, the propagation environment is defined with N_{BS} transmit lobes and K receive lobes at each UE constituting a beamforming pattern [14], [18]. The system model expressed by equation (1) is represented in the angle-frequency domain as:

$$\boldsymbol{Y}^a_{i,n}(j) = \boldsymbol{H}^a_{i,n}(j)\boldsymbol{X}^a(j) + \boldsymbol{W}^a_{i,n}(j), \quad (8)$$

where $\boldsymbol{H}^a_{i,n}(j) = \boldsymbol{U}_r^H \boldsymbol{H}_{i,n}(j)\boldsymbol{U}_t \in \mathbb{C}^{K \times N_{BS}}$ is the angle-frequency domain channel that between the BS and the i^{th} UE at the n^{th} cluster over the j^{th} subcarrier, while the transmitted and received angular domain pilot symbols over the j^{th} subcarrier, $\boldsymbol{X}^a(j)$ and $\boldsymbol{Y}^a_{i,n}(j)$, can be expressed as $\boldsymbol{U}_t^H \boldsymbol{X}(j) \in \mathbb{C}^{N_{BS} \times N_p}$, and $\boldsymbol{U}_r^H \boldsymbol{Y}_{i,n}(j) \in \mathbb{C}^{K \times N_p}$, respectively. $\boldsymbol{W}^a_{i,n}(j) = \boldsymbol{U}_r^H \boldsymbol{W}_{i,n}(j) \in \mathbb{C}^{K \times N_p}$ is the received AWGN. The unitary matrices $\boldsymbol{U}_r \in \mathbb{C}^{K \times K}$, and $\boldsymbol{U}_t \in \mathbb{C}^{N_{BS} \times N_{BS}}$ are the angular domain transformation basis at UEs and BS respectively.

The layout of the UPA equipped at the BS estimates the entire values $\boldsymbol{U}_t$. From [3], $\boldsymbol{U}_t$ can be formulated as:

$$\boldsymbol{U}_t^T = \boldsymbol{\mathcal{R}}_h \otimes \boldsymbol{\mathcal{R}}_v, \quad (9)$$

where the vertical and the horizontal transformation matrices $\boldsymbol{\mathcal{R}}_v \in \mathbb{C}^{N_{co} \times N_{co}}$ and $\boldsymbol{\mathcal{R}}_h \in \mathbb{C}^{N_{ro} \times N_{ro}}$, respectively, can be expressed from [13], [14] as follows:

$$\boldsymbol{\mathcal{R}}_v = [\boldsymbol{\mathcal{V}}_v(0), \boldsymbol{\mathcal{V}}_v(\frac{1}{N_{co}}), \ldots, \boldsymbol{\mathcal{V}}_v(\frac{N_{co}-1}{N_{co}})] \quad (10)$$

$$\boldsymbol{\mathcal{R}}_h = [\boldsymbol{\mathcal{V}}_h(0), \boldsymbol{\mathcal{V}}_h(\frac{1}{N_{ro}}), \ldots, \boldsymbol{\mathcal{V}}_h(\frac{N_{ro}-1}{N_{ro}})] \quad (11)$$

For the ULA antennas at UEs' side, the received basis $\boldsymbol{U}_r$ can be an inverse discrete Fourier transform (IDFT) matrix [14]. By accumulating all subcarriers, the concatenated angular domain channel for the 3D mMIMO can be rewritten as:

$$\boldsymbol{H}^a_{i,n} = [\boldsymbol{H}^a_{i,n}(1), \boldsymbol{H}^a_{i,n}(2), \ldots, \boldsymbol{H}^a_{i,n}(N_s)]^T \in \mathbb{C}^{KN_s \times N_{BS}}. \quad (12)$$

Although the angular domain is a proper sparse domain for the mMIMO channel, a deep sparse representation for the frequency-selective channels can be obtained by further describing the angular domain channels in the time domain which constitute the so-called angle-time domain [6]. In the case of limited influential scatterers, the channel delay spread is usually larger than the sampling interval that has a bigger influence on most of the channel coefficients driving them to be almost zero. The channel in the time domain can thus be turned into a sparse tapped delay line (TDL) [19]. From [13], [20], the stacked angle-time domain channel over N_h taps, $\boldsymbol{\mathcal{H}}^a_{i,n}$, can expressed from the concatenated angle-frequency domain, $\boldsymbol{H}^a_{i,n}$, as:

$$\boldsymbol{H}^a_{i,n} = \boldsymbol{\mathcal{F}}_{N_u} \boldsymbol{\mathcal{H}}^a_{i,n}, \quad (13)$$

where $\boldsymbol{\mathcal{F}}_{N_u} = \boldsymbol{F}_{N_s \times N_h} \otimes \boldsymbol{I}_{N_u}$ is the concatenated matrix of the truncated Fourier matrix $F_{N_s \times N_h}$ as in [18], and $\boldsymbol{\mathcal{H}}^a_{i,n} = [\boldsymbol{\mathcal{H}}^a_{i,n}(1), \ldots, \boldsymbol{\mathcal{H}}^a_{i,n}(N_h)]^T \in \mathbb{C}^{KN_h \times N_{BS}}$, while $\boldsymbol{\mathcal{H}}^a_{i,n}(m) \in \mathbb{C}^{K \times N_{BS}}$ is the angle-time domain channel at the m^{th} tap for the i^{th} UE in the n^{th} cluster.

The number of channel coefficients that need to be estimated is decreased from $N_s N_{BS} N_u$ to $N_s N_{sp}$ by describing the channel in the angle domain only, where $N_{sp} << N_{BS} N_u$ is the number of angle domain's non-zero bins. However, the number of unknowns N_s can be decreased to N_q, where $N_q << N_s$ is the number of non-zero taps in the time domain. Thus, exploiting the angle-time as a sparsifying domain reduces the number of the channel coefficients that need to be estimated to $N_q N_{sp}$.

In this work, we propose that the relative distance between any two adjacent antennas at UE's, d_r, is small such that the distance between the first and last antennas at UE, $d_{max} = K \times d_r$, is less than $c/(10BW)$. Thus, the CIR taps are non-resolvable since the paths' times of arrivals at the antennas of each UE are quite close which constituting the so-called space invariant antenna array (SIA) structure. For the SIA structure of i^{th} UE, the dominant taps of the CIR for all K antennas have the same indices set Ω_{t_i}: $0 < \Omega_{t_i} << N_h$. Consequently, the angle-time domain channel has a sparsity structure of the following two properties:

- Property 1 (individual local supports): As the K antenna elements of each UE are adjacent to each other, they practically experience the same scatterers. Therefore, for all the K antennas of a certain UE, their angular domain channels share the same location for the angular domain bins and have equivalent sparsity pattern of S non-zero coefficients with distinct values; this leads to what is called the multiple measurement vectors (MMV) structure. For the i^{th} UE in the n^{th} cluster, there is a local supports set $\Omega_{A_{i,n}}$: $0 < |\Omega_{A_{i,n}}| = S << N_{BS} N_q$ and such that:

$$supp(\boldsymbol{\hbar}^a_{1,i,n} |_{\Omega_{t_i}}) = \cdots = supp(\boldsymbol{\hbar}^a_{K,i,n} |_{\Omega_{t_i}}) = \Omega_{A_{i,n}}, \quad (14)$$

where $\boldsymbol{\hbar}^a_{k,i,n} \in \mathbb{C}^{1 \times N_{BS} N_h}$ is the angle-time domain channel vector between the BS and the k^{th} antenna element of the i^{th} UE in the n^{th} cluster, and $\Omega_{A_{i,n}} = \{\Omega_{A_{i,n}}(1), \Omega_{A_{i,n}}(2), \ldots, \Omega_{A_{i,n}}(N_h)\}$, where $\Omega_{A_{i,n}}(m)$ is the individual local support for the i^{th} UE at the m^{th} tap.

- Property 2 (shared common supports): Although the UEs inside the cluster experience various sparsity patterns, there is a cross-channel correlation among UEs in the same cluster, especially when they are physically near to each other [9]. And thus, there are S_c joint indices that are commonly shared among all UEs within the cluster, that are termed as shared common supports. Within the n^{th} cluster, there exists a common indices set $\Omega_{A_{C,n}} : \Omega_{A_{C,n}} \subset \Omega_{A_{i,n}}$ and $|\Omega_{A_{C,n}}| = S_c$ such that:

$$\bigcap_{i=1}^{M} \Omega_{A_{i,n}} = \Omega_{A_{C,n}} \tag{15}$$

III. PROPOSED COOPERATIVE CHANNEL ESTIMATION SCHEME AND PARAMETRIC FEEDBACK

In this section, we propose a cooperative channel estimation scheme to improve the channel estimation performance and decrease the feedback of mMIMO system by exploiting the two property of angle-time domain channels. In the proposed cooperative channel estimation scheme, the UEs within the cell are clustered, and a cluster head (CH) is chosen as shown in Fig.1. To ensure the existence of the joint sparsity structure, the clusters are formed according to the relative distance between UEs. Also we assume that the CH is predetermined. The transmitted pilot signals $X(j)$ are observed distributively at each UE and sent via D2D communication to the CH (red-dashed lines) who initiates the estimation process.[2] Using the WFISTA algorithm proposed in the following section, the CH can exploit the common indices $\Omega_{A_{C,n}}$ and feed them back to the UEs of the corresponding cluster using D2D signaling (blue-solid lines). Using these estimated common indices, each UE can easily estimate the rest of its individual channel coefficients by applying a second step of WFISTA algorithm. One advantage of the proposing cooperative estimation is its ability to decrease the feedback overhead by feeding back a limited version of the estimated coefficients 'parametric feedback' rather than sending all the coefficients back as conventional FDD mode. In the proposed parametric feedback scheme, we can feed only the individual coefficients of the sparse angle-time channel matrix, $\mathcal{H}^{a}_{i,n}$, and their corresponding indices back to the BS. The proposed cooperative estimation and parametric feedback scheme are summarized in Table 1.

Using the proposed parametric feedback scheme, the feedback overhead per UE can be reduced from $\mu KSLog(N_{BS})$, in the distributed CS feedback schemes, to $K(S + (S - S_c)/K)$ in the proposed cooperative scheme, which gives us the opportunity to install more antennas at the BS side as shown in Fig.2(a). Besides, it is also worth noting that the decreasing percentage in the feedback overhead becomes more noticeable as the number of the common supports S_c or the number of antennas per UE K increase as shown in Fig.2(b). In particular, for the case of $N_u = 20$ and $K = 1$, i.e. number of UEs is equal to 20, we can install $N_{BS} = 200$ antennas at the BS. However, for the case of $N_u = 20$ and $K = 2$, i.e., number of UEs is equal to 100, we can install $N_{BS} = 300$ antennas at the BS.

Our proposed cooperative channel estimation and feedback schemes are applicable to any channel model that has a sparse representation that contains jointly non-zero indices between UEs' channels like the 3D-MIMO channel in the angle-time domain as explained in Sect. II. However, for the proposed schemes, to be deployed in the mMIMO cellular system, a MMV-CS algorithm is needed to exploit the common and individual indices. In the following section, we will explain the CS model formulation for the estimation at the CH and the other UEs in the cluster and how we can solve the formulated model with the proposed WFISTA algorithm.

TABLE 1. Proposed cooperative estimation and parametric feedback scheme.

Step 1: The mMIMO BS sends the compressed pilot signal $\boldsymbol{X}(j)$ of N_p symbols, $N_p << MK$, to the UEs.
Step 2: Each UE within the cluster observes its compressed transmitted pilots and sends its received symbols, $\boldsymbol{Y}^{a}_{i,n}(j)$, to the CH via D2D communication.
Step 3: Using the proposed WFISTA algorithm, the CH will exploit the indices of the common supports $\Omega_{A_{C,n}}$ and resend it back to the other UEs of its cluster.
Step 4: Using the indices $\Omega_{A_{C,n}}$ and by applying a second step of WFISTA at UE sides, each UE will estimate its common and individual coefficients and feed these coefficients and their corresponding indices back to the BS.

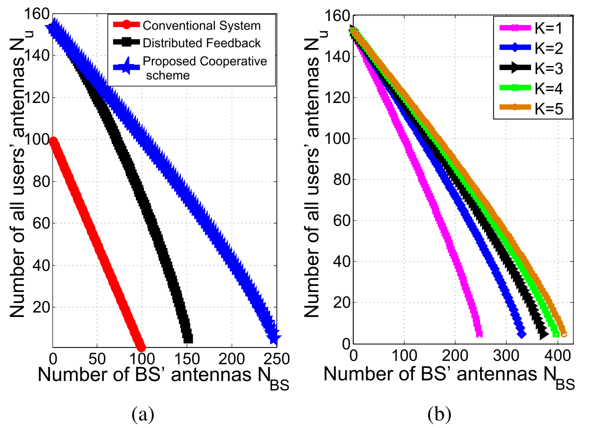

FIGURE 2. Illustration of the typical maximum number of UEs that can be served at different number of BS's antennas N_{BS} (a) for the proposed cooperative scheme versus distributed and conventional schemes, and (b) for the proposed cooperative scheme at different number of antennas per UEs K.

[2]In this work, we assume that the D2D channel links between the UEs in the same clusters and the CH are known, and the received pilots of all UEs in the cluster are perfectly detected at the CH.

IV. PROBLEM FORMULATION AND PROPOSED WFISTA ALGORITHM

A. PROBLEM FORMULATION

According to the proposed cooperative channel estimation scheme, the CH will exploit the common indices, while the other UEs within the cluster, will exploit the rest of individual indices and the channel coefficients. Thus, we have two different CS model, i.e., at the CH and the other UEs in the cluster.

1) CS MODEL FORMULATION FOR CHANNEL ESTIMATION AT THE CLUSTER HEAD (CH)

By assuming that the signals from the other UEs in the cluster are perfectly received at the CH, the received signal over the j^{th} subcarrier at the CH from $(M-1)$ UEs in the cluster via D2D communication in addition to the CH received signal can be expressed in the standard form as:

$$\boldsymbol{Y}_n(j) = \boldsymbol{H}_n(j)\boldsymbol{X}(j) + \boldsymbol{W}_n(j), \tag{16}$$

where $\boldsymbol{Y}_n(j) = [\boldsymbol{Y}_{1,n}(j), \boldsymbol{Y}_{2,n}(j), \ldots, \boldsymbol{Y}_{M,n}(j)]^T \in \mathbb{C}^{MK\times N_p}$ is the received N_p pilot symbols for the combined M UEs over the j^{th} subcarrier at CH, and $\boldsymbol{H}_n(j) = [\boldsymbol{H}_{1,n}(j), \boldsymbol{H}_{2,n}(j), \ldots, \boldsymbol{H}_{M,n}(j)]^T \in \mathbb{C}^{MK\times N_{BS}}$ is the M concatenated channel matrix, while $\boldsymbol{W}_n(j) \in \mathbb{C}^{MK\times N_p}$ is AWGN. By considering the angle-time representation proposed in sec (II-B) as the sparse transformation domain, the CS form for the system transmission model in (16) can be expressed as:

$$\bar{\boldsymbol{Y}}_n^a = \boldsymbol{A}\bar{\boldsymbol{\mathcal{H}}}_n^a + \bar{\boldsymbol{W}}_n^a, \tag{17}$$

where $\bar{\boldsymbol{Y}_n}^a = [\bar{\boldsymbol{Y}}_{1,n}{}^a, \ldots, \bar{\boldsymbol{Y}}_{M,n}{}^a] \in \mathbb{C}^{N_sN_p\times KM}$, is the stacked N_s transposed received signal concatenated over M UEs at the CH, whereas $\bar{\boldsymbol{Y}}_{i,n}^a = [\boldsymbol{Y}_{i,n}^{a^T}(1), \ldots, \boldsymbol{Y}_{i,n}^{a^T}(N_s)]^T \in \mathbb{C}^{N_sN_p\times K}$. The channel matrix $\bar{\boldsymbol{\mathcal{H}}}_n^a$ is M stacked of the transposed angle-time channel, $\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a$, as $\bar{\boldsymbol{\mathcal{H}}}_n^a = [\bar{\boldsymbol{\mathcal{H}}}_{1,n}^a, \bar{\boldsymbol{\mathcal{H}}}_{2,n}^a, \ldots, \bar{\boldsymbol{\mathcal{H}}}_{M,n}^a]\mathbb{C}^{N_{BS}N_h\times MK}$, and $\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a = [\boldsymbol{\mathcal{H}}_{i,n}^{a^T}(1), \ldots, \boldsymbol{\mathcal{H}}_{i,n}^{a^T}(N_h)]^T \in \mathbb{C}^{N_{BS}N_h\times K}$. The multiplication of the measurement pilot matrix and the Fourier transformation basis can form the sensing matrix $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \bar{\boldsymbol{X}}^a\boldsymbol{\mathcal{F}}_{N_{BS}}, \tag{18}$$

where $\bar{\boldsymbol{X}}^a = Blkdiag[\boldsymbol{X}^{a^T}(1), \boldsymbol{X}^{a^T}(2), \ldots, \boldsymbol{X}^{a^T}(N_s)] \in \mathbb{C}^{N_sN_p\times N_{BS}}$ is the concatenated transmitted signal in the angle-frequency domain over N_s subcarrier, while $\boldsymbol{\mathcal{F}}_{N_{BS}} = \boldsymbol{F}_{N_s\times N_h} \otimes \boldsymbol{I}_{N_{BS}}$ is the stacked N_{BS} truncated Fourier matrix $F_{N_s\times N_h}$. From [8], [9], exploiting the joint sparsity property of the MMV channel, $\bar{\boldsymbol{\mathcal{H}}}_n^a$, is associated with solving $l_{1,2}$ norm optimization problem as follow:

$$\begin{aligned} &\mathit{minimize}_{\bar{\boldsymbol{\mathcal{H}}}_n^a} \left\|\bar{\boldsymbol{\mathcal{H}}}_n^a\right\|_{1,2} = \Big(\sum_{k=1}^{K}\sum_{i=1}^{M}\left\|\bar{\boldsymbol{\hbar}}^{a}_{k,i,n}\right\|_1^2\Big)^{1/2} \\ &\mathit{subject\ to}\ \bar{\boldsymbol{Y}}_n^a = \boldsymbol{A}\bar{\boldsymbol{\mathcal{H}}}_n^a, \end{aligned} \tag{19}$$

where $\bar{\boldsymbol{\hbar}}^{a}_{k,i,n}$ is the column of the k^{th} antenna of $\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a$. The $l_{1,2}$ mixed norm, $\left\|\bar{\boldsymbol{\mathcal{H}}}_n^a\right\|_{1,2}$, can be interpreted as combination of l_1 and l_2 norms. Thus, we can exploit their convexity feature. In $l_{1,2}$ mixed norm, the sparsity of each individual column in $\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a$ is exploited using the l_1-norm, while the joint sparsity among columns is extracted by deploying l_2-norm on the resultant vector [21].

2) CS MODEL FORMULATION FOR CHANNEL ESTIMATION AT INDIVIDUAL UES

The next step is estimating the remaining individual coefficients separately at each UE after receiving the estimated common supports indices from the CH. At each UE in the n^{th} cluster, the CS model for individually estimating the sparse channels' coefficients in the angle-time domain can be extracted from the system model equation (1) as:

$$\bar{\boldsymbol{Y}}_{i,n}^a = \boldsymbol{A}\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a + \bar{\boldsymbol{W}}_{i,n}^a, \tag{20}$$

As in the case of estimation at the CH, the optimization problem that exploit the joint sparsity of MMV structure can be represented as follow:

$$\begin{aligned} &\mathit{minimize}_{\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a} \left\|\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a\right\|_{1,2} = \Big(\sum_{k=1}^{K}\left\|\bar{\boldsymbol{\hbar}}^{a}_{k,i,n}\right\|_1^2\Big)^{1/2} \\ &\mathit{subject\ to}\ \bar{\boldsymbol{Y}}_{i,n}^a = \boldsymbol{A}\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a, \end{aligned} \tag{21}$$

B. PROPOSED WFISTA ALGORITHM

To exploit the common and individual indices from the $l_{1,2}$-norm optimization problems represented by equation (19) and (21), a MMV-CS algorithm is needed. In this subsection, we propose WFISTA algorithm to utilize the joint sparsity property of (19) and (21).

1) PROPOSED WFISTA ALGORITHM FOR SMV CASE

FISTA algorithm proposed in [22] is considered as a sort of iterative shrinkage-thresholding algorithms (ISTA) with better convergence rate and same computational complexity. However, FISTA algorithm fails to trade-off between the degree of sparsity and the undersampling reconstruction, which we called sparsity-undersampling trade-off [23]. Some works have been introduced to improve FISTA by reshaping the l_1-norm problem using a smoothing step as in [24]. However, the work in [24] has not considered the sparsity-undersampling trade-off problem of FISTA.

Unlike [24], this work proposes a weighted FISTA (WFISTA)[3] to enhance the estimation performance and the sparsity-undersampling trade-off of original FISTA for the case of single measurement vector CS problem. One way to achieve this goal is to increase the search space and avoiding the FISTA to fall into a local minimum. Thus, in our proposed WFISTA, we adopt the following three modifications to the vector-updating step of the original FISTA.

[3]It is worth noting that our proposed WFISTA algorithm is completely different from the works proposed in [24], in which the original FISTA is applied to a smoothing $l_1 - l_1$ problem. Thus, algorithm proposed in [24] has modified the objective function not the FISTA algorithm itself.

TABLE 2. Comparison between FISTA and proposed WFISTA algorithms [12].

"Original FISTA"

$$\bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} = \eta_\tau(\boldsymbol{z}^t_{k,i,n}, \tau)$$
$$r^{t+1} = \frac{1+\sqrt{1+4r^t}}{2}$$
$$\boldsymbol{z}^{t+1}_{k,i,n} = \bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} + \frac{r^k-1}{r^{k+1}}(\bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} - \bar{\boldsymbol{\hbar}}^{a^{t-1}}_{k,i,n}, \tau)$$

"Proposed WFISTA"

$$\bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} = \eta_\tau(\boldsymbol{z}^t_{k,i,n}, \tau)$$
$$r^{t+1} = \frac{1+\sqrt{1+4r^t}}{2}$$
$$\boldsymbol{z}^{t+1}_{k,i,n} = \bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} + \frac{r^k-1}{r^{k+1}}\underbrace{\eta_\tau(\bar{\boldsymbol{\hbar}}^{a^t}_{k,i,n} - \boldsymbol{W}_1\bar{\boldsymbol{\hbar}}^{a^{t-1}}_{k,i,n} - \boldsymbol{W}_2\bar{\boldsymbol{\hbar}}^{a^{t-2}}_{k,i,n}, \tau)}_{\text{Weighted Regularization Term}}$$

where:

$$\eta_\tau(x,\tau) = sign(x)max(0, |x| - \tau)$$

- To explore more in the search space, we propose to adjust the vector-updating step to depend on two previously estimated vectors instead of only one previously estimated vector in the original FISTA. Increasing the search space by adopting more previously estimated vectors is proposed based on the mutant strategies studied in some heuristic algorithms as in [25].
- Two weights have been added, $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$, such that a weighted sum of the two previous estimated vectors is subtracted from the current estimated one instead of the direct subtraction of one previous estimated vector in original FISTA. Adding weights gives the proposed WFISTA algorithm the capability of tracking the fast change in the values of the vector coefficients [26]. To prevent our algorithm from falling into a local minimum point, the weights are adapted every iteration according to the inverse of the previously estimated vector.
- A soft thresholding function η_τ is applied to the modified term inspiring from the approximate message passing algorithm proposed in [23]. The threshold function regularizes the value of the modified crucial term and obliges the minimized function to descend to a minimization point during the iterative process.

For the case of estimating the channel of each antenna individually, the objective function can be written as:

$$F = \left\|\bar{\boldsymbol{y}}^a_{k,i,n} - \boldsymbol{A}\bar{\boldsymbol{\hbar}}^a_{k,i,n}\right\|_2^2 + \lambda\left\|\bar{\boldsymbol{\hbar}}^a_{k,i,n}\right\|_1 \tag{22}$$

where $\bar{\boldsymbol{y}}^a_{k,i,n}$ is the column of the k^{th} antenna of $\bar{\boldsymbol{Y}}^a_{i,n}$. The proposed WFISTA algorithm to estimate $\bar{\boldsymbol{\hbar}}^a_{k,i,n}$ compared to original FISTA is demonstrated in Table 2, where η_τ is a soft thresholding operator with τ threshold value, and λ is a regularization multiplier.

From Fig. 3(a), the enhancement in the proposed WFISTA performance upon original FISTA, in terms of the mean square error (MSE) [27] between the estimated and the actual channel coefficients, becomes more obvious at higher values of SNRs. Moreover, the WFISTA shows superior performance for different number of measurements, $N_m = N_s N_p$, in Fig. 3(b). In general, the estimation performance becomes worse as the number of non-zeros elements, S, increased as shown in Fig. 3(c). However, the instability in the original FISTA curve in Fig. 3(c) proves its sparsity-undersampling trade-off problem.

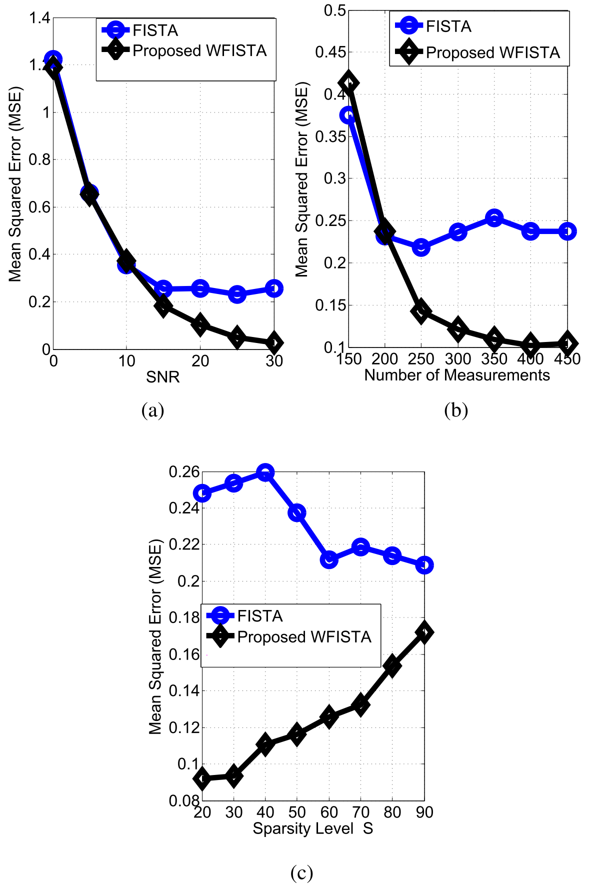

FIGURE 3. Proposed WFISTA vs. original FISTA for a channel vector, $\boldsymbol{\hbar}^a_i$, of length 800 coefficients, where (a) MSE vs. SNR for $S = 70$, and $N_m = 400$, (b) MSE vs. number of measurements 'N_m' for $SNR = 20dB$, and $S = 70$, and (c) MSE vs. sparsity level 'S' for $SNR = 20\,dB$, and $N_m = 400$.

On the other hand, the improved performance of proposed WFISTA over original FISTA comes at the expense of increasing the required memory storage as we need to store two previously estimated vectors rather than one. Also, as the proposed WFISTA explores more in the search space, it requires more iteration to converge compared to original FISTA. Thus, there is a trade-off relationship between the estimation accuracy improvement and the convergence speed.

2) PROPOSED WFISTA ALGORITHM FOR MMV CASE

Both FISTA and the proposed WFISTA have the ability to recover only an individual vector from its measurements solving the SMV CS problem. However, in the proposed cooperative estimation scheme, utilizing the property of joint sparsity structure between the antennas of each UE or between different UEs in the cluster represented by $l_{1,2}$ mixed norm optimization problems (19) and (21) needs the SMV algorithms to be extended to the MMV case. In addition, the $l_{1,2}$ mixed norm is a semi-definite or a second order programming which is computationally costly to be solved by utilizing existing standard algorithms [21].

In this paper, ReMBo (reduce MMV and boost) strategy is adopted to the proposed WFISTA to solve the MMV optimization problems (19) and (21). ReMBo is a low complex strategy in which the MMV problem can be turned into the SMV case. Then, the proposed WFISTA algorithm can be applied. By employing the following three steps, the MMV problem can be solved using the proposed WFISTA.

- *Step 1:* Transforming the MMV problem to the SMV problem of the same sparsity pattern by multiplying the received pilot matrix by a random vector "z" as $\boldsymbol{y}_n^z = \bar{\boldsymbol{Y}}_n^a \boldsymbol{z}$ in the case of estimation at the CH, and $\boldsymbol{y}_{i,n}^z = \bar{\boldsymbol{Y}}_{i,n}^d \boldsymbol{z}$ in the case of estimation at the other UEs in the cluster.
- *Step 2:* Finding the common and individual indices by applying WFISTA algorithm. The resultant vector, in the case of CH and in the case of other UEs, preserves the position of the joint indices of the original matrix as:

$$supp(\boldsymbol{\hbar}_n^z) = supp(\bar{\boldsymbol{\mathcal{H}}}_n^a) = \Omega_{A_{C,n}} \quad (23a)$$

$$supp(\boldsymbol{\hbar}_{i,n}^z) = supp(\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a) = \Omega_{A_{i,n}} \quad (23b)$$

- *Step 3:* Given the common and individual indices exploited by the WFISTA and the received pilot matrices $\bar{\boldsymbol{Y}}_{i,n}^a$, the channel coefficients can be estimated by employing LSE estimator individually at each UE on the estimated indices only.

By applying these three steps, the proposed WFISTA can recover any matrix of MMV structure from its compressed measurements. Table 3 describes the three steps of the proposed WFISTA algorithm to solve (19) at CH and (21) at other UEs in the cluster for the proposed cooperative scheme, which is referred as WFISTA-Coop.

3) PROPOSED WEIGHTS

To track the all possible sparsity pattern for the vectors $\boldsymbol{\hbar}_n^{z^t}$ and $\boldsymbol{\hbar}_{i,n}^{z^t}$, the weights $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ can be adapted every iteration according to its previous solution $\boldsymbol{\hbar}_n^{z^{t-1}}$ [28]. The entire values of the diagonal of $\boldsymbol{W}_1$ are designed such that small weights are assigned to the non-zero coefficients and larger weights elsewhere as in [28]. This adaptation strategy for the weights gives us the opportunity to explore more in the available search space. However, to ensure that the variable sequence of $\boldsymbol{\hbar}_n^{z^t}$ and $\boldsymbol{\hbar}_{i,n}^{z^t}$ will converge to a minimum point the sum of the weights must equal to one. For the case of the proposed WFISTA in the CH, in each iteration, the introduced

TABLE 3. The proposed WFISTA algorithm with cooperative scheme (WFISTA-Coop).

"Estimation at the CH"

- Initializing: $\boldsymbol{\hbar}_n^{z^0} = \boldsymbol{0}$, $\boldsymbol{q}_n^0 = 0$, $\boldsymbol{z}$ is a random vector
  - Step 1: Converting the MMV problem to SMV problem
    $\boldsymbol{y}_n^z = \bar{\boldsymbol{Y}}_n^a \boldsymbol{z}$
  - Step 2: while $t < \mathcal{T}_{max}$ OR $\left\|F_n^t - F_n^{t-1}\right\|_2 > \epsilon$, where $\epsilon \simeq 10^{-8}$, repeat the following steps:

$$\boldsymbol{\hbar}_n^{z^t} = \eta_\tau(\boldsymbol{q}_n^t, \tau^t)$$
$$r^{t+1} = \frac{1+\sqrt{1+4r^t}}{2}$$
$$\boldsymbol{q}_n^{t+1} = \boldsymbol{\hbar}_n^{z^t} + \frac{r^k - 1}{r^{k+1}} \eta_t(\boldsymbol{\hbar}_n^{z^t} - \boldsymbol{W}_1 \boldsymbol{\hbar}_n^{z^{t-1}} - \boldsymbol{W}_2 \boldsymbol{\hbar}_i^{z^{t-2}})$$
$$F_n = \left\|\boldsymbol{y}_n^z - \boldsymbol{A}\boldsymbol{\hbar}_n^{z^t}\right\|_2^2 + \lambda \left\|\boldsymbol{\hbar}_n^{z^t}\right\|_1$$

Update $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ using (24) and (25).

End While
$\Omega_{A_c} = supp(\boldsymbol{\hbar}_n^z)$

"Estimation at the UEs"

- Input $\boldsymbol{G}_{i,n} = \left\langle \boldsymbol{I} - (\boldsymbol{A}\,|_{\Omega_{A_c}})(\boldsymbol{A}\,|_{\Omega_{A_c}})^\dagger \right\rangle \bar{\boldsymbol{Y}}_{i,n}^a$
- Initializing: $\boldsymbol{\hbar}_{i,n}^{z^0} = \boldsymbol{0}$, $\boldsymbol{q}_{i,n}^0 = 0$, $\boldsymbol{z}$ is a random vector
- For $i : 1 \rightarrow M$
  - Step 1: Converting the MMV problem to SMV problem
    $\boldsymbol{g}_{i,n}^z = \boldsymbol{G}_{i,n}\boldsymbol{z}$
  - Step 2: while $t < \mathcal{T}_{max}$ OR $\left\|F_{i,n}^t - F_{i,n}^{t-1}\right\|_2 > \epsilon$, where $\epsilon \simeq 10^{-8}$, repeat the following steps:

$$\boldsymbol{\hbar}_{i,n}^{z^t} = \eta_\tau(\boldsymbol{q}_{i,n}^t, \tau^t)$$
$$r^{t+1} = \frac{1+\sqrt{1+4r^t}}{2}$$
$$\boldsymbol{q}_{i,n}^{t+1} = \boldsymbol{\hbar}_{i,n}^{z^t} + \frac{r^k - 1}{r^{k+1}} \eta_t(\boldsymbol{\hbar}_{i,n}^{z^t} - \boldsymbol{W}_1 \boldsymbol{\hbar}_{i,n}^{z^{t-1}} - \boldsymbol{W}_2 \boldsymbol{\hbar}_{i,n}^{z^{t-2}})$$
$$F_{i,n} = \left\|\boldsymbol{g}_{i,n}^z - \boldsymbol{A}\boldsymbol{\hbar}_{i,n}^{z^t}\right\|_2^2 + \lambda \left\|\boldsymbol{\hbar}_{i,n}^{z^t}\right\|_1$$

Update $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ using (24) and (25).

End While
$\Omega_i = supp(\boldsymbol{\hbar}_i^z)$
$\Omega_{A_{i,n}} = \Omega_{i,n} \cup \Omega_{A_{c,n}}$
  - Step 3: Estimate the MMV channel
    $\bar{\boldsymbol{\mathcal{H}}}_{i,n}^a\,|_{\Omega_{A_{i,n}}} = (\boldsymbol{A}\,|_{\Omega_{A_{i,n}}})^\dagger \bar{\boldsymbol{Y}}_{i,n}^a$

- where: $\tau^t = \frac{\tau^{t-1}}{\delta} \langle \eta_{t-1}(\boldsymbol{A}^H \boldsymbol{y}_{i,n}^z + \boldsymbol{\hbar}_{i,n}^{z^{t-1}}; \tau^{t-1}) \rangle$
  $\eta_\tau(x, \tau^t) = sign(x)max(0, |x| - \tau^t)$

weights are adapting according to:

$$diag(\boldsymbol{W}_1) = min\left(\zeta, \frac{\rho}{|\boldsymbol{h}_n^z| + \varepsilon}\right) \tag{24}$$

$$\boldsymbol{W}_2 = \boldsymbol{I} - \boldsymbol{W}_1 \tag{25}$$

where $\rho, \varepsilon > 0$ are tuning parameters to provide stability and to prohibit a zero value in the denominator, while $0 < \zeta < 1$ is an upper bound for the weights.

4) COMPLEXITY ANALYSIS

The overall computational complexity of the proposed WFISTA algorithm either for SMV or MMV is the same as the FISTA algorithm and is calculated as $\mathcal{O}(N_h N_{BS} N_p N_s)$. The complexity computation comes from the matrix-vector multiplication part in the objective function in (22) rather than matrix multiplication or matrix inversion as the other joint recovery algorithms. The complexity of the WFISTA at the CH is still $\mathcal{O}(N_h N_{BS} N_p N_s)$. However, the WFISTA complexity at other UEs in the cluster is reduced to $\mathcal{O}(N_h(N_{BS} - S_c)N_p N_s)$, as the length of the recovered vector is reduced from $N_h N_{BS}$ to $N_h(N_{BS} - S_c)$ due to the step of estimating the common supports at the CH.

In contrast with other joint estimation algorithms [26], [27], the proposed WFISTA is independent of either the number of UEs per cluster M or the number of antenna per UE K. However, the proposed WFISTA needs more memory storage than original FISTA. The bottleneck complexity order of different joint algorithm in the case of solving the optimization problem (19) is listed in Table 4.

V. ANALYSIS OF THE FEEDBACK DECREASING PROBABILITY FOR THE PROPOSED COOPERATIVE SCHEME

Although the probability of determining the common support indices is tightly related to the CSI estimation quality [9], it also related to the feedback decreasing probability. The higher probability recovery for the common support indices, the higher feedback overhead decreasing probability. In this section, we will analyze the decreasing in the feedback probability in the case of applying the proposed cooperative scheme. First, we suppose the following events:

TABLE 4. The complexity order of the joint estimation algorithms for problem (19).

Algorithm	Complexity
SAMUSIC [29]	$\mathcal{O}(N_{BS}^3 N_h^3 + N_{BS}^2 N_h^2 N_s N_p)$
M-OMP [26]	$\mathcal{O}(N_h N_{BS} N_p N_s K)$
M-FOCUSS [30]	$\mathcal{O}(N_{BS}^3 N_h^3 + N_{BS}^2 N_h^2 N_s N_p)$
SPGL1 [27]	$\mathcal{O}(N_h N_{BS} N_p N_s K)$
AMP-MMV [31]	$\mathcal{O}(N_h N_{BS} N_p N_s K)$
GGAMP-SBL [32]	$\mathcal{O}(N_h N_{BS} N_p N_s K)$
WFISTA (CH)	$\mathcal{O}(N_h N_{BS} N_p N_s)$
WFISTA (UE)	$\mathcal{O}(N_h (N_{BS} - S_c) N_p N_s)$

- Θ_{df}: is the event of decreasing in feedback.
- Θ_{co}: is the event of UEs cooperation.
- Θ_{cs}: is the event that the l^{th} row of $\bar{\mathcal{H}}_n^a$ is a common support for a given sparsity pattern.
- Θ_{Ecs}: is the event that the l^{th} row of $\bar{\mathcal{H}}_n^a$ is estimated as a common support row.

The probability of decreasing in the feedback given that the UEs cooperate with each others is upper bounded by the existence of common supports and the ability of estimating these supports as:

$$\mathcal{P}r(\Theta_{df} \mid \Theta_{co}) \leq \mathcal{P}r(\Theta_{cs})\mathcal{P}r(\Theta_{Ecs}) \tag{26}$$

where $\mathcal{P}r(\Theta_{cs})$ is the probability that the l^{th} row of $\bar{\mathcal{H}}_n^a$ is common non-zero support and can be expressed as:

$$\mathcal{P}r(\Theta_{cs}) = \frac{S_c}{S} \tag{27}$$

From the proposed algorithm in table 2 and equations (19) and (21), a common non-zero row in the channel matrix $\bar{\mathcal{H}}_n^a$ can be detected by comparing the Frobenius norm value of this row $\left\|\bar{\mathcal{H}}_n^a(l,:)\right\|_F^2$ with a threshold. The squared of the Frobenius norm can be considered as a chi-squared χ distribution [9]. The chi-squared χ distribution with degree of freedom 2κ, $\chi_{2\kappa}$, can be confined with Chernoff bounds as:

$$\mathcal{P}r(\chi_{2\kappa} < 2a\kappa) \leq \exp(-\kappa(-1 + a - \ln a)) \tag{28}$$

Suppose λ_i is the probability that the l^{th} row of $\bar{\mathcal{H}}_n^a$ is estimated as a common zero row, and using Chernoff bounds in equation (28), λ_i can be upper bounded by:

$$\begin{aligned}\mathcal{P}r(\chi_{2\kappa} < 2a\kappa) &= \mathcal{P}r(\left\|\bar{\mathcal{H}}_n^a(l,:)\right\|_F^2 < KM\tau) = \lambda_i \\ &\leq \exp(-\frac{KM}{2}(-1+\tau-\ln\tau)),\end{aligned} \tag{29}$$

From equation (29), the probability that the l^{th} row of $\bar{\mathcal{H}}_n^a$ is estimated as a common non-zero support can be lower bounded by:

$$\begin{aligned}\mathcal{P}r(\left\|\bar{\mathcal{H}}_n^a(l,:)\right\|_F^2 > KM\tau) &= 1 - \lambda_i \\ &\geq 1-\exp(-\frac{KM}{2}(-1+\tau-\ln\tau)),\end{aligned} \tag{30}$$

Hence, the probability $\mathcal{P}r(\Theta_{df} \mid \Theta_{co})$ is lower bounded by:

$$\mathcal{P}r(\Theta_{df} \mid \Theta_{co}) \geq \frac{S_c}{S}(1 - \exp(-\frac{KM}{2}(-1 + \tau - \ln\tau))) \tag{31}$$

From equation (31), we conclude that $\mathcal{P}r(\Theta_{df} \mid \Theta_{co})$ increases as K and M increase. Therefore, in the proposed cooperative estimation and feedback schemes, larger number of antennas per UE K or larger number of UEs per cluster M will result in a more decreasing in the feedback overhead. Also, increasing the number of common support S_c compared to the total sparsity level S will give the same result of reducing the feedback overhead as referred in Sec III.

VI. NUMERICAL RESULTS AND DISCUSSION

A. SIMULATION PARAMETERS

In terms of SNR, the degree of sparsity S, and the number of pilots N_p, the performance of the proposed FDD cooperative estimation scheme with WFISTA, which is referred as WFISTA-Coop, is compared with various joint channel estimation schemes. As we estimate the channel at UE side that has a limited power source, we target in our comparison low-complex fast convergence CS algorithms that do not need prior information on the estimated channels. We compare the proposed scheme with the following state-of-the-art baselines:

- Conventional channel estimation: LSE in [17], [33] is considered as a benchmark estimator
- Conventional CS channel estimation at UE side: SS-MUSIC [34], M-OMP [35], AMP-MMV [31], GGAMP-SBL [32], [36], and SPGL1 [27] are adopted to individually recover $\bar{\mathcal{H}}^a_{i,n}$ by solving the optimization problem (21) at each UE. Also, for fair comparison, the proposed WFISTA without the cooperative scheme is also applied to solve (21).
- Distributed CS channel estimation: J-OMP [9] is simulated as an example of distributed algorithms where the channel of all UEs are estimated at the BS taking into account the common and individual supports.

To satisfy the restricted isometry property (RIP) of CS [5], Gaussian distribution is used to produce the pilot matrix. The sparsity level S refers to the S-nonzero rows in the channel $\bar{\mathcal{H}}^a_{i,n}$ (Individual supports). Out of S individual supports, each $\bar{\boldsymbol{h}}^a_{k,i,n}$ has S_c common supports. The cell is divided into four clusters, $N_c = 4$, each cluster has $M = 20$ UEs, while the other common simulation parameters are $N_{BS} = 100$, $K = 2$, $N_q = 5$, $N_h = 100$. The normalized mean square error [37] is applied to evaluate the performance of each estimator and is defined as:

$$NMSE = 10log(\frac{\left\|\boldsymbol{H}_{i,n} - \hat{\boldsymbol{H}}_{i,n}\right\|_F^2}{\left\|\boldsymbol{H}_{i,n}\right\|_F^2}) \tag{32}$$

B. SIMULATION RESULTS AND DISCUSSION

In compared with other algorithms, Fig.4 shows the estimation performance of the proposed WFISTA with cooperative estimation scheme, referred as WFISTA-Coop, in terms of NMSE versus different values of SNR for $S = 21$, $S_c = 15$, and $N_p = 45$. Thanks to the introduced adapted weights and the cooperative scheme, the proposed WFISTA and WFISTA-Coop shows improved performance over either joint algorithms or distributed algorithms.

In Fig.5, the performance of estimation is studied under different lengths of pilot symbols, N_p, for SNR=10, $S = 21$, and $S_c = 15$. Estimating the channel depending on a history of two previous estimated values rather than one improves the estimation performance of both WFISTA and WFISTA-Coop even with a small number of transmitted pilots. However, as the number of N_P increases towards $N_C \times M \times K$, all

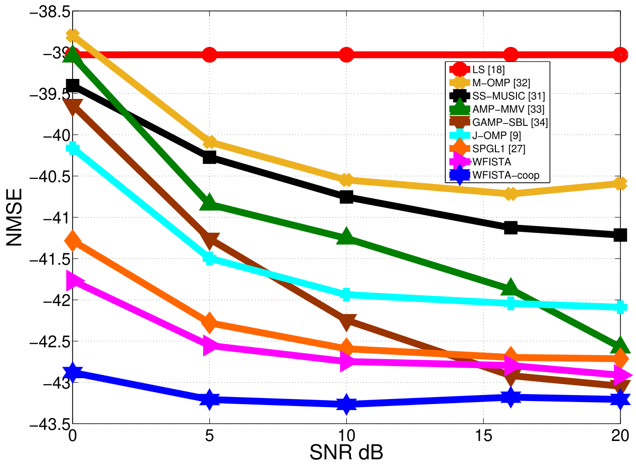

FIGURE 4. The performance of the proposed WFISTA and WFISTA-Coop versus state-of-the-art joint estimation algorithms in terms of *NMSE* at several values of SNR for $M = 40$, $K = 2$, $N_{BS} = 100$, $S_c = 15$, $S = 21$, $S_c = 15$, $N_c = 10$, $N_p = 45$, $N_q = 5$, and $N_h = 100$.

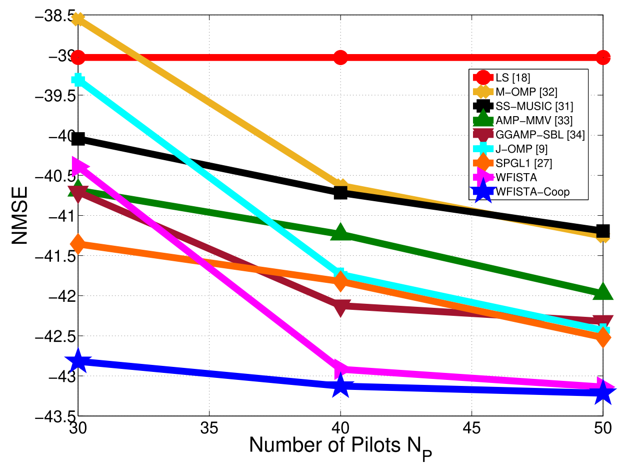

FIGURE 5. The performance of the proposed WFISTA and WFISTA-Coop versus state-of-the-art joint estimation algorithms in terms of *NMSE* at several lengths N_p of pilot symbols for $M = 40$, $K = 2$, $N_{BS} = 100$, $S_c = 15$, $S = 21$, $SNR = 10\,dB$, $N_c = 10$, $N_q = 5$, and $N_h = 100$.

algorithms act the same performance as the problem moves from unsolved to solved problem.

Under different challenging sparsity levels S, Fig.6 shows the performance in terms of NMSE for SNR=10, and $N_p = 45$. Thanks to the ability of WFISTA to exploit the joint individual supports in addition to extracting the common supports by proposing the cooperative scheme, the proposed algorithm provides robust estimation performance over other joint algorithms.

From the simulation figures, the proposed WFISTA algorithm improves the channel estimation performance even without the cooperative scheme. Although the proposed cooperative scheme improves the estimation process and reduces the feedback overhead, it needs more time than the other techniques for channel estimate process due to the D2D communication steps. Thus, a trade-off exists in the proposed cooperative scheme between the required time for estimation and the decreasing in the feedback overhead.

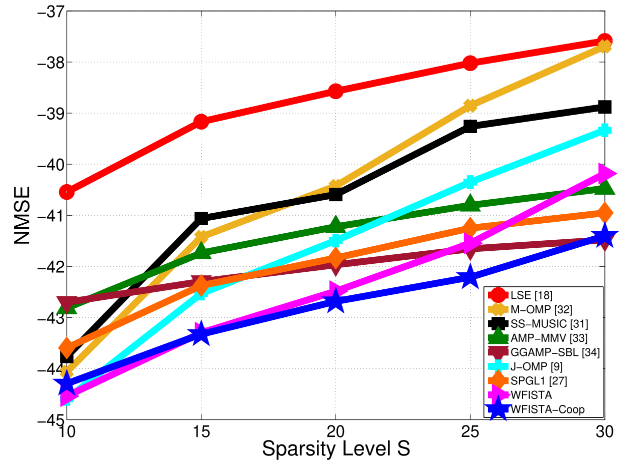

FIGURE 6. The performance of the proposed WFISTA and WFISTA-Coop versus state-of-the-art joint estimation algorithms in terms of *NMSE* at several sparsity levels S for $M = 40$, $K = 2$, $N_{BS} = 100$, $S_c = 15$, $S = 21$, SNR $= 10\,dB$, $N_c = 10$, $N_p = 45$, $N_q = 5$, and $N_h = 100$.

VII. CONCLUSION

In this paper, a FDD cooperative channel estimation scheme has been proposed for downlink 3D-OFDM mMIMO system where UEs within the cell are clustered and via D2D communication they can cooperate to exploit the sparsity structure of the channels. Also, WFISTA algorithm has been proposed to exploit the channel sparsity structure of UEs within the clusters. In WFISTA algorithm, the FISTA algorithm has been extended to the MMV case using REMBO strategy, while new adaptive weights are introduced to improve the overall estimation performance. Complexity and probability analysis also have been discussed, and the improved performance of the proposed scheme over various joint channel estimation techniques are validated by the simulation results.

REFERENCES

[1] 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.

[2] 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.

[3] K. Liu, H. Feng, T. Yang, and B. Hu, "Structured sparse channel estimation for 3D-MIMO Systems," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.

[4] 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.

[5] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[6] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, Jun. 2014.

[7] A. Nasser and M. Elsabrouty, "Adaptive split Bregman for sparse and low rank massive MIMO channel estimation," in *Proc. 23rd Int. Conf. Telecommun. (ICT)*, May 2016, pp. 1–5.

[8] A. Nasser, M. Elsabrouty, and O. Muta, "Alternative direction for 3D orthogonal frequency division multiplexing massive MIMO FDD channel estimation and feedback," *IET Commun.*, vol. 12, no. 11, pp. 1380–1388, Jul. 2018.

[9] 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, Jun. 2014.

[10] 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.

[11] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Nov. 2014.

[12] A. Nasser, M. Elsabrouty, and O. Muta, "Weighted fast iterative shrinkage thresholding for 3D massive MIMO channel estimation," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[13] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 5, Feb. 2007.

[14] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[15] S. Wu, L. Kuang, Z. Ni, D. Huang, Q. Guo, and J. Lu, "Message-passing receiver for joint channel estimation and decoding in 3D massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8122–8138, Dec. 2016.

[16] Y. Zhu, L. Liu, and J. Zhang, "Joint angle and delay estimation for 2D active broadband MIMO-OFDM systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3300–3305.

[17] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Mar. 2013.

[18] L. Huang, C. K. Ho, J. W. M. Bergmans, and F. M. J. Willems, "Pilot-aided angle-domain channel estimation techniques for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 906–920, Mar. 2008.

[19] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

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

[21] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," *Proc. SPIE*, vol. 8858, Sep. 2013, Art. no. 88580R.

[22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[23] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[24] T. Sun and L. Cheng, "Reweighted fast iterative shrinkage thresholding algorithm with restarts for l_1-l_1 minimisation," *IET Signal Process.*, vol. 10, no. 1, pp. 28–36, Feb. 2016.

[25] S.-M. Guo, C.-C. Yang, P.-H. Hsu, and J. S. H. Tsai, "Improving differential evolution with a successful-parent-selecting framework," *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 717–730, Oct. 2015.

[26] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[27] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[28] M. S. Asif and J. Romberg, "Fast and accurate algorithms for re-weighted ℓ_1-norm minimization," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5905–5916, Dec. 2013.

[29] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.

[30] E. van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1201–1229, 2011.

[31] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, 2013.

[32] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294–308, Jan. 2018.

[33] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[34] Z. Wen, B. Hou, and L. Jiao, "Joint sparse recovery with semisupervised MUSIC," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 629–633, May 2017.
[35] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, Jr., "Compressed channel feedback for correlated massive MIMO systems," *IEEE/KICS J. Commun. Netw.*, vol. 18, no. 1, pp. 95–104, Feb. 2016.
[36] J. Zhu, L. Han, and X. Meng, "An AMP-based low complexity generalized sparse Bayesian learning algorithm," *IEEE Access*, vol. 7, pp. 7965–7976, 2019.
[37] M. Masood, L. H. Afify, 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. doi: 10.1109/TSP.2014.2369005.

AHMED NASSER received the B.Sc. degree (Hons.) in electronics and communications engineering from Suez Canal University, Egypt, in 2012, and the M.Sc. degree in electronics and communications engineering from Egypt Japan University of Science and Technology, Alexandria, Egypt, in 2016. He is currently pursuing the double Ph.D. degree with the Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan, and the Department of Electronics and Communication, Egypt-Japan University of Science and Technology. He joined Suez Canal University as a Teaching Assistant. His research interests include interference mitigation in wireless networks, massive MIMO, heterogeneous networks, channel estimation, compressive sensing, and emerging technologies for 5G wireless communications.

MAHA ELSABROUTY received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from Cairo University, Egypt, and the M.Sc. and Ph.D. degrees, both in electrical engineering, from the University of Ottawa. Currently, she is with Egypt-Japan University of Science and Technology. Her current research interests include massive MIMO techniques, interference management in HetNets, cognitive radio, intelligent techniques for wireless communications, and green communication systems.

OSAMU MUTA received the B.E. degree from Ehime University, in 1996, and the M.E. and Ph.D. degrees from Kyushu Institute of Technology, Japan, in 1998 and 2001, respectively. In 2001, he joined the Graduate School of Information Science and Electrical Engineering, Kyushu University, as an Assistant Professor. Since 2010, he has been an Associate Professor with the Center for Japan–Egypt Cooperation in Science and Technology, Kyushu University. His current research interests include signal processing techniques for wireless communications and powerline communications, MIMO, and nonlinear distortion compensation techniques for high-power amplifiers. He received the 2005 Active Research Award from IEICE Technical Committee of Radio Communication Systems, the 2014, the 2015, and the 2017 Chairman's Awards for excellent research from IEICE Technical Committee of Communication Systems. He is a senior member of the IEICE.