

Received 13 August 2022, accepted 15 September 2022, date of publication 30 September 2022, date of current version 10 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3211076

RESEARCH ARTICLE

Efficient Distributed Multi-Task Schemes for mmWave MIMO Channel Estimation

MARIA TRIGKA¹, CHRISTOS MAVROKEFALIDIS¹, (Member, IEEE),
AND KOSTAS BERBERIDIS¹, (Senior Member, IEEE)

Computer Engineering and Informatics Department, University of Patras, 26504 Patras, Greece

Corresponding author: Maria Trigka (trigka@ceid.upatras.gr)

This work was supported in part by the Greece and the European Union (European Social Fund-ESF) through the Operational Programme 'Human Resources Development, Education and Lifelong Learning' in the context of the project 'Strengthening Human Resources Research Potential via Doctorate Research,' implemented by the State Scholarships Foundation (IKY) under Grant MIS-5000432; and in part by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation under Project INFRASTRUCTURES/1216/0017 IRIDA. The publication of the article has been financed by the Research Committee of the University of Patras.

ABSTRACT In this study, the problem of sparse channel estimation is investigated with the employment of a fully distributed approach. We exploit the spatially joint sparsity structure of the involved channels to formulate the channel estimation problem in the angular domain. The devices collaboratively estimate their channel sparsity support sets before the local estimation of the channel values, assuming the existence of global and common support subsets. The combination of the proposed distributed scheme with the classical Simultaneous Orthogonal Matching Pursuit (SOMP) algorithm is called Weighted Distributed Simultaneous Orthogonal Matching Pursuit (WDiSOMP). The performance of WDiSOMP is assessed under a multi-task scheme considering a new weighted voting method. The new mechanism is applied at each support set index and enhances its estimation accuracy and, thus, the eventual estimation of the overall channel. The mean squared error (MSE) is utilized to derive the performance bounds and assess the efficacy of the WDiSOMP estimator. Finally, the performance and the theoretical findings are evaluated via the comparison of WDiSOMP with Distributed Simultaneous Orthogonal Matching Pursuit (DiSOMP), local SOMP, and a centralized approach based on Structured SOMP (SSOMP) in terms of the MSE.

INDEX TERMS Compressive sensing, SOMP, distributed estimation, channel estimation, mmWave.

I. INTRODUCTION

The millimeter wave (mmWave) spectrum plays a key role in the implementation of 5G (and upcoming generations) communication systems. The available bands are between 30 GHz and 300 GHz (i.e., wavelengths between 1 mm to 10 mm), and the large unoccupied frequencies in these bands can be exploited to meet the increasing demand for transmission capacity [1], [2], [3], [4]. However, the transmitting signals in the mmWave bands confront high penetration loss and propagation attenuation, restricting the communication range [4]. Massive Multi-Input, Multi-Output (MIMO) and directional beamforming (facilitated by the short wavelength of mmWave signals) have been leveraged in the relevant

literature to moderate this loss and ensure reliability in communication.

The incorporation of mmWave MIMO antenna systems in the physical layer of 5G networks [5] has increased their transmission efficiency (spectral and power), throughput and network coverage, ensuring the successful support of various evolving technologies, such as Vehicular-to-Vehicular (V2V) or Vehicular-to-Everything (V2X), Device-to-Device (D2D), and Device-to-Everything (D2E) communication systems, Machine-to-Machine (M2M)/ Internet-of-Things (IoT), Augmented Reality (AR), Virtual Reality (VR), Smart City, etc. [6]. MmWave is a key technology for the deployment of Heterogeneous Networks (HetNets) in 5G IoT (such as smart homes, smart cities, e-health) [7].

To fully exploit the potential of these applications, channel knowledge is required among all the involved antenna pairs of the participating devices. In practical scenarios, the

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Khawaja¹.

mmWave (massive) MIMO channel is sparse in the angular domain due to the narrow-angle spread and limited scattering in the propagation environment at the base station (BS) side [8]. Thus, a few significant paths are actually active (and, essentially, fewer than the number of transmitting/receiving antennas) and contribute to the transmission. In recent years, Compressed Sensing (CS) techniques have been widely used for channel estimation tasks [9], [10]. CS theory can be leveraged to estimate the channel with reduced training overhead by identifying the parameters involved in the commonly adopted geometric channel model (channel gains, Angles of Departure (AoD) and Angles of Arrival (AoA)), related only to those paths, rather than estimating the channel matrix. In the following sections, related works in the domain of channel estimation with CS-based approaches, the motivation and the contribution of this research will be described.

A. RELATED WORKS

Numerous state-of-the-art solutions for mmWave channel estimation are outlined by Hassan *et al.* in [10]. In the relevant literature, the problem is mainly addressed either in point-to-point [8], [11], [12], [13] or in point-to-multiple points frameworks [14]. In addition, the sparse channel recovery approaches [15] can be employed either individually [11], [16] or centrally (e.g., at the BS) [8], [17], [18]. In the centralized solutions, the CS algorithms exploit the common sparsity patterns among the channels of nearby devices that are manifested in their measurements due to the transmission environment and, thus, enhance the estimation accuracy or reduce the training overhead. The downlink pilot training overhead would be prohibitively in MIMO systems with large-scale antenna arrays. Exploiting already established CS techniques and the hidden sparsity in sparse channel matrices, efficient channel estimation can be accomplished with fewer training symbols and improved performance due to the limited local scattering at the BS.

More specifically, to further enhance the estimation performance and reduce pilot demand, they leverage the so-called joint sparsity either among devices or among different sub-carriers of single devices due to the shared transmission environment. [8] is one of the first works to establish this property with devices of multiple antennas. In [13], [17], [19], [20], [21], and [22], channel estimation algorithms exploit common sparsity patterns via a distributed-aided approach employed over different orthogonal frequency division multiplexing (OFDM) sub-carriers of single devices. It should be noted that in previous algorithms for channel estimation, such as in the Joint Orthogonal Matching Pursuit (J-OMP) in [8] and the Distributed Sparsity Adaptive Matching Pursuit (DSAMP) in [17], the term “distributed” refers to geographically distributed devices that send their data to a central node for processing (e.g., the BS), or engage distributed algorithms executed, however, locally over different sub-carriers of a single device. Here, this term refers to a fully distributed channel estimation approach in which many devices collaborate and exchange information with their neighborhood.

TABLE 1. Comparison with previous works.

Work	Channel model	Multi-task sparsity induced	Architecture	Processing
[23]	geometric, clustered, multi-path	no	MISO, Multi-user	local
[14]	geometric, clustered, multi-path	no	MISO, Multi-user	centralized
[24]	geometric, clustered, single-path	no	MIMO, Multi-user	local
[26]	geometric, clustered, multi-path	no	MISO, Multi-user	local
[27]	geometric, clustered, single-path	no	MIMO, Single-user	local
[25]	non-parametric, MIMO channel description	no	MIMO, Single-user	local
proposed	geometric, clustered, multi-path	yes	MIMO, Multi-user	distributed

Moreover, in [23], Dai *et al.* impose a non-uniform size burst-sparsity structure in the sparse channel without considering any correlations among the channels of different users. They propose a generic Sparse Bayesian Learning (SBL) framework that jointly detects unknown outliers and bursts and estimates the channel. Furthermore, the authors in [14] introduce a common sparsity model, where the individual non-zero values are treated as outliers, and propose an SBL method that jointly estimates the channel and finds the users’ groups. In [24], the authors exploit common sparsity to build a sparse Bayes tensor and Direction of Arrival (DoA) tracking-based channel estimation algorithm for V2X mmWave massive hybrid MIMO systems.

Recent studies see channel estimation as multi-task, which, however, is interpreted differently from this study. Specifically, in [25], the MIMO channel prediction is based on a multi-task least square support vector machine that exploits the spatial correlations between the channels of different antenna pairs instead of predicting the channel per antenna pair independently. Also, in [26], a joint channel estimation scheme is presented for a reconfigurable intelligent surface-assisted mmWave communication system. In particular, it estimates the direct and cascaded channels at the same coherence time instead of learning them separately. Different from the above works, in [27], a multi-task deep learning-based network is proposed to simultaneous design the hybrid precoding and combining matrices under the geometric mmWave channel model. In Table 1, we demonstrate a brief comparative description of our work with recent previous studies. The methods listed in this table are compared, focusing on some key traits, i.e., the assumed properties of the channel, whether multi-task is considered from a sparse representation perspective, the architecture and the type of processing (centralized, distributed, local).

Moving on to the distributed joint sparsity recovery literature [10], [28], several algorithms have been developed for the

joint sparsity models JSM-1 (common support and individual values) and JSM-2 (common support), and their generalizations, which, first, were introduced by Baron *et al.* in [29]. These algorithms either follow a greedy and iterative process (e.g., OMP) or solve an optimization problem (e.g., Least absolute shrinkage and selection operator - Lasso) to estimate the desired sparse vectors or matrices. Here, we will focus on a greedy method, especially OMP, due to its simplicity, fast implementation and lower computational complexity than optimization techniques. However, other sparse coding algorithms can be used, such as matching pursuit (MP) [30]. The distributed versions of subspace pursuit (SP) and orthogonal matching pursuit (OMP), the so-called DiSP and DiOMP, respectively, were first introduced in [31], and they apply to both JSM-1 and JSM-2 models (assuming a priori knowledge of the sparsity structure). Besides, two versions of Decentralized and Collaborative Orthogonal Matching Pursuit (DC-OMP1, DC-OMP2) for JSM-1 have been suggested in [32], taking into account some communication constraints. Compared to DiOMP, DC-OMP 1 recovers the support set with higher accuracy by utilizing a voting mechanism incorporated in the iterations of the OMP. DC-OMP 2 is based on SOMP and is much more accurate than DC-OMP 1 but at the cost of a higher communication burden. It is noted that in the framework of distributed collaborative algorithms, our recent works [33], [34] constitute state-of-the-art solutions for a multi-task joint sparse representation model. Compared to their predecessors, they follow an adaptive process via the employment of adaptive weights majority voting and only need to know the total sparsity level.

B. MOTIVATION

The constantly increasing use of smart devices and multimedia services, as well as the demand for high data rates, high energy efficiency, low latency and high quality of service (QoS), have made fifth-generation (5G) networks and beyond more imperative than ever. 5G networks address the emerging challenges by considering i) massive MIMO systems in the mmWave range, ii) full-duplex (FD) transmissions, iii) D2D communication protocols and iv) network densification with heterogeneous networking architectures, where various types of cells (e.g., small, large) and access technologies are expected to coexist. In the near future, direct D2D communication is regarded as an essential and promising technology, e.g., for extending capacity, enabling scalability and traffic offloading in various short-range applications, including IoT, M2M, and emergency networks [35]. In the vision of 5G networks, users organized into small groups will exchange information, e.g., about the network and channel status, cooperate in various operations of common interest and employ new technologies in favour of coordinated communication [36].

A thorough search of relevant literature shows that the trend toward distributed and cooperative schemes also emerges in the physical layer, with some indicative, yet notable, examples being beamforming [37], [38] and spectrum sensing [39]. For the first time, the present study focuses

on the problem of fully-distributed channel estimation for mmWave MIMO systems and proposes novel and efficient schemes that enable devices to cooperate for estimating the involved channels.

C. OUR CONTRIBUTION

As observed in Table 1, our method is distinct from the sparsity-induced multi-task and processing perspectives. In particular, the main innovations of this research are the following:

- The measurements model allows the proposed solution to be directly applicable to scenarios with either single-antenna devices, following a Single Measurement Vector method, e.g., OMP, or multi-antenna devices, utilizing a Multiple Measurement Vector method, e.g., SOMP.
- A multi-task framework is elaborated under which the mmWave MIMO sparse channel is modeled and estimated, exploiting the joint sparsity among the channels of proximal devices surrounded by global and some common scattering clusters.
- A *fully* distributed cooperative channel estimation algorithm, WDi(S)OMP, is suggested that undertakes the multi-task commonalities among the sparsity patterns of adjacent devices in an efficient manner. WDi(S)OMP remains well-performing even when non-uniform overlapping sparsity groups occur in the network. It is achieved by incorporating a new voting mechanism which estimates an individual weight per support set index, not a global one as in our prior scheme [34], and further improves the channel estimation accuracy and reduces training time by the BS.
- A performance evaluation of WDi(S)OMP is presented from various aspects. More specifically:
 - The new algorithm requires a low number of iterations, as it properly exploits the channel sparsity level, which, in mmWave communications, is small.
 - The overall complexity is discussed at each stage of the algorithm per iteration - communication round.
 - The associated communication overhead due to collaboration is analytically captured. Despite the overhead, the distributed approaches, such as the one presented here, might be preferable in many cases encountered in practice (e.g., cases with low channel sparsity appearing in the mmWave range and/or multi-antenna end-users) in contrast to adopting centralized processing at BS.
 - A theoretical analysis is conducted where the MSE is analytically computed assuming an oracle estimator as a benchmark, from which we evaluate in the experiments how well the suggested estimator behaves and identify the parameters that most affect our solution to attain the minimum error values.

The evaluation of the algorithm is demonstrated via simulations, and it turns out that it performs significantly better than the existing ones. Several experiments have been

conducted based on the elaborated multi-task sparsity model to evaluate the channel matrix recovery performance in terms of the number of measurements (i.e., training overhead) and noise conditions (i.e., Signal-to-Noise-Ratio (SNR)).

The rest of this paper is organized as follows. In Section II, the data model is described. In Section III, the joint sparse channel model is elaborated in detail, under which our estimator will be designed. Then, Section IV presents the problem formulation for the mmWave channel estimation. In the following, Section V shows the proposed distributed MIMO downlink channel estimation algorithm with the aid of the CS theory. Moreover, Section VI analyzes the proposed method's required iterations and computational complexity, the respective transmission overhead, and the impact of support set and angular grid resolution on reconstruction quality. In Section VII, several experiments are presented to verify the efficiency of the proposed approach, accompanied by the appropriate discussion. Finally, Section VIII outlines conclusions and future directions of our research.

The following notations are used in this paper. Uppercase bold letters are matrices, lowercase bold letters are vectors, letters with a hat are estimations, $(\cdot)^T$ denotes the transposition, $(\cdot)^H$ denotes the complex conjugate transposition, $E[\cdot]$ means statistical expectation, $Tr(\cdot)$ denotes the trace of a matrix, \mathbf{I} is the identity matrix, $\lambda_{\min}(\cdot)$, $\lambda_{\max}(\cdot)$ are the minimum and maximum eigenvalue of a matrix; $\text{supp}(\mathbf{h})$, $\text{supp}(\mathbf{H})$ are the index set of the non-zero entries of vector \mathbf{h} and the non-zero entries columns of matrix \mathbf{H} , respectively; $\mathbf{A}_{\mathcal{S}}$ denotes the sub-matrix formed by collecting the columns of \mathbf{A} whose indexes are in set \mathcal{S} ; $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$ and $\|\mathbf{a}\|$ denote the Frobenius norm, spectrum norm of \mathbf{A} and Euclidean norm of vector \mathbf{a} respectively.

II. THE SYSTEM MODEL

In this paper, we will focus on downlink transmission. BS of N_t antennas communicates with K devices of N_r antennas to transmit (broadcasts) T training symbol vectors which are concatenated in $\mathbf{X} \in \mathbb{C}^{N_r \times T}$ in T time slots via a narrowband, block-fading mmWave MIMO channel. Each device receives the signal

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k, \quad (1)$$

where matrix $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ captures the mmWave downlink MIMO channel from the BS to the k -th device. Also, $\mathbf{Y}_k \in \mathbb{C}^{N_r \times T}$ and $\mathbf{N}_k \in \mathbb{C}^{N_r \times T}$ denote the concatenated received signal and additive noise vectors, respectively. The elements of the latter are modeled as random vectors with elements being independent and identically distributed (i.i.d.) and complex Gaussian distribution with zero mean and variance equal to σ^2 .

Since mmWave channels are expected to have limited scattering on the BS side, a geometric channel model is adopted where each cluster of scatterers is assumed to contribute multiple propagation paths between the BS and device k . Under this model, the physical channel matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$

between the BS and each device can be given as [40]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L_{AoD} L_{AoA}}} \sum_{p=1}^{L_{AoD}} \sum_{q=1}^{L_{AoA}} \alpha_{p,q} \mathbf{a}_r(\theta_{p,q}) \mathbf{a}_t^H(\phi_p), \quad (2)$$

where L_{AoD} represents the number of AoDs, each of which contributes L_{AoA} arrival paths on the device side, $\alpha_{p,q}$ is the complex gain of the q -th path on the device side that departs from the p path on the BS side, ϕ_p , $\theta_{p,q}$ are the p -th AoD and its corresponding q -th AoA, and $\mathbf{a}_t(\phi_p)$, $\mathbf{a}_r(\theta_{p,q})$ are the antenna array response vectors at the BS and devices, respectively. The matrix form of the channel model in (2) can be written as

$$\mathbf{H} = \mathbf{A}_r \mathbf{H}_a \mathbf{A}_t^H \quad (3)$$

where $\mathbf{A}_t = [\mathbf{a}_t(\phi_1), \mathbf{a}_t(\phi_2), \dots, \mathbf{a}_t(\phi_{L_{AoD}})]$, $\mathbf{A}_r = [\mathbf{a}_r(\theta_1), \mathbf{a}_r(\theta_2), \dots, \mathbf{a}_r(\theta_{L_{AoA}})]$, $\mathbf{H}_a \in \mathbb{C}^{L_{AoA} \times L_{AoD}}$ is a sparse matrix whose non-zero complex values are associated with the channel path gains.

Assuming Uniform Linear Arrays (ULAs), the transmit response vector $\mathbf{a}_t(\phi_p)$ is defined as

$$\frac{1}{\sqrt{N_t}} \left[1 \quad e^{j \frac{2d\pi}{\lambda} \cos \phi_p} \quad \dots \quad e^{j \frac{2(N_t-1)d\pi}{\lambda} \cos \phi_p} \right]^T, \quad (4)$$

while the receive response vector $\mathbf{a}_r(\theta_{p,q})$ is given as

$$\frac{1}{\sqrt{N_r}} \left[1 \quad e^{j \frac{2d\pi}{\lambda} \cos \theta_{p,q}} \quad \dots \quad e^{j \frac{2(N_r-1)d\pi}{\lambda} \cos \theta_{p,q}} \right]^T, \quad (5)$$

where λ is the signal wavelength at mmWave frequency and d the inter-element spacing.

Since the quality of CS-based channel estimation is limited by the quantization error in the angular domain representation, the AoDs and AoAs (or the corresponding atoms) are taken from high-density grids of $G_r \gg L_{AoA}$ and $G_t \gg L_{AoD}$ points, obtaining overcomplete dictionaries $\mathbf{D}_t \in \mathbb{C}^{N_t \times G_t}$, $\mathbf{D}_r \in \mathbb{C}^{N_r \times G_r}$ [40], correspondingly. Then, the channel matrix \mathbf{H} can be represented in terms of a non-diagonal sparse matrix $\mathbf{H}^v \in \mathbb{C}^{G_r \times G_t}$, containing the path gains of the quantized angles, as follows

$$\mathbf{H} = \mathbf{D}_r \mathbf{H}^v \mathbf{D}_t^H. \quad (6)$$

The representation in (6) constitutes a discretized approximation of the channel matrix that can be exploited to cast the estimation of \mathbf{H} to that of detecting some non-zero coefficients in the sparse channel matrix \mathbf{H}^v . Focusing on the rows of the channel matrix \mathbf{H}^v , they will contain only a few non-zero values corresponding to the AoDs, due to the sparsity of mmWave channels on the BS side. Notice that the number of non-zero values in each of those columns is exactly equal to L_{AoA} .

In the special case of ULAs, with $d = \frac{\lambda}{2}$ and quantized AoDs and AoAs selected from uniform grids of resolution equal to transmitting and receiving antennas ($G_t = N_t$ and $G_r = N_r$), respectively, the dictionaries $\mathbf{D}_r \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{D}_t \in \mathbb{C}^{N_t \times N_t}$ result in unitary Discrete Fourier Transform (DFT) matrices. This case is known as the *Virtual Channel Model* [41].

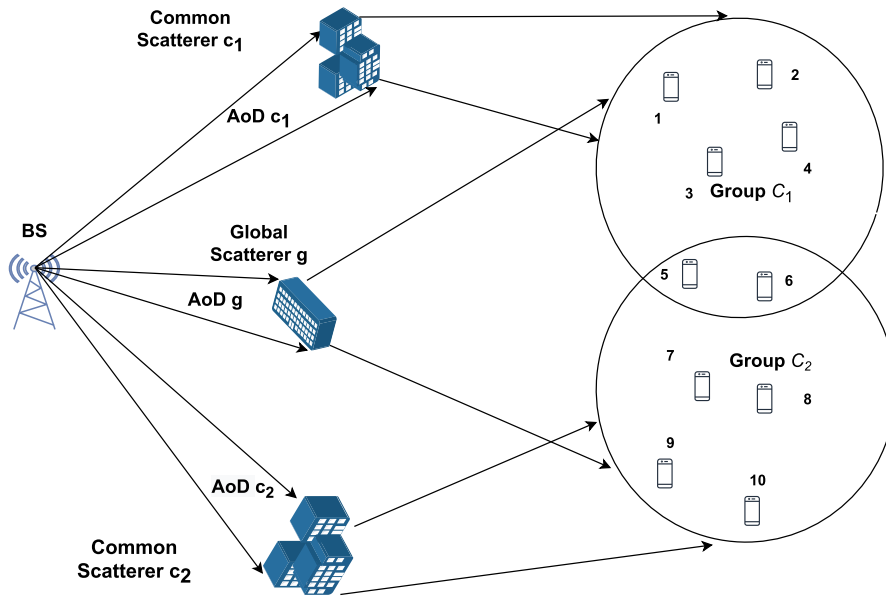


FIGURE 1. Downlink transmission scenario: $C_1 = \{1, 2, 3, 4, 5, 6\}$, $C_2 = \{5, 6, 7, 8, 9, 10\}$ [33].

III. JOINT SPARSE CHANNEL MODELING

In this section, the adopted joint sparsity model for the involved channels will be presented. This model captures the different tasks in the network, which in [42] is called multi-tasking with overlapping parameters.

A set of K geospatially distributed devices are interconnected by a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, K\}$ is the set of devices (or nodes) and $\mathcal{E} = \{(i, j) : \text{Node } i \in \mathcal{V} \text{ is connected with node } j \in \mathcal{V}\}$.

As illustrated in Figure 1, the scattering nature of the transmission environment may lead to the devices grouping according to the scatterers getting involved in their channels. As stated in [8], the devices that belong to the same group may face and thus be affected by the same scatterer. It means that the sparse channel matrices \mathbf{H}_k^v share the same sparsity support set on the BS side, but the corresponding path gains may differ due to the surrounding area of each device. To improve the downlink channel estimation performance, we exploit the previously described relationship on the sparsity support sets among the involved channels of the devices [8], [17].

More specifically, the scatterers on the BS side can be clustered into global and common, from now on referred to as g and c_j , which translates the corresponding AoDs into global to all devices in the network and common to particular subgroups. From this observation, we conclude that both the columns (otherwise atoms) of the employed dictionary \mathbf{D}_t are also global and common, respectively. Following [28], the physical channel matrix structural sparsity is discriminated into

- 1) **Sparsity Support within Channel Matrix:** The rows within \mathbf{H}_k^v usually have the same column sparsity support, namely there exists support set \mathcal{S}_k in \mathbf{H}_k^v , such that $\mathcal{S}_k = \text{supp}(\mathbf{H}_k^v)$, $|\mathcal{S}_k| = L_{AoD}$.
- 2) **Global and Multiple Common Supports between Different Channel Matrices:** The channel matrices

of different devices are correlated, especially when the devices are physically close to each other. Based on this condition, different devices and their respective (sparse) channel matrices tend to share some global and, partially, common local scatterers on the BS part.

Based on the previously established sparsity cases, the sparse representation matrix \mathbf{H}_k^v of device k can be written as

$$\mathbf{H}_k^v = \mathbf{H}_k^g + \sum_{j \in I_k} \mathbf{H}_k^{c_j}, \quad (7)$$

where \mathbf{H}_k^g represents the sparse matrix whose row support set is globally shared among all devices in the network and $\mathbf{H}_k^{c_j}$ captures the sparse matrix whose support set is commonly shared among different subgroups of devices in the network. Also, I_k is a set of indices $j = \{1, 2, \dots, J\}$ indicating the maximum number of common interest support sets per device k .

Moreover, it is assumed that the rows of \mathbf{H}_k^g and $\mathbf{H}_k^{c_j}$ s of each device k share L_g non-zero columns and L_{c_j} non-zero columns, respectively. So, the sparsity pattern of a device k is divided into global and multiple common interest support sets. The column sparsity is induced by the assumption that the resolution of the departing angular grid points is much greater than the number of the active transmission paths, i.e., $L_{AoD} \ll G_t$.

Considering a scenario of multiple devices, the AoDs of the devices' channels impacted by the same scattering clusters are similar [8], [14]. It means that they share the same non-zero columns (i.e., global and common, respectively) in their sparse channels, as shown in Figure 2, which motivated us to investigate the channel estimation via a distributed algorithm that estimates the shared support sets in a collaborative manner. In the following, some definitions will be provided.

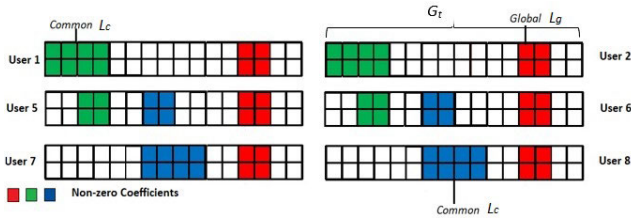


FIGURE 2. Sparse Channel matrices of the sided devices $\bar{\mathbf{H}}_1^v, \bar{\mathbf{H}}_2^v, \bar{\mathbf{H}}_7^v, \bar{\mathbf{H}}_8^v$ and devices in the overlapping area $\bar{\mathbf{H}}_5^v, \bar{\mathbf{H}}_6^v$ assuming $N_r = 2, G_r = N_r$ and $G_t = 256$ where red, green and blue bins stand for the global and partially common atoms with total sparsity level $L_{AoD} = 6$.

$\Omega = \{1, 2, \dots, G_t\}$ denotes the set of atom indices (namely, the number of AoDs) in the overcomplete dictionary \mathbf{D}_t .

Definition 1 (Global Interest Support Set): Let the sparse representation matrix \mathbf{H}_k^g , whose columns are globally shared with only L_g nonzero entries. The global interest support set is defined as

$$\mathcal{S}^g = \{i \in \Omega, s.t. \text{supp}(\mathbf{H}_m^g) = \text{supp}(\mathbf{H}_n^g), \forall m, n \in \mathcal{V}\}, \quad (8)$$

where $\mathcal{S}^g \subset \Omega$ and $|\mathcal{S}^g| = L_g$ (l_0 sparsity).

Definition 2 (Common Interest Support Set): Let the sparse representation matrix $\mathbf{H}_k^{c_j}$ with only L_{c_j} nonzero entries. The common interest support set is defined as

$$\mathcal{S}^{c_j} = \{i \in \Omega, s.t. \text{supp}(\mathbf{H}_m^{c_j}) = \text{supp}(\mathbf{H}_n^{c_j}), \forall m, n \in \mathcal{C}_j\}, \quad (9)$$

where $\mathcal{S}^{c_j} \subset \Omega$, $|\mathcal{S}^{c_j}| = L_{c_j}$ (l_0 sparsity) and $\mathcal{C}_j \subset \mathcal{V}$ a set of devices indices that are impacted by the common scattering cluster c_j , for $j = 1, 2, \dots, J$, and are concerned about \mathcal{S}^{c_j} .

Based on the previous definitions, the support set \mathcal{S}_k of device k is formulated as $\mathcal{S}_k = \mathcal{S}^g \cup \{\mathcal{S}_k^{c_j}\}$, for all $j \in \mathbf{I}_k$, meaning that the joint sparsity $L_{AoD,k} \leq L_g + \sum_{j \in \mathbf{I}_k} L_{c_j}$. Here, despite \mathcal{S}^g being globally shared among all devices in the network or \mathcal{S}^{c_j} being the same among particular devices in the same subgroup \mathcal{C}_j , it doesn't hold for the respective non-zero values of \mathbf{H}_k^g and $\mathbf{H}_k^{c_j}$ which remain individual and possibly independent among the devices.

Thus, taking into account the aforementioned sharing of support sets, the channel \mathbf{H}_k of device k can be written as

$$\begin{aligned} \mathbf{H}_k &= \mathbf{D}_r(\mathbf{H}_k^g + \sum_{j \in \mathbf{I}_k} \mathbf{H}_k^{c_j})\mathbf{D}_t^H \\ &= (\mathbf{D}_r\mathbf{H}_k^g + \sum_{j \in \mathbf{I}_k} \mathbf{D}_r\mathbf{H}_k^{c_j})\mathbf{D}_t^H \\ &= ((\bar{\mathbf{H}}_k^g)^H + \sum_{j \in \mathbf{I}_k} (\bar{\mathbf{H}}_k^{c_j})^H)\mathbf{D}_t^H \\ &= (\bar{\mathbf{H}}_k^v)^H \mathbf{D}_t^H \end{aligned} \quad (10)$$

From (10), we see that the column sparsity in \mathbf{H}_k^v translates into row sparsity in $(\bar{\mathbf{H}}_k^v)^H = \mathbf{D}_r\mathbf{H}_k^v$. According to the adopted formulation, \mathbf{D}_r is hidden in the sparse matrix $\bar{\mathbf{H}}_k^v$ which will be the matrix of interest in the proposed algorithm. Here, the columns of \mathbf{D}_r are fixed to L_{AoA} , i.e., as if we

had the receiver response matrix $\mathbf{A}_r \in \mathbb{C}^{N_r \times L_{AoA}}$, and the joint sparsity is only observed in the AoDs. In summary, considering (10), our aim is to exploit the joint sparsity sets of neighboring devices to estimate the sparse channel per device in a collaborative and distributed manner. Such an approach will be elaborated on in the following.

IV. MULTI-TASK MMWAVE CHANNEL ESTIMATION

In this section, the problem formulation for the multi-task mmWave channel estimation will be presented. Given the geometric channel model representation in (10), the problem of interest will be addressed as a sparse reconstruction based on compressed sensing, assuming an overcomplete \mathbf{D}_t for the AoDs sparsifying matrix.

Employing per device k the representation (10), we can rewrite the measurements model in (1), in compact form, as

$$\bar{\mathbf{Y}}_k = \Phi \bar{\mathbf{H}}_k^v + \bar{\mathbf{N}}_k, \quad (11)$$

where $\bar{\mathbf{Y}}_k = [\bar{y}_{k,1} | \bar{y}_{k,2} | \dots | \bar{y}_{k,N_r}] \in \mathbb{C}^{T \times N_r}$, $\Phi = \mathbf{X}^H \mathbf{D}_t \in \mathbb{C}^{T \times G_t}$ and $\bar{\mathbf{N}}_k = (\mathbf{N}_k)^H$. Also, each column of $\bar{\mathbf{Y}}_k$, which constitutes the Single Measurement Vectors (SMV) analogue, is captured by

$$\bar{y}_{k,r} = \Phi \bar{\mathbf{h}}_{k,r}^v + \bar{\mathbf{n}}_{k,r}, \quad r = 1, 2, \dots, N_r. \quad (12)$$

According to the Multiple Measurement Vectors (MMV) model [43], in (11), each device k exploits simultaneously the low dimensional measurement vectors, $\{\bar{y}_{k,r}\}_{r=1}^{N_r}$, from all antenna elements, that share the same measurement matrix Φ and sparsity support sets \mathcal{S}_k , to recover the high dimensional $G_t \times 1$ sparse channel vectors between all receiving antennas and BS.

Let $\{\bar{\mathbf{H}}_1^v, \bar{\mathbf{H}}_2^v, \dots, \bar{\mathbf{H}}_K^v\}$ be the collection of parameter matrices across the network. Through (11), the objective of each device k is to determine the sparse AoDs, which is equivalent to estimating the rows support set of $\bar{\mathbf{H}}_k^v \in \mathbb{C}^{G_t \times N_r}$. Each device k , ignoring the joint sparsity channel modelling, could minimize $\mathcal{J}_k(\bar{\mathbf{H}}_k^v) = \|\bar{\mathbf{Y}}_k - \Phi \bar{\mathbf{H}}_k^v\|_F^2$ on its own, locally (without cooperation) or the aggregated optimization function $\sum_{k=1}^K \mathcal{J}_k(\bar{\mathbf{H}}_k^v)$ could be used in a centralized manner, as in [20]. However, in this paper, we follow a third direction that leads to a fully distributed approach. More specifically, by taking into account the joint row sparsity between the $\bar{\mathbf{H}}_k^v$ s of different devices, individual but related cost functions $\mathcal{J}_k(\bar{\mathbf{H}}_k^v)$ can be adopted, thus, allowing us to formulate the problem as a multi-task one [42]. The relatedness lies in the fact that the support set of a device k is $\mathcal{S}_k = \mathcal{S}^g \cup \{\mathcal{S}_k^{c_j}\}$, e.g. with global and partially common elements. The common elements with the \mathcal{S}_l s of its neighborhood \mathcal{L}_k^{in} will be taken into consideration in the collaboration step of the proposed approach, where the (joint) support set is estimated through adaptive weights voting. Hence, the optimization problem can be formulated locally as

$$\text{minimize}_{\mathcal{S}_k, \bar{\mathbf{H}}_k^v} \mathcal{J}_k(\bar{\mathbf{H}}_k^v) \quad (13)$$

$$\text{subject to } \mathcal{S}_k \cap \mathcal{S}_l = \mathcal{S}^g \cup \mathcal{S}_k^{c_j}, \quad \forall l \in \mathcal{L}_k^{in} \quad (14)$$

$$|\mathcal{S}_k| = \|\text{diag}((\bar{\mathbf{H}}_k^v)^H \bar{\mathbf{H}}_k^v)\|_0 = L_{AoD} \quad (15)$$

Accurate retrieval of the sparse channel matrix $\bar{\mathbf{H}}_k^v$ is only guaranteed when the measurement matrix Φ satisfies specific properties from CS like the Restricted Isometry Property (RIP) and mutual coherence. As this matrix is dependent on both the training symbols and the transmit (overcomplete) sparsifying matrix \mathbf{D}_t , the symbols can be properly designed to acquire a more optimized measurement matrix to ensure that the channel is reconstructed successfully with a high probability. Such issues have been addressed, for example, in [16] and [44], but, here, the symbols are selected randomly, which still makes the measurement matrix “well-behaved”.

V. VOTING-BASED MULTI-TASK DISTRIBUTED SOLUTION

In this section, the main stages of the proposed algorithm, WDiSOMP, are analysed. Since the dictionary \mathbf{D}_t in (10) is predefined, the channel estimation is equivalent to the estimation of $\bar{\mathbf{H}}_k^v$ in (11), assuming, in the transmission environment, the existence of sparsity division captured in (7) or equivalently in (10). More specifically, WDiSOMP was designed to estimate the sparse channel in a collaborative distributed manner. The collaboration stage concerns only the support sets \mathcal{S}_k s, while the non-zero values of the path gains are individual and, thus, estimated via Least-Squares (LS) locally. Our solution employs a weighted majority voting mechanism with adaptive weights that can be estimated either at the support set or index level. With the incorporation of this step, the proposed algorithm only needs to know the total number of departing paths or angles L_{AoD} .

A. PROPOSED SOLUTION

The algorithm operates in stages (see Algorithm 1). The first stage initializes the process, and the second one is executed iteratively, consisting of a collaborative (through voting) and a local step (where the SOMP runs to make an individual estimate of the sparsity support set) and the final one that concerns channel estimation is executed locally.

In the *Initialization* stage, the BS, for T time slots, transmits to all devices in the network a number of training symbols, represented by matrix \mathbf{X} . Then, each device k collects all measurements in \mathbf{Y}_k and executes SOMP (assuming the overcomplete sparsifying matrix Φ and an empty support set) to get an initial local estimate of its complete support set $\hat{\mathcal{S}}_k$.

In the *Iterative* stage, each device k , $\forall k$, transmits its $\hat{\mathcal{S}}_k$ to its neighbors denoted by the set $\mathcal{L}_k^{\text{out}}$, and receives the estimates $\hat{\mathcal{S}}_l$'s from its neighbors, $\forall l \in \mathcal{L}_k^{\text{in}}$ (*communication step*). At iteration i of the algorithm, each device exploits the received $\hat{\mathcal{S}}_l$'s and engages a *weighted majority voting* mechanism to determine the i most voted indices. Then, SOMP is executed locally to acquire the remaining non-zero indices, up to L_{AoD} , and again $\hat{\mathcal{S}}_k$ s are communicated over the network. These steps are repeated for L_{AoD} iterations. It is assumed, here, that all devices communicate via a connected network and can exchange their support sets indices by engaging in a suitable D2D communication protocol [45]).

Each device k exploits the estimated set $\hat{\mathcal{S}}_k$ on (11) to retain only these specific columns of Φ , apply LS to find the associated non-zero values of $\hat{\mathbf{H}}_k^v$ and through (10) its channel matrix.

The above stages and the corresponding steps are also illustrated in Figure 3, where the information flow is depicted per device k . The BS broadcasts the same training data to all devices in the network. Then, each device exploits the acquired measurements to initialize the process and iteratively (via sharing - collaboration) estimate the desired support set. Finally, each device k utilizes the estimated support set to recover the sparse channel representation and, thus, the channel matrix.

Algorithm 1: WDiSOMP

Result: $\{\hat{\mathbf{H}}_k\}$

Input: $\Phi = \mathbf{X}^H \mathbf{D}_t$, $\bar{\mathbf{Y}}_k = (\mathbf{Y}_k)^H$; L_{AoD} , $b_{lk,-1} = 0$, $\forall k \in \{1, \dots, K\}$ and $l \in \mathcal{L}_k^{\text{in}}$;

Stage I: Initialization Part

Local step

for $k = 1:K$ **do**

$[\hat{\mathcal{S}}_{k,0}] = \text{SOMP}(\Phi, \bar{\mathbf{Y}}_k, L_{AoD}, \emptyset)$;

end

Transmit: $\hat{\mathcal{S}}_{k,0}$ to all nodes $\mathcal{L}_k^{\text{out}}$

Receive: $\hat{\mathcal{S}}_{l,0}$ s from all nodes $\mathcal{L}_k^{\text{in}}$

for $k=1:K$ **do**

Exploit $\hat{\mathcal{S}}_{l,0}$ s to estimate the weights $a_{lk,0}$

 Estimate $a_{lk,0}$ with either (17), (18) or (19), (20),

$\forall l \in \mathcal{L}_k^{\text{in}}$

end

Stage II: Iterative Part

for $i = 1 : L_{AoD}$ **do**

for $k=1:K$ **do**

Weighted Majority Voting

$\hat{\mathcal{S}}_{k,v} = \text{WMVoting}(\{\hat{\mathcal{S}}_{l,i-1}\}_{l \in \mathcal{L}_k^{\text{in}}, i}, G_t, a_{lk,i-1})$;

Local Step

$[\hat{\mathcal{S}}_{k,i}] = \text{SOMP}(\Phi, \bar{\mathbf{Y}}_k, L_{AoD} - i, \hat{\mathcal{S}}_{k,v})$;

end

Transmit: $\hat{\mathcal{S}}_{k,i}$ to all nodes $\mathcal{L}_k^{\text{out}}$

Receive: $\hat{\mathcal{S}}_{l,i}$ s from all nodes $\mathcal{L}_k^{\text{in}}$

Weights $a_{lk,i}$ *Estimation*

for $k=1:K$ **do**

 Estimate $a_{lk,i}$ with either (17), (18) or (19), (20),

$\forall l \in \mathcal{L}_k^{\text{in}}$

end

end

Stage III: Final

LS: $\hat{\mathbf{H}}_k^v = \Phi^\dagger(:, \hat{\mathcal{S}}_k) \bar{\mathbf{Y}}_k$;

$\hat{\mathbf{H}}_k = (\hat{\mathbf{H}}_k^v)^H \mathbf{D}_t^H$, $\forall k = 1, \dots, K$

B. ADAPTIVE WEIGHTS VOTING MECHANISM

The aim of adopting a weighted voting mechanism was to make the algorithm WDiSOMP robust to different conditions.

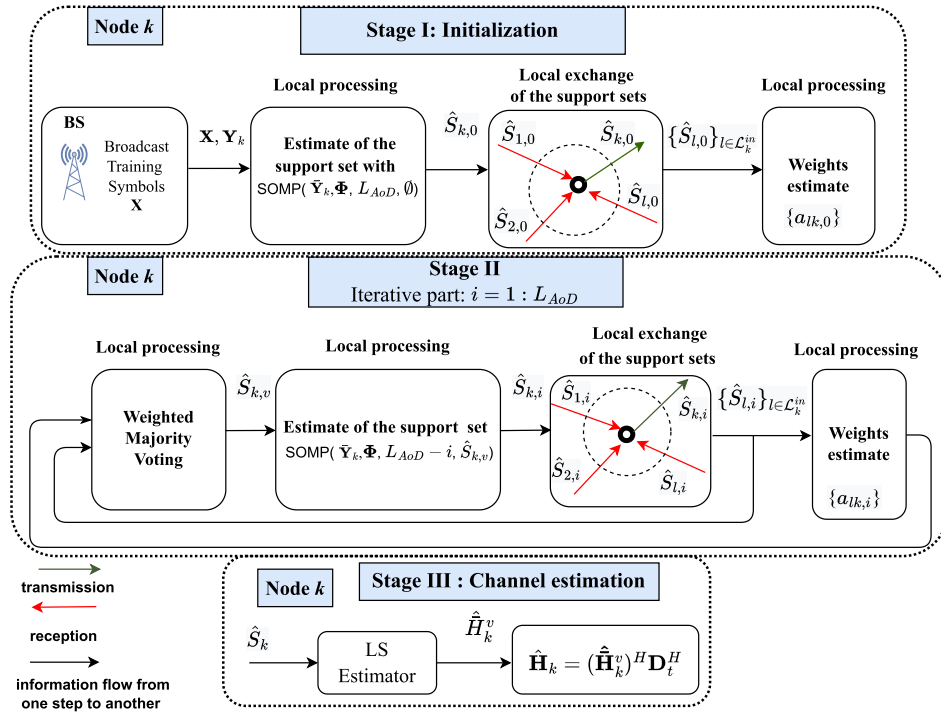


FIGURE 3. Local information sharing and processing per device k .

In particular, this step (step 1 of stage II) was motivated by the fact that a device may not have the same impact on the support set retrieval, which can happen for several reasons. First, even if the devices have the same sparsity profile, they may face different communication conditions (i.e., noise level). Also, different common support sets among device groups may exist on the unknown sparsity patterns. Finally, the (hidden) sparsity order of the support sets and the grouping of devices are not usually known. For these reasons, two weighting methods are selected in the majority voting mechanism. The first method can cope with these cases with satisfactory performance [33] except for a case [34] where part of the support set of a device in the network is partially shared with different subgroups of devices, such that all devices' support sets have the same cardinality. For this purpose, the second method is suggested.

Algorithm 2: WMVoting

Result: $\hat{S}_{k,v}$
 $i, \{\hat{S}_{l,i}, l \in \mathcal{L}_k^{\text{in}}, \mathbf{z}_k = \mathbf{0}_{G_t \times 1}, \{a_{lk,i}\};$
for $l \in \mathcal{L}_k^{\text{in}}$ **do**
 for $j \in \hat{S}_{l,i}$ **do**
 $\mathbf{z}_k(j) = \mathbf{z}_k(j) + a_{lk,i}(j);$
 end
end
 $\hat{S}_{k,v} = \text{bestindices}(\mathbf{z}_k, i)$

More specifically, at each iteration i of the algorithm, device k employs a zero elements vector $\mathbf{z}_{k,i}$ of size G_t and

activates the positions indexed by each received $\hat{S}_{l,i}$, adding the weight a_{lk} at the relevant positions for all $l \in \mathcal{L}_k^{\text{in}}$. Afterwards, the device k , through function 'bestindices', finds the set of indices corresponding to the i largest amplitude components of the final \mathbf{z}_k . These steps are captured in Algorithm 2. The weights a_{lk} may be updated assuming either of the two following adaptive weights methods similar to the one in [46].

1) WEIGHT METHOD BASED ON SUPPORT SET AS SINGLE ENTITY

We introduce the instantaneous metric

$$b_{lk} \triangleq |\hat{S}_{l,i} \setminus \hat{S}_{k,i-1}| \quad (16)$$

that measures the number of different elements in the involved sets. We further smooth the quantity $b_{lk,i}$ through a first-order filter, say,

$$b_{lk,i} = (1 - \nu)b_{lk,i-1} + \nu|\hat{S}_{l,i} \setminus \hat{S}_{k,i-1}| \quad (17)$$

where $\nu \in (0, 1)$ is a small positive factor. Then, the combination weight assigned for each index j by device k to device l is approximated by

$$a_{lk,i}(j) \approx \frac{b_{lk,i}^{-1}}{\sum_{m \in \mathcal{L}_k^{\text{in}}} b_{mk,i}^{-1}}, \quad l \in \mathcal{L}_k^{\text{in}}, k \in K, \forall j \in \hat{S}_l. \quad (18)$$

At this point, it should be noted that the total number of common indices could be measured to determine the $b_{lk,i}$ s and the corresponding weight $a_{lk,i}$. In either case, all indices of a neighbour are graded with the same weight assigning the same significance to all indices even if some of them

aren't included in the device's support set. This mechanism can efficiently identify devices' groups with i) similar sparsity patterns and different noise levels (per group) and ii) non-overlapping but common sparsity patterns inside the groups [33]. To manage the existence of overlapping sparsity profiles, like the one depicted in Figure 1, we assumed a per index weighting method in the majority voting mechanism.

2) WEIGHT METHOD BASED ON PER-INDEX SUPPORT SET SIMILARITIES

Here, our aim is to give an individual weight for each index of the support set (L_{AoD} in number) of all devices. To achieve this, each device k votes with '1' the support set indices that belong to its neighbor l and its own support set. Then, a candidate index approved by the majority of neighboring devices wins. At each iteration i , each device k builds a matrix $\mathbf{Z}_k \in \mathbb{R}^{G_r \times |\mathcal{L}_k^{in}|}$ (initialized with zero values), which is updated by ones only in positions dictated by $\hat{\mathcal{S}}_{l,i} \cap \hat{\mathcal{S}}_{k,i-1} \neq \emptyset$ and fed as input in (19). At iteration i , the weight that device k assigns to its neighbor l per support set index j is estimated as follows

$$b_{lk,i}(j) = (1 - v)b_{lk,i-1}(j) + vZ_k(j, l) \quad (19)$$

$$a_{lk,i}(j) \approx \frac{b_{lk,i}(j)}{\sum_{m \in \mathcal{L}_k^{in}} b_{mk,i}(j)} \quad (20)$$

for $l \in \mathcal{L}_k^{in}$, $k \in K$ and all $j \in \hat{\mathcal{S}}_k \cap \hat{\mathcal{S}}_l$.

The elaborated weight methods are exploited in the voting mechanism in Algorithm 1 to enhance the channel estimation performance. In the following section, we will analytically estimate the error bounds of the proposed estimator.

VI. PERFORMANCE EVALUATION

In this section, we analyze the performance of the proposed estimator from various aspects. More specifically, the analysis will concern i) the iterations of the algorithm, ii) the computational complexity, iii) the communication overhead of WDiSOMP compared to the rest considered schemes, iv) the channel estimation error in terms of the support set reconstruction quality and, finally, v) the impact of angular grid resolution assuming an overcomplete dictionary to minimize the channel quantization error.

A. REQUIRED ITERATIONS BEFORE TERMINATION AND COMPUTATIONAL COMPLEXITY

In Algorithm 1, the iterations of WDiSOMP are fixed, constant by construction, and equal to L_{AoD} . Also, they show how many times the devices exchange their support set indices (communication step) in the network. The communication step includes local processing, which is highly dependent on the available energy of the device, especially for devices with limited resources (e.g., battery-dependent) to run an energy harvesting process. In this study, we assume that the devices satisfy the minimum power requirements to solve the assigned tasks. SOMP is executed locally at each iteration of WDiSOMP, exploiting the obtained part of the support after

voting to build an estimate of the full support set until its size is equal to L_{AoD} . The number of internal iterations of SOMP is exactly $L_{AoD} - i$ (since $|\hat{\mathcal{S}}_{k,voting}| = i$). As a result, in WDiSOMP running time, the total iterations of SOMP is $\sum_{i=1}^{L_{AoD}} (L_{AoD} - i) = \frac{L_{AoD}(L_{AoD}-1)}{2}$.

Moving on to the computational complexity of the proposed algorithm, we consider that operations like addition, subtraction, division and multiplication of any two numbers have computational complexity $O(1)$. Also, the complexity for two matrices multiplication of size $M \times N$ and $N \times P$ is $O(MPN)$.

Let us focus on communication round i . As we see in Figure 3, the WDiSOMP algorithm executed locally at each device k has two specific processing steps that actually contribute to the overall computational complexity. In particular, first, the complexity of SOMP is on the order of $O(N_r G_r T s_i)$ [47], where $s_i = L_{AoD} - i$ is the size of the support set that is estimated at iteration i . Second, the suggested voting mechanism contributes a small cost to the complexity of the algorithm due to the adaptive weights estimate. Both weighting methods should first compute the involved $b_{lk,i}$ s. In the first method, the cost mainly derives from the fact that each device k should measure the number of different indices between its support set and one of its neighbors ($|\hat{\mathcal{S}}_{l,i} \setminus \hat{\mathcal{S}}_{k,i-1}|$). In the second, per index method, the respective cost stems from i) the intersection operation ($\hat{\mathcal{S}}_{l,i} \cap \hat{\mathcal{S}}_{k,i-1}$) among the support set of device k and its neighbors, and ii) the term $vZ_k(j, l)$, $j = 1, 2, \dots, L_{AoD}$ whose activation depends on the previous step. In either case, the cost is in the order of the support set size or, else, sparsity level, i.e., $O(L_{AoD})$. After this sub-step, the cost of estimating the desired $a_{lk,i}$ for each one of its $K - 1$ neighbors (where $K - 1$ captures the $|\mathcal{L}_k^{in}|$) and itself involves simple operations, $K - 1$ sums (denominator) and 1 division, thus the cost is $O(K)$. In the former method, the weights $a_{lk,i}$ are the same for all indices, while the latter estimates a weight per index.

Having available the weights $a_{lk,i}$, the computational complexity of the majority voting (Algorithm 2) is determined by: i) the number of additions to the zero-elements vector \mathbf{z}_k , which are equal to L_{AoD} (the support set size or sparsity level) for each of the $K - 1$ neighbors and the device itself, or, else, KL_{AoD} in total, and ii) the sorting method used in the function 'bestindices' to find the i most voted indices. In our case, the number of elements to be sorted is small and equal to the support set level L_{AoD} .

Finally, at Stage III of Algorithm 1, each device k uses an LS estimator to acquire the non-zero involved coefficients (i.e., the channel gains). This stage adds a cost which depends on the computation of the pseudo-inverse of matrix Φ , keeping only that columns dictated by $\hat{\mathcal{S}}_k$, i.e. $O(L_{AoD}T^2 + L_{AoD}^3)$, and then by its multiplication with the measurements' matrix imposing an additional cost of $O(L_{AoD}TN_r)$.

It is noted here that, apart from the voting mechanism, the complexity of the other two local processing methods (namely, SOMP and LS) depends on the adopted algorithms. As this paper aims to study, for the first time,

a fully-distributed channel estimation algorithm enhanced by carefully designed voting mechanisms that improve the cooperation among the devices, two well-known simple algorithms have been chosen. However, it should be clear that other candidates with lower complexity can also be selected.

Before moving on to the next section, let us compare the complexity of the proposed algorithm with the ones that will be compared in the simulations section. WDiSOMP has a slightly higher complexity than DiSOMP due to the employed voting mechanism. However, as demonstrated in the experiments, its performance is greatly improved, while there is no requirement for knowing beforehand the sparsity structure of the involved channels. As for the locally executed SOMP and the centralized SSOMP (namely, the one executed after all data are collected to a central point), the computational complexity is on the order of $O(N_r G_t T L_{AoD})$ and $O(K N_r G_t T L_{AoD})$, respectively. Again, as it will be demonstrated, their performance is worse.

B. TRANSMISSION OVERHEAD

In this section, we compare the transmission overhead of the distributed schemes DiSOMP and WDiSOMP with the centralized scheme, SSOMP, in terms of the number of transmitted bits, denoted as \mathcal{T}_o^d and \mathcal{T}_o^c , respectively. In the distributed schemes, \mathcal{T}_o^d captures the amount of information exchanged between a device k and each neighbor. In the centralized method, \mathcal{T}_o^c stands for the communication overhead between BS and each device k . The transmitted bits are analytically evaluated for all algorithms, as follows.

In the case of the centralized solution SSOMP, we assume that each device k transmits to the BS separately the real and complex part of the measurements in $\bar{\mathbf{Y}}_k \in \mathbb{C}^{T \times N_r}$, assigning q bits per value, which corresponds to $2N_r T q$ bits from each device to BS. Then the BS, by employing SSOMP, utilizes the measurements from all devices to estimate the joint support set, whose indices are fed back to the devices delivering $L_{AoD} \lceil \log_2(G_t) \rceil$ bits. Hence, the total number of bits amounts to

$$\mathcal{T}_o^c = K(2N_r T q + L_{AoD} \lceil \log_2(G_t) \rceil). \quad (21)$$

Moving on to the distributed algorithms, let us, first, assume a network topology in which each node k transmits to its \mathcal{L}_k^{out} neighbors the L_{AoD} indices of the candidate support set derived by applying SOMP locally. In (W)DiSOMP, only L_{AoD} indices from $\{1, \dots, G_t\}$ (where G_t denotes the number of atoms) are transmitted, and each index is supposed to be encoded with $\lceil \log_2(G_t) \rceil$ bits, which entails $L_{AoD} \lceil \log_2(G_t) \rceil$ transmitted bits. Each device transmits in each iteration the full support set to its neighbors, and this process is repeated L_{AoD} times until the full sparsity level is achieved. Given that the proposed approach is fully distributed, $|\mathcal{L}_k^{out}| = K - 1$, the total transferred bits amount to

$$\mathcal{T}_o^d = K(K - 1) \lceil \log_2(G_t) \rceil L_{AoD}^2. \quad (22)$$

From (22), it is shown that the transmission overhead of the distributed schemes is a quadratic function of the sparsity parameter L_{AoD} , considering the rest parameters constant.

TABLE 2. Transmission overhead in Kbits of centralized SSOMP for $K = 10$, $G_t = 512$, $T = 30$, $|\mathcal{L}_k| = K - 1$, $L_{AoD} = 6$.

	$N_r = 1$	$N_r = 2$	$N_r = 4$	$N_r = 8$
$q = 8$	5.34	10.14	19.74	38.94
$q = 16$	10.14	19.74	38.94	77.34
$q = 32$	19.74	38.94	77.34	154.14
$q = 64$	38.94	77.34	154.14	307.74

In Table 2, we show the \mathcal{T}_o^c in terms of the feedback quantization bits q , for $N_r = \{1, 2, 4, 8\}$. It is observed that the \mathcal{T}_o^c cost is significantly impacted by the increase in receiving antennas and feedback quantization bits. In Figure 4, we capture the impact of training symbols T in the \mathcal{T}_o^c , \mathcal{T}_o^d of SSOMP and WDiSOMP, respectively, assuming devices of single, two and four antennas. We have set the resolution of the AoDs dictionary (e.g., the number of atoms) equal to $G_t = 512 (\gg N_t = 128)$. The communication overhead of the distributed solutions is independent of the receiving antennas and thus kept at a constant level. Also, it is observed that the distributed solutions prevail for $N_r > 2$. These parameters only affect the complexity of the centralized solution, making the required communication cost prohibitive when the involved devices have many antenna elements.

TABLE 3. Transmission overhead in Kbits of (W)DiSOMP and centralized SSOMP for $K = 10$, $N_r = 2$, $G_t = \{256, 512, 1024\}$, $T = 30$, $|\mathcal{L}_k^{out}| = K - 1$, $L_{AoD} = 6$.

G_t	256		512		1024	
	(W)DiSOMP	SSOMP	(W)DiSOMP	SSOMP	(W)DiSOMP	SSOMP
$q = 8$	25.92	10.08	29.16	10.14	32.40	10.20
$q = 16$	25.92	19.68	29.16	19.74	32.40	19.80
$q = 32$	25.92	38.88	29.16	38.94	32.40	39.00
$q = 64$	25.92	77.28	29.16	77.34	32.40	77.40

Moreover, in Table 3, the centralized and distributed solutions are assessed in terms of the atoms G_t for various q . These outcomes show a subtle increase in the overhead of SSOMP and a higher increase in the one of (W)DiSOMP. In Table 4, the overhead of the distributed solutions is evaluated in terms of the sparsity level, selecting proper L_{AoD} compatible for mmWave systems [17], [19], [48] and setting fixed values at the rest parameters. The same table presents the performance of the centralized solution. Although the increase in the number of receiving antennas is accompanied by an essential increase in \mathcal{T}_o^c , this isn't verified with the G_t and L_{AoD} that are involved in the second term of (21).

From the above, we conclude that the distributed approaches, such as the one presented here, might be preferable in many cases encountered in practice (e.g., cases with low sparsity and/or non-single antenna users).

C. IMPACT OF SUPPORT SET ON RECONSTRUCTION QUALITY

The greedy approaches are based on the LS solution to estimate the complex values corresponding to the estimated

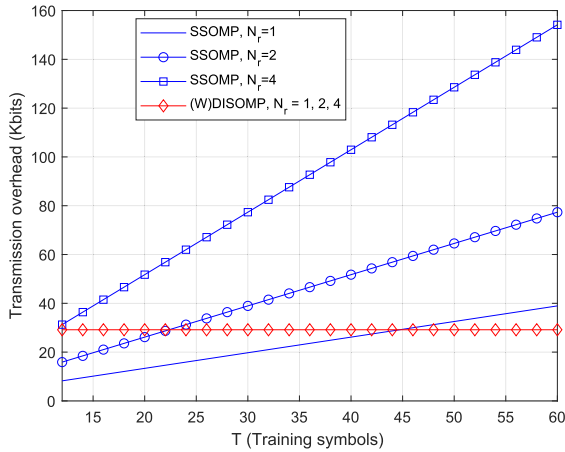


FIGURE 4. Transmission overhead in Kbits of SSOMP, (W)DiSOMP for $K = 10$, $T = [12 : 60]$, $L_{AoD} = 6$, $N_r = \{1, 2, 4\}$, $G_t = 512$, $q = 32$ bits.

TABLE 4. Transmission overhead in Kbits of SSOMP, (W)DiSOMP for $K = 10$, $T = 30$, $L_{AoD} = [5 : 10]$, $N_r = 2$, $G_t = 512$, $q = 32$ bits.

L_{AoD}	5	6	7	8	9	10
(W)DiSOMP	20.25	29.16	39.69	51.84	65.61	81.00
SSOMP $N_r = 2$	38.85	38.94	39.03	39.12	39.21	39.30
SSOMP $N_r = 4$	77.25	77.34	77.43	77.52	77.61	77.70

support set. Their accuracy may deteriorate if part of the support set isn't correctly identified. To evaluate the proposed CS-based channel estimator and determine a lower error bound, we formulate the problem for the support set reconstruction assuming an oracle estimator, which considers perfect knowledge of the true AoDs.

First, we reform the measurements model assuming (3) to acquire the following

$$\bar{\mathbf{Y}}_k = \Phi_a \bar{\mathbf{H}}_a^k + \bar{\mathbf{N}}_k, \quad (23)$$

with $\bar{\mathbf{H}}_a^k = (\mathbf{A}_r \mathbf{H}_a^k)^H \in \mathbb{C}^{L_{AoD} \times N_r}$ and $\Phi_a = \mathbf{X}^H \mathbf{A}_t \in \mathbb{C}^{T \times L_{AoD}}$ the measurement matrix for the Least Squares (LS) problem. The problem in (23) is similar to the one in (11) if we keep the L_{AoD} known atoms (or, else, AoDs) involved in the channel, \mathbf{D}_t is reduced to \mathbf{A}_t (response matrix structure). Each column of (23) is captured by $\bar{\mathbf{y}}_{k,r}$ for $r = 1, 2, \dots, N_r$ and is written as

$$\bar{\mathbf{y}}_{k,r} = \Phi_a \bar{\mathbf{h}}_a^{k,r} + \bar{\mathbf{n}}_{k,r}, \quad (24)$$

with $\bar{\mathbf{h}}_a^{k,r} = \bar{\mathbf{H}}_a^k(:, r) \in \mathbb{C}^{L_{AoD} \times 1}$, $\bar{\mathbf{n}}_{k,r} = \bar{\mathbf{N}}_k(:, r) \in \mathbb{C}^{T \times 1}$. Given (24), the oracle LS approach estimates the L_{AoD} entries of $\bar{\mathbf{h}}_a^{k,r}$, requiring $\Phi_a^H \Phi_a$ be full rank, i.e., $T \geq L_{AoD}$ and assuming that AoDs are exactly known, by

$$\bar{\mathbf{h}}_a^{k,r,o} = (\Phi_a^H \Phi_a)^{-1} \Phi_a^H \bar{\mathbf{y}}_{k,r} \quad (25)$$

Then, the oracle estimate of \mathbf{H}_k , if in (10) we keep the L_{AoD} rows of $\bar{\mathbf{H}}_k^v$ and \mathbf{D}_t is substituted by \mathbf{A}_t , is given by

$$\mathbf{H}_k^o = (\bar{\mathbf{H}}_a^{k,o})^H \mathbf{A}_t^H \quad (26)$$

with $\bar{\mathbf{H}}_a^{k,o} = [\bar{\mathbf{h}}_a^{k,o,1}, \bar{\mathbf{h}}_a^{k,o,2}, \dots, \bar{\mathbf{h}}_a^{k,o,N_r}]$.

An accurate recovery of the L_{AoD} -row matrix $\bar{\mathbf{H}}_a^k$ is equivalent to simultaneously recovering N_r vectors $\{\bar{\mathbf{h}}_a^{k,r}\}_{r=1}^{N_r}$. Considering the oracle estimator, which estimates the channel gain under the assumption that \mathbf{A}_t is known, the MSE can be written as

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{H}_k - \mathbf{H}_k^o\|_F^2 \right] &= \mathbb{E} \left[\sum_{r=1}^{N_r} \|\mathbf{h}_{k,r} - \mathbf{h}_{k,r}^o\|_2^2 \right] \\ &= \mathbb{E} \left[\sum_{r=1}^{N_r} \|\mathbf{A}_t (\bar{\mathbf{h}}_a^{k,r} - \bar{\mathbf{h}}_a^{k,o,r})\|_2^2 \right] \\ &= \sum_{r=1}^{N_r} \mathbb{E} \left[\|\mathbf{A}_t \bar{\mathbf{h}}_e^{k,r}\|_2^2 \right], \end{aligned} \quad (27)$$

where $\bar{\mathbf{h}}_e^{k,r} \triangleq \bar{\mathbf{h}}_a^{k,r} - \bar{\mathbf{h}}_a^{k,o,r}$. Then, keeping the equivalent form of channel estimation error in (27), the MSE can be written as

$$\mathbb{E} \left[\|\mathbf{A}_t \bar{\mathbf{h}}_e^{k,r}\|_2^2 \right] \geq \lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t) \mathbb{E} \left[\|\bar{\mathbf{h}}_e^{k,r}\|_2^2 \right] \quad (28)$$

The inequality holds due to the Rayleigh-Ritz ratio [49].

Combining (24) with (25) and assuming that $\mathbb{E}[\bar{\mathbf{n}}_{k,r} \bar{\mathbf{n}}_{k,r}^H] = \sigma^2 \mathbf{I}_T$, $\bar{\mathbf{h}}_a^{k,o,r} = \bar{\mathbf{h}}_a^{k,r} - \mathbf{Q} \bar{\mathbf{n}}_{k,r}$, the $\mathbb{E}[\|\bar{\mathbf{h}}_e^{k,r}\|_2^2]$ can be rewritten as

$$\begin{aligned} \mathbb{E} \left[\|\bar{\mathbf{h}}_e^{k,r}\|_2^2 \right] &= \mathbb{E} \left[\|\mathbf{Q} \bar{\mathbf{n}}_{k,r}\|_2^2 \right] \\ &= \mathbb{E} \left[\text{Tr}(\mathbf{Q} \bar{\mathbf{n}}_{k,r} \bar{\mathbf{n}}_{k,r}^H \mathbf{Q}^H) \right] \\ &= \text{Tr}(\mathbf{Q} \mathbb{E}[\bar{\mathbf{n}}_{k,r} \bar{\mathbf{n}}_{k,r}^H] \mathbf{Q}^H) \\ &= \sigma^2 \text{Tr}(\mathbf{Q} \mathbf{Q}^H) \end{aligned} \quad (29)$$

with $\mathbf{Q} = (\Phi_a^H \Phi_a)^{-1} \Phi_a^H$. Since $\mathbf{Q} \mathbf{Q}^H = (\Phi_a^H \Phi_a)^{-1}$, the $\text{Tr}(\mathbf{Q} \mathbf{Q}^H)$ is further processed, exploiting the arithmetic and harmonic means [50] and trace properties, as follows

$$\begin{aligned} \text{Tr}(\mathbf{Q} \mathbf{Q}^H) &= \sum_{l=1}^{L_{AoD}} \lambda_l \left((\Phi_a^H \Phi_a)^{-1} \right) \\ &= L_{AoD} \left(\frac{1}{L_{AoD}} \sum_{l=1}^{L_{AoD}} \lambda_l \left((\Phi_a^H \Phi_a)^{-1} \right) \right) \\ &\geq \frac{(L_{AoD})^2}{\sum_{l=1}^{L_{AoD}} 1/\lambda_l \left((\Phi_a^H \Phi_a)^{-1} \right)} \\ &= \frac{(L_{AoD})^2}{\sum_{l=1}^{L_{AoD}} \lambda_l (\Phi_a^H \Phi_a)} \\ &= \frac{(L_{AoD})^2}{\text{Tr}(\Phi_a^H \Phi_a)} \end{aligned} \quad (30)$$

From (27)-(30), we conclude that the MSE is lower-bounded as

$$\mathbb{E} \left[\|\mathbf{H}_k - \mathbf{H}_k^o\|_F^2 \right] \geq N_r \lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t) \frac{\sigma^2 L_{AoD}^2}{\text{Tr}(\Phi_a^H \Phi_a)}. \quad (31)$$

Assuming an orthonormal basis, and specifically the DFT matrix, to construct the AoDs dictionary \mathbf{A}_t in Φ_a , it holds that $\mathbf{A}_t^H \mathbf{A}_t = \mathbf{I}_{L_{AoD}}$ and thus $\lambda_{\min}(\mathbf{A}_t^H \mathbf{A}_t) = 1$, simplifying (31) to

$$\mathbb{E} \left[\|\mathbf{H}_k - \mathbf{H}_k^o\|_F^2 \right] \geq N_r \frac{\sigma^2 L_{AoD}^2}{\text{Tr}(\Phi_a^H \Phi_a)}. \quad (32)$$

Denominator $\text{Tr}(\Phi_a^H \Phi_a)$ is equivalently written as $\sum_{i=1}^{L_{AoD}} \lambda_i(\Phi_a^H \Phi_a)$. To attain the lower bound, $\Phi_a^H \Phi_a$ should be a diagonal matrix with identical diagonal elements. Previous condition holds if and only if the eigenvalues λ_i 's are equal. In particular, if $\mathbf{X}\mathbf{X}^H = \mathbf{T}\mathbf{I}$ and exploiting $\mathbf{A}_t^H \mathbf{A}_t = \mathbf{I}_{L_{AoD}}$, $\text{Tr}(\Phi_a^H \Phi_a) = L_{AoD}T$ and therefore

$$\mathbb{E} \left[\|\mathbf{H}_k - \mathbf{H}_k^o\|_F^2 \right] \geq N_r \sigma^2 \frac{L_{AoD}}{T}. \quad (33)$$

The bound in (33) increases linearly with the sparsity level L_{AoD} and the receive antennas N_r , while it is inversely proportional to the number of pilots T .

The oracle LS exploits the exact support-AoDs information; thus, under the previous conditions, its MSE will be the lower bound of WDiSOMP and every other estimator, which ignores the true AoDs and identifies them through a learning process under noisy conditions. Furthermore, when part of the estimated support set isn't correct, there will be a term in the error that captures the impact of the incorrect indices. This term is expected to become powerful in high SNR conditions adding a systematic deviation from the real support or channel and a respective floor. As part of erroneous indices is shrunk, the floor will gradually disappear, and the estimators' error will tend towards the lower bound. Such an approach could lead to designing more effective weighting methods than the ones elaborated here, suitable for multi-task joint sparsity to correct the imposed error (e.g., due to quantization). Although this study is out of scope here, it constitutes a challenging task which would deserve further investigation.

D. IMPACT OF DICTIONARY RESOLUTION

The aim of this subsection is to study the impact of the dictionary resolution, exploiting its mutual coherence, which will help us determine an upper bound for the error of the quantized channel matrix $\hat{\mathbf{H}}$ in (6). The rationale behind the estimation of the upper-bound is to quantify it based on the quality of \mathbf{D}_t and Φ , which depends on the number of grid points G_t and is measured by the mutual coherence. Assuming an overcomplete dictionary \mathbf{D}_t to minimize the quantization error, the symbols matrix should be properly designed to ensure the successful retrieval of the sparse channel matrix $\hat{\mathbf{H}}_k^v$ by the SOMP (locally).

For this purpose, we exploit (10) to write the error, in an equivalent form, as follows

$$\begin{aligned} \|\hat{\mathbf{H}}_k - \mathbf{H}_k\|_F^2 &= \left\| \left(\hat{\mathbf{H}}_k^v \right)^H \mathbf{D}_t^H - \left(\bar{\mathbf{H}}_k^v \right)^H \mathbf{D}_t^H \right\|_F^2 \\ &= \left\| \left(\mathbf{D}_t \hat{\mathbf{H}}_k^v \right)^H - \left(\mathbf{D}_t \bar{\mathbf{H}}_k^v \right)^H \right\|_F^2 \end{aligned}$$

$$\begin{aligned} &= \left\| \mathbf{D}_t \left(\hat{\mathbf{H}}_k^v - \bar{\mathbf{H}}_k^v \right) \right\|_F^2 \\ &= \sum_{r=1}^{N_r} \left\| \mathbf{D}_t \left(\hat{\mathbf{h}}_{k,r}^v - \bar{\mathbf{h}}_{k,r}^v \right) \right\|_2^2, \end{aligned} \quad (34)$$

where $\bar{\mathbf{H}}_k^v = [\bar{\mathbf{h}}_{k,1}^v | \bar{\mathbf{h}}_{k,2}^v | \dots | \bar{\mathbf{h}}_{k,N_r}^v]$.

An upper bound for the quantized channel estimation error is derived, assuming that the supports of $\hat{\mathbf{h}}_{k,r}^v$ and $\bar{\mathbf{h}}_{k,r}^v$ are similar. Hence, the channel estimation error in (34), based on the Cauchy Schwarz inequality, satisfies

$$\begin{aligned} \|\hat{\mathbf{H}}_k - \mathbf{H}_k\|_F^2 &= \sum_{r=1}^{N_r} \left\| \left(\mathbf{D}_t \right)_{S_k} \left(\left(\hat{\mathbf{h}}_{k,r}^v \right)_{S_k} - \left(\bar{\mathbf{h}}_{k,r}^v \right)_{S_k} \right) \right\|_2^2 \\ &\leq \sum_{r=1}^{N_r} \left\| \left(\mathbf{D}_t \right)_{S_k} \right\|_2^2 \left\| \left(\hat{\mathbf{h}}_{k,r}^v \right)_{S_k} - \left(\bar{\mathbf{h}}_{k,r}^v \right)_{S_k} \right\|_2^2. \end{aligned} \quad (35)$$

First, we will focus to find an upper bound for the term $\left\| \left(\mathbf{D}_t \right)_{S_k} \right\|_2^2$. For this purpose, we define $\mathbf{G}_{S_k} \triangleq \left(\mathbf{D}_t \right)_{S_k}^H \left(\mathbf{D}_t \right)_{S_k}$. Then, the squared spectral norm satisfies $\left\| \left(\mathbf{D}_t \right)_{S_k} \right\|_2^2 = \lambda_{\max}(\mathbf{G}_{S_k})$. Focusing on matrix \mathbf{G}_{S_k} , otherwise, it is written as $\left(\mathbf{D}_t^H \mathbf{D}_t \right)_{S_k}$ and given that $\mathbf{D}_t = [\mathbf{a}_{t,1} | \dots | \mathbf{a}_{t,G_t}]$, $\mathbf{a}_{t,i} \in \mathbb{C}^{N_r \times 1}$ with exponential elements as in (4), is equal to

$$\begin{aligned} \mathbf{G}_{S_k} &= \begin{bmatrix} \mathbf{a}_{t,1}^H \mathbf{a}_{t,1} & \dots & \mathbf{a}_{t,1}^H \mathbf{a}_{t,G_t} \\ \vdots & \dots & \vdots \\ \mathbf{a}_{t,G_t}^H \mathbf{a}_{t,1} & \dots & \mathbf{a}_{t,G_t}^H \mathbf{a}_{t,G_t} \end{bmatrix}_{S_k} \\ &= \begin{bmatrix} 1 & \dots & \mathbf{a}_{t,1}^H \mathbf{a}_{t,G_t} \\ \vdots & \dots & \vdots \\ \mathbf{a}_{t,G_t}^H \mathbf{a}_{t,1} & \dots & 1 \end{bmatrix}_{S_k} \end{aligned} \quad (36)$$

Based on (36), $\mathbf{G}_{S_k} = \mathbf{I} + \tilde{\mathbf{G}}_{S_k}$ where

$$\tilde{\mathbf{G}}_{S_k} = \begin{bmatrix} 0 & \dots & \mathbf{a}_{t,1}^H \mathbf{a}_{t,G_t} \\ \vdots & \dots & \vdots \\ \mathbf{a}_{t,G_t}^H \mathbf{a}_{t,1} & \dots & 0 \end{bmatrix}_{S_k}. \quad (37)$$

Exploiting the properties of the mutual coherence, $\mu(\mathbf{G}_{S_k}) = \mu(\tilde{\mathbf{G}}_{S_k})$, thus the off-diagonal entries of $\tilde{\mathbf{G}}_{S_k}$ are upper-bounded by $\mu(\mathbf{D}_t)$. From this observation, all the off-diagonal entries of $\tilde{\mathbf{G}}_{S_k}$ can be replaced by the $\mu(\mathbf{D}_t)$, constituting $\tilde{\mathbf{G}}_{S_k}'$. Then, the Gershgorin circle theorem is utilized to find an upper-bound of the $\lambda_{\max}(\mathbf{I} + \tilde{\mathbf{G}}_{S_k})$. In particular, it holds that $\lambda_{\max}(\mathbf{G}_{S_k}) \leq \lambda_{\max}(\mathbf{I} + \tilde{\mathbf{G}}_{S_k}') \leq 1 + (L_{AoD} - 1)\mu(\mathbf{D}_t)$. From this inequality, (35) is rewritten as

$$\|\hat{\mathbf{H}}_k - \mathbf{H}_k\|_F^2 \leq \sum_{r=1}^{N_r} \bar{\mu} \cdot \left\| \hat{\mathbf{h}}_{k,r}^v - \bar{\mathbf{h}}_{k,r}^v \right\|_2^2, \quad (38)$$

where $\bar{\mu} = 1 + (L_{AoD} - 1)\mu(\mathbf{D}_t)$. In $\bar{\mu}$, the term $(L_{AoD} - 1)\mu(\mathbf{D}_t)$ denotes the imposed overhead in the sparse channel estimation error due to the discretization with resolution G_t higher than the transmitter antennas N_t .

It depends on the mutual coherence of the dictionary \mathbf{D}_t , which, in turn, depends on the selected resolution G_t . When $G_t = N_t$, $\mu(\mathbf{D}_t) = 0$, and $\bar{\mu} = 1$. However, for $G_t \gg N_t$, $\mu(\mathbf{D}_t)$ attains values close to 1, making $\bar{\mu}$ and thus the upper bound of the channel error in (38) on the order of L_{AoD} .

In (38), $\|\hat{\mathbf{h}}_{k,r}^v - \bar{\mathbf{h}}_{k,r}^v\|_2^2$ represents the sparse channel vector estimation error of the SMV problem in (12). The proposed algorithm (WDiSOMP) aims to estimate the MIMO channel by exploiting the commonalities in the support sets of the sparse channel matrices of nearby devices. All devices execute SOMP locally (using the measurements collected from all antenna elements), exchange the estimated support sets with their neighbors' and apply the weighted voting mechanism to estimate their support set. To ensure support recovery, the amplitudes of all nonzero elements must be large enough to overcome the noise. Hence, we define $|\bar{\mathbf{h}}_{k,r,\min}^v| = \min_{m \in \mathcal{S}_k} |\bar{\mathbf{h}}_{k,r}^v(m)|$ to derive a condition that the (S)OMP correctly identifies the support \mathcal{S}_k . Suppose that the event $\mathcal{B}_r = \{ \max_{1 \leq m \leq G_t} |\Phi^H(m) \bar{\mathbf{n}}_{k,r}| < \sigma \sqrt{(1 + \beta) \log G_t^2} \}$ (where constant $\beta > 0$) occurs simultaneously for all antenna elements and thus $|\bar{\mathbf{h}}_{k,r,\min}^v| (1 - (2L_{AoD} - 1)\mu(\Phi)) \geq 2 \left(\sigma \sqrt{(1 + \beta) \log G_t^2} \right)$ holds. An upper-bound of this estimation error is derived by considering the Lemma 3 for OMP performance guarantees in [51]. When the support \mathcal{S}_k is successfully recovered, our estimator operates as the oracle from which the corresponding complex gains are estimated as $(\bar{\mathbf{h}}_{k,r}^{v,o})_{\mathcal{S}_k} = \Phi_{\mathcal{S}_k}^\dagger \bar{\mathbf{y}}_{k,r}$ and its error is written as

$$\begin{aligned} & \left\| (\bar{\mathbf{h}}_{k,r}^{v,o})_{\mathcal{S}_k} - (\bar{\mathbf{h}}_{k,r}^v)_{\mathcal{S}_k} \right\|_2^2 \\ &= \left\| \Phi_{\mathcal{S}_k}^\dagger \Phi_{\mathcal{S}_k} (\bar{\mathbf{h}}_{k,r}^v)_{\mathcal{S}_k} - \Phi_{\mathcal{S}_k}^\dagger \bar{\mathbf{n}}_{k,r} - (\bar{\mathbf{h}}_{k,r}^v)_{\mathcal{S}_k} \right\|_2^2 \\ &= \left\| \Phi_{\mathcal{S}_k}^\dagger \bar{\mathbf{n}}_{k,r} \right\|_2^2 \\ &= \left\| \left(\Phi_{\mathcal{S}_k}^H \Phi_{\mathcal{S}_k} \right)^{-1} \Phi_{\mathcal{S}_k}^H \bar{\mathbf{n}}_{k,r} \right\|_2^2 \\ &\leq \left\| \left(\Phi_{\mathcal{S}_k}^H \Phi_{\mathcal{S}_k} \right)^{-1} \right\|_2^2 \sum_{m \in \mathcal{S}_k} \left(\Phi^H(m) \bar{\mathbf{n}}_{k,r} \right)^2 \\ &\leq \frac{L_{AoD} \left((1 + \beta) \sigma^2 \log G_t^2 \right)}{\left(1 - (L_{AoD} - 1) \mu(\Phi) \right)^2}, \end{aligned} \tag{39}$$

where the last inequality is derived by using the definition of \mathcal{B}_r and $\left\| \left(\Phi_{\mathcal{S}_k}^H \Phi_{\mathcal{S}_k} \right)^{-1} \right\|_2^2 = \frac{1}{\lambda_{\min}(\Phi_{\mathcal{S}_k}^H \Phi_{\mathcal{S}_k})} \leq \frac{1}{(1 - (L_{AoD} - 1)\mu(\Phi))^2}$ by applying the Gershgorin circle theorem to $\lambda_{\min}(\Phi_{\mathcal{S}_k}^H \Phi_{\mathcal{S}_k})$. Combining (38) with (39), the upper bound is as

$$\begin{aligned} \left\| \hat{\mathbf{H}}_k - \mathbf{H}_k \right\|_F^2 &\leq \frac{N_r L_{AoD} \left(1 + (L_{AoD} - 1) \mu(\mathbf{D}_t) \right)}{\left(1 - (L_{AoD} - 1) \mu(\Phi) \right)^2} \\ &\quad \times \left((1 + \beta) \sigma^2 \log G_t^2 \right) \end{aligned} \tag{40}$$

The upper bound of the squared error in (40) shows that the sparse channel recovery performance is affected by the mutual coherence of the measurement matrix Φ , for which $\mu(\Phi) < 1/(2L_{AoD} - 1)$ should hold in order to ensure reliable estimation of the support set in noisy conditions. The upper limit relates to the fact that some of the L_{AoD} involved AoDs in the channel matrix may occupy off-diagonal entries of the Gram matrix \mathbf{G} , with magnitudes, at most, equal to $\mu(\mathbf{D}_t)$. Given an overcomplete dictionary, the training symbols matrix should be properly designed to reduce the magnitudes of these off-diagonal entries.

Table 5 lists the right term in (40) (upper bound) of the instantaneous channel error in terms of the mutual coherence $\mu(\Phi)$ assuming an overcomplete DFT-based dictionary and $\beta = 0.5, N_t = 128, L_{AoD} = 6$. Assuming that the training pilots' design achieves $\mu(\Phi) = 0.05$, the obtained upper bound is higher in low SNR regimes and decreases linearly as the noise variance reduces (high SNR). Also, the upper bound increased by 48%, when the $G_t = 256$ with $\mu(\mathbf{D}_t) = 0.637$ increased to $G_t = 512$ with $\mu(\mathbf{D}_t) = 0.9$. This bound is maintained at low levels as low as possible, and close to zero is the $\mu(\Phi)$. To attain $\mu(\Phi)$ values close to zero, \mathbf{X} should be appropriately constructed, which is out of scope in this work. In our study, the symbols are designed as random Gaussian variables, thus ensuring a "well-behaved" measurement matrix and, hence, correct channel reconstruction with high probability.

TABLE 5. Right term in (40) of the instantaneous channel error.

$G_t = 512, \mu(\mathbf{D}_t) = 0.900$				
$\mu(\Phi)$	0.1	0.05	0.0025	0.00125
$\sigma^2 = 0.01$	49.42	21.97	12.67	12.51
$\sigma^2 = 0.001$	4.94	2.19	1.26	1.25
$G_t = 256, \mu(\mathbf{D}_t) = 0.637$				
$\mu(\Phi)$	0.1	0.05	0.0025	0.00125
$\sigma^2 = 0.01$	33.40	14.85	8.56	8.46
$\sigma^2 = 0.001$	3.34	1.48	0.86	0.85

VII. SIMULATIONS AND DISCUSSION

In this section, the elaborated algorithm is evaluated for the problem of channel estimation in a ULA mmWave massive MIMO system with fully-digital transmitters and receivers by employing SOMP locally.

To be more specific, the proposed distributed scheme is evaluated under two different voting mechanisms in a multi-tasking scenario. Moreover, its behavior will be compared to the individual SOMP (which is executed locally by each device), DiSOMP (which engages a simple majority voting mechanism) and the Structured SOMP (SSOMP), which is executed in a centralized manner at the BS. The performance measure, namely, Normalized Mean Squared Error (NMSE), is defined as

$$NMSE = \frac{1}{M} \sum_{m=1}^M \frac{\sum_{k=1}^K \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|_F^2}{\sum_{k=1}^K \|\mathbf{H}_k\|_F^2} \tag{41}$$

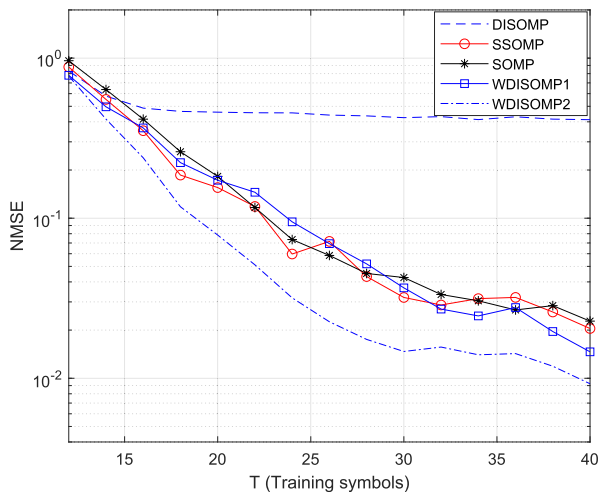


FIGURE 5. Average NMSE over all nodes for $T = [12 : 2 : 40]$, $SNR = 20$ dB.

where \mathbf{H}_k , $\hat{\mathbf{H}}_k^m$ are the correct and reconstructed channel vectors of device k at the m -th channel realization. In the simulations, $M = 50000$ captures the number of Monte Carlo iterations and $\nu = 0.1$ (which is utilized in the recursive equation of the adaptive weighting step in (17), (19)).

In the following, the NMSE behavior of these methods will be assessed i) in terms of the number of training symbols $T = [12, 40]$ for a specific Signal-to-Noise-Ratio (SNR) value, i.e., $SNR = 20$ dB, and ii) for a fixed $T = 30$ in terms of SNR in the range $[0, 30]$.

We assume a network of $K = 10$ devices and a BS with $N_r = 2$ and $N_t = 128$ antenna elements, respectively. Following the relevant literature, e.g. [20], [40], a DFT-based overcomplete, predefined dictionary \mathbf{D}_t is assumed (due to the similarity of the array response vectors with its columns) with resolution $G_t = 512$ atoms. Note that the AoDs and AoAs were selected from $[0, \pi]$, and the AoAs were randomly assigned. We assume that there are $L_g = 2$ global interest AoDs that contribute to all devices' channels and two different groups of $L_c = 4$ commonly shared AoDs (or channel paths) that contribute to devices' channels 1-5 and 7-10, respectively. Also, 1 out of 10 devices, and specifically device 6, partially shares its $L_c = 4$ AoDs with different subgroups. Table 6 presents the AoDs values that were considered to create the devices' channels and the corresponding grouping. The elements of the training matrix $\mathbf{X} \sim \mathcal{CN}(0, \frac{1}{T})$ are i.i.d. random variables [20]. Each device has no prior knowledge of the sparsity order apart from the total sparsity level ($L_{AoD} = 6$), which is similar to all devices.

Figure 5 depicts the average NMSE over all involved devices. It is noticed that WDiSOMP1, which assumes voting on support set level, approaches the behavior of local SOMP. However, WDiSOMP2 achieves superior performance than SOMP, and particularly DiSOMP, due to the employment of the voting method on the index level. For the latter, the floor in the performance is attained for low values of T and remains unchanged even when measurements T are high enough. This behavior is justified by the unknown sparsity

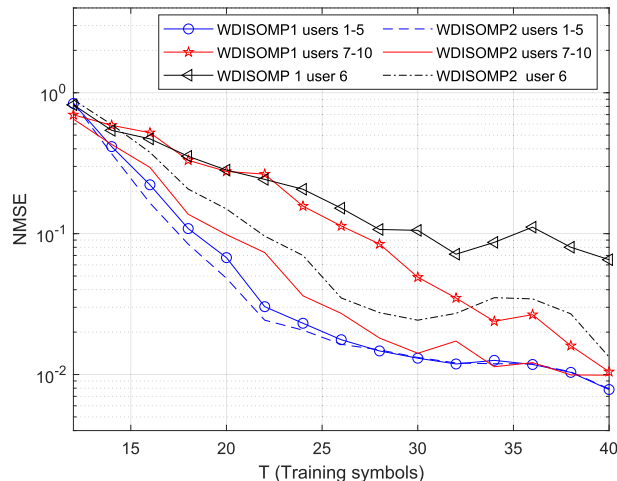


FIGURE 6. NMSE of devices 1-5, devices 7-10 and device 6 for $T = [12 : 2 : 40]$, $SNR = 20$ dB.

structure, which WiDSOMP 1, 2 can learn via the suggested voting mechanisms. Also, WDiSOMP, in either case, attains the performance of SSOMP (the centralized solution), which, nonetheless, needs to know the number of global L_g and common L_c indices in the sparsity pattern division, as in [20].

Figure 6 demonstrates the channel estimation performance at the two subgroups and device 6, focusing on WDiSOMP 1, 2. It is reminded that these subgroups comprise devices 1-5 and 7-10, respectively, whose support set is not partially shared and device 6, which shares 4 out of 6 indices in its support set with the devices from different subgroups. It is observed that the behavior of the algorithm follows, as expected, the one observed in Figure 5, namely, follows the average NMSE. The performance of WDiSOMP2 in devices 1-5 enhanced slightly, while device 6, which captures the challenging multi-tasking condition, has considerably benefited from the new weighting mechanism that operates at the index level. The same holds for devices 7-10.

Now, we will focus on the second experiment of this study. In Figure 7, we show the average performance of all methods in terms of SNR for $T = 30$. Then, in Figure 8, the performance of WDiSOMP1 and WDiSOMP2 for the devices 1-5, 7-10 and 6 is captured. Both in this case, it is verified that the behavior of the proposed algorithm, under the per-index weighting method before voting, is superior to exploiting the support set as a global entity (i.e., not per index).

In Figure 9, we isolate the average behavior of WDiSOMP and local SOMP in terms of the noise level for $T = \{30, 60\}$. The theoretical lower error bound, presented in Section VI, is also plotted to validate the performance of the proposed scheme for the mmWave channel estimation. More specifically, the lower bound was derived based on (31). For $T = 30$, WDiSOMP1 has a similar performance to SOMP with a performance gap by WDiSOMP2 for high SNR. When the available measurements per device were doubled ($T = 60$), the two versions of WDiSOMP presented comparable behavior and, in high SNR, were better than SOMP (with a smaller

TABLE 6. Multitask device grouping according to AoDs sparsity patterns for $L_g = 2, L_c = 4$.

AoD Type	AoD Name	device 1	device 2	device 3	device 4	device 5	device 6	device 7	device 8	device 9	device 10
Global	AoD 1	144.34	144.34	144.34	144.34	144.34	144.34	144.34	144.34	144.34	144.34
Global	AoD 2	128.68	128.68	128.68	128.68	128.68	128.68	128.68	128.68	128.68	128.68
Common	AoD 3	112.02	112.02	112.02	112.02	112.02	112.02	93.58	93.58	93.58	93.58
Common	AoD 4	91.79	91.79	91.79	91.79	91.79	91.79	73.66	73.66	73.66	73.66
Common	AoD 5	64.05	64.05	64.05	64.05	64.05	53.57	53.57	53.57	53.57	53.57
Common	AoD 6	55.77	55.77	55.77	55.77	55.77	28.95	28.95	28.95	28.955	28.95

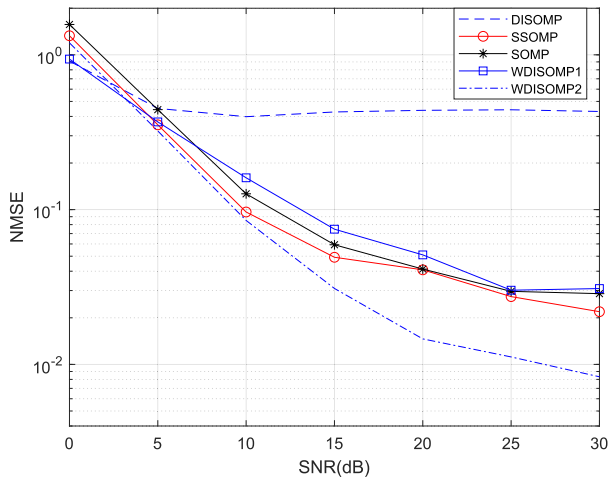


FIGURE 7. Average NMSE of all devices for $SNR = [0 : 5 : 30]$ dB, $T = 30$.

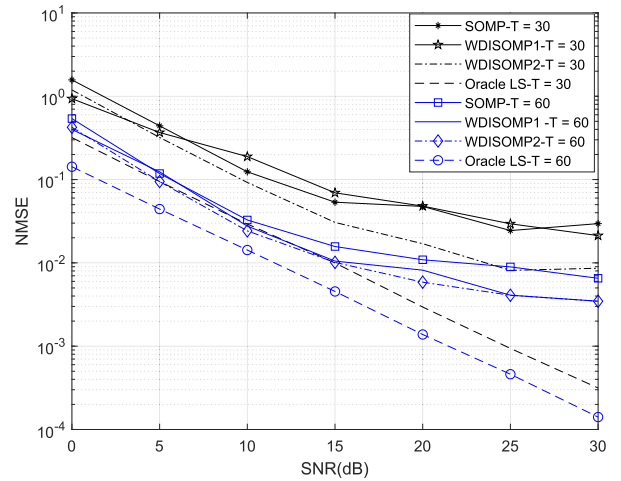


FIGURE 9. Average NMSE of all devices for $SNR = [0 : 5 : 30]$, $T = \{30, 60\}$.

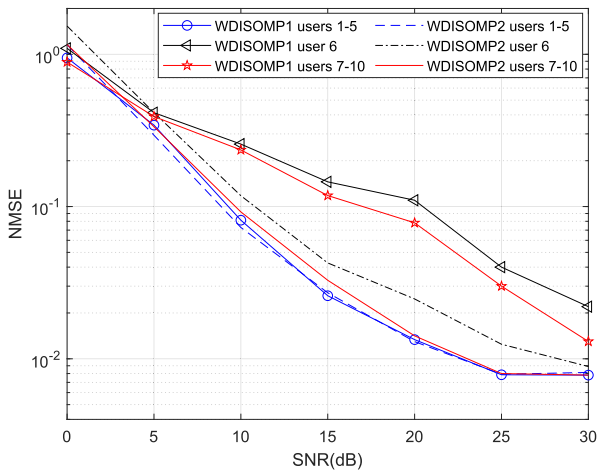


FIGURE 8. Average NMSE of devices 1-5, 7-10 and device 6 for $SNR = [0 : 5 : 30]$ dB, $T = 30$.

gap in their performances than previously). Also, in high SNR, the floor in the performance of WDiSOMP relates to the reconstruction error imposed by the mistaken support indices that the devices estimate (to find then the corresponding AoDs) compared to the oracle LS estimator, which assumes prior knowledge of them to estimate the channel. This experiment indicated that, in medium-to-high SNR, WDiSOMP benefited from the per index weighting method when the number of measurements was not enough. In low SNR, it was considerably reduced the distance between WDiSOMP1,

WDiSOMP2 and the lower bound. However, they didn't approach the lower bound since the matrix \mathbf{X} isn't properly designed, making the eigenvalues λ_{iS} of $\Phi_a^H \Phi_a$ different, even if it still holds that $\text{Tr}(\Phi_a^H \Phi_a) = \sum_{i=1}^{L_{AoD}} \lambda_i(\Phi_a^H \Phi_a) = L_{AoD}$. As mentioned in Section VI.C, the eigenvalues equality is a necessary condition to approach the lower bound, which in our case isn't satisfied.

From the above simulations, it is noted that WDiSOMP, with the use of either weighting method, can efficiently exploit, under different conditions, the inherent common sparsity structures and its performance is not impacted when various related device groupings coexist.

VIII. CONCLUSION

In conclusion, the channel estimation problem is addressed by a distributed approach that exploits the joint sparsity among proximal devices' sparse channel matrices to estimate the desired parameters collaboratively. The down-link multi-device MIMO channel estimation is solved via WDiSOMP, a fully distributed method that exploits a generalized model for the sparsity support of the involved channel matrices. The enhanced performance of WDiSOMP has been evaluated under the per index weighted voting mechanism when the structure of the sparsity support is unknown and non-uniform under a multi-tasking framework. Also, the transmission overhead of the proposed algorithm was investigated and compared to the corresponding cost of the centralized solution. The outcomes revealed the efficacy of

the method under specific configurations of the involved parameters.

A promising future direction is to examine and broaden the proposed method in hybrid architectures [16], [41] since, in this work, fully-digital transmitters and receivers are considered. Furthermore, another goal is to enhance the performance of the proposed algorithm by focusing on the proper training matrix design. This analysis will help to attain the performance bound of the oracle LS estimator and ensure reliable channel recovery under the overcomplete dictionary (as already mentioned in the performance analysis). Another extension of the study could focus on alternative weighting methods in the voting process (quantifying the part of the error related to the deviation of the MSE from the lower bound) to optimize the mmWave channel estimation performance (under such a complicated joint sparsity framework). Finally, a well-known challenging direction is to reformulate the suggested fully distributed cooperative approach under a CS-based federated learning framework in order to benefit from the advantages and capabilities (such as privacy) that this technique provides, exploiting the recent research effort in this field.

REFERENCES

- [1] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [2] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.
- [3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Feb. 2016.
- [4] N. Al-Falahy and O. Y. K. Alani, "Millimetre wave frequency band as a candidate spectrum for 5G network architecture: A survey," *Phys. Commun.*, vol. 32, pp. 120–144, Feb. 2019.
- [5] S. He, Y. Zhang, J. Wang, J. Zhang, J. Ren, Y. Zhang, W. Zhuang, and X. Shen, "A survey of millimeter-wave communication: Physical-layer technology specifications and enabling transmission technologies," *Proc. IEEE*, vol. 109, no. 10, pp. 1666–1705, Oct. 2021.
- [6] K. Sakaguchi, T. Haustein, S. Barbarossa, E. C. Strinati, A. Clemente, G. Destino, A. Pärssinen, I. Kim, H. Chung, and J. Kim, "Where, when, and How mmWave is used in 5G and beyond," *IEICE Trans. Electron.*, vol. 100, no. 10, pp. 790–808, Apr. 2017.
- [7] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 16–32, Jan. 2020.
- [8] 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.
- [9] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 40–58, May 2018.
- [10] K. Hassan, M. Masarra, M. Zwingelstein, and I. Dayoub, "Channel estimation techniques for millimeter-wave communication systems: Achievements and challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1336–1363, 2020.
- [11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [12] S. Sun and T. S. Rappaport, "Millimeter wave MIMO channel estimation based on adaptive compressed sensing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 47–53.
- [13] W. Huang, Y. Huang, W. Xu, and L. Yang, "Beam-blocked channel estimation for FDD massive MIMO with compressed feedback," *IEEE Access*, vol. 5, pp. 11791–11804, 2017.
- [14] J. Dai, A. Liu, and V. K. N. Lau, "Joint channel estimation and user grouping for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 622–637, Feb. 2019.
- [15] E. C. Marques, N. Maciel, L. Naviner, H. Cai, and J. Yang, "A review of sparse recovery algorithms," *IEEE Access*, vol. 7, pp. 1300–1322, 2019.
- [16] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [17] 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.
- [18] A. Rajoriya, R. Budhiraja, and L. Hanzo, "Centralized and decentralized channel estimation in FDD multi-user massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7325–7342, Jul. 2022.
- [19] Z. Gao, L. Dai, and Z. Wang, "Channel estimation for mmWave massive MIMO based access and backhaul in ultra-dense network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
- [20] R. Zhang, H. Zhao, and J. Zhang, "Distributed compressed sensing aided sparse channel estimation in FDD massive MIMO system," *IEEE Access*, vol. 6, pp. 18383–18397, 2018.
- [21] I. Khan, J. J. P. C. Rodrigues, J. Al-Muhtadi, M. I. Khattak, Y. Khan, F. Altaf, S. S. Mirjavadi, and B. J. Choi, "A robust channel estimation scheme for 5G massive MIMO systems," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–8, Aug. 2019.
- [22] F. Kulsoom, A. Vizziello, H. N. Chaudhry, and P. Savazzi, "Joint sparse channel recovery with quantized feedback for multi-user massive MIMO systems," *IEEE Access*, vol. 8, pp. 11046–11060, 2020.
- [23] J. Dai, A. Liu, and H. C. So, "Non-uniform burst-sparsity learning for massive MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1075–1087, Feb. 2019.
- [24] K. Luo, X. Zhou, B. Wang, J. Huang, and H. Liu, "Sparse Bayes tensor and DOA tracking inspired channel estimation for V2X millimeter wave massive MIMO system," *Sensors*, vol. 21, no. 12, p. 4021, Jun. 2021.
- [25] J. Li, D. Sun, and Z. Liu, "A MIMO channel prediction scheme based on multi-task learning," *Wireless Pers. Commun.*, vol. 115, no. 3, pp. 1869–1880, Dec. 2020.
- [26] W. Xie, J. Xiao, P. Zhu, and C. Yu, "Multi-task learning-based channel estimation for RIS assisted multi-user communication systems," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 577–581, Mar. 2022.
- [27] Z. Li, W. Gao, M. Zhang, and J. Xiong, "Multi-task deep learning based hybrid precoding for mmWave massive MIMO system," *China Commun.*, vol. 18, no. 10, pp. 96–106, Oct. 2021.
- [28] R. Middy, N. Chakravarty, and M. K. Naskar, "Compressive sensing in wireless sensor networks: A survey," *IETE Tech. Rev.*, vol. 34, no. 6, pp. 642–654, 2017.
- [29] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Depart. Elect. Comput. Eng., Rice Univ., Houston, TX, USA, Tech. Rep. TREE-0612, 2006.
- [30] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
- [31] D. Sundman, S. Chatterjee, and M. Skoglund, "Distributed greedy pursuit algorithms," *Signal Process.*, vol. 105, pp. 298–315, Dec. 2014.
- [32] T. Wimalajeewa and P. K. Varshney, "OMP based joint sparsity pattern recovery under communication constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5059–5072, Oct. 2014.
- [33] M. Trigka, C. Mavroukafalidis, and K. Berberidis, "A distributed sparse channel estimation technique for mmWave massive MIMO systems," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2021, pp. 31–35.
- [34] M. Trigka, C. Mavroukafalidis, and K. Berberidis, "An efficient decentralized approach for mmWave MIMO channel estimation," in *Proc. Int. Balkan Conf. Commun. Netw. (BalkanCom)*, Sep. 2021, pp. 31–35.
- [35] R. I. Ansari, C. Chrysostomou, S. A. Hassan, M. Guizani, S. Mumtaz, J. Rodriguez, and J. J. Rodrigues, "5G D2D networks: Techniques, challenges, and future prospects," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3970–3984, Dec. 2018.
- [36] G. I. Tsiropoulos, A. Yadav, M. Zeng, and O. A. Dobre, "Cooperation in 5G HetNets: Advanced spectrum access and D2D assisted communications," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 110–117, Oct. 2017.

- [37] S. Jayaprakasam, S. K. A. Rahim, and C. Y. Leow, "Distributed and collaborative beamforming in wireless sensor networks: Classifications, trends, and research directions," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2092–2116, 4th Quart., 2017.
- [38] O. B. Smida, S. Zaidi, S. Affes, and S. Valaee, "Robust distributed collaborative beamforming for wireless sensor networks with channel estimation impairments," *Sensors*, vol. 19, p. 1061, Mar. 2019.
- [39] H. He and H. Jiang, "Deep learning based energy efficiency optimization for distributed cooperative spectrum sensing," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 32–39, Jun. 2019.
- [40] C.-R. Tsai, Y.-H. Liu, and A.-Y. Wu, "Efficient compressive channel estimation for millimeter-wave large-scale antenna systems," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2414–2428, May 2018.
- [41] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [42] R. Nassif, S. Vlaski, C. Richard, J. Chen, and A. H. Sayed, "Multitask learning over graphs: An approach for distributed, streaming machine learning," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 14–25, May 2020.
- [43] S. Narayanan, S. K. Sahoo, and A. Makur, "Greedy pursuits assisted basis pursuit for reconstruction of joint-sparse signals," *Signal Process.*, vol. 142, no. 1, pp. 485–491, Jan. 2018.
- [44] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.
- [45] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2133–2168, 3rd Quart., 2018.
- [46] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.
- [47] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Jan. 2008.
- [48] J. Iscar, I. Guevenc, S. Dikmese, and A. S. Ibrahim, "Optimal angular spread of the multipath clusters in mmWave systems under pilot contamination," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–6.
- [49] R. A. Horn, *CR Johnson Matrix Analysis*. New York, NY, USA: Cambridge UP, 1985.
- [50] P. S. Bullen, *Handbook Means Their Inequalities*, vol. 560. Cham, Switzerland: Springer, 2013.
- [51] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.



publications in the area of signal processing, wireless communications, and machine learning.

MARIA TRIGKA received the Diploma degree from the Electrical and Computer Engineering Department, University of Patras, and the M.Sc. and Ph.D. degrees from the Computer Engineering and Informatics Department, University of Patras. She is currently working as a Researcher at the University of Patras. As a Ph.D. candidate, she served as a research scholar under the project funded by the State Scholarships Foundation (IKY). She is the author and coauthor of



CHRISTOS MAVROKEFALIDIS (Member, IEEE) received the Diploma degree in computer engineering and informatics, the master's degree in signal processing systems, and the Ph.D. degree in signal processing for wireless communications from the University of Patras, Patras, Greece, in 2004, 2006, and 2011, respectively. Since 2006, he has been a Research Associate with the Signal Processing and Communications Laboratory, Computer Engineering and Informatics Department, University of Patras. Since 2006, he has also been with the Signal Processing and Communication Research Unit, Computer Technology Institute and Press "Diophantus," Patras. Since 2019, he has been with the Multimedia Information Processing Systems Group, Industrial Systems Institute, Athena Research Center, Patras. He has also been involved in numerous national, European, and bilateral research projects in application areas like wireless communications and sensor networks, smartgrids, and computer vision. In the past, he was involved in the design of integrated circuits for microprocessors that support multimedia operations. His research interests include scientific areas of statistical signal processing and learning with a focus on estimation theory, adaptive/distributed signal processing, and sparse representations. He is a member of the Technical Chamber of Greece. He is a regular reviewer in various journals and conferences in the general area of signal processing.

He is a regular reviewer in various journals and conferences in the general area of signal processing.



KOSTAS BERBERIDIS (Senior Member, IEEE) received the Diploma degree in electrical engineering from DUTH, Greece, in 1984, and the Ph.D. degree in signal processing and communications from the University of Patras, Patras, Greece, in 1990. In 1991, he worked at the Signal Processing Laboratory, National Defense Research Center. From 1992 to 1994 and from 1996 to 1997, he was a Researcher at the Computer Technology Institute (CTI), Patras. From 1994 to 1995, he was

a Postdoctoral Fellow at CCETT/CNET, Rennes, France. Since December 1997, he has been with the Computer Engineering and Informatics Department (CEID), University of Patras, where he is currently a Professor and the Head of the Signal Processing and Communications Laboratory. Since 2008, he has also been the Director of the Signal Processing and Communications Research Unit, Computer Technology Institute and Press "Diophantus." Since 2018, he has been collaborating faculty with the Industrial Systems Institute, Athena Research Center, Greece. His research interests include distributed signal and information processing and learning, adaptive learning algorithms, signal processing for communications, wireless communications and sensor networks, array signal processing, and smart grid. He is a member of the Technical Chamber of Greece and EURASIP. Since August 2015, he has been a member of the EURASIP Technical Area Committee "Signal Processing for Multisensor Systems." Since January 2017, he serves as a member of the Board of Directors (BoD) for EURASIP (second term started January 2021). He has served or has been serving as a member of scientific and organizing committees of several international conferences, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and *EURASIP Journal on Advances in Signal Processing*, and a Guest Editor for the *EURASIP Journal on Advances in Signal Processing*. From February 2010 to December 2017, he has served as the Chair of the Greece Chapter for the IEEE Signal Processing Society. He has also served in the "Signal Processing Theory and Methods" Technical Committee of the IEEE SPS, the "Signal Processing for Communications and Electronics" Technical Committee of the IEEE COMSOC, and the EURASIP Special Area Team "Theoretical and Methodological Trends for Signal Processing."

• • •