IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 6 August 2023, accepted 1 September 2023, date of publication 11 September 2023,
date of current version 14 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3313736

RESEARCH ARTICLE

Sparsifying Dictionary Learning for Beamspace Channel Representation and Estimation in Millimeter-Wave Massive MIMO

MEHMET ALI AYGÜL[1,2], (Student Member, IEEE), MAHMOUD NAZZAL[3], (Member, IEEE), AND HÜSEYIN ARSLAN[3,4], (Fellow, IEEE)

[1]Department of Electronics and Communications Engineering, Istanbul Technical University, 34467 Istanbul, Turkey
[2]Department of Research and Development, Vestel, 45030 Manisa, Turkey
[3]Department of Electrical and Electronics Engineering, Istanbul Medipol University, 34810 Istanbul, Turkey
[4]Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA

Corresponding author: Mehmet Ali Aygül (mehmetali.aygul@ieee.org)

This work was supported in part by the Scientific and Technological Research Council of Turkey (TÜBTAK) under Grant 116E078, and in part by Vestel Elektronik Sanayi ve Ticaret Anonim Şirketi.

ABSTRACT Millimeter-wave (mmWave) massive multiple-input-multiple-output (mMIMO) is reported as a key enabler in fifth-generation communication and beyond. It is customary to use a lens antenna array to transform a mmWave mMIMO channel into a beamspace where the channel exhibits sparsity. This beamspace transformation is equivalent to performing a Fourier transformation of the channel. Still, a Fourier transformation is not necessarily optimal for many reasons. For example, it can cause a power leakage problem. Accordingly, this paper proposes using a learned sparsifying dictionary as the transformation operator leading to another beamspace for channel representation. Since a dictionary is obtained by training over actual channel measurements in an end-to-end manner, this transformation is shown to yield two immediate advantages. First is enhancing channel sparsity, thereby leading to more efficient pilot reduction. Second is improving the channel representation quality, thus reducing the underlying power leakage phenomenon. Consequently, this allows for improved channel estimation and facilitates beam selection in mmWave mMIMO. In addition, a learned dictionary is used as the channel estimation operator for the same reasons. Extensive simulations under various operating scenarios and environments validate the added benefits of using learned dictionaries in improving the channel estimation quality and beam selectivity, thus improving spectral efficiency.

INDEX TERMS Beamspace channel, channel estimation, channel representation, dictionary learning, lens antenna array, massive MIMO, millimeter-wave.

I. INTRODUCTION

Massive multiple-input-multiple-output (mMIMO) is widely considered a key enabler for wireless communication in the era of the fifth generation and beyond. This is because of its ability to improve the system data rate [1]. Especially, when it operates at millimeter-wave (mmWave) frequencies, it has crucial importance. This allows for increased data rates due to the higher spectral efficiency [2] and wider bandwidth [3]. However, the main challenge with mmWave mMIMO is the hardware and power requirements.

Beamforming techniques are used to reduce power consumption and cost by suppressing the co-channel interference and improving the signal-to-noise ratio (SNR) at the receiver end [4]. These techniques can be divided into three categories; analog, digital, and hybrid. Analog beamforming is cost and power-effective but only supports one data stream at a time [5]. On the other hand, digital precoding uses a radio frequency (RF) chain per antenna element and thus requires high power consumption, complexity, and cost. Therefore,

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

a hybrid precoding technique has been introduced as a compromise to both settings [6].

Hybrid precoding connects hundreds of antennas to a small number of RF chains through analog phase shifters [7]. However, the design of precoding matrices is usually based on channel state information (CSI) [7], and it is difficult to obtain CSI in mmWave mMIMO due to large numbers of antennas [8] and RF chains [9]. Thus, developing low-complexity channel estimation techniques is crucial for the mmWave mMIMO system operation. On the other hand, mMIMO channels show strongly directional propagation with low dimensionality properties at mmWave frequencies. This motivates a beamspace representation [10] where channel sparsity can be exposed. The channel sparsity can be exploited with the advent of compressive sensing (CS), allowing for reduced channel training and feedback overheads.

A. RELATED WORKS AND MOTIVATION

CS-based channel estimation algorithms exploit angle domain sparsity of mmWave mMIMO channels [11], [12], [13]. CS allows for sub-Nyquist sampling by enabling sparse signal recovery at a sampling rate below the Nyquist rate. However, these algorithms are designed with high-resolution phase shifters for hybrid precoding systems. Still, a phase shifter network can be replaced by a lens antenna array (LAA) [14] to further reduce the hardware cost and power consumption. Hence, an LAA is widely used to expose a beamspace channel representation in mmWave mMIMO. Therefore, the dimension of a mmWave mMIMO channel can be reduced by beam selection over the sparse beamspace channel [15], [16].

A promising channel estimation technique for the case of using an LAA is sparsity mask detection [17]. In this setting, the beams of large power are determined initially. Then, the dimension of the beamspace channel is reduced, and it is estimated in this reduced dimension. However, scanning over all the beams is a time-consuming process. Another algorithm to reduce the number of antennas is the support detection (SD) algorithm for sparse coding. This algorithm divides the channel estimation problem into a series of subproblems, each of which only considers one channel path component [18]. To this end, this multitude of beamspace channel estimation algorithms models the impact of the LAA by a discrete Fourier transform (DFT) matrix. DFT discretizes the continuous angular channel parameter space into a finite set of predefined spatial angles. This set covers the whole angular beam range and emphasizes sparsity. Thus, the performance of these algorithms largely depends on how accurately this discretization can model the true sparsity of the channel, i.e., it depends on the representation power of this sparsifying basis/transform.

Despite achieving state-of-the-art performance in mmWave mMIMO channel estimation, a DFT sparsifying basis is known to have several inherent shortcomings. Specifically, the actual angles of departure (AoDs) of paths are continuously distributed since the spatial sampling points of the LAA are not finite and fixed in practice. Therefore, the AoD of a path will not necessarily match the spatial sample points of the LAA [19] modeled by DFT, as illustrated in Fig. 1. Consequently, the power of a beam will leak onto multiple beams in the beamspace (*off-grid* problem) [14]. This *power leakage* effect is serious even for the simplest cases and incurs obvious SNR losses [16]. References [20], [21], [22], [23], [24], [25], and [26] estimate a beamspace channel using a machine learning framework to improve estimation quality. However, these works do not address the sparsity of the representation. Therefore, their performance improvement is limited.

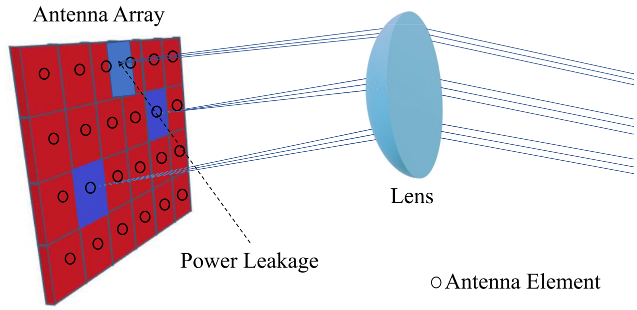


FIGURE 1. The concept of power leakage in LAAs.

The problem of obtaining an efficient sparsifying transform is studied in the context of signal representation. It has been shown that one can use a *redundant* (over-complete) DFT basis aiming at a finer discretization of the channel signal space. A redundant dictionary is tailored for LAA to mitigate the power leakage caused by the continuous angles of multipath components in [27]. Nonetheless, there are certain limits to the redundancy of this basis, as it substantially increases the computational cost. Besides, high redundancy creates the side effect of more similarity between the columns of the basis matrix, thereby degrading the representation quality. Therefore, recent research necessities a limited degree of redundancy on the sparsifying basis, while trying to tackle the off-grid effects. Although [17], [28], [29], and [30] use moderate degrees of redundancy, their computational complexity is still prohibitively large.

Rather than expanding the quantity of discretized points, recent literature calls for developing new beamspace transformation operators to combat off-grid effects. Such operators are not restricted to having the DFT character. For example, the Fourier domain is shown to overlook the Dirichlet structure inherent in mmWave channels [31]. Thus, the authors proposed using a set of Dirichlet kernels to serve as a sparsifying dictionary. Besides, the DFT is shown to be sub-optimal as a sparsifying transform [32]. Thus, the authors proposed using a Karhunen-Loève transform (a.k.a. principal component analysis) as a data-dependent optimal basis. Alternatively, enhanced and more general dictionaries are anticipated to offer a better alternative to DFT bases [10].

A dictionary that is generated by a finer-grained point further improves the approximation of the continuous points and also the estimation quality [33]. In this context, dictionary learning offers dictionaries with enhanced sparsity and representation quality leading to lowering the severity of power leakage, more efficient pilot reduction, and improved channel estimation quality.

B. CONTRIBUTIONS

In view of the above discussion, a preliminary version of this work [34] showing the advantage of using a learned dictionary for channel estimation in mmWave mMIMO is extended. This paper proposes algorithms for beamspace channel representation and estimation based on sparse coding over a learned dictionary. This paper presents the following contributions.

- As opposed to standard beamspace channel sparsification, such as using the DFT to represent LAA operation, a learned dictionary enhances the channel sparsity. This leads to a further reduction in hardware, cost, and power consumption. Besides, this allows for easier beam selection at the receiver end. Along with this line, using a learned dictionary obtained by training over previous channel observations is proposed as the channel sparsifying transform operator.
- The success of channel estimation not only depends on the sparsity of the channel, but also depends on how this sparsity can be exposed and utilized. Along with this line, the usage of a different learned dictionary is also proposed as the channel estimation operator. Such a dictionary is obtained by training over example precoding matrix realizations.

Extensive simulations in various operating scenarios and environments validate that using learned dictionaries for beamspace channel sparsification and estimation improves the channel estimation quality and enhances the beam selectivity by improving the spectral efficiency. Therefore, the paper introduces the concept of learned dictionaries for channel representation and estimation, offering valuable insights into the field of mmWave mMIMO. By demonstrating their benefits through simulations, it presents novel contributions that advance this area. These ideas contribute to enhancing sparsity, resulting in reduced power leakage and improved spectral efficiency.

C. ORGANIZATION AND NOTATION

This paper is organized as follows. Section II revises the preliminaries and presents the system model. The proposed algorithms for channel representation and estimation are detailed in Section III. Section IV presents experiments conducted to evaluate the performance of the proposed algorithm, and challenges and future work are discussed in Section V. Finally, the paper is concluded in Section VI.

Plain-faced letters demonstrate scalars. Bold-faced lower-case and bold-faced upper-case letters represent vectors and matrices, respectively. In a matrix $\boldsymbol{X}$, the symbol $\boldsymbol{X}_i$ represents its ith column. Similarly, $\boldsymbol{x}_i$ is the ith element in a vector $\boldsymbol{x}$. The conjugate transpose symbol is denoted by $\dagger$. $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. The $\|.\|_2$, $\|.\|_0$, and tr symbols signify the 2-norm, the number of nonzero elements in a vector, and the trace operation, respectively.

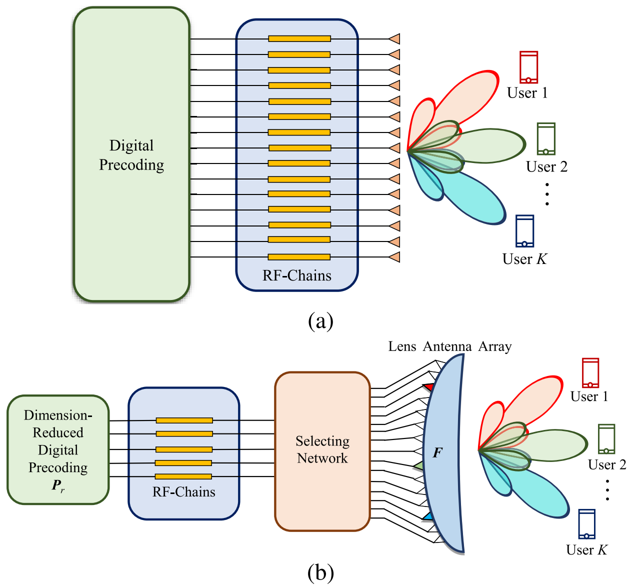


FIGURE 2. Antenna array configuration in mMIMO; (a) conventional and (b) with an LAA controlled by sparse coding beam selection [18].

II. PRELIMINARIES AND SYSTEM MODEL

A. SYSTEM MODEL

This paper considers a mmWave mMIMO system running in time division duplexing (TDD). The base station (BS) uses N antennas with $N_{\rm RF}$ RF chains to serve K single-antenna users (UEs) [35], [36], [37]. In this subsection, the downlink (DL) model is considered to explain the main rationality of the mmWave m-MIMO, while, Section III considers the uplink (UL) model for channel estimation, which is a transposition of the DL model according to the TDD channel reciprocity.

1) TRADITIONAL mmWave mMIMO

Figure 2 (a) shows a conventional mmWave mMIMO setting. The $K \times 1$ received signal vector $\boldsymbol{y}_{\rm DL}$ of all K UEs in the DL for the conventional MIMO systems in the spatial domain can be presented as

$$\boldsymbol{y}_{\rm DL} = \boldsymbol{H}^{\dagger}\boldsymbol{P}\boldsymbol{s} + \boldsymbol{n}, \tag{1}$$

where the DL channel matrix is denoted by $\boldsymbol{H}^{\dagger} \in \mathbb{C}^{K\times N}$, $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K]$ is the UL channel matrix according to the channel reciprocity [12], $\boldsymbol{h}_k$ of size $N \times 1$ is the channel between the kth UE and the BS, $\boldsymbol{s}$ of size $K \times 1$ is the data signal vector for all K UE with normalized power $\mathbb{E}(\boldsymbol{s}\boldsymbol{s}^{\dagger}) = \boldsymbol{I}_K$, $\boldsymbol{P} \sim N \times K$ is the precoding matrix. This matrix satisfies the total transmit power constraint ρ as $tr(\boldsymbol{P}\boldsymbol{P}^{\dagger}) \leq \rho$. Finally, $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2_{\rm DL}\boldsymbol{I}_K)$ is the $K\times 1$ additive white Gaussian noise vector, where σ_{DL}^2 is the DL noise power. Figure 2 (a) shows that the number of RF chains needed in conventional MIMO systems is equal to the number of antennas. i.e., $N_{\mathrm{RF}} = N$, which is mostly large for mmWave mMIMO systems, e.g., $N_{\mathrm{RF}} = N = 256$ [8].

Two channel models are used in this paper; the Saleh-Valenzuela (SV) and the geometry-based stochastic channel model (GSCM). Despite their similarity, the SV model is primitive and widely used in mmWave channel modeling, whereas the GSCM better reflects the operation of antenna arrays and can form the benchmark for mMIMO channel modeling, as a more advanced model [38]. Therefore, we opt to use both channel models to represent mmWave mMIMO.

2) THE SALEH-VALENZUELA CHANNEL MODEL

The SV channel model is customarily used to model mmWave channels as it accounts for their low-rank nature. According to this model, the channel is expressed as follows [7], [39]

$$\boldsymbol{h}_k = \sqrt{\frac{N}{L+1}} \sum_{i=0}^{L} \beta_k^{(i)} \boldsymbol{a}\left(\psi_k^{(i)}\right) = \sqrt{\frac{N}{L+1}} \sum_{i=0}^{L} \boldsymbol{c}_i, \tag{2}$$

where the line-of-sight (LoS) component of the kth UE is $\boldsymbol{c}_0 = \beta_k^{(0)} \boldsymbol{a}(\psi_k^{(0)})$. Also, $\beta_k^{(0)}$ represents the complex gain and $\psi_k^{(0)}$ denotes the spatial direction. The non-LoS (NLoS) component of the kth UE is $c_i = \beta_k^{(i)} \boldsymbol{a}(\psi_k^{(i)})$ for $1 \leq i \leq L$ and the total number of NLoS components, denoted by L, is usually obtained by channel measurement [40]. Besides, $\boldsymbol{a}(\psi)$ is the $N \times 1$ array steering vector. For a typical linear array with N antennas, the steering vector can be represented as follows [41].

$$\boldsymbol{a}(\theta) = \frac{1}{\sqrt{N}}[1,\ e^{-j2\pi\psi_i(\theta)}, \ldots, e^{-j2\pi\psi_i(\theta)}(N-1)]^{\dagger}, \tag{3}$$

where the direction of physical propagation is denoted by θ and the spatial direction is defined as $\psi_i \triangleq \frac{d_i}{\lambda}\sin(\theta)$ [39], λ denotes the wavelength, and d_i represents the antenna spacing in the ith column and it is usually $\lambda/2$ for a linear antenna array.

3) GEOMETRY-BASED STOCHASTIC CHANNEL MODEL

The GSCM is also used as it is a more realistic channel model. For this model, the DL channel vector is considered from the BS to the kth UE. This can be represented as [42]

$$\boldsymbol{h}_k = \sum_{i=1}^{N_c} \sum_{l=1}^{N_s} \beta_k^{(i,l)} \boldsymbol{a}({\theta_k}^{(i,l)}), \tag{4}$$

where the complex gain of the lth scattering cluster is denoted by $\beta^{i,l}$, the number of scattering clusters is denoted by N_c, and the number of sub-paths per scattering cluster is denoted by N_s. The symbol $\theta^{i,l}$ denotes the angle-of-arrival/ angle-of-departure (AoA/AoD) of the lth subpath in the ith scattering cluster. The steering vector $\boldsymbol{a}(\theta^{(i,l)})$ represents the normalized array response at the UE.

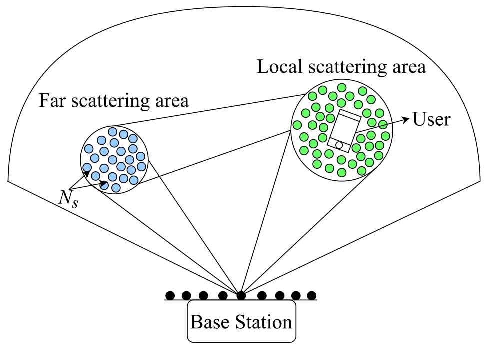


FIGURE 3. The GSCM concept [43]. In this configuration, local scatterers are centered around the UE and far scatterers are far away from both UE and BS.

For scattering, the principles of GSCM are adopted as in Fig. 3. In this figure, far scatterers represent mountains, high-rise buildings, etc. Also, they determine the locations of the dominant scattering clusters for a specific cell and are common to all UEs irrespective of UEs' position. We assume that these are far away from the BS. Thus, the subpaths associated with a specific scattering cluster will be concentrated in a small range, i.e., having a small angular spread. While modeling the scattering effects that are UE-location dependent (e.g., the ground reflection close to the UE or some moving physical scatterers near the UE), we assume the UE is far from the BS. Thus, subpaths associated with the UE-location-dependent scattering cluster also have a small angular spread. Since the BS is far away and is commonly assumed to be mounted at a height, the number of scattering clusters contributing to the channel responses is limited, i.e., N_c is small.

4) MmWave mMIMO CHANNELS IN THE BEAMSPACE

Transforming the conventional channel [9] to a beamspace representation can be done conveniently using an LAA [39], as demonstrated in Fig. 2 (b). Particularly, a well-designed LAA plays the role of a spatial DFT matrix $\boldsymbol{U}$ that comprises the array steering vectors of N orthogonal directions (beams) covering the entire angle space. This matrix can be represented as [39]

$$\boldsymbol{U} = [\boldsymbol{a}(\bar{\psi}_1), \boldsymbol{a}(\bar{\psi}_2), \cdots, \boldsymbol{a}(\bar{\psi}_N)]^{\dagger}, \tag{5}$$

where $\bar{\psi}_n = \frac{1}{N}(n - \frac{N+1}{2})$ for $n = 1, 2, \ldots, N$ are previously defined spatial directions by LAA. Then, the system model of mmWave mMIMO with an LAA can be represented by

$$\tilde{\boldsymbol{y}}^{\mathrm{DL}} = \boldsymbol{H}^{\dagger}\boldsymbol{U}^{\dagger}\boldsymbol{B}\boldsymbol{P}_r\boldsymbol{s} + \boldsymbol{n} = \tilde{\boldsymbol{H}}^{\dagger}\boldsymbol{B}\boldsymbol{P}_r\boldsymbol{s} + \boldsymbol{n}, \tag{6}$$

where the received DL signal in the beamspace is $\tilde{\boldsymbol{y}}_{\mathrm{DL}}$, $\tilde{\boldsymbol{H}}^{\dagger} = \boldsymbol{H}^{\dagger}\boldsymbol{U}^{\dagger} = (\boldsymbol{U}\boldsymbol{H})^{\dagger}$ represents the DL beamspace channel matrix, in which N columns being N orthogonal beams, $\boldsymbol{B}$ of size $N \times K$ form the selecting matrix whose entries belong to $\{0, 1\}$. As an example, when the nth beam is selected by the kth UE, the element of $\boldsymbol{B}$ at the nth row and the kth column would be one. After that, $\boldsymbol{P}_r$ of size $K \times K$ is the dimension-reduced digital precoding matrix.

It should be noted that under the limited number of dominant scatters in the mmWave prorogation environments [8], a beamspace channel $\tilde{\boldsymbol{H}}^\dagger$ (or evenly $\tilde{\boldsymbol{H}}$) has a sparse structure [35], [39]. Consequently, it is obvious from Fig. 2 (b) that a small number of beams can be selected to decrease the effective channel dimension, without causing an evident wastage in the performance. Moreover, a small number of RF chains is needed since a small-size digital precoder $\boldsymbol{P}_r$ is needed. However, in practice, obtaining a beamspace channel in a large size with a limited number of RF chains is challenging. Specifically, the channel dimension is large while the number of RF chains is limited, and the signals on all antennas cannot be sampled simultaneously. Therefore, it can be advantageous to obtain sparser representation of a signal for both channel representation and estimation.

B. DICTIONARY LEARNING FOR SPARSE RECOVERY

A signal $\boldsymbol{r} \in \mathbb{C}^N$ is said to have a sparse representation in $\boldsymbol{\Psi}$ if it can be approximated as $\boldsymbol{r} \approx \boldsymbol{\Psi w}$, where $\boldsymbol{\Psi} \in \mathbb{C}^{N \times K}$ and $\boldsymbol{w} \in \mathbb{C}^K$ denote a sparsifying transform operator and a sparse coding coefficient vector composed mainly of zeros, respectively. For a given $\boldsymbol{r}$ and $\boldsymbol{\Psi}$, $\boldsymbol{w}$ can be obtained through the following sparse recovery process.

$$\arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_0 \; s.t. \; \|\boldsymbol{r} - \boldsymbol{\Psi w}\|_2^2 < \epsilon, \tag{7}$$

where ϵ is an error tolerance.

It is noted that the problem in (7) is NP-hard as one has to solve for the positions and magnitudes of the nonzero elements in $\boldsymbol{w}$. Still, there are two main approaches to approximately solve this problem [44]. The first approach is the family of greedy pursuit algorithms that offer efficient approximate solutions by iteratively minimizing the number of nonzeros in $\boldsymbol{w}$. This is known as the matching pursuit (MP) algorithm. Second is the ℓ_1-relaxation approach that relax ℓ_0 to the ℓ_1 norm. This relaxation offers a loose bound on sparsity but achieves a significant reduction in computational cost. These are known as basis pursuit algorithms. The orthogonal MP (OMP) [45] is a widely used benchmark sparse representation technique of the MP algorithms.

A sparsifying dictionary represents the transformation matrix to a domain in which the signal of interest is sparse. To this end, there are two main families of dictionaries. First is mathematically-defined basis functions, such as the DFT and discrete-cosine transform matrix. These are easy to prepare. However, they may not necessarily transform into the domain that exhibits signal sparsity. Second is learned dictionaries. A learned dictionary, especially if redundant, promotes sparsity, enhances the representation quality, and is locally adaptive to the signals of interest. In essence, this dictionary is composed of prototype signals as its columns. These signals are rich in structure as compared to fixed basis vectors.

In a learned dictionary, one learns a dictionary by training over a set of example training signals $\boldsymbol{R} \in \mathbb{C}^{N \times M}$ through a machine learning procedure referred to as *dictionary learning*, described as follows.

$$\arg\min_{\boldsymbol{W},\boldsymbol{D}} \|\boldsymbol{W}_i\|_0 \; s.t. \; \|\boldsymbol{R}_i - \boldsymbol{D}\boldsymbol{W}_i\|_2^2 < \epsilon \; \forall \; i, \tag{8}$$

where $\boldsymbol{D} \in \mathbb{C}^{N \times K}$ denotes a learned dictionary.

The K-singular value decomposition (K-SVD) algorithm [46] is one of the widely used algorithms for a dictionary learning process. In this algorithm, first, the parameter Λ_i of nonzero elements of the ith row of $\boldsymbol{W}$ is determined for each dictionary atom $\boldsymbol{D}_i$. Then, a partial residual matrix is calculated, and its columns are restricted to the active set of signals that use the ith atom for their sparse approximation. Finally, the atom $\boldsymbol{D}_i$ and the coefficients $\boldsymbol{W}^i_{\Lambda_i}$ are updated using the solution of the best rank-one approximation of the matrix, which can be calculated using its singular value decomposition. More explanation can be found in [44].

III. LEARNED DICTIONARIES FOR BEAMSPACE CHANNEL REPRESENTATION AND ESTIMATION

A. THE PROPOSED DICTIONARY LEARNING ALGORITHM FOR BEAMSPACE CHANNEL REPRESENTATION

The AoDs in an mMIMO system are distributed continuously in the angular domain. However, modeling the lens operator with a DFT basis limits the angular spread to include specific sample points. Thus, an AoD of a specific propagation path should not necessarily be matched by the given sample points. This causes the power of a path to leak onto multiple beams in the beamspace channel [39], as known as power leakage [16]. For a single-UE single-path scenario, when a uniform linear array (ULA) is used, the worst power leakage is [16]

$$\eta_{ULA} = 1 - \frac{1}{2\sum_{i=1}^{N/2} \frac{\sin^2(\pi/2N)}{\sin^2((2i-1)\pi/2N)}}. \tag{9}$$

With the system models considered in this paper, the worst power leakage is around 0.60, according to (9), which is quite high.

Power leakage can be viewed as an imperfection in the sparse representation obtained with a given sparsifying basis. Thus, we compare the quality of a sparse representation over a DFT basis $\boldsymbol{F} \in \mathbb{C}^{N \times N}$ to that over a learned dictionary $\boldsymbol{D} \in \mathbb{C}^{N \times K}$, where $K > N$. In this setting, the signal of interest is a (mmWave mMIMO) beamspace channel $\boldsymbol{h} \in \mathbb{C}^N$. Here, we compare these representations with a sparsity level s.

First, an exact representation of $\boldsymbol{h}$ over $\boldsymbol{F}$ can be obtained using the whole N basis functions (columns) in $\boldsymbol{F}$, as follows.

$$\boldsymbol{h} = \boldsymbol{F}_1 a_1 + \boldsymbol{F}_2 a_2 + \cdots + \boldsymbol{F}_N a_N, \tag{10}$$

where a_1 through a_N signify the representation coefficients of $\boldsymbol{h}$ with respect to $\boldsymbol{F}$. Note that these are obtained by performing an inner product between $\boldsymbol{h}$ and $\boldsymbol{F}$. Then, an s-sparse representation of $\boldsymbol{h}$ over $\boldsymbol{F}$ can be obtained by selecting the most dominant s coefficients. Let us assume that they happen to be the first s coefficients, as follows.

$$\hat{\boldsymbol{h}}_F = \boldsymbol{F}_1 a_1 + \boldsymbol{F}_2 a_2 + \cdots + \boldsymbol{F}_s a_s. \quad (11)$$

Second, with respect to $\boldsymbol{D}$, an s-sparse representation of $\boldsymbol{h}$ is

$$\hat{\boldsymbol{h}}_D = \boldsymbol{D}_1 b_1 + \boldsymbol{D}_2 b_2 + \cdots + \boldsymbol{D}_s b_s. \quad (12)$$

Let us assume that the first s atoms (columns) of $\boldsymbol{D}$ are selected, with the corresponding coefficients b_1 through b_s.

Each dictionary atom is a prototype signal that is rich in structure, as shown in the motivating example of Section III-B. Thus, one can assume that it can be expanded spanning many DFT basis functions. Therefore, it can be written as:

$$\boldsymbol{D}_1 = \boldsymbol{F}_1 c_1 + \boldsymbol{F}_2 c_2 + \cdots + \boldsymbol{F}_K c_K. \quad (13)$$

where K is the number of DFT columns required to represent the dictionary atom $\boldsymbol{D}_1$ with coefficients c_1 through c_K. Also, the atoms $\boldsymbol{D}_2$ through $\boldsymbol{D}_s$ can be expanded using $K+1$ through $K+s-1$ columns from F.

Now, (12) can be rewritten as follows.

$$\begin{aligned}\hat{\boldsymbol{h}}_D = (&\boldsymbol{F}_1 c_1 + \cdots + \boldsymbol{F}_K c_K) b_1 + \ldots \\ &+ (\boldsymbol{F}_1 d_1 + \cdots + \boldsymbol{F}_K d_K) b_s. \end{aligned} \quad (14)$$

(14) is evident that using the same sparsity level, the sparse representation of $\boldsymbol{h}$ over $\boldsymbol{D}$ is s-sparse, in terms of sparsity. On the other hand, it is richer in terms of the structure as it is equivalent to using many columns from $\boldsymbol{F}$ [47]. Said conversely, one can obtain a sparser representation over $\boldsymbol{D}$ with almost the same representation quality.

In view of the above-mentioned motivation, this paper proposes a learned dictionary-based algorithm for channel representation. This algorithm consists of training and testing stages. In the training stage, a set of UL channel realizations is obtained to learn a dictionary. Here note that a set of channel realizations can be obtained by classical channel estimation algorithms in the literature [48]. In these techniques, a training signal is sent from the receiver, and its response is observed at the transmitter end.[1] It is noted that this process will be done periodically (for example, every night) by the BS to learn any far scatterer changes[2] in the environment. After a set of channel realizations is obtained, a dictionary is trained over this set, and the learned dictionary ($\boldsymbol{D}_H$) is obtained.

The dictionary learning (training stage) for the channel representation algorithm begins by using a set of channel

[1]Based on the channel reciprocity, training signal also can be sent by the transmitter and observed at the receiver, but in that case, the channel information should be shared with the transmitter as well since the dictionary learning will be done at the transmitter.

[2]Channel measurements of the signals reflected from the same far scatterers contain signals with similar incident angles [49], [50]. In fact, local scattering changes do not affect the representation of the dictionary. This is because machine learning algorithms (e.g., a dictionary learning algorithm) are powerful for denoising [51] and so they reduce the effect of local scatterers.

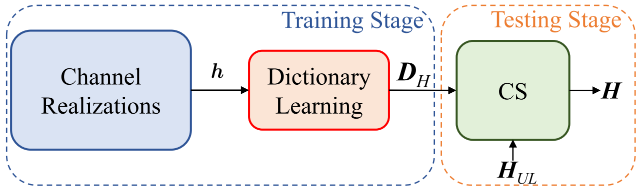


FIGURE 4. The diagram of the proposed dictionary learning-based algorithm to convert the channel to beamspace channel.

information as a dictionary initialization. Then, a training set is further tuned to the dictionary. Here note that this part is a sparse coding process where one calculates the sparse coding coefficient vectors of the given training data based on the current dictionary estimate. In other words, the algorithm implicitly approximates the solution to the ℓ_0-constrained least-squares problem. The main principle behind this iterative algorithm is to use the residual error from the previous iteration to successfully approximate the position of nonzero entries and estimate their values. More details of this part can be found in [52] and [53].

In the testing stage, a CS algorithm is applied with the learned dictionary ($\boldsymbol{D}_H$) and UL channel ($\boldsymbol{H}_{UL}$)[3] to generate a beamspace channel ($\boldsymbol{H}$). Here note that the UL channel only represents the environment (without the effect of the LAA and dictionary learning). This environment can be learned with classical channel estimation algorithms [55]. Then, a beamspace channel can be created based on the learned environment and a CS algorithm. The block diagrams of the training and testing stages are represented in Fig. 4. However, implementing sparse coding over a dictionary requires a particular type of LAA that synthesizes the sparse coding over the dictionary atoms. For this purpose, one can employ a set of classical LAAs. In this setting, the aggregate effect of the composite LAAs can translate to the intended sparse coding over the dictionary. This setting is motivated by the idea that a dictionary atom can be expanded in terms of several DFT basis vectors. Therefore, the dictionary, as a whole, can be cast as a composition of multiple DFT transformations of different orientations. Thus, dictionary learning can be configured with the traditional function of beam selection.

Analogous to the way a dictionary selectively chooses specific basis vectors (called atoms) to represent such a signal, one can use a phase shifter network to selectively choose certain phase shifters to create a specific beam. In other words, beam selection attained by controlling the switches in this network mimics atom selection in a given dictionary based on the atom's similarity to the signal of interest. Several phase shifter networks can be combined to realize dictionary learning. In these networks, some phase shifters are turned off to realize "unselect" [56] and set some phase shifters to shift the phase 0 degree to realize "select" in beam selection. Besides that, the adaptive selecting network [18]

[3]The channel between the receiver and transmitter can be measured at the UL training mode in mmWave mMIMO TDD systems [54].

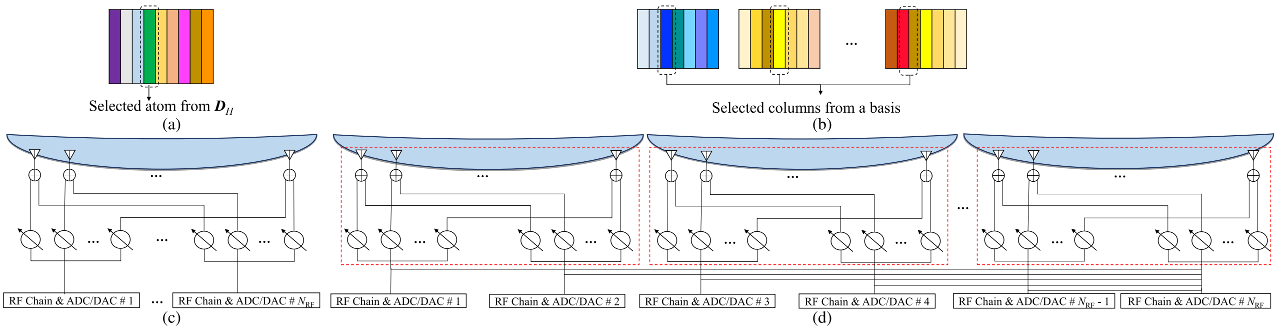


FIGURE 5. Realizing dictionary learning-based beamforming: (a) selecting a certain atom in a learned dictionary of a specific direction can be achieved by (b) selecting multiple basis functions each belonging to a certain DFT basis of a specific general directionality. Correspondingly, (c) achieving certain narrow beamforming can be realized in practice by (d) aggregating the multiple LAAs.

can be directly utilized to design an analog precoder for data transmission, which can further improve performance. One possible way is to extend the simple conjugate analog precoder [57] to scenarios where only one-bit phase shifters are used.

The realization of the proposed algorithm with a phase shifter network for channel representation is illustrated in Fig. 5. In this figure, a phase shifter network is obtained with one-bit phase shifters as in [18] and a certain atom in a learned dictionary of a specific direction is selected thanks to the selection of multiple basis functions, each belonging to a certain DFT basis of a specific general directionality. Therefore, certain narrow beamforming with a learned dictionary can be realized in practice by aggregating the multiple LAAs. Thus, the proposed algorithm necessitates the aggregation of multiple LAAs to achieve specific narrow beamforming using a learned dictionary. This indicates that for the application of the proposed algorithm, there should be the presence and coordination of multiple LAAs.

B. THE PROPOSED DICTIONARY LEARNING ALGORITHM FOR BEAMSPACE CHANNEL ESTIMATION

For the channel estimation, the pilot transmission strategy used in [18], [58], and [59] is applied. All of the UEs transmit pilot sequences to the BS over Q instants to estimate the beamspace channel in the UL of TDD systems. Besides that, the beamspace channel remains unchanged within such channel coherence time as in [60], Q instants are divided into M blocks, and each block consists of K instants such as MK^2. For the mth block, $\mathbf{\Omega}$ pilot matrix with a size of $K \times K$ is used.

For the mth block, we define $\mathbf{\Omega}_m$ of size $K \times K$ as the pilot matrix, which contains K mutually orthogonal pilot sequences transmitted by K UEs over K instants [60], [61]. To normalize the UL pilot power to unit, we apply $\mathbf{\Omega}_m \mathbf{\Omega}_m^H = \mathbf{I}_K$ and $\mathbf{\Omega}_m^H \mathbf{\Omega}_m = \mathbf{I}_K$. Afterward, based on the channel reciprocity in TDD systems [12], the received UL signal matrix can be represented as

$$\tilde{\mathbf{Y}}_m^{\text{UL}} = \mathbf{U}\mathbf{H}\mathbf{\Omega}_m + \mathbf{N}_m = \tilde{\mathbf{H}}\mathbf{\Omega}_m + \mathbf{N}_m, \tag{15}$$

where $m = 1, 2, \ldots, M$ and $\mathbf{N}_m$ is the $N \times K$ noise matrix in the mth block, whose entries are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and σ_{UL}^2 variance, which is also the UL noise power. Since the UL pilot power is normalized to 1, UL SNR is $1/\sigma_{UL}^2$.

During the pilot transmission, the BS employs a combiner $\mathbf{W}_m$ of size $K \times N$ to combine the received UL signal matrix $\mathbf{Y}_m^{\text{UL}}$ (15). Afterward, $\mathbf{S}_m$ of size $K \times K$ is obtained in the baseband sampled by $N_{RF} = K$ RF chains as follows.

$$\mathbf{S}_m = \mathbf{W}_m \tilde{\mathbf{Y}}_m^{\text{UL}} = \mathbf{W}_m \tilde{\mathbf{H}}\mathbf{\Omega}_m + \mathbf{W}_m \mathbf{N}_m. \tag{16}$$

Then, by multiplying the known pilot matrix $\mathbf{\Omega}_m^H$ on the right side of (16) the $K \times K$ measurement matrix $\mathbf{Z}_m$ of the beamspace channel $\tilde{\mathbf{H}}$ is obtained by

$$\mathbf{Z}_m = \mathbf{S}_m \mathbf{\Omega}_m^H = \mathbf{W}_m \tilde{\mathbf{H}} + \mathbf{N}_m^{\text{eff}}, \tag{17}$$

where $\mathbf{N}_m^{\text{eff}}$ is the effective noise matrix. Afterward, we obtain an $Q \times 1$ measurement vector $\bar{z}_k$ for $\tilde{\mathbf{h}}_k$[4] as follows.

$$\bar{z}_k = \begin{bmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{M,k} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_M \end{bmatrix} \tilde{\mathbf{h}}_k + \begin{bmatrix} \mathbf{n}_{1,k}^{\text{eff}} \\ \mathbf{n}_{2,k}^{\text{eff}} \\ \vdots \\ \mathbf{n}_{M,k}^{\text{eff}} \end{bmatrix} \triangleq \bar{\mathbf{W}}\tilde{\mathbf{h}}_k + \bar{\mathbf{n}}_k, \tag{18}$$

where $z_{m,k}$, $\tilde{\mathbf{h}}_k$, and $\mathbf{n}_{1,k}^{\text{eff}}$ are the kth column of $\mathbf{Z}_m$, $\tilde{\mathbf{H}}$, and $\mathbf{N}_m^{\text{eff}}$ in (17), respectively. $\bar{z}_k$, $\bar{\mathbf{W}}$, and $\bar{\mathbf{n}}_k$ are of size $Q \times 1$, $Q \times N$, and $Q \times 1$, respectively. The target is to reliably reconstruct $\tilde{\mathbf{h}}_k$ based on z_k with the number of pilot symbols Q as low as possible. On the other hand, if we directly utilize the selecting network technique [36], [58] to design $\bar{\mathbf{W}}$ (or equivalently $\mathbf{W}_m$ for $m = 1, 2, \ldots, M$), each row of $\bar{\mathbf{W}}$ will have one and only one nonzero element. Therefore, to guarantee that the measurement vector $\bar{z}_k$ contains the complete information of the beamspace channel $\tilde{\mathbf{h}}_k$, the number of pilot symbols Q must be larger than N, which is still high in mmWave

[4]Here, we focus on estimating the beamspace channel $\tilde{\mathbf{h}}_k$ of the kth UE without loss of generality, and a similar method can be directly applied to other UEs to obtain the complete beamspace channel $\tilde{\mathbf{H}}_k$.

mMIMO systems. Therefore, we used the adaptive selecting network [18] for mmWave mMIMO systems with LAA, where the selecting network with switches is replaced by one-bit phase shifters. During the data transmission, the adaptive selecting network is configured to realize the traditional function of beam selection. Moreover, during the beamspace channel estimation, this adaptive selecting network is adaptively used as an analog combiner $\boldsymbol{W}_m$ to combine the UL signals. Here note that $\tilde{\boldsymbol{h}}_k$ is a sparse vector, as the number of dominant scatterers in the mmWave prorogation environments is limited [35]. Thus, by utilizing the adaptive selecting network, $\bar{z}_k$ (18) has the complete information of $\tilde{\boldsymbol{h}}_k$ even if $Q < N$. Afterward, (18) can be formulated as a typical sparse signal recovery problem [62]. Our next goal is to design the analog combiner $\bar{\boldsymbol{W}}$. In CS, to achieve satisfying recovery accuracy, $\bar{\boldsymbol{W}}$ is designed to make the mutual coherence

$$\mu \triangleq \max_{i \neq j} \left| \bar{\boldsymbol{w}}_i^H \bar{\boldsymbol{w}}_j \right| \tag{19}$$

as small as possible, where $\bar{\boldsymbol{w}}_i$ is the ith column of $\bar{\boldsymbol{W}}$. There are already some matrices that enjoy small μ, such as the i.i.d. Gaussian random matrix and Bernoulli random matrix [62]. In this paper, we select the Bernoulli random matrix as the combiner $\bar{\boldsymbol{w}}$, i.e., each element of $\bar{\boldsymbol{w}}$ is randomly selected from $\frac{1}{\sqrt{Q}}\{-1, +1\}$ with equal probability. This is because all elements of $\bar{\boldsymbol{w}}$ share the same normalized amplitude, which phase shifters can realize and the resolution of phase shifter can be only 1 bit since we only need to shift the phase by 0 or π. After $\bar{\boldsymbol{w}}$ designed by the adaptive selecting network, (18) can be solved by the classical CS algorithms, such as OMP and SD, using the DFT basis. However, a DFT basis is essentially a mathematically defined basis function where its basis vectors are defined to quantize the directions in the vector space uniformly. Thus, it is a generic basis, and the success of its representation directly depends on the extent to which a given signal is aligned to the (fixed) basis functions that span the directionality in the vector space. Conversely, a learned dictionary has learned vectors as its columns. These vectors are trainable parameters over a comprehensive set of example signals in a machine learning operation referred to as the dictionary learning/training process. Therefore, each dictionary vector forms a prototype signal, and it is thus commonly referred to as an atom. Hence, dictionary atoms are obtained by learning over training data rather than uniformly sampling the space based on a certain criterion. This learning enjoys the generalization properties of machine learning, i.e., a learned dictionary is expected to work well with new and unforeseen data points. Therefore, this inherent data-fitting property empowers dictionaries to better represent signals of the same class of its training set more sparsely and compactly, as opposed to generic bases like the DFT. In essence, signals of any type may belong to a specific subspace and may not necessarily be spread all over the vector space. Therefore, a custom-made basis like the dictionary is tailored to best-fit data of a certain type, for example, images, channel responses, or beams. Accordingly, a dictionary learning-based algorithm is proposed for channel estimation.

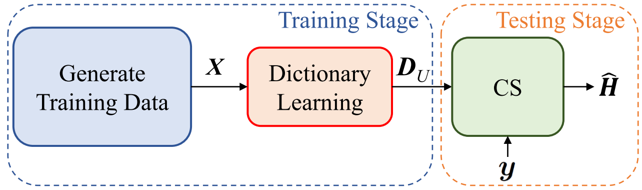


FIGURE 6. The diagram of the proposed dictionary learning-based algorithm for channel estimation.

A block diagram of the proposed channel estimation through a learned dictionary is represented in Fig. 6. This algorithm has training and testing stages as in the previous algorithm. In the training stage, a training set of DFT-based precoding matrices is generated. Afterward, a dictionary is trained over this set, and a learned dictionary is created for channel estimation.

C. COMPRESSIVE SENSING FOR BEAMSPACE CHANNEL ESTIMATION AND REPRESENTATION

Let $\boldsymbol{r} \in \mathbb{C}^n$ denote a vector signal. The notion of CS considers obtaining a compressed measurement $\boldsymbol{r}_c = \boldsymbol{\Phi r}$ where $\boldsymbol{\Phi} \in \mathbb{C}^{m \times n}$ is a measurement/sensing matrix, with $m < n$, rather than measuring every element in $\boldsymbol{s}$. Clearly, an n-to-m dimensionality reduction is made possible by this undersampling operation. It is noted that CS is only applicable to compressible signals, those being sparse explicitly or have a sparse representation in a certain domain [63]. Since $\boldsymbol{s}$ is not necessarily sparse in its own shape, its sparse representation is typically obtained using a sparsifying transform/basis ($\boldsymbol{\Psi}$); either a fixed basis or a redundant (overcomplete) learned dictionary ($\boldsymbol{D}$). In this context, for the case of a learned dictionary, the signal can be approximated as $\boldsymbol{r} = \boldsymbol{Dw}$, where $\boldsymbol{w}$ is a sparse coding coefficient vector having only $s \ll n$ nonzero elements. Obtaining $\boldsymbol{w}$ from $\boldsymbol{y}_c$ can be formulated as follows.

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{w}\|_0 \ s.t. \ \boldsymbol{s}_c = \boldsymbol{\Phi r} = \boldsymbol{\Phi D w}. \tag{20}$$

The inverse problem in (20) is inherently ill-posed. Still, the sparsity of the solution lends itself as an efficient regularizer to this problem under mild conditions. In this regard, the restricted isometry property (RIP) [64] of $\boldsymbol{\Phi}$ assures a unique solution with high probability. In addition, a number of compressed measurements m being at least equal to $(cs \log n/m)$ for some small constant $c > 0$ assures exact recovery according to the robust uncertainty principle [64]. Technically, a variety of sparse recovery techniques can be applied to obtain $\boldsymbol{w}$ given $\boldsymbol{r}_c$, $\boldsymbol{\Phi}$, and $\boldsymbol{D}$. To this end, the fundamental intuition behind CS is measuring only the nonzero elements in $\boldsymbol{w}$. Hence, it resembles a compressed measurement of the original signal. Finally, the original signal can be reconstructed as $\hat{\boldsymbol{r}} = \boldsymbol{Dw}$.

Algorithm 1 Beamspace Channel Representation and Estimation

Input: UL channel $\boldsymbol{H}_{UL}$, channel sparsity s_c, precoding sparsity s_p, a learned dictionary $\boldsymbol{D}_H$ for channel representation and $\boldsymbol{D}_U$ for channel estimation.

Output: A channel impulse response estimate $\hat{\boldsymbol{H}}_U$.

1: Solve: $\boldsymbol{w}_e = \underset{\boldsymbol{w}}{\text{argmin}} \|\boldsymbol{H}_{UL} - \boldsymbol{D}_H \boldsymbol{w}\|_2^2 \ s.t. \ \|\boldsymbol{w}\|_0 < s_c$

2: Obtain a beamspace channel: $\boldsymbol{H} = \boldsymbol{D}_H \boldsymbol{w}_e$

3: Send the signal through the $\boldsymbol{H}$.

4: Obtain $\boldsymbol{Y}$ in the receiver.

5: Solve: $\boldsymbol{w}_u = \underset{\boldsymbol{w}}{\text{argmin}} \|\boldsymbol{Y} - \boldsymbol{D}_U \boldsymbol{w}\|_2^2 \ s.t. \ \|\boldsymbol{w}\|_0 < s_p$

6: Obtain a channel estimate: $\hat{\boldsymbol{H}} = \boldsymbol{D}_U \boldsymbol{w}_u$

The CS algorithm represents the testing stage (run-time operation) for both the proposed channel representation and estimation algorithms. Along with this line, the CS algorithm is applied with the learned dictionary ($\boldsymbol{D}_H$) and UL channel ($\boldsymbol{H}_{UL}$) for the case of channel representation and with the learned dictionary ($\boldsymbol{D}_U$) and received signal ($\boldsymbol{y}$) for the case of channel estimation. The overall testing steps for both channel representation and estimation are given in Algorithm 1. In this stage, a sparse coding vector is obtained according to the UL channel and the learned dictionary of channel representation (Step 1 of Algorithm 1). Then, the beamspace channel is obtained according to this sparse coding vector and the learned dictionary of the channel representation (Step 2 of Algorithm 1). Afterward, the signal is sent through this beamspace channel, and the signal is received by the receiver (Steps 3 and 4 of Algorithm 1). Finally, the channel is estimated according to the learned dictionary of channel estimation and received signal (Steps 5 and 6 of Algorithm 1). Also, the antenna array configuration in mMIMO when the proposed dictionary learning algorithm is used for only channel representation, only channel estimation, and both channel representation and estimation are represented in Figs. 7 (a), (b), and (c), respectively.

D. DISCUSSION ON COMPUTATIONAL COMPLEXITY

The computational complexity of the proposed algorithms mainly depends on sparse coding and dictionary learning. Let us consider the naive OMP algorithm as an example of sparse coding, where it is working on sparse coding of a signal $\boldsymbol{x} \in \mathbb{C}^N$ over a given dictionary $\boldsymbol{D} \in \mathbb{C}^{N\times K}$. Its computational complexity at the kth iteration is $\mathcal{O}(NK+Ks+Ks^2+s^3)$ [65]. With sparsity s, the overall complexity of the OMP algorithm is $\mathcal{O}(NKs+Ns^2+Ns^3+s^4)$. Note that sparse coding is used both during the training and testing stages. The K-SVD [46] algorithm can be considered an example of the dictionary learning process. The total complexity of K-SVD working on a training set $\boldsymbol{X} \in \mathbb{C}^{N\times L}$, with sparsity s and *Num* iterations is $\mathcal{O}(Num(s^2+N)KL)$ [66]. Therefore, the complexity of the OMP algorithm for sparse coding and the

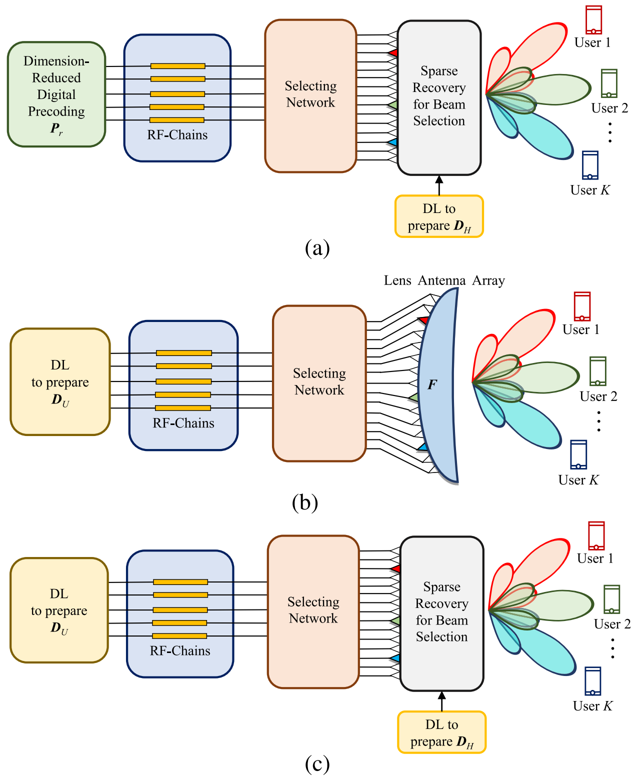


FIGURE 7. Antenna array configuration in mMIMO when (a) a channel representation is made by a learned dictionary, (b) a sparse coding beam selection controlled by a learned dictionary, and (c) a channel representation and a sparse coding beam selection made by learned dictionaries.

K-SVD algorithm for dictionary learning depends on factors such as signal dimensionality, dictionary size, sparsity level, and the number of iterations. On the other hand, for the case of computational complexity, the dictionary learning process is performed only during the training stage and not during the testing stage. Therefore, the computational complexity associated with dictionary learning is not a concern during the run-time operation of the proposed algorithm.

IV. SIMULATIONS AND RESULTS

A. PARAMETER SETTING

This paper considers a mmWave mMIMO system with $N = 256$ antennas and $N_{\text{RF}} = 16$ RF chains. This system simultaneously serves 16 UEs at the receiver end. Two different channel models are used, and the proposed algorithm is tested in both models. These channel models are SV and GSCM. Here note that all the channel samples used for training and testing datasets are generated uniquely using SV and GSCM channel models with the following parameters, as in [67].

With the SV channel model, similar to the experimental setup in [67], the kth UE spatial channel is obtained as a composition of one LoS component and two NLoS components. These are set to have $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$ and $\beta_k^{(i)} \sim \mathcal{CN}(0, 10^{-0.5})$ for $i = 1, 2; 3$. $\psi_k^{(0)}$ and $\psi_k^{(i)}$ follow the i.i.d. uniform distribution within $\psi \in [-0.5, 0.5]$.

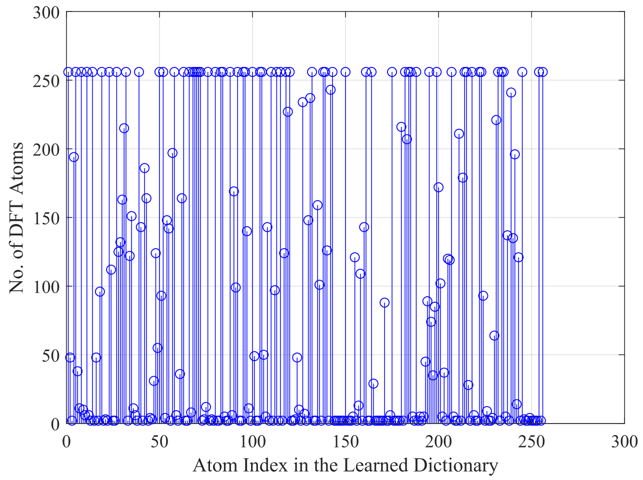


FIGURE 8. Beamspace sampling: a dictionary beam corresponds to a composite of DFT-modeled beams.

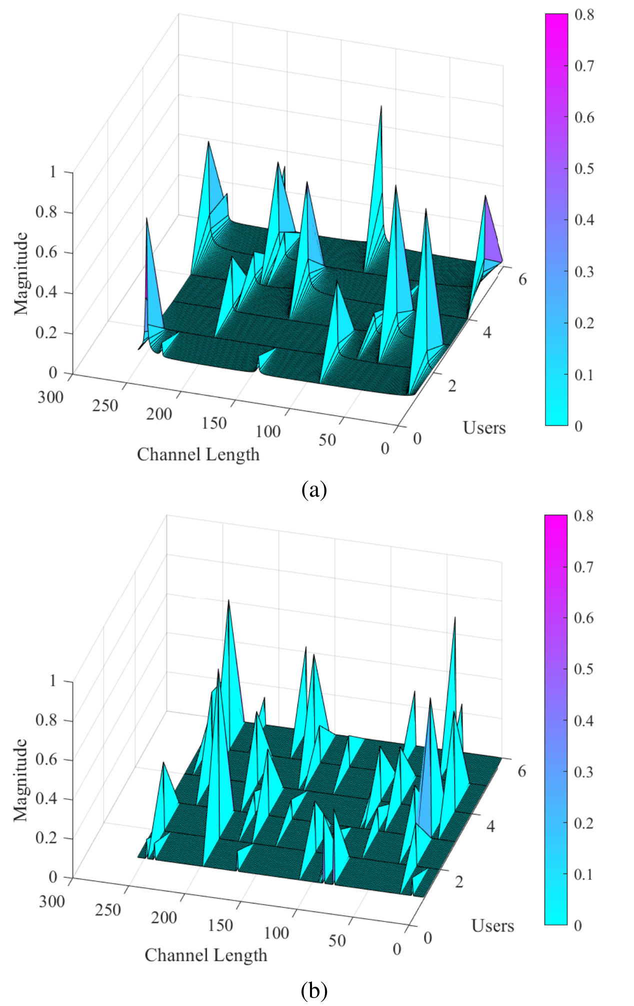


FIGURE 9. Magnitudes of beamspace channel coefficients obtained with (a) the DFT and (b) a learned dictionary for a multiple-UEs multiple-paths scenario.

For simulating the GSCM, the experimental setup used in [42] is used. This setup considers a system made up of a single urban cell with a radius of 1200 meters, with the BS at its center. The DL channel is generated according to the GSCM principles [43] with coefficients provided by the spatial channel model [68]. Also, the azimuth angle θ ranges between $-\pi/2$ and $\pi/2$. As for the scattering environment, the cell has seven fixed-location scattering clusters. The distance between each cluster and the BS is selected randomly in ranges between 300 meters and 800 meters. Four scattering clusters are used for each channel modeling; one is at the UE location, and the remaining three clusters are the closest to the UE from the previously mentioned seven scattering clusters. The UE location is spanned consistently to be between 500 meters and 1200 meters. Under the GSCM guidelines, each scattering cluster has 20 effective propagation subpaths with a 4-degree angular spread.

For dictionary learning, we use a training set of 10000 training vectors using the K-SVD algorithm [46] with 50 iterations, and a sparsity level of 16. Also, the OMP algorithm [45] is used as a sparse approximation algorithm. All the experiments are made with 5000 trials. Also, it is assumed that the true values of the channel realizations are known by the transmitter in the dictionary learning stage for the sake of simplicity.

B. PERFORMANCE EVALUATION

Due to the power leakage and the many nonzero elements, a beamspace channel is not ideally sparse [69]. Therefore, using a better sparsifying transformation allows for revealing the sparsity of the channel in a better fashion. We propose the use of a learned dictionary as a better alternative. Such a learned dictionary is trained in a data-driven manner over example channel realizations and is thus better able to expose intrinsic sparsity patterns of channel responses. This suggests that sparse representation with a learned dictionary is sparser than that with a DFT basis. As an empirical investigation of this proposition, Fig. 8 quantifies how many DFT columns are required to represent each atom in a given learned dictionary. It is seen that on many occasions, a single dictionary atom would require a large number of DFT columns to be represented. This shows that such an atom in a prototype signal with a rich structure as many DFT columns are required in its representation. This result hints at the added benefit of dictionary atoms in achieving high-quality yet much sparser transformation. This is further illustrated in Fig. 9. This figure shows the magnitudes of beamspace channel coefficients obtained by the DFT resembling the space defined by using an LAA and a learned dictionary for multiple-UEs multiple-path scenarios. In both cases, the dictionary is obtained by training it over a set of channel realizations. Any standard dictionary learning algorithm can be used for this purpose, such as the K-SVD algorithm used in this paper. Besides, the dictionary size is 96×256. Also, the SV channel model is used where it has four multiple-path components and is

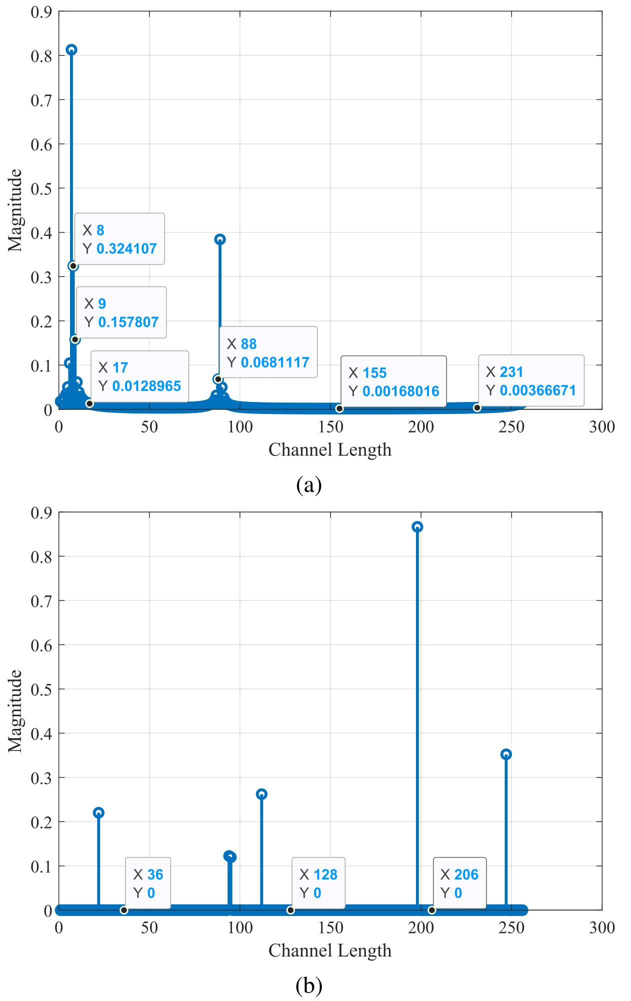


FIGURE 10. **Magnitudes of beamspace channel coefficients obtained with (a) the DFT and (b) a learned dictionary for a single-UE multiple-path scenario.**

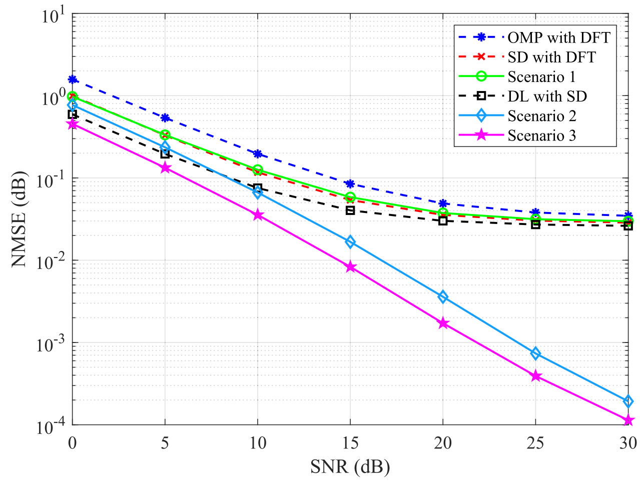


FIGURE 11. **ULA NMSE performance comparison versus SNR with the SV channel model.**

generated according to the specification presented in Section IV. It can be seen from Fig. 9 that DFT magnitudes exhibit side lobes around the nonzero elements, which are the smaller shapes just next to the main shapes. Besides them, even far elements from the main lobes are nonzero. On the other, far elements from the main lobes are zero, and there are no side lobes in the dictionary learning-based algorithm. These are evident beamspace sparsity is enhanced in the space defined by the dictionary. On the contrary, one that was created with dictionary learning does not have such a thing. This is further illustrated in Fig. 10, where analyses are made for a single UE. The figure clearly shows that the DFT-based channel has side lobes, and the dictionary learning-based channel is more sparse. The improved sparsification obtained with the learned dictionary is expected to improve the channel estimation quality.

The channel estimation performance is evaluated in terms of the normalized mean-square error (NMSE) quality metric. Then, the sum-rate performance is considered a secondary quality metric. In this text, we compare the following algorithms.

- *OMP with DFT*: OMP channel estimation when DFT bases are used for channel representation and estimation
- *SD with DFT*: SD algorithm ([18], [67] where OMP-based estimation is followed by a least-squares update exploiting the structure of mmWave mMIMO channels in beamspace) when DFT bases are used for channel representation and estimation
- *Scenario* 1: OMP channel estimation when a DFT basis and a learned dictionary are used for channel representation and estimation, respectively
- *DL with SD*: SD algorithm when a DFT basis and a learned dictionary are used for channel representation and estimation, respectively
- *Scenario* 2: OMP channel estimation when a learned dictionary and DFT basis are used for channel representation and estimation, respectively
- *Scenario* 3: OMP channel estimation when learned dictionaries are used for channel representation and estimation

1) THE QUALITY OF CHANNEL ESTIMATION IN TERMS OF NMSE

The NMSE performance of the aforementioned channel representation and estimation settings versus SNR is investigated. This experiment is first performed with the SV channel model and then with the GSCM. A ULA is considered for both models. The results of these settings are shown in Figs. 11 and 12, respectively. For SD-based channel estimation, we keep the strongest $V = 9$ elements for each channel component and assume that the sparsity level of the beamspace channel for the OMP-based channel estimation is equal to $V(L+1) = 16$. We also assume that all channel estimation algorithms use $Q = 96$, training pilots.

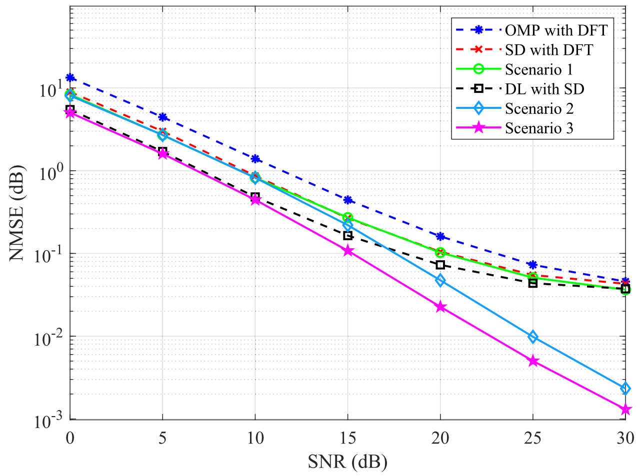


FIGURE 12. ULA NMSE performance comparison versus SNR values with the GSCM.

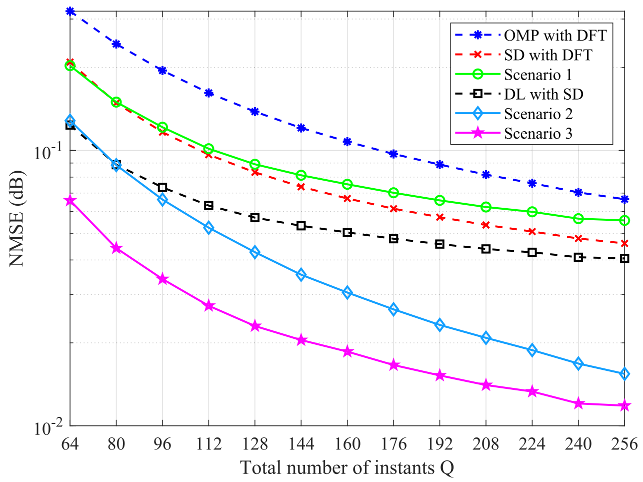


FIGURE 13. ULA NMSE performance comparison against the total number of instants Q for pilot transmission in SV channel model.

In view of Figs. 11 and 12, it is evident that using a learned dictionary in the channel estimation improves the channel estimation quality. This is the case for both OMP-based reconstruction and the SD algorithm. Also, using a learned dictionary channel representation further improves the performance, especially for high SNR values.

Next, the previous experiment is repeated with the difference that SNR is fixed at 10 dB and the number of training pilots (Q) is varied. The results are depicted in Figs. 13 and 14 for the SV channel model and GSCM, respectively. In view of these figures, it is shown that for the same Q, using learned dictionaries for channel representation and estimation improves the NMSE performance. Said equivalently, using a learned dictionary allows for reducing the training overhead for having the same NMSE performance attained with a DFT basis.

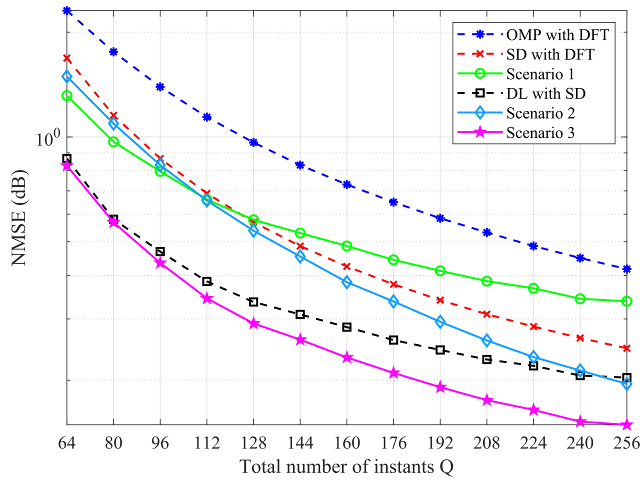


FIGURE 14. ULA NMSE performance comparison against the total number of instants Q for pilot transmission in GSCM.

2) THE QUALITY OF CHANNEL ESTIMATION IN TERMS OF BEAMS SELECTION

The quality of channel estimation is measured in terms of beam selection. The following scenarios are compared for this purpose.

- *FD*: Fully digital zero-forcing (ZF) precoders, included as a benchmark when a DFT basis (*FD with H-beam*1) and a learned dictionary (*FD with H-beam*2) are used for channel representations
- *IA*: Interference-aware (IA) beam selection algorithm [37] which assumes perfect beamspace channel knowledge when a DFT basis (*IA with H-beam*1) and a learned dictionary (*IA with H-beam*2) are used for channel representations
- *IA with SD*: IA fed with a beamspace channel estimate obtained with SD when DFTs are used for channel representation and estimation
- *IA with Scenario* 1: IA fed with a beamspace channel estimate obtained with SD when a DFT basis and a learned dictionary are used for channel representation and estimation, respectively
- *IA with Scenario* 3: IA fed with a beamspace channel estimate obtained with OMP when learned dictionaries are used for channel representation and estimation

Here, the previously mentioned parameter setting is used. The results are shown in Fig. 15.

As clearly seen in Fig. 15, fully digital ZF algorithms achieve the best sum-rate. Next, are the IA algorithms with perfect beamspace channel knowledge. In both cases, dictionary learning-based algorithms are superior to DFT-based algorithms. This verifies the improvement of the channel representation quality when a learned dictionary is used. Also, the performance of the proposed *IA with Scenario* 3 algorithm is very close to the perfect IAs. Besides, the IA algorithm fed with the beamspace channel

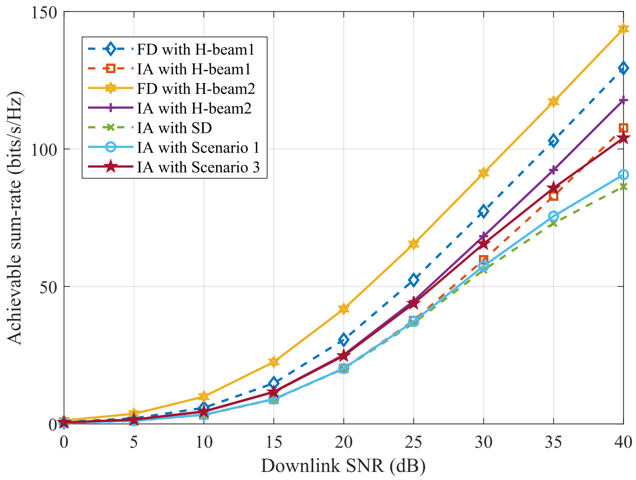


FIGURE 15. ULA sum-rate comparison for DFT and dictionary learning-based algorithms.

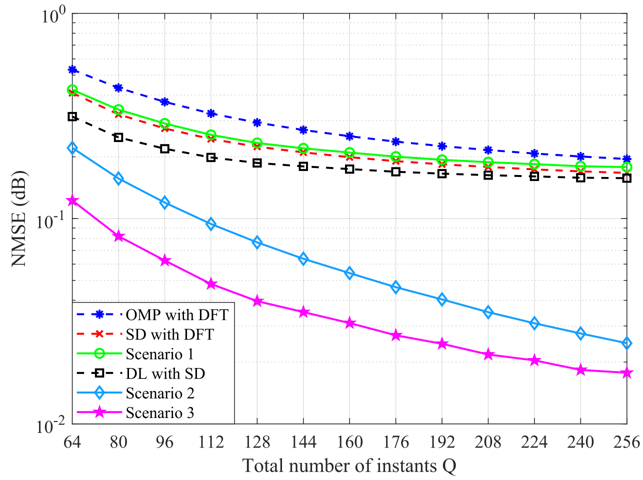


FIGURE 17. NULA NMSE performance comparison against the total number of instants *Q* for pilot transmission for SV channel model.

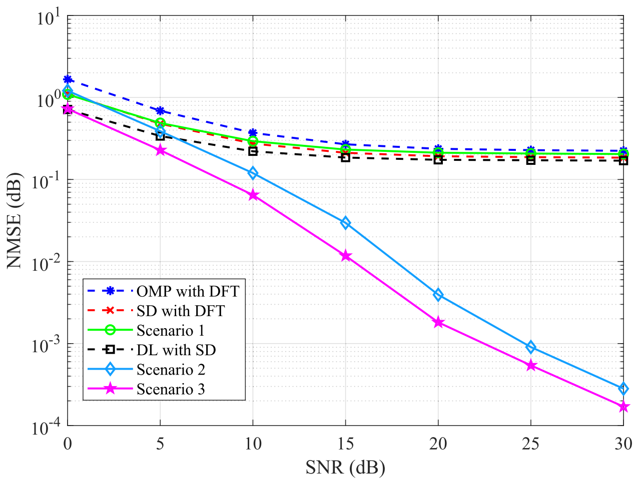


FIGURE 16. NULA NMSE performance comparison against different SNR values for SV channel model.

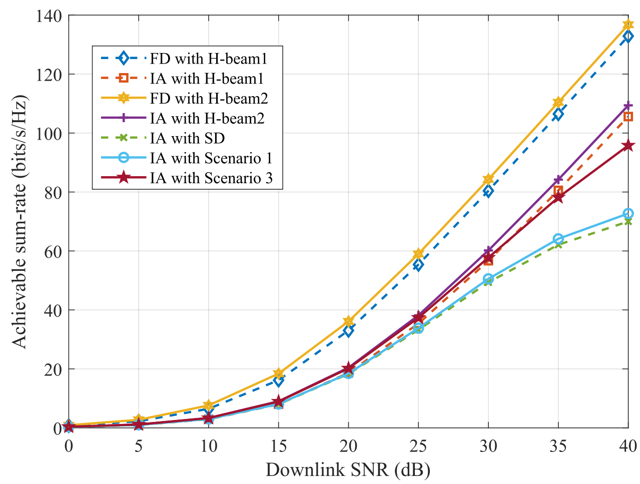


FIGURE 18. NULA sum-rate comparison for DFT and dictionary learning-based algorithms.

estimate obtained with a learned dictionary is consistently better than the case of the SD algorithm.

3) THE QUALITY OF CHANNEL ESTIMATION IN NON-ULA CASE

All the simulations are also done with a non-ULA (NULA).[5] Here, we provide only simulations with the SV channel model to avoid repetition. For the GSCM, similar behavior in the graphs is observed. Figures 16, 17, and 18 show that the behaviors are similar to the case of using a ULA. However, the advantages of using learned dictionaries are more strongly pronounced in the NULA case. This is especially the case with high SNR values. However, in the low SNR regime, the improvement is not significant.

[5]NULA case defines the manufacturing error and evaluates the irregular array geometries is made by assuming that the antenna spacing is uniformly distributed within 0.45λ and 0.55λ, where λ is the carrier wavelength.

V. CHALLANGES AND FUTURE WORK

Implementing the proposed algorithms in a practical setting requires considering various factors beyond the theoretical model. While the theoretical model allows potential effectiveness, practical implementation introduces additional challenges and considerations that should be addressed:

- *Hardware Constraints*: Practical implementation must consider the limitations and constraints of the hardware platform. The proposed algorithm relies on multiple LAAs being available and coordinated, so it is crucial to investigate the feasibility of integrating these components into a real-world communication system. This investigation should include factors such as the cost of phase shifters and power consumption considerations for optimal decision-making.
- *Channel Variations*: Real-world wireless channels can experience variations and uncertainties. To effectively handle different channel conditions, including changes in channel sparsity, it is crucial for the algorithm to

be robust and adaptable. Evaluating the algorithm's performance across various channel scenarios becomes significant to ensure its effectiveness in practical environments.

- *Computational Complexity*: Computational complexity poses a crucial consideration in practical implementation. To ensure efficiency, the algorithm should be designed with suitable computational capabilities and memory resources in mind. In some cases, optimizations like simplifications, vectorization techniques, or hardware accelerations can be necessary to achieve real-time performance. Also, parallel computing techniques, such as distributed computing, can be used to divide the workload among multiple processing units (multiple LAAs).
- *Training and Calibration*: The proposed algorithm may require training and calibration procedures for practical implementation. These procedures involve collecting and processing training data to learn the dictionary and other parameters used in the algorithm. It is crucial that the training process is manageable in terms of time, resources, and scalability. Additionally, calibration techniques might be necessary to address hardware imperfections and achieve accurate beamforming.

To validate the practical feasibility of the proposed algorithm, comprehensive simulations and experimental studies that consider the mentioned factors should be conducted. Real-world measurements and performance evaluations are crucial in assessing the algorithm's effectiveness, limitations, and potential areas for improvement.

VI. CONCLUSION

This paper proposed the use of learned dictionaries as the sparsifying transform operators used in creating beamspace channels in mmWave mMIMO. This corresponded to the use of composite LAAs that enhance the beamspace sparsity. This enhancement led to a more efficient pilot reduction in comparison to the standard case of using LAAs corresponding to fixed basis functions. Dictionary atoms were shown to possess riches structures compared to DFT basis functions. A learned dictionary was shown to reduce the phenomenon of power leakage in mmWave mMIMO due to the use of such atoms. Similarly, we proposed the use of a learned dictionary to function as the precoding operator matrix, meeting the same objective of channel sparsity enhancement. To realize beamspace mmWave mMIMO hardware by a learned dictionary a set of classical LAAs were used. Numerical experiments showed that the proposed algorithms lead to improving the quality of channel estimation and spectral efficiency, as validated in terms so the NMSE and sum-rate performance measure. It was noted that the performance improvement was especially strong in the cases on a NULA. Although there is high-performance improvement using the proposed algorithm, there are several challenges for the effective usage of the proposed algorithms. This should be considered in future work.

REFERENCES

[1] N. Al-Falahy and O. Y. Alani, ''Technologies for 5G networks: Challenges and opportunities,'' *IT Prof.*, vol. 19, no. 1, pp. 12–20, Jan. 2017.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, ''Scaling up MIMO: Opportunities and challenges with very large arrays,'' *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] Z. Pi and F. Khan, ''An introduction to millimeter-wave mobile broadband systems,'' *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[4] S. Kutty and D. Sen, ''Beamforming for millimeter wave communications: An inclusive survey,'' *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.

[5] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin, and A. Ghosh, ''Millimeter wave communications for future mobile networks,'' *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[6] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, ''Spatially sparse precoding in millimeter wave MIMO systems,'' *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, ''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.

[8] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, ''Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,'' *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[9] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, ''MIMO precoding and combining solutions for millimeter-wave systems,'' *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[10] 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, Apr. 2016.

[11] A. Alkhateeb, G. Leus, and R. W. Heath, ''Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[12] T. Kim and D. J. Love, ''Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels,'' in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun./Jul. 2015, pp. 146–150.

[13] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, ''mmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network,'' *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[14] A. Sayeed and N. Behdad, ''Continuous aperture phased MIMO: Basic theory and applications,'' in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2010, pp. 1196–1203.

[15] Y. Zeng and R. Zhang, ''Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm,'' *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[16] T. Xie, L. Dai, D. W. K. Ng, and C.-B. Chae, ''On the power leakage problem in millimeter-wave massive MIMO with lens antenna arrays,'' *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4730–4744, Sep. 2019.

[17] L. Yang, Y. Zeng, and R. Zhang, ''Efficient channel estimation for millimeter wave MIMO with limited RF chains,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[18] X. Gao, L. Dai, S. Han, and X. Wang, ''Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,'' *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.

[19] G. Tang, B. Bhaskar, P. Shah, and B. Recht, ''Compressed sensing off the grid,'' *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[20] H. He, C.-K. Wen, S. Jin, and G. Y. Li, ''Deep learning-based channel estimation for beamspace mmWave massive MIMO systems,'' *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[21] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, ''Deep CNN-based channel estimation for mmWave massive MIMO systems,'' *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 989–1000, Sep. 2019.

[22] Y. Jin, J. Zhang, S. Jin, and B. Ai, ''Channel estimation for cell-free mmWave massive MIMO through deep learning,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10325–10329, Oct. 2019.

[23] Y. Zhang, Y. Mu, Y. Liu, T. Zhang, and Y. Qian, ''Deep learning-based beamspace channel estimation in mmWave massive MIMO systems,'' *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2212–2215, Dec. 2020.

[24] X. Wei, C. Hu, and L. Dai, ''Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems,'' *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 182–193, Jan. 2021.

[25] V. Baranidharan, N. Hariprasath, K. Tamilselvi, S. Vignesh, P. Chandru, A. Srinigha, and V. Yashwanthi, ''Modified Gaussian mixture distribution-based deep learning technique for beamspace channel estimation in mmWave massive MIMO systems,'' in *High Performance Computing and Networking*. Berlin, Germany: Springer, 2022, pp. 383–397.

[26] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, ''Beamspace channel estimation for wideband millimeter-wave MIMO: A model-driven unsupervised learning approach,'' *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1808–1822, Mar. 2023.

[27] Z. Wan, Z. Gao, B. Shim, K. Yang, G. Mao, and M.-S. Alouini, ''Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array,'' *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2337–2342, Feb. 2020.

[28] 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, Dec. 2007.

[29] A. C. Gurbuz, Y. Yapici, and I. Guvenc, ''Sparse channel estimation in millimeter-wave communications via parameter perturbed OMP,'' in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[30] H. Tang, J. Wang, and L. He, ''Off-grid sparse Bayesian learning-based channel estimation for mmWave massive MIMO uplink,'' *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 45–48, Feb. 2019.

[31] C. K. Anjinappa, Y. Zhou, Y. Yapici, D. Baron, and I. Guvenc, ''Channel estimation in mmWave hybrid MIMO system via off-grid Dirichlet kernels,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[32] Y. Gwon, H. T. Kung, and D. Vlah, ''Compressive sensing with optimal sparsifying basis and applications in spectrum sensing,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 5386–5391.

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

[34] M. Nazzal, M. A. Aygul, A. Gorcin, and H. Arslan, ''Dictionary learning-based beamspace channel estimation in millimeter-wave massive MIMO systems with a lens antenna array,'' in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 20–25.

[35] A. Sayeed and J. Brady, ''Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3679–3684.

[36] X. Gao, L. Dai, and A. M. Sayeed, ''Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications,'' *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[37] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, ''Near-optimal beam selection for beamspace mmWave massive MIMO systems,'' *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[38] M. Nazzal, M. A. Aygül, and H. Arslan, ''Channel modeling for 5G and beyond,'' in *Flexible Cognitive Radio Access Technologies 5G Beyond*. 2020, p. 341.

[39] J. Brady, N. Behdad, and A. M. Sayeed, ''Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements,'' *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[40] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, ''State of the art in 60-GHz integrated circuits and systems for wireless communications,'' *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[41] A. M. Sayeed, ''Deconstructing multiantenna fading channels,'' *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[42] 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.

[43] A. F. Molisch, A. Kuchar, J. Laurila, K. Hugl, and R. Schmalenberger, ''Geometry-based directional model for mobile radio channels?Principles and implementation,'' *Eur. Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359, 2003.

[44] M. Nazzal, ''Structural dictionary learning and sparse representation with signal and image processing applications,'' Ph.D. dissertation, Dept. Elect. Electron. Eng., Eastern Mediterranean Univ. (EMU), Turkish Republic Northern Cyprus, 2015.

[45] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, ''Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,'' in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA,1993, pp. 40–44.

[46] M. Aharon, M. Elad, and A. Bruckstein, ''K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,'' *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[47] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge, U.K.: Cambridge Univ. Press, Oct. 2015.

[48] 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.

[49] 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.

[50] A. F. Molisch and F. Tufvesson, ''Propagation channel models for next-generation wireless communications systems,'' *IEICE Trans. Commun.*, vol. 97, no. 10, pp. 2022–2034, 2014.

[51] P. Kaur, G. Singh, and P. Kaur, ''A review of denoising medical images using machine learning approaches,'' *Current Med. Imag. Rev.*, vol. 14, no. 5, pp. 675–685, Sep. 2018.

[52] S. Bahmani, B. Raj, and P. T. Boufounos, ''Greedy sparsity-constrained optimization,'' *J. Mach. Learn. Res.*, vol. 14, pp. 807–841, Mar. 2013.

[53] B. Dumitrescu and P. Irofti, *Dictionary Learning Algorithms and Applications*. London, U.K.: Bantam, 1988.

[54] X. Sun, C. Qi, and G. Y. Li, ''Beam training and allocation for multiuser millimeter wave massive MIMO systems,'' *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.

[55] H. Arslan and G. E. Bottomley, ''Channel estimation in narrowband wireless communication systems,'' *Wireless Commun. Mobile Comput.*, vol. 1, no. 2, pp. 201–219, 2001.

[56] P.-Y. Chen, C. Argyropoulos, and A. Alu, ''Terahertz antenna phase shifters using integrally-gated graphene transmission-lines,'' *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1528–1537, Apr. 2013.

[57] L. Liang, W. Xu, and X. Dong, ''Low-complexity hybrid precoding in massive multiuser MIMO systems,'' *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[58] J. Hogan and A. Sayeed, ''Beam selection for performance-complexity optimization in high-dimensional MIMO systems,'' in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2016, pp. 337–342.

[59] J. H. Kotecha and A. M. Sayeed, ''Transmit signal design for optimal estimation of correlated MIMO channels,'' *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

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

[61] H. Xie, F. Gao, S. Zhang, and S. Jin, ''A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,'' *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

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

[63] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, ''Signal processing with compressive measurements,'' *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[64] E. J. Candes, J. Romberg, and T. Tao, ''Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,'' *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[65] B. L. Sturm and M. G. Christensen, ''Comparison of orthogonal matching pursuit implementations,'' in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 220–224.

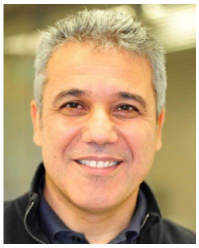
[66] K. Skretting and K. Engan, ''Recursive least squares dictionary learning algorithm,'' *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2121–2130, Apr. 2010.

[67] L. Dai, X. Gao, S. Han, I. Chih-Lin, and X. Wang, ''Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,'' in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Jul. 2016, pp. 1–6.

[68] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*, document 25.996 Version 12.0.0, Release 12, Sep. 2014.
[69] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

MEHMET ALI AYGÜL (Student Member, IEEE) received the B.S. degree in electronics and telecommunication engineering from Yildiz Technical University, in 2018, and the M.S. degree in electrical and electronics engineering and cyber systems from Istanbul Medipol University, in 2020. He is currently pursuing the Ph.D. degree in telecommunication engineering with Istanbul Technical University. He has been a member of the Communications, Signal Processing, and Networking Center (CoSINC) Group, as a Research Assistant. He is currently a Research Engineer with Vestel. His research interests include wireless communication, channel estimation, spectrum prediction, physical layer security, Wi-Fi sensing, and machine learning. He is a Voting Member of the IEEE 802.11 Standards and a Delegate to 3GPP Standards.

MAHMOUD NAZZAL (Member, IEEE) received the B.Sc. degree in electrical engineering from Birzeit University, in 2009, and the M.Sc. and Ph.D. degrees in electrical and electronic engineering from Eastern Mediterranean University, in 2010 and 2015, respectively. He was a Lecturer with the Electrical and Electronic Engineering Department, Eastern Mediterranean University, from 2015 to 2016. He was also a Lecturer with the Izmir University of Economics, from 2016 to 2017. Since July 2017, he has been a Postdoctoral Researcher with Istanbul Medipol University. His research interests include sparse coding, compressive sensing, computer vision, and signal processing for wireless communications.



HÜSEYIN ARSLAN (Fellow, IEEE) received the B.S. degree from Middle East Technical University (METU), Ankara, Turkey, in 1992, and the M.S. and Ph.D. degrees from Southern Methodist University (SMU), Dallas, TX, USA, in 1994 and 1998, respectively.

From January 1998 to August 2002, he was with the Research Group of Ericsson, where he was involved with several projects related to 2G and 3G wireless communication systems. In December 2013, he joined Istanbul Medipol University, to found the engineering college, where he was the Dean of the School of Engineering and Natural Sciences. Since 2022, he has been a Professor with the Department of Electrical Engineering, University of South Florida. He is a Distinguished Lecturer and a member of the Turkish Academy of Sciences. In addition, he was a part-time Consultant for various companies and institutions, including Anritsu Company and the Scientific and Technological Research Council of Turkey. He conducts research in wireless systems, with an emphasis on the physical and medium access layers of communications. He has been collaborating extensively with key national and international industrial partners. His research has generated significant interest in companies, such as InterDigital, Anritsu, NTT DoCoMo, Raytheon, Honeywell, and Keysight Technologies. Collaborations and feedback from industry partners has significantly influenced his research. In addition to his research activities, he has also contributed to wireless communication education. He has integrated the outcomes of his research into education, which lead him to develop a number of courses with the University of South Florida and Istanbul Medipol University. He has developed a unique "Wireless Systems Laboratory" course (funded by the National Science Foundation and Keysight Technologies), where he was able to teach not only the theory but also the practical aspects of wireless communication systems with the most contemporary test and measurement equipment. His current research interests include 6G and beyond radio access technologies, physical layer security, interference management (avoidance, awareness, and cancelation), cognitive radio, multi-carrier wireless technologies (beyond OFDM), dynamic spectrum access, co-existence issues, non-terrestrial communications (high altitude platforms), joint radar (sensing), and communication designs.

Dr. Arslan has served as the general chair, the technical program committee chair, a session and symposium organizer, the workshop chair, and a technical program committee member for several IEEE conferences. He is currently a member of the editorial board of the IEEE Communications Surveys and Tutorials and *Sensors*. He has also served as a member of the editorial board for the IEEE Transactions on Communications, the IEEE Transactions on Cognitive Communications and Networking, and several other scholarly journals by Elsevier, Hindawi, and Wiley.