Convolutional Sparse Coding Based Channel Estimation for OTFS-SCMA in Uplink

Anna Thomas, *Student Member, IEEE*, Kuntal Deka, Patchava Raviteja, and Sanjeev Sharma, *Senior Member, IEEE*

Abstract—Orthogonal time frequency space (OTFS) has emerged as the most sought-after modulation technique in a high mobility scenario. Sparse code multiple access (SCMA) is an attractive code-domain non-orthogonal multiple access (NOMA) technique. Recently a code-domain NOMA approach for OTFS, named OTFS-SCMA, is proposed. OTFS-SCMA is a promising framework that meets the demands of high mobility and massive connectivity. This paper presents a channel estimation technique based on the convolutional sparse coding (CSC) approach for OTFS-SCMA in the uplink. The channel estimation task is formulated as a CSC problem following a careful rearrangement of the OTFS input-output relation. We use an embedded pilot-aided sparse-pilot structure that enjoys the features of both OTFS and SCMA. The existing channel estimation techniques for OTFS in multi-user scenarios for uplink demand extremely high overhead for pilot and guard symbols, proportional to the number of users. The proposed method maintains a minimal overhead equivalent to a single user without compromising on the estimation error. The results show that the proposed channel estimation algorithm is very efficient in bit error rate (BER), normalized mean square error (NMSE), and spectral efficiency (SE).

Index Terms—OTFS, SCMA, NOMA, channel estimation, compressive sensing, convolutional sparse coding.

I. INTRODUCTION

A. Motivation

IN CURRENT communication standards for 4G and 5G, orthogonal frequency division multiplexing (OFDM) has been undisputedly recommended for modulation. OFDM was particularly designed to eliminate the inter-symbol-interference (ISI) caused by the time dispersion of the channel. The success of OFDM depends on the orthogonality of the sub-carriers. If the channel introduces frequency dispersion too in the form of Doppler shifts, the orthogonality of

the sub-carriers is destroyed. This results in inter-carrier-interference (ICI), which hinders the application of OFDM in high Doppler scenario. The coming generation of wireless networks implicates high Doppler due to increased mobility in environments like vehicle-to-everything (V2X) and high carrier frequency such as in mmWave communication. Orthogonal time frequency space (OTFS) has turned up as a solution to boost up the potential of such communication scenario [1]. The remarkable performance of OTFS is attributed to the two-dimensional (2D) modulation technique in which both the data and the channel are represented in the delay-Doppler (DD) domain, by exploiting the quasi-periodicity property in this domain. OTFS modulation benefits from the existence of 2D localised pulses in the DD domain. These pulses subsequently occupy the entire time-frequency (TF) grid, thereby achieving full diversity. All the symbols in an OTFS DD frame experience nearly the same channel. Moreover, fewer parameters are required to describe the channel in the DD domain facilitating a sparse representation, which makes the channel estimation an easy task.

The superior performance of OTFS is evident in multi-user scenarios also. There can be two approaches: (1) orthogonal multiple access (OMA) and (2) non-orthogonal multiple access (NOMA). In NOMA, unlike in OMA, various users share resources. The spectral efficiency of NOMA is significantly better than that of OMA. OTFS-OMA with different interleaving patterns was proposed in [2], [3]. OTFS with power-domain NOMA was explored in [4], [5]. Recently, we proposed a code-domain NOMA approach for OTFS based on sparse code multiple access (SCMA), named OTFS-SCMA [6]. The simulation results and diversity analysis showed the superiority of OTFS-SCMA over other multi-user techniques of OTFS. Any channel estimation technique for OTFS can be easily extended to OTFS-SCMA in downlink, unlike in uplink. Thus, it is highly relevant to devise a channel estimation technique for OTFS-SCMA in uplink.

B. Related Prior Works

The detection of an OTFS-based system necessitates an accurate channel estimation. OTFS-SCMA integrates the techniques of modulation of OTFS and the multiple access scheme of SCMA. Hence, to devise the channel estimation algorithm of OTFS-SCMA, the conventional methods used for both OTFS and SCMA are to be analyzed. The commonly-adopted

approach of channel estimation for OTFS is the pilot-aided method presented in [7]. With the help of the pilot symbols placed in the DD grid, the channel coefficients are estimated. Following this approach, the authors in [8] presented a systematic method for channel estimation using an embedded single QAM pilot symbol with a guard band. By applying thresholding in the observation region, paths are identified, and the channel coefficients are estimated by element-by-element division. The same method was also extended to the MIMO and the multiuser cases [8]. In [9], the authors adopted a sparse signal recovery approach using the sparse Bayesian learning (SBL) algorithm. In this method, the embedded pilot structure does not contain a guard band; instead, it has multiple QAM pilot symbols. Reference [10] also follows a similar pilot-aided technique for channel estimation of OTFS-MIMO. Channel estimation techniques for massive MIMO-OTFS are developed in [11]–[13], where a 3D channel model of delay-Doppler-angle is considered. For channel estimation, 3D-structured Newtonised Orthogonal Matching Pursuit (NOMP), OMP, and SBL algorithms are used in [11], [12], and [13], respectively. The channel estimation for OTFS-OMA is carried out in [14] using a sparse signal recovery approach based on subspace pursuit (SP) and OMP. An embedded QAM pilot symbol with a guard band is used for OTFS-OMA channel estimation in [2]. In the context of the SCMA channel estimation, the reference [15] considered both pilot-aided and data-aided approaches and analyzed the trade-off between the spectral efficiency and pilot overheads. Active user detection followed by channel estimation for SCMA was investigated in [16], [17]. The method in [18] considered sparse pilot vectors with the non-zero pilot symbols placed according to the pattern of the corresponding codebook of a user.

If the existing OTFS-OMA channel estimation techniques are extended to OTFS-SCMA, we face either of the two following major issues in uplink. First, we must rely on a dedicated OTFS frame to train pilot symbols, resulting in higher complexity. Secondly, a significantly large pilot and guard band overhead are required if we insert the pilots in the same frame, along with the data symbols, followed by a subsequent channel estimation based on thresholding. This loss in resources is proportional to the number of transmitting antennas or users. To deal with these issues, we propose a channel estimation technique where the pilot symbols are non-orthogonal, embedded in the data frame itself, and follow a sparse structure. Additionally, we formulate the problem as a CSC model to reduce the complexity of the sparse signal recovery algorithm.

C. Contributions

This paper proposes a channel estimation method for OTFS-SCMA based on convolutional sparse coding (CSC). This method is found to provide impressive bit-error-rate (BER) performance with minimal overhead and complexity. The special features and the main contributions of the proposed method are summarized below:

- The work presented in this paper is the first attempt to perform channel estimation for OTFS in any NOMA

environment, let alone an SCMA system. Although OTFS was studied in the context of power-domain NOMA in [4], [5], [19], no channel estimation techniques were proposed.
- The proposed pilot structure is non-orthogonal and data-embedded and thus offers low pilot and guard band overhead, which is the same as a single user. This aspect of the technique means all users' pilot vectors and guard bands share the same resources, and the data symbols are also transmitted in the same frame. While in [8] and [2], the overhead needed for channel estimation is proportional to the number of users.
- In addition, the proposed pilot vectors have a sparse structure. They follow the same sparse and resource sharing pattern in accordance with the factor matrix of the underlying SCMA system. The sparsity reduces the interference among the pilot vectors of multiple users, owing to the lower degree of the factor nodes [6].
- The input-output relation for OTFS-SCMA in uplink is developed to formulate the CE as a CSC problem. This formulation significantly reduces the dimensionality of the problem facilitating the use of low-complexity SP-based recovery algorithms.
- For solving the CSC-based channel estimation problem, SP is considered. For the successful operation of SP, the knowledge of sparsity is a prerequisite. To meet this requirement, we convert the unknown sparsity of paths to a known sparsity of the number of users by exploring the properties of the standard propagation channel models and the CSC model.
- The initialization of the estimates needed for SP is done sequentially for each user in a greedy manner, resulting in fast convergence of the algorithm. While in [20] and [14], no prior information is considered to obtain the initial estimates.
- Instead of using the random Gaussian sequences, the pilot vectors are designed to minimize the dictionary's mutual coherence. For this optimization, differential evolution [21] is considered.

D. Outline

Section II describes the preliminaries of OTFS and SCMA. This section also presents various features of OTFS-SCMA and the concept of CSC. The proposed channel estimation technique and its analysis are presented in Section III, highlighting the embedded pilot-aided structure and the CSC modeling. The CSC-based sparse signal recovery algorithm for channel estimation is presented in Section IV. Section V presents the simulation results and their analysis. Finally, the paper is concluded in Section VI.

Notations: Boldface upper-case, boldface lower-case, and lower-case letters denote the matrices, the vectors and the scalars respectively. For an $m \times n$ matrix $\mathbf{A}$, vec$(\mathbf{A})$ denotes the $mn \times 1$ column vector which is obtained by vertical concatenation of the n columns of $\mathbf{A}$. $\mathbb{I}_N$ denotes the identity matrix of size $N \times N$. The all zero matrix of size $n \times m$ is denoted by $\mathbf{0}_{n \times m}$. For a matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^\dagger$

represent the transpose, the Hermitian transpose, and the pseudo-inverse of $\mathbf{A}$ respectively. For any real number x, $\lceil x \rceil$ is the smallest integer that is not smaller than x. For any integers k and N, the notation $[k]_N$ refers to $(k \mod N)$. $\circledast$ denotes circular convolution and $\odot$ denotes Hadamard product. $A = |\mathbb{A}|$ denotes the cardinality of the modulation alphabet $\mathbb{A}$. $\mathbb{C}$ denotes the set of complex numbers. The notation $\mathcal{CN}(0, \sigma^2)$ denotes a zero-mean complex Gaussian random number with variance σ^2. Tx and Rx denote transmitter and receiver, respectively.

II. PRELIMINARIES

A. OTFS

We discuss the basic operations involved in OTFS without delving into the conceptual details. OTFS modulation considers an $N \times M$ DD grid (N Doppler bins, M delay bins) for the input data and the corresponding output data. The delay and the Doppler bins are considered in the horizontal and the vertical directions of the DD grid respectively. As the input and the output data are perceived in the DD domain rather than the conventional TF domain, OTFS modulation includes an additional pre-processing block of inverse symplectic finite Fourier transform (ISFFT) and a post-processing block of SFFT. This version of OTFS makes it compatible with the existing OFDM system. At first, the OTFS modulator converts the input data $x[k, l]$ in the DD domain to the symbols $X[n, m]$ in the TF domain using ISFFT operation: $X[n, m] = \frac{1}{MN} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k, l] e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}$. The TF data is converted to time-domain signal $s(t)$ by applying Heisenberg transform: $s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n, m] e^{j2\pi m \triangle f(t - nT)} g_{\text{tx}}(t - nT)$ where $g_{\text{tx}}(t)$ is the transmit basis pulse. The signal $s(t)$ is transmitted through a wireless communication channel whose DD-domain response is $h(\tau, \nu)$. Hence the received signal in time domain is given by $r(t) = \iint h(\tau, \nu) e^{j2\pi\nu(t-\tau)} s(t - \tau) \, d\tau d\nu + z(t)$, where $z(t)$ is the additive white Gaussian noise (AWGN). At the receiver, by applying Wigner transform, the time-domain signal is converted to TF domain: $Y[n, m] = \int e^{-j2\pi\nu(t-\tau)} g_{\text{rx}}^*(t - \tau) r(t) \, dt \mid_{\tau = nT, \, \nu = m\triangle f}$, where $g_{\text{rx}}(t)$ is the receive basis pulse. Finally, SFFT converts the TF signal back to the DD-domain: $y[k, l] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y[n, m] e^{-j2\pi(\frac{nk}{N} - \frac{ml}{M})}$.

The benefit of DD-domain processing is fully realized if the pulses $g_{\text{tx}}(t)$ and $g_{\text{rx}}(t)$ satisfy the so-called *bi-orthogonality* property [22], under which case they are called *ideal* pulses. For ideal pulse-shaping, the input-output relation in the DD domain is given by

$$y[k, l] = \sum_{i=1}^{P} h_i x[[k - k_i]_N, [l - l_i]_M] + z[k, l] \quad (1)$$

where, $k = 0, 1, \ldots, N-1$, $l = 0, 1, \ldots, M-1$; P is the total number of paths; $h_i = h_i' e^{-j2\pi\nu_i\tau_i}$, $h_i' \sim \mathcal{CN}(0, \frac{1}{P})$, k_i, and l_i denote the complex channel gain, integer Doppler, and integer delay tap, respectively of the ith path; and $z[k, l]$ is the complex AWGN. The input-output relation in (1) can be compactly expressed as $\mathbf{y} = \mathbf{Hx} + \mathbf{z}$, where $\mathbf{H} \in \mathbb{C}^{NM \times NM}$; $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are the $NM \times 1$ input, output, and noise vectors

formed by $\text{vec}(\cdot)$ of the corresponding $N \times M$ grids. Note that $\mathbf{H}$ follows a sparse and circulant-block structure [23].

B. SCMA

SCMA is a code-domain NOMA technique in which the available K orthogonal resources (time-slot/frequency-band/code) are shared among J users ($J > K$) [24], [25]. This structure is denoted by (J, K) SCMA system with an overloading factor of $\lambda = \frac{J}{K} > 100\%$. Each user has a specific codebook having A codeword vectors of length K. The codebooks are designed in such a way that each of the K resources is shared by d_f users and each codeword has only d_v non-zero components. As an example, consider a ($J = 6, K = 4$) SCMA system with $\lambda = 150\%$. The set of codebooks can be represented by $\{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_6\}$. For $A = 4$, the codebook for the jth user is given by $\mathbf{C}_j = [\mathbf{c}_{j1} | \mathbf{c}_{j2} | \mathbf{c}_{j3} | \mathbf{c}_{j4}]$ where $\mathbf{c}_{ji} \in \mathbb{C}^{4 \times 1}$, $i = 1, \ldots, 4$. If the input data of the jth user is i, the corresponding codeword vector $\mathbf{c}_{ji}$ is selected. In this way, 6 codeword vectors are identified and they are simultaneously transmitted over the 4 available resources. Due to the sparse structure of the SCMA codewords, message passing algorithm (MPA) can be successfully used for the data detection [26], [27].

C. OTFS-SCMA

OTFS-SCMA is a code-domain NOMA approach for OTFS recently proposed in [6]. The key features of OTFS-SCMA are described in the following.

1) Codeword Allocation Schemes: The significant difference of OTFS-SCMA from the other related works is that it uses SCMA vector codewords as data symbols instead of QAM symbols. In [6], three schemes of SCMA codeword allocation are presented. In *Scheme-1*, the codewords are placed along the Doppler axis as $K \times 1$ vectors and hence $[N]_K = 0$. *Scheme-2* allocates the codewords as $1 \times K$ vectors along the delay axis with $[M]_K = 0$. Also, a third scheme, *Scheme-3* is analyzed where the non-zero components of the codewords are swapped in a particular fashion after allocating the codewords as per *Scheme-1* or *Scheme-2*. Note that, in all the three schemes, the overall overloading factor of OTFS-SCMA is the same as that of the underlying basic (J, K) SCMA system, i.e., $\lambda = J/K$.

2) Downlink and Uplink: In the downlink scenario, the codewords from J SCMA encoders are superimposed first, followed by an OTFS modulator. The input-output relationship of the jth user is given by

$$\mathbf{y}_j = \mathbf{H}_j \mathbf{x}_{\text{sum}} + \mathbf{z}_j \quad (2)$$

where $\mathbf{x}_{\text{sum}}$ is the superimposed input and $\mathbf{H}_j$ is the channel matrix for the jth user. For a particular user, the received data is first passed through an OTFS detector (LMMSE detector using $\mathbf{H}_j$) to resolve the DD interference. The output of the OTFS detector is a noisy version of $\mathbf{x}_{\text{sum}}$. Finally, an SCMA detector (MPA for AWGN channel) acts upon the noisy $\mathbf{x}_{\text{sum}}$ to detect the user's data by removing the multi-user interference.
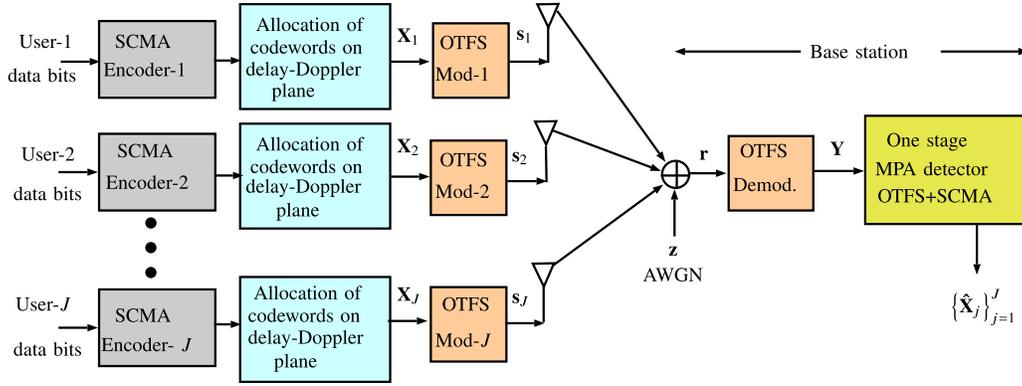
Fig. 1.   Block diagram of OTFS-SCMA in uplink.

Fig. 1 depicts an uplink scenario of OTFS-SCMA. The input-output relation is governed by

$$y = \sum_{j=1}^{J} H_j x_j + z = Hx + z \qquad (3)$$

where $H = [H_1, \ldots, H_J] \in \mathbb{C}^{NM \times JNM}$ and $x = [x_1^T, \ldots, x_J^T]^T \in \mathbb{C}^{JNM \times 1}$. Observe that H has the combined effect of both OTFS and SCMA. It is not possible to segregate the DD interaction and the multi-user fusion. Thus the sequential OTFS and SCMA detection is not feasible in uplink. A combined detector for OTFS-SCMA is proposed in [6] using MPA, which resolves the DD and the multi-user interference in single stage.

3) Channel Estimation in Downlink: The downlink scenario explained above indicates that the multipath channel values are required only for the OTFS detector. For the SCMA detector, only the effect of the AWGN channel remains. Hence, the embedded QAM-pilot-based channel estimation technique with a guard band from [8] is successfully extended to OTFS-SCMA in downlink. The results presented in [6] demonstrate that a single QAM pilot is sufficient for the channel estimation with a pilot power of 35 dB so that the BER performance closely follows that of the perfect CSI case. This strategy is not applicable for OTFS-SCMA in uplink due to the very high guard band overhead. Hence, in this paper, we propose an efficient channel estimation technique suitable for uplink scenarios of OTFS-SCMA.

4) Diversity and BER Analysis: The diversity analysis in [6] shows that OTFS-SCMA can achieve a significantly higher asymptotic diversity order than OTFS-OMA. The use of vector codewords instead of the conventional QAM symbols paves the way for the diversity gain of OTFS-SCMA. *Theorem 1* presented in [6] shows the diversity orders of *Scheme-1* and *Scheme-2* for both uplink and downlink. While the diversity gain of *Scheme-1* depends on the number of distinct mod-K Doppler taps, for *Scheme-2*, it is related to the number of distinct mod-K delay taps. Note that *Scheme-1* and *Scheme-2* perform equally for high values of N and M. *Scheme-3* can achieve a higher diversity order by interleaving the non-zero components of the SCMA codewords based on the channel information at the transmitter. The simulation results obtained

for the practical EVA channel model [28] and different overloading factors agree with the diversity analysis, which shows that OTFS-SCMA can perform remarkably better than other multi-user OTFS schemes.

D. Convolutional Sparse Coding

Convolutional sparse coding (CSC) is a structured coding technique that has found applications in many signal processing problems. CSC was initially applied to one-dimensional signals [29] and then extended to two-dimensional ones [30]. In CSC, a signal $y \in \mathbb{C}^{U \times 1}$ is represented as the sum of v convolutions:

$$\min_{x_i} f(x_i) \text{ s.t } \sum_{i=1}^{v} d_i \circledast x_i = y \qquad (4)$$

where $f(\cdot)$ is usually ℓ_1-norm, $\{d_i\}_{i=1}^{v} \in \mathbb{C}^{u \times 1}$ are support filters $(p \ll U)$, and $\{x_i\}_{i=1}^{v} \in \mathbb{C}^{U \times 1}$ are vectors of varying sparsity. The CSC problem of (4) can be compactly represented as [31], [32]

$$\min_{\tilde{x}} f(\tilde{x}) \text{ s.t } \tilde{D}\tilde{x} = y \qquad (5)$$

where $\tilde{D} = \begin{bmatrix} \tilde{D}_1 & \tilde{D}_2 & \ldots & \tilde{D}_v \end{bmatrix}$, $\tilde{D}_i \in \mathbb{C}^{U \times U}$ is the circulant matrix formed by d_i and its $U-1$ circularly shifted vectors and $\tilde{x} = \begin{bmatrix} x_1^T & x_2^T & \ldots & x_v^T \end{bmatrix}^T$. The sparsity of the signal $\tilde{x}$ makes it possible to use sparse signal recovery algorithms.

Recently, less complex CSC algorithms are developed based on *working locally; thinking globally* [33]. Under this local paradigm, the global dictionary is broken down into smaller local dictionaries D_L of dimension $p \times v$, where $D_L = [d_1 \ d_2 \ \ldots \ d_v]$. Through a simple permutation of its columns, $\tilde{D}$ can be represented as the concatenation of circularly shifted versions of D_L. Fig. 2 depicts the process of converting the global problem into a local one.

The CSC problem from (5) can now be written as

$$\min_{\alpha_i} f(\alpha_i) \text{ s.t } \sum_{i=1}^{U} R_i^T D_L \alpha_i = y \qquad (6)$$

where α_i is the local $v \times 1$ sparse vector and $R_i = \begin{bmatrix} 0_{p \times (i-1)} & \mathbb{I}_p & 0_{p \times (U-i-p+1)} \end{bmatrix}$ is a $p \times U$ matrix to extract a $p \times 1$ patch from y. The CSC applications discussed in [34]
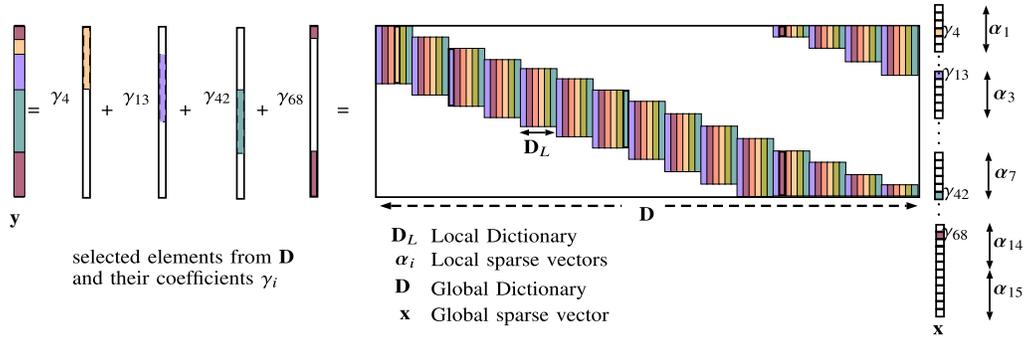
Fig. 2.   Dictionary structure of convolutional sparse coding.

and [35] establish that the sparse signal recovery algorithms based on convex relaxation (e.g. gradient descent) and greedy approach (e.g. matching pursuit, OMP, stage-wise OMP) are efficient when working with local dictionary also. The use of local dictionary helps to consider the global CSC of $\mathbf{y}$ in terms of independent CSC problems of smaller vectors of length p. It is interesting to note that the input-output relation for the proposed method with the DD grid containing embedded pilot vectors can be modeled similarly to (4), and its simplified form in Fig. 2. This observation motivated us to develop a channel estimation technique for OTFS-SCMA using the CSC approach, which is explained thoroughly in the next section.

III. FORMULATION OF CHANNEL ESTIMATION AS A CSC PROBLEM

This section comprehensively describes the CSC-based channel estimation for OTFS-SCMA, which is organized as follows. In Section III-A, we present the arrangement of the data and the pilot vectors in the transmit DD grid and analyze the corresponding received grid. Section III-B formulates the channel estimation task as a CSC problem, based on the data and the pilot arrangement.

We initially develop the channel estimation technique for the ideal pulse-shaping in which the input-output relation follows a simple 2D circular convolution as shown in (1). Later, the proposed method is extended to the rectangular pulse-shaping case.

A. Pilot and Data Arrangement in Delay-Doppler Grid

Although the proposed method applies to any codeword allocation schemes, for simplicity, we consider *Scheme-1* to describe the channel estimation technique. For a given channel, l_τ and k_ν denote the maximum integer delay and integer Doppler tap, respectively. For maximum delay $\tau_{\max}$ and maximum Doppler $\nu_{\max}$, we must have $\tau_{\max} < \frac{l_\tau}{M\triangle f}$ and $\nu_{\max} < \frac{k_\nu}{NT}$ where $\triangle f$ is the sub-carrier bandwidth, and T is the symbol duration satisfying $\triangle f = 1/T$. Let $\mathbf{X}_u \in \mathbb{C}^{N\times M}$ denote the uth user's input symbol matrix which is placed on the respective $N \times M$ Tx DD grid. If $\mathbf{x}_{u,l} \in \mathbb{C}^{N\times 1}$ denotes the column vector for delay tap l, the input symbol matrix can be written as $\mathbf{X}_u = [\mathbf{x}_{u,0}\ \mathbf{x}_{u,1}\ \ldots\ \mathbf{x}_{u,M-1}]$. The element in the location $[k,l]$ of $\mathbf{X}_u$ is denoted by $x_u[k,l]$ with $k = 0, 1, \ldots, N-1$ and $l = 0, 1, \ldots, M-1$. We consider column

vectors of specific length for pilots. Suppose the pilot vector of the uth user is denoted by $\mathbf{p}_u = \begin{bmatrix} p_{u,0}\ p_{u,1}\ \cdots\ p_{u,L_p-1} \end{bmatrix}^T$ and is placed at the delay index $\bar{l}$ such that $\bar{l} + l_\tau = M - 1$ where, l_τ is the maximum integer delay tap. Thus a just adequate guard band from the data part is maintained to minimize the overhead. The total number of data symbols that can be transmitted over the DD grid is $N_{\text{data}} = \frac{N}{K}(\bar{l} - l_\tau)$. The nth SCMA codeword transmitted by the uth user is given by $\mathbf{c}_n^u = \begin{bmatrix} c_{n,0}^u\ c_{n,1}^u\ \cdots\ c_{n,K-1}^u \end{bmatrix}^T$, $n = 1, 2, \ldots, N_{\text{data}}$. The symbols in the Tx grid of the uth user can be described as

$$x_u[k,l] = \begin{cases} c_{n,i}^u & 0 \le l < \bar{l} - l_\tau,\ k = [(n-1)K + i]_N, \\ p_{u,i} & l = \bar{l}\ \ k = i < L_p \\ 0 & l = \bar{l}\ \ k > L_p - 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Example 1: Consider an OTFS-SCMA system with $M = 5$, $N = 8$, $J = 6$, $K = 4$, $d_v = 2$, $l_\tau = k_\nu = 1$, and $L_p = 4$. The codewords follow the sparsity pattern as per the following factor matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{8}$$

The pilot vector is placed at the delay tap $\bar{l} = 3$. This example will be revisited multiple times to illustrate various concepts and procedures.

Fig. 3(a) shows the embedded arrangement of the data and the pilot vectors over the Tx DD grid for Example 1. The codewords and the pilot vectors follow the sparsity pattern as per (8). The Tx grids for only User 1 and User 6 are shown. The received symbols $\mathbf{Y} \in \mathbb{C}^{N\times M}$ is the collection of M columns as $[\mathbf{y}_0\ \mathbf{y}_1 \ldots \mathbf{y}_{M-1}]$. As a result of the sufficient guard band, there are two non-overlapping regions: (1) data detection region $\mathbf{Y}_d \in \mathbb{C}^{N\times \bar{l}}$ formed by the columns $\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_{\bar{l}-1}$ and (2) pilot observation region $\mathbf{Y}_p \in \mathbb{C}^{N\times(l_\tau+1)}$ formed by the columns $\mathbf{y}_{\bar{l}}, \mathbf{y}_{\bar{l}+1}, \cdots, \mathbf{y}_{\bar{l}+l_\tau}$. For Example 1, the Rx grid with multi-user and DD interference is shown in Fig. 3(b).

B. CSC Model for Channel Estimation

The received delay-Doppler frame with pilot and data embedded in a single frame is shown in Fig. 4. Also,

(a) Tx DD grids of User 1 ($\tilde{\mathbf{X}}_1$) and User 6 ($\tilde{\mathbf{X}}_6$) (b) Rx DD grid ($\mathbf{Y}$)
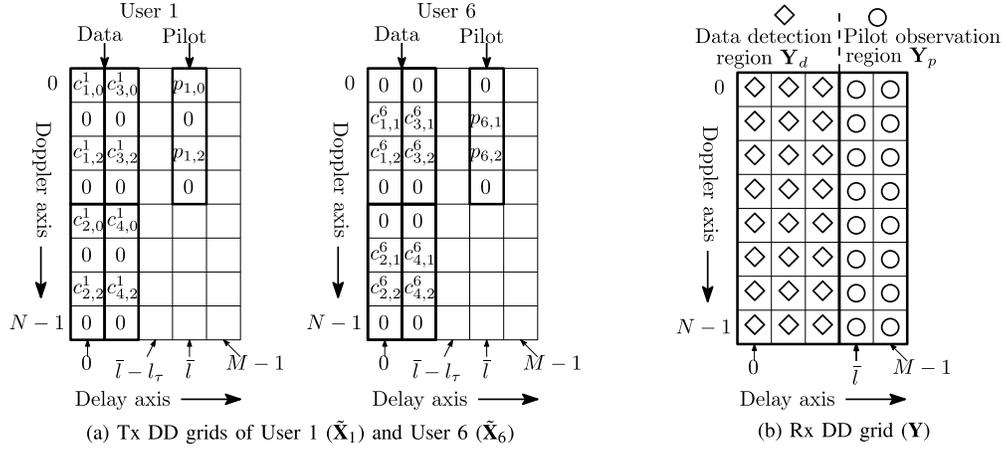
Fig. 3. Pilot-data arrangement[1] in DD domain for Example 1 (For convenience, pilot vectors of length 4 are shown. The selection of the length of pilot vectors is discussed in Section IV-B.1).

an overview of the proposed method is presented. The channel estimation task can be carried out by considering it as a CSC problem. For that, the input-output relation of OTFS-SCMA is analyzed for the single-user case first and then for the multi-user one.

• *Single user case:* Ignoring the AWGN at the BS, the input-output relation for the uth user from (1) can be written as

$$y_u[k,l] = \sum_{i=1}^{P} h_u[i] x[[k-k_i]_N, [l-l_i]_M] \quad (9)$$

where P is the number of paths and $h_u[i]$ is the channel coefficient of the ith path having the delay and Doppler tap l_i and k_i respectively, for the user. Note that P, $h_u[i]$, and $[k_i, l_i]$ are unknown in the channel estimation problem. Hence, we have to express (9) in terms of all the possible values of $[k_i, l_i]$, given by

$$y_u[k,l] = \sum_{k'=0}^{N-1} \sum_{l'=0}^{M-1} h_u[k',l'] x_u [[k-k']_N, [l-l']_M] \quad (10)$$

where $h_u[k',l']$ denotes the channel coefficient of the uth user for the path having delay tap l' and Doppler tap k'. Out of all the (k',l') pairs, there will be P pairs for which $h_u[k',l'] \neq 0$, such that $h_u[i] = h_u[k',l']$ for $(k_i, l_i) = (k',l')$. Let $\mathbf{y}_{u,l} = [y_u[0,l] \; y_u[1,l] \; \ldots \; y_u[N-1,l]]^T$ denote the contribution of the pilot vector of the uth user, received at the delay index l. Also, let $\mathbf{h}_{u,l} = [h_u[0,l] \; h_u[1,l] \; \ldots \; h_u[N-1,l]]^T$ denote the channel coefficients of the uth user for all possible Doppler taps at the delay tap l. Then, for any specific delay tap l, the input-output relation in (10) can be written in the vector form as

$$\mathbf{y}_{u,l} = \sum_{l'=0}^{M-1} \mathbf{x}_{u,[l-l']_M} \circledast \mathbf{h}_{u,l'} \quad \text{for } l = 0, 1, \ldots, M-1$$

$$(11)$$

where, l denotes the delay index of the received DD grid and l' denotes the delay tap in the channel. Observe from Fig. 4(b) that for the purpose of channel estimation, we have to consider only the pilot observation region $\mathbf{Y}_p$. Hence we can apply the following limits for l' and l in (11):

• $\bar{l} \leq l \leq M-1$, pilot observation region $\mathbf{Y_p}$ as shown in Fig. 4(b).
• $0 \leq l' \leq l_\tau$, since $\mathbf{h}_{u,l'} = \mathbf{0}$ for $l' > l_\tau$ where, l_τ is the maximum delay tap.

Considering the above two ranges, (11) can be simplified as:

$$\mathbf{y}_{u,l} = \sum_{l'=0}^{l_\tau} \mathbf{x}_{u,[l-l']_M} \circledast \mathbf{h}_{u,l'} \quad \text{for } l = \bar{l},\ \bar{l}+1, \ldots, M-1.$$

$$(12)$$

For $l = \bar{l}$, the relation in (12) can be written as

$$\begin{aligned}
\mathbf{y}_{u,\bar{l}} &= \sum_{l'=0}^{l_\tau} \mathbf{x}_{u,[\bar{l}-l']_M} \circledast \mathbf{h}_{u,l'} \\
&= \mathbf{x}_{u,[\bar{l}]_M} \circledast \mathbf{h}_{u,0} + \mathbf{x}_{u,[\bar{l}-1]_M} \\
&\quad \circledast \mathbf{h}_{u,1} + \ldots + \mathbf{x}_{u,[\bar{l}-l_\tau]_M} \circledast \mathbf{h}_{u,l_\tau}.
\end{aligned} \quad (13)$$

From (7) and also from Fig. 3(a), we have $\mathbf{x}_{u,[\bar{l}]_M} = \mathbf{x}_{u,\bar{l}} = [p_{u,0} \; p_{u,1} \; \cdots \; p_{u,L_p-1} \; \mathbf{0}_{1\times(N-L_p)}]^T$, denoted by $\mathbf{p}_u$. In addition, we can see that each of $\mathbf{x}_{u,[\bar{l}-1]_M}$, $\mathbf{x}_{u,[\bar{l}-2]_M}$, $\cdots \mathbf{x}_{u,[\bar{l}-l_\tau]_M}$ belongs to the guard band, which are $\mathbf{0}_{N\times 1}$ only. Hence, (13) is further simplified as

$$\begin{aligned}
\mathbf{y}_{u,\bar{l}} &= \mathbf{x}_{u,\bar{l}} \circledast \mathbf{h}_{u,0} \\
&= \begin{bmatrix}
p_{u,0} & 0 & \cdots & 0 & \cdots & p_{u,1} \\
p_{u,1} & p_{u,0} & \cdots & p_{u,L_p-1} & \cdots & p_{u,2} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \cdots & p_{u,0}
\end{bmatrix}_{N\times N} \\
&\quad \times \begin{bmatrix}
h_u[0,0] \\
h_u[1,0] \\
\vdots \\
h_u[N-1,0]
\end{bmatrix}
\end{aligned}$$

(a) Transmitted delay-Doppler grids of all users.

$\mathbf{Y}_d$: overlapped data received

$\mathbf{Y}_p$: overlapped pilot received

$\mathbf{P}_L$: pilot vectors of all users (local dictionary)

(b) Received delay-Doppler grid.
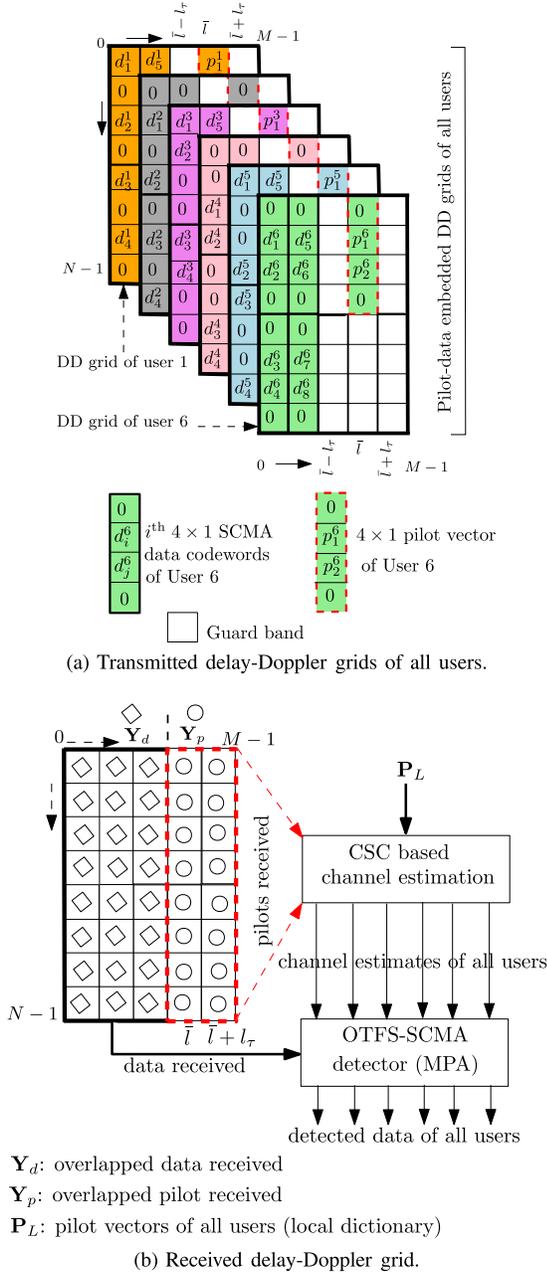
Fig. 4. Overview of the proposed channel estimation method for $\Gamma_{N,M}$, $N = 8, M = 5, J = 6, K = 4, l_\tau = 1,$ and $L_p = 4$.

$$= \begin{bmatrix} \mathbf{p}_u^{(0)} & \mathbf{p}_u^{(1)} & \cdots & \mathbf{p}_u^{(N-1)} \end{bmatrix} \mathbf{h}_{u,0}$$
$$= \mathbf{P}_u \mathbf{h}_{u,0} \qquad (14)$$

where, $\mathbf{p}_u^{(k)}$ is the transpose of the kth forward circular shift of $\begin{bmatrix} \mathbf{p}_u^T & \mathbf{0}_{1 \times (N-L_p)} \end{bmatrix}$, corresponding to the Doppler tap k. Similarly, considering (12) for all values of l and repeating the analysis as in (13)-(14), we have $l_\tau + 1$ independent equations as

$$\mathbf{y}_{u,l} = \mathbf{P}_u \mathbf{h}_{u,l-\bar{l}} \quad \text{for } l = \bar{l}, \ \bar{l}+1, \ldots, M-1. \qquad (15)$$

Remark 1: For the pilot and data arrangement in Section III-A, (15) is applicable to all users as sufficient guard

band is reserved considering l_τ and k_ν as the maximum delay and Doppler taps of the channels of all users.

● *Multi user case:* Now, we extend (15) to the OTFS-SCMA uplink with J users. The pilot symbols in the observation region of Fig. 4(b) is the superposition of the pilot symbols received from all J users. Hence, the received pilot observation signal at the delay index l is given by

$$\mathbf{y}_l = \mathbf{y}_{1,l} + \mathbf{y}_{2,l} + \cdots + \mathbf{y}_{J,l}$$
$$= \mathbf{P}_1 \mathbf{h}_{1,l-\bar{l}} + \mathbf{P}_2 \mathbf{h}_{2,l-\bar{l}} + \ldots + \mathbf{P}_J \mathbf{h}_{J,l-\bar{l}} \qquad (16)$$
$$= [\mathbf{P}_1 \ \mathbf{P}_2 \ldots \mathbf{P}_J] \begin{bmatrix} \mathbf{h}_{1,l-\bar{l}} \\ \mathbf{h}_{2,l-\bar{l}} \\ \vdots \\ \mathbf{h}_{J,l-\bar{l}} \end{bmatrix} = \sum_{u=1}^{J} \left(\sum_{k'=0}^{N-1} \mathbf{p}_u^{(k')} h_u[k', l-\bar{l}] \right)$$
$$= \sum_{k'=0}^{N-1} \left(\sum_{u=1}^{J} \mathbf{p}_u^{(k')} h_u[k', l-\bar{l}] \right)$$
$$\text{for } l = \bar{l}, \ \bar{l}+1, \ldots, M-1. \qquad (17)$$

Expressing (17) in terms of one summation, we get

$$\mathbf{y}_l = \sum_{k'=0}^{N-1} \boldsymbol{\mathcal{P}}^{k'} \mathbf{h}_{l-\bar{l}}^{k'} \quad \text{for } l = \bar{l}, \ \bar{l}+1, \ldots, M-1 \qquad (18)$$

where $\boldsymbol{\mathcal{P}}^{k'} = \begin{bmatrix} \mathbf{p}_1^{(k')} & \mathbf{p}_2^{(k')} & \cdots & \mathbf{p}_J^{(k')} \end{bmatrix}$ and $\mathbf{h}_{l-\bar{l}}^{k'} = \begin{bmatrix} h_1 \begin{bmatrix} k', l-\bar{l} \end{bmatrix}, h_2 \begin{bmatrix} k', l-\bar{l} \end{bmatrix}, \ldots, h_J \begin{bmatrix} k', l-\bar{l} \end{bmatrix} \end{bmatrix}^T$. Hence, (18) can be expressed in the matrix form as

$$\mathbf{y}_l = \mathbf{P}\mathbf{h}_{l-\bar{l}} \quad \text{for } l = \bar{l}, \ \bar{l}+1, \ldots, M-1 \qquad (19)$$

where $\mathbf{P} \in \mathbb{C}^{N \times JN}$ and $\mathbf{h}_{l-\bar{l}} \in \mathbb{C}^{JN \times 1}$ are given by

$$\mathbf{P} = \begin{bmatrix} \boldsymbol{\mathcal{P}}^0 & \boldsymbol{\mathcal{P}}^1 & \cdots & \boldsymbol{\mathcal{P}}^{N-1} \end{bmatrix} \quad \text{and}$$
$$\mathbf{h}_{l-\bar{l}} = \begin{bmatrix} (\mathbf{h}_{l-\bar{l}}^0)^T & (\mathbf{h}_{l-\bar{l}}^1)^T & \cdots & (\mathbf{h}_{l-\bar{l}}^{N-1})^T \end{bmatrix}^T, \quad \text{respectively.}$$

Since the maximum Doppler tap is k_ν, we have $\mathbf{h}_{l-\bar{l}} = \begin{bmatrix} (\mathbf{h}_{l-\bar{l}}^0)^T & \cdots & (\mathbf{h}_{l-\bar{l}}^{k_\nu})^T (\mathbf{0}_{1 \times J(N-2k_\nu-1)}) (\mathbf{h}_{l-\bar{l}}^{N-k_\nu})^T \cdots (\mathbf{h}_{l-\bar{l}}^{N-1})^T \end{bmatrix}^T$. Note that we have put 0s in $\mathbf{h}_{l-\bar{l}}$ by taking into account both positive and negative Doppler values.

The local dictionary associated with $\mathbf{P}$ is given by $\mathbf{P}_L = [\mathbf{p}_1 \ \mathbf{p}_2 \cdots \mathbf{p}_J]$. Fig. 5 shows the structures of the global dictionary $\mathbf{P}$ and the local dictionary $\mathbf{P}_L$ for Example 1. Clearly (19) can be considered as a CSC problem similar to that of (5) and (6). The channel estimation is now formulated as

$$\left. \begin{array}{l} \min_{\mathbf{h}_{l-\bar{l}}} f(\mathbf{h}_{l-\bar{l}}) \text{ s.t } \mathbf{P}\mathbf{h}_{l-\bar{l}} = \mathbf{y}_l \qquad \text{or} \\ \min_{\mathbf{h}_{l-\bar{l}}^k} f(\mathbf{h}_{l-\bar{l}}^k) \text{ s.t } \sum_{k \in \mathcal{K}} \mathbf{R}_k^T \mathbf{P}_L \mathbf{h}_{l-\bar{l}}^k = \mathbf{y}_l \end{array} \right\} \qquad (20)$$

where $l = \bar{l}, \ \bar{l} + 1, \ldots, M - 1$, $\mathcal{K} = \{0, \ldots, k_\nu, N - k_\nu, \ldots, N-1\}$, and $\mathbf{R}_k$ is the same patch extraction matrix as in (6).
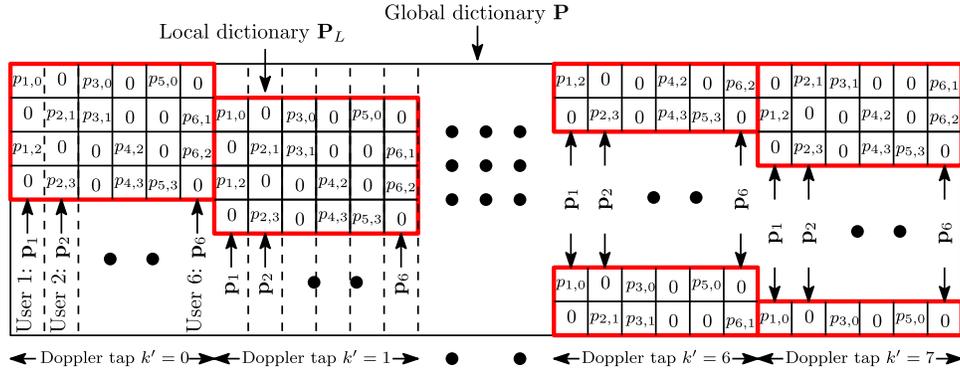
Fig. 5. Structure of global dictionary $\mathbf{P}$ and local dictionary $\mathbf{P}_L$ for Example 1.

IV. ALGORITHM FOR CHANNEL ESTIMATION BASED ON CSC

In this section, we propose an efficient sparse signal recovery algorithm for the CSC-based approach of channel estimation problem given in (20). We explore the features of sparse pilot vector design to ensure fast convergence and improved performance. **Algorithm 1** shows the steps for the CSC-based channel estimation technique.

A. Low Complexity and Fast Convergence Sparse Signal Recovery Algorithm

The sparse signal recovery algorithms based on greedy approach are extensively used for channel estimation problems. **Algorithm 1** is designed based on SP with a modified method for obtaining the initial estimates to achieve faster convergence. The steps of **Algorithm 1** are discussed next, highlighting the factors crucial for effective channel estimation.

1) Obtaining the Global Dictionary Using the Local Dictionary: From (20), it is observed that the sensing matrix $\mathbf{P}$ remains the same for all the delay taps and is used as the global dictionary throughout **Algorithm 1**. Hence we initially form $\mathbf{P}$, which is simply the block circulant version of the local dictionary input $\mathbf{P}_L$ as depicted in Fig. 2 for Example 1.

2) Shrinking the Global Dictionary: Although $\mathbf{P}$ is formed considering the complete Doppler taps from 0 to $N-1$, the search range can be limited to the maximum spread of the integer Doppler taps $2k_\nu$ of the channel. We shrink the dimension of $\mathbf{P}$ from JN to $J(2k_\nu + 1)$ columns (line 2 of **Algorithm 1**). This shrinkage in the dimension of the sensing matrix reduces the complexity of the sparse signal recovery algorithm.

3) Thresholding of Pilot Observation Region: The pilot observation region accommodates the maximum delay tap l_τ of the channel but not every delay tap starting sequentially from 0 to l_τ is present. Only a few dominant delay taps will be present in the channel. If M_{dom} is the number of dominant delay taps in the channel, then in (20), we need to apply sparse signal recovery algorithm only to M_{dom} equations ($M_{\text{dom}} \ll M$). A thresholding of $||\mathbf{y}_l||_2 > \mathcal{T}$ can easily find out these M_{dom} equations, since all other $\mathbf{y}_l$ have only the noise

---

Algorithm 1 SP algorithm with modified initial estimate

input : $\mathbf{Y}_p$: pilot observation region; $\mathbf{P}_L$: dictionary of pilots; J: total number of users; k_ν: maximum integer Doppler tap; $\mathcal{T}$: Threshold for dominant path detection;

Obtain $\mathbf{P}$ using the input $\mathbf{P}_L$

Limit the search range of Doppler to the maximum spread of Doppler taps of channel;

foreach *delay tap l in the pilot observation region* **do**

Thresholding: Check whether l is a dominant delay path based on $||\mathbf{y}_l||_2 > \mathcal{T}$. If l is not a dominant path then skip the following part and go to line 3 for the next delay tap $l + 1$;

Channel estimation:

if $||\mathbf{y}_l||_2 > \mathcal{T}$ **then**

Initialization: iteration $i = 0$;

Initial estimates: Estimate the dominant Doppler tap per user and the residue ;

repeat

$i = i + 1$;

Find the J columns of $\mathbf{P}$ with high correlation with the residue as in (**A1**);

Update the current temporary list $\tilde{S}^{(i)}$ of selected columns by adding these J columns to the previous list $S^{(i-1)}$ as in (**A2**) ;

Find the pseudoinverse solution for the $2J$ selected columns of $\mathbf{P}$ as in (**A3**);

Find $S^{(i)}$ and refine the channel estimates finding the J selected columns of significant coefficients as in (**A4**) and (**A5**);

Update the residue using the refined channel estimates as in (22) of (**A6**);

until *stopping criteria, as in (23)*;

output: Channel estimates of all users: $\hat{\mathbf{h}}_{l-\bar{l}}$ for $l = \bar{l}, \bar{l} + 1, \ldots, M - 1$.

---

component (σ_p). We select $\mathcal{T} = c\sigma_p$, where $\sigma_p = \sqrt{\frac{N}{\text{SNR}_p}}$, $\text{SNR}_p = \frac{\mathbb{E}(|p_{u,i}|^2)}{\sigma^2}$ and the factor c is decided based on the observations from simulations. For the simulation parameters used, we fix $c = 2$.

4) Obtaining the Initial Estimates Sequentially: Instead of selecting the initial J elements from the overall set of JN elements, we can strategically choose the initial values based on a group-by-group greedy approach. There are J groups of N channel coefficients corresponding to J users and N Doppler taps of every user at a fixed delay tap. We select the dominant Doppler tap per user in the initial estimate and execute the remaining iterations. Hence, we can ensure that one path per user is present in the initial estimate making the convergence faster (line 8 of **Algorithm 1**).

Initialization for $i = 0$: For each jth user, form $\mathbf{P}_j$ collecting all the columns corresponding to that user from $\mathbf{P}$. From each $\mathbf{P}_j$, we find the column that has the highest correlation with $\mathbf{y}_l$ and, considering all users, form $\mathbf{P}_{(S^{(0)})}$. The initial J estimates are $\mathbf{h}_{(S^{(0)})} = \mathbf{P}^{\dagger}_{(S^{(0)})} \mathbf{y}_l$. Repeat the following iterations until the stopping criteria are satisfied.

(A1): Obtain the J elements which are highly correlated to the residue of the previous iteration, i.e., the J maximum correlation coefficients in the vector $|\mathbf{P}^H \mathbf{r}^{(i-1)}|$, and collect these J column indices in $\mathbf{s}^{(i)}$.

(A2): Form the list of $2J$ elements by including the new J elements of the current iteration to the list of the previous iteration: $\tilde{S}^{(i)} = S^{(i-1)} \cup \mathbf{s}^{(i)}$.

(A3): Find the nearest solution for $2J$ elements using the psuedoinverse method:

$$\mathbf{h}_{(\tilde{S}^{(i)})} = \mathbf{P}^{\dagger}_{(\tilde{S}^{(i)})} \mathbf{y}_l. \tag{21}$$

(A4): Find the J maximum values in the vector $\left| \mathbf{h}_{(\tilde{S}^{(i)})} \right|$ and put those J indices in the selected list $S^{(i)}$ of J indices. Collect the columns of $\mathbf{P}$ corresponding to $S^{(i)}$ in $\mathbf{P}_{(S^{(i)})}$.

(A5): Refine the J channel estimates as: $\mathbf{h}_e = \mathbf{P}^{\dagger}_{(S^{(i)})} \mathbf{y}_l$.

(A6): Using the J estimated coefficients, update the residue for the next iteration:

$$\mathbf{r}^{(i)} = \mathbf{y}_l - \mathbf{P}_{(S^{(i)})} \mathbf{h}_e. \tag{22}$$

5) Stopping Criteria: The algorithm is considered to converge to the optimal solution when the norm of the residue is minimal ([20], Section III). For SP algorithms, in each iteration, the J estimated coefficients are updated with J more elements. The final J estimates are determined from the updated list of $2J$ elements having the highest correlation. Once the residual norm reaches the minimum, a further update of the J new elements results in divergence from the minimum. Hence, if the residual norm starts increasing for the next iteration, the optimal solution is attained, i.e., the iteration stops when

$$||\mathbf{r}^{(i)}||_2 > ||\mathbf{r}^{(i-1)}||_2. \tag{23}$$

The estimated channel coefficients $\hat{\mathbf{h}}_{l-\bar{l}}$ of the current delay tap are set to the refined estimate $\mathbf{h}_e$ of the previous iteration $i - 1$.

The features of **Algorithm 1** that result in fast convergence are highlighted next.

- *Sparse pilot vector and sequential initial estimates*: The pilot vector of each user follows the same sparse structure of the user's codewords. SCMA codebook is designed such that only d_f users share a resource. The similar sparse structure for the pilot vectors favors the sparse signal recovery since those columns of $\mathbf{P}$ will be highly correlated with $\mathbf{y}_l$. Also, it ensures that all the d_v non-zero components of all pilot vectors are available for channel estimation. The sparseness of the pilot vectors of the multiple users provides minimal interference amongst them. In addition, the sequential selection of the initial estimates offers uniform treatment to all users without prioritizing any user.

- *Dealing with unknown sparsity*: The knowledge of sparsity is a prerequisite for SP algorithm. The dictionary $\mathbf{P} \in \mathbb{C}^{N \times JN}$ is used to recover the channel coefficients at the all Doppler taps of all users for a single delay tap as shown in (20). If the sparse signal recovery algorithm is used directly, then we need to estimate P_{eff} coefficients, where P_{eff} is unknown. By estimating the Doppler taps of all delay taps independently, the sparse signal recovery in (20) aims to recover J components from the available JN components of $\mathbf{P}$. The unknown sparsity of P_{eff} is converted to a known sparsity of the number of users, J.

B. Pilot Pattern Design

1) Length of Pilot Vector: Sparse signals can be successfully recovered if the dictionary elements follow specific criteria [36]. For a sparse signal recovery problem with the sparsity of s/V, the number U of observations must satisfy $U > s \log V$. In addition, for CSC-based sparse signal recovery, the global sparsity s must be approximately $(s_L/(2p - 1))V$, where s_L is the local sparsity, and p is the length of the local dictionary vector [33]. For the SP algorithm to recover $2s$ estimates in each iteration, we should have rank$(P^{\dagger}) \geq 2s$. To attain the maximum spectral efficiency, we need to consider the minimum pilot vector length that can guarantee successful sparse signal recovery. Considering these conditions and the requirements for OTFS-SCMA, we state a condition for the length of the pilot vector in the following lemma.

Lemma 1: For successful channel estimation using the proposed method, the length L_p of pilot vector must satisfy the following condition:

$$L_p \geq \max \left\{ 2J, \lceil cJ \log(J(2k_\nu + 1)) \rceil - 2k_\nu, k_\nu + 1 \right\} \tag{24}$$
$$\text{with } [L_p]_K = 0$$

where c is a constant satisfying $1 < c \leq 2$.

Proof: The proof is given in APPENDIX A. ∎

2) Pilot Vector Sequence: The design of the pilot sequence plays a crucial role in successful channel estimation. For compressive-sensing-based techniques, the primary factor of concern is the cross-correlation among the columns of sensing matrix (dictionary) $\mathbf{P}$. In that regard, one important figure of merit is the mutual coherence of $\mathbf{P}$ which is defined as $\mu(\mathbf{P}) = \max_{i \neq j} |\mathbf{p}^H_{(i)} \mathbf{p}_{(j)}|$ for any two columns $\mathbf{p}_{(i)}$ and

$\mathbf{p}_{(j)}$; and $||\mathbf{p}_{(i)}||_2 = 1, \forall\ i$. As $\mu(\mathbf{P})$ becomes smaller, the probability of successful sparse signal recovery increases. Hence, we consider the following optimization problem to design $\mathbf{P}$:

$$\min\ \mu(\mathbf{P}) = \min\ \max_{i \neq j} |\mathbf{p}_{(i)}^H \mathbf{p}_{(j)}|$$
$$\text{s.t.}\quad ||\mathbf{p}_{(i)}||_2 = 1,\quad 1 \leq i, j \leq JN \qquad (25)$$

To carry out the optimization in (25), we consider the method of differential evolution [21]. Differential evolution is a robust evolutionary algorithm to solve arbitrary optimization problems with real-valued parameters. The problem (25) involves complex variables which are converted to the equivalent real ones of twice lengths to apply differential evolution. Specific to the CSC model, there exists a lower bound for $\mu(\mathbf{P})$ depending on the local dictionary's dimensions [33]. Considering the local dictionary $\mathbf{P}_L$ shown in Fig. 5, we have $\mathbf{P}_L \in \mathbb{C}^{L_p \times J}$ and the condition is obtained as $\sqrt{\frac{J-1}{J(2L_p-1)-1}} \leq \mu(\mathbf{P}) < 1$. We consider the particular case of $J = 6$ and $L_p = 20$ to verify the effectiveness of the proposed dictionary learning method based on the mutual coherence condition. The resultant dictionary optimized by differential evolution is referred to as 'Learned pilot' and it has $\mu(\mathbf{P}) \approx 0.36$. Note that for this case, we have $0.15 \leq \mu(\mathbf{P}) < 1$ which shows that 'Learned pilot' yields a mutual coherence value close to its lower limit. Additionally, we analyze the following pilot sequences: (1) 'Gaussian pilot' where the pilot symbols are i.i.d. complex Gaussian, (2) 'Zadoff-Chu pilot' where $\mathbf{p}_u[i] = \exp\left(-j\frac{\pi r i^2}{L_p}\right)$ for the uth user with r being the order of the sequence and for each user, it is selected as a distinct prime number with $1 \leq r < L_p$ and $0 \leq i < L_p$, and (3) 'SCMA cw pilot' where we consider SCMA codewords as the pilot vectors.

Fig. 6 shows the histogram of the cross-correlation of the columns of $\mathbf{P}$ for the corresponding pilot vectors. Observe that the mutual coherence values for 'Learned pilot', 'Gaussian pilot', 'Zadoff-Chu pilot', and 'SCMA cw pilot' are 0.36, 0.5, 0.78, and 0.86 respectively. Moreover, the spread of the cross-correlation of the columns of $\mathbf{P}$ is limited to a small range for 'Learned pilot', compared to those of the other pilot sequences. For 'SCMA cw pilot', the spread of cross-correlation to higher values is even more. This analysis justifies the use of optimized dictionary for channel estimation.

C. Rectangular Pulse Shaping

So far, ideal pulse shaping has been assumed for the proposed channel estimation technique. When rectangular pulses replace the ideal ones, the changes are reflected only through an additional multiplicative phase factor with the channel coefficients [22], as shown below:

$$y_u[k, l]$$
$$= \sum_{k'=0}^{N-1} \sum_{l'=0}^{M-1} \beta_{(k',l')}[k, l] h_u[k', l'] x_u \left[[k - k']_N, [l - l']_M\right] \quad (26)$$
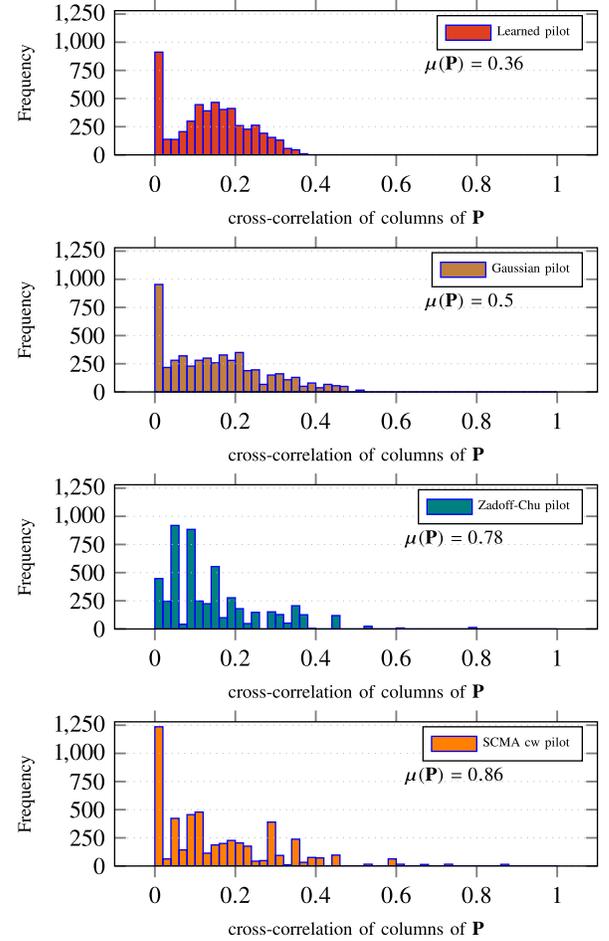
Fig. 6. Histogram of cross-correlation of columns of $\mathbf{P}$.

where

$$\beta_{(k',l')}[k, l] = \begin{cases} (\frac{N-1}{N}) e^{j2\pi \frac{k'}{N}(\frac{l-l'}{M})} e^{-j2\pi(\frac{[k-k']_N}{N})} & 0 \leq l < l' \\ e^{j2\pi \frac{k'}{N}(\frac{l-l'}{M})} & l' \leq l < M \end{cases} \quad (27)$$

Observe that neither the sparsity nor the circulant block property is affected by the rectangular pulses. Therefore, the method proposed for ideal pulse shaping can be easily extended to rectangular pulse shaping by considering the phase factor matrix $\boldsymbol{\Upsilon}_l \in \mathbb{C}^{N \times NM}$ defined as

$$\boldsymbol{\Upsilon}_l[n, m] = \beta_{(k',l')}[k, l],\ \text{for } n = k,\ m = Nl' + k'. \quad (28)$$

Now (26) can be written in vector form as

$$\mathbf{y}_{u,l} = (\mathbf{P}_u \odot \boldsymbol{\Upsilon}_l)\, \mathbf{h}_{u,l-\bar{l}},\quad \text{for } 0 \leq l - \bar{l} \leq l_\tau,\ l = \bar{l} + l - \bar{l}. \quad (29)$$

Note that $\boldsymbol{\Upsilon}_l$ depends only on N and M and is independent of a user's channel conditions. Therefore, the proposed method can be directly applied in the case of rectangular pulse, by replacing $\mathbf{P}_u$ by $(\mathbf{P}_u \odot \boldsymbol{\Upsilon}_l)$ for all users in (15).

D. Complexity Analysis

The independent delay tap processing and the reduced search range in the Doppler domain result in remarkable

TABLE I
COMPLEXITY ANALYSIS OF **ALGORITHM 1**

Dimension of sensing matrix	$N \times J(2k_\nu + 1)$
Iterations(i)	$\mathcal{O}(\log(J))$
Computational-complexity	$\mathcal{O}(M_{\text{dom}} N(J(2k_\nu + 1) + J^2)i)$

dimension reduction for all matrix operations of **Algorithm 1**. The complexity is analyzed with reference to the general SP algorithm, given in Section IV.C of [20]. For the sparse signal recovery problem with a sensing matrix of dimension $U \times V$ and sparsity of s, the SP algorithm's complexity is $\mathcal{O}(U(V + s^2))$ per iteration. The number of iterations needed for convergence is $\mathcal{O}(\log s)$. Here we have $U = N, V = J(2k_\nu + 1)$, and $s = J$. In TABLE I, $M_{\text{dom}} \ll M$, which is the number of dominant delay taps in the channel.

Remark 2: Channel estimation based on the sparse signal recovery algorithms using one complete OTFS frame of full-length pilot vector has a sensing matrix of dimension $NM \times JNM$. This non-embedded scheme results in higher complexity than Algorithm 1, in terms of the number of complex multiplications.

E. Cramer-Rao Lower Bound

The Cramer-Rao lower bound (CRLB) for the estimates of the channel coefficients is presented here. We consider the pilot symbols' arrangement as discussed in Section III-A. Writing (19) in matrix form, we get

$$
\begin{bmatrix} \mathbf{y}_{\bar{l}} \\ \mathbf{y}_{\bar{l}+1} \\ \vdots \\ \mathbf{y}_{M-1} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \cdots & \mathbf{0} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{M-1-\bar{l}} \end{bmatrix} \tag{30}
$$

Writing (30) compactly, we get

$$
\mathbf{y}^p = \mathbf{P}^{\text{diag}} \mathbf{h} \tag{31}
$$

where $\mathbf{y}^p = \text{vec}(\mathbf{Y}_p) \in \mathbb{C}^{N(M-\bar{l}) \times 1}$ contains the received symbols in the pilot observation region, $\mathbf{P}^{\text{diag}} = \text{diag}(\mathbf{P}) \in \mathbb{C}^{N(M-\bar{l}) \times JN(M-\bar{l})}$, and $\mathbf{h} \in \mathbb{C}^{JN(M-\bar{l})}$ contains the channel coefficients of all users. As the number of multi paths of each user is P, $\mathbf{h}$ contains JP non-zero elements. Suppose the location of these non-zeros are $c_1, c_2, \ldots, c_{JP}$. For deriving the CRLB, we assume that these locations are known. Let the eth non-zero element of $\mathbf{h}$ be denoted by θ_e (c_e^{th} element of $\mathbf{h}$). The parameter vector is given by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{JP}]^T$ which are to be estimated. Let $p^{\text{diag}}[n, c]$ denote the element in the nth row and the cth column of $\mathbf{P}^{\text{diag}}$ and y_n^p denote the nth element of $\mathbf{y}^p$. Then considering the AWGN at the BS, (31) can be alternatively written as

$$
y_n^p = s_n + w_n, \qquad n = 1, 2, \ldots, N(M - \bar{l})
$$

where $w_n \sim \mathcal{CN}(0, \sigma^2)$ and s_n is given by

$$
s_n = \sum_{e=1}^{JP} p^{\text{diag}}[n, c_e] \theta_e. \tag{32}
$$

For rectangular pulses, we have

$$
s_n = \sum_{e=1}^{JP} p^{\text{diag}}[n, c_e] \gamma[n, c_e] \theta_e \tag{33}
$$

where $\gamma[n, c_e]$ is given by $\beta_{(k', l')}[k, l]$, for $n = Nl + k + 1$ and $c_e = J(Nl' + k') + 1$. As per CRLB, the variance of any unbiased estimator $\hat{\theta}_i$ is lower bounded as [37]

$$
\text{var}\left(\hat{\theta}_i\right) \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{ii} \tag{34}
$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the $JP \times JP$ Fisher information matrix whose element in the ith row and the jth column is given by

$$
[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=1}^{N(M-\bar{l})} \left\{ \frac{\partial s_n}{\partial \theta_i} \frac{\partial s_n^*}{\partial \theta_j} + \frac{\partial s_n^*}{\partial \theta_i} \frac{\partial s_n}{\partial \theta_j} \right\}. \tag{35}
$$

From (32) and (33), it can be easily found that

$$
\frac{\partial s_n}{\partial \theta_i} = \begin{cases} p^{\text{diag}}[n, c_i] & \text{for ideal pulse} \\ p^{\text{diag}}[n, c_i] \gamma[n, c_i] & \text{for rectangular pulse} \end{cases} \tag{36}
$$

Similarly $\frac{\partial s_n^*}{\partial \theta_i}$ can be found out. From (35), the Fisher information matrix can be found out and then by using (34), CRLB for the estimates of the channel coefficients can be obtained. Finally, for the normalized mean square error (NMSE) analysis, we consider the average normalized CRLB: $\left\{ \sum_{i=1}^{JP} \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{ii} \right\} / \|\boldsymbol{\theta}\|_2^2$.

V. RESULTS AND DISCUSSIONS

This section presents the simulation results for the proposed method of channel estimation and analyzes them. The OTFS-SCMA scheme uses an $N \times M$ DD grid of different dimensions and a basic $(J = 6, K = 4)$ SCMA system with 150% overloading factor. *Scheme-1* is considered where the $K \times 1$ codewords are placed along the Doppler axis. The codebook design follows the technique described in [38]. The channel conditions of all users are assumed such that only a single Doppler value is associated with a delay path [22]. Integer delay and Doppler taps are assumed in simulations. Note that the proposed method applies to the fractional cases also, albeit with an increase in complexity. For the data detection in uplink, we use the powerful single-stage MPA [6]. The average pilot and data SNRs are denoted by $\text{SNR}_p = \frac{\mathbb{E}(|p_{u,i}|^2)}{\sigma^2}$ and $\text{SNR}_d = \frac{\mathbb{E}(|c_{n,i}^u|^2)}{\sigma^2}$ respectively, where σ^2 is the variance of AWGN. The simulation observations are presented in terms of various performance indicators like BER, spectral efficiency, NMSE, and complexity analysis. The proposed pilot symbols are 'Learned pilot' as discussed in Section IV-B.2. We extend the methods of OTFS-OMA in [8] and [14] to OTFS-SCMA for the comparison purpose and are referred to as 'QAM-pilot' and 'MSP' (modified subspace pursuit), respectively. We assume that all users transmit at the same power.[2]

---

[2]An exciting future research direction is to design an optimal power allocation strategy by considering resource management, codebook design, and user grouping.

(a) BER performance for different SNR_p.

(b) BER performance for different user velocities.

Fig. 7. OTFS-SCMA uplink channel estimation using proposed method ($\Gamma_{32,32}$; $J = 6, K = 4$; $P = 2$).

Fig. 8. BER of proposed and conventional methods.

In Fig. 8, we compare the BER performance of the proposed method to those of 'QAM-pilot' and 'MSP'. The results are presented for $SNR_p = 40$ dB. Observe that the proposed method gives a slightly better BER over 'QAM-pilot'. The technique of 'QAM-pilot' [8] uses a single QAM-pilot symbol and a thresholding-based method for detecting the paths and estimating the channel coefficients one at a time in a scalar form. On the other hand, the proposed scheme is devised based on a sparse signal recovery algorithm with non-orthogonal pilot vectors, where the channel coefficients at a particular delay tap are estimated as a vector. In contrast to 'MSP' [14], which is based on sparse signal recovery, we use sparse pilot vectors which are non-orthogonal and embedded in the data frame, rather than reserving a frame dedicated for the orthogonal pilot vectors. The comparison in Fig. 8 shows that the proposed method of channel estimation gives an improved BER performance over 'MSP'.

The sparse signal recovery of 'MSP' considers the channel coefficients of all users as a single vector. Since JP coefficients have to be estimated in each iteration, the path missed for one user will result in an additional path detected for another user to maintain the sparsity. Two users will be disturbed by a single path detection error, affecting the BER performance. The initial estimate is taken sequentially in the proposed method, considering one path per user. Estimating the channel coefficients at each delay path is considered an independent sparse signal recovery problem. Hence, all users and delay paths are given equal preference in channel estimation, reducing the probability of error in path detection.

A. BER Analysis

For an initial investigation, we consider a DD grid with $N = M = 32$. The number of multi-paths is $P = 2$. The sub-carrier spacing is $\triangle f = 15$ KHz. The delay values are taken as $\tau_i = (0, 1.04, 2.08, 3.12, 4.16, 5.20, 6.25, 7.29)\mu s$ and the maximum delay tap $l_\tau = 4$ [39]. For each delay tap l_i, the Doppler shift is generated using Jakes' formula $\nu_i = \nu_{max}\cos\theta_i$, for $\theta_i \sim U(-\pi, \pi)$. The maximum Doppler shift is given by $\nu_{max} = \frac{f_c v_u}{(3 \times 10^8)}$, where v_u is the user's speed in m/s. Moreover, we take $k_\nu = \lceil \nu_{max}NT \rceil$. We have $k_\nu = 1$ and $k_\nu = 4$ for $v_u = 120$ Kmph and 500 Kmph, respectively. As per Lemma 1, we take $L_p = 12$ and $L_p = 20$ for the respective velocities.

Fig. 7(a) shows the BER for three different SNR_p values of 30, 35, and 40 dB. As expected, when SNR_p increases, the BER also improves. At $SNR_p = 40$ dB, the BER nearly approaches to that of the perfect CE. The BER performance for different user velocities at $SNR_p = 40$ dB are presented in Fig. 7(b). Observe that the user velocity has no significant impact on the channel estimation performance. Since the guard band is allocated considering the maximum Doppler spread among all the users, the pilot-data interference is avoided for any user velocity.

B. Guard-Band Overhead and Spectral Efficiency Analysis

Low guard band overhead is a desirable feature for any channel estimation technique. Since the proposed method uses an embedded pilot-aided structure, we compare its overhead with that of 'QAM-pilot' [8] as displayed in TABLE II. The overheads is presented for different user speeds considering the EVA propagation model given in TABLE III.

Observe from TABLE II that the guard band overhead for the proposed method is independent of the number J of users.

UE speed (Kmph)	Reduced guard band[a]						Full guard band[b]	
	30		120		500			
Max. Doppler tap k_ν	$k_\nu = 1$		$k_\nu = 4$		$k_\nu = 16$			
	Prop.	[8]	Prop.	[8]	Prop.	[8]	Prop.	[8]
Pilot+Guard symbols[1] (N_g)	984	1168	1476	2920	3444	9928	5248	18688
Data symbols[2] (N_d)	96828	96552	96090	93924	93138	83412	90432	70272
Guard band overhead $\left(\frac{N_g}{NM}\right)$	1.5%	1.7%	2.2%	4.5%	5.2%	15.2%	8%	28%

[1][a] $N_g : (4k_\nu + L_p)(2l_\tau + 1)$ [Prop] ; $(4k_\nu + 4)(J(l_\tau + 1) + l_\tau)$ 'QAM-pilot' [8]
[1][b] $N_g : N(2l_\tau + 1)$ [Prop] ; $N(J(l_\tau + 1) + l_\tau)$ [8]; [2]$N_d : (NM - N_g)\frac{J}{K}$

TABLE III
OTFS PARAMETERS FOR EVA CHANNEL MODEL

Parameter	Value
Carrier frequency, f_c	4 GHz
Subcarrier spacing, $\triangle f$	15 KHz
Number of Doppler bins, N	128
Number of delay bins, M	512
Maximum delay tap, l_τ	20

Moreover, the overhead for the proposed method is less than that of 'QAM-pilot' [8]. The contrast in the overheads of the two methods increases as the user speed increases. In the case of full guard band, the proposed method offers distinctively lower overhead than 'QAM-pilot'.

The reduction in guard band overhead can also be interpreted in terms of spectral efficiency (SE). We analyze the effective SE (SE_{eff}) of the proposed OTFS-SCMA uplink channel estimation following [15]. Since MPA is a non-linear detector, instead of directly using SINR for the spectral efficiency, SE_{eff} is formulated as the number of bits that can be correctly decoded and is expressed as $SE_{eff} = (1 - BER)S$. The nominal SE is given as $S = \frac{\text{Number of data bits transmitted}(N_b)}{\text{Number of resources}(N_{RE})}$. Note that in the ideal case (BER=0), we have $SE_{eff} = S$. For an uncoded SCMA system, we have $S = \left(\frac{J}{K}\right)\log_2(|\mathbb{A}|)$. Extending this concept to the channel estimation of uncoded OTFS-SCMA uplink system, for QAM modulation ($\log_2(|\mathbb{A}|) = 2$), we get

$$S = \frac{2N_d}{NM}; \text{ where } N_d = (NM - N_g)\frac{J}{K} \quad (37)$$

where N_g is the number of REs reserved for channel estimation. Hence the effective SE for the OTFS-SCMA uplink case becomes

$$SE_{eff} = (1 - BER)\frac{2(NM - N_g)}{NM}\frac{J}{K}. \quad (38)$$

Here, N_g depends on the channel estimation technique adopted, whose generalized expression is given as the footer of TABLE II. For perfect CSI, we have $N_g = 0$, which gives the upper bound on SE_{eff}. Fig. 9 presents SE_{eff} for the proposed method and the existing methods. The parameters used in Fig. 9 are the same as those of the BER comparison in Fig. 8. Based on (38), for an overloading factor of 150% and $|\mathbb{A}| = 4$, we have $SE_{eff} \approx 3$ bits/s/Hz. Observe from Fig. 9 that the proposed channel estimation technique gives $SE_{eff} \approx 2.75$ bits/s/Hz at $SNR_p = 40$ dB. The non-orthogonal
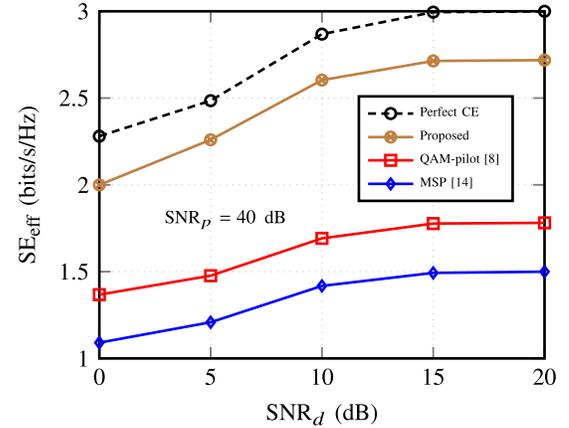
Fig. 9.   Effective spectral efficiency in uplink.

pilot pattern improves SE_{eff} by almost 1 bits/s/Hz compared to the orthogonal 'QAM-pilot' technique [8]. The method 'MSP' in [14] has a dedicated pilot frame for channel estimation alone, and a guard band is not required. Hence, the overhead is always 50% irrespective of the channel conditions. Correspondingly, $SE_{eff} \approx 1.5$ bits/s/Hz at $SNR_p = 40$ dB. Note that for a practical EVA channel model, the proposed method gives even higher SE_{eff} advantage as observed from the guard band overhead presented in TABLE II.

C. Pilot Analysis

1) Length Analysis: For the proposed method to be successful, the pilot vectors must have a minimum length L_p satisfying the condition given in Lemma 1. We consider $M = 512, N = 128, SNR_d = 20$ dB, $SNR_p = 40$ dB, and EVA model parameters of TABLE III for the analysis of the pilot vector's length. Fig. 10 shows the probability of success, defined as the ratio of the number of trials that resulted in successful sparse signal recovery to the total number of trials for different lengths of pilot vectors. Observe that the choice of L_p influences the successful sparse signal recovery. Depending on the Doppler spread of the channel, the lower bound of L_p also decreases. However, a minimum of $L_p = 12$ is to be maintained irrespective of the channel conditions. The channel estimation error can be analyzed by calculating the NMSE, which is given by $10\log_{10}\left(\frac{\mathbb{E}(||\mathbf{h} - \hat{\mathbf{h}}||_2^2)}{\mathbb{E}(||\mathbf{h}||_2^2)}\right)$ dB. We plot the NMSE values against the pilot lengths in Fig. 11. For the successful channel estimation, even at a maximum user velocity of 500 Kmph, we can select $L_p = 20$. Further, in Fig. 11, we consider a user velocity of 500 Kmph and other parameters as given in the figure. It is evident that to achieve a tolerable NMSE, a minimum pilot vector length of 20 is necessary. For $L_p \geq 20$, the NMSE remains almost constant while the guard band overhead increases by around 2.5%. Lemma 1 presents the minimum length L_p of the pilot vector needed for successful sparse signal recovery. Once it is maintained, we attain the minimum NMSE. Further increase in L_p will not give any advantage to sparse signal recovery. Also, from the CRLB curve, it is clear that for $L_p \geq 20$, the proposed method performs close to the lower bound
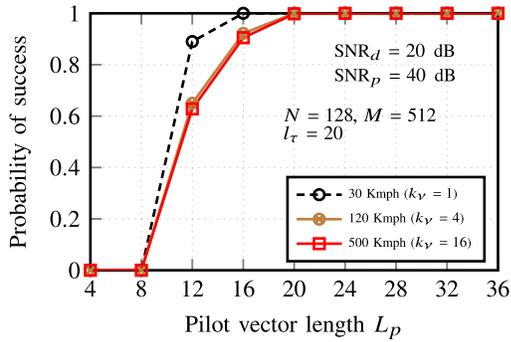
Fig. 10. Probability of success for different user velocities.
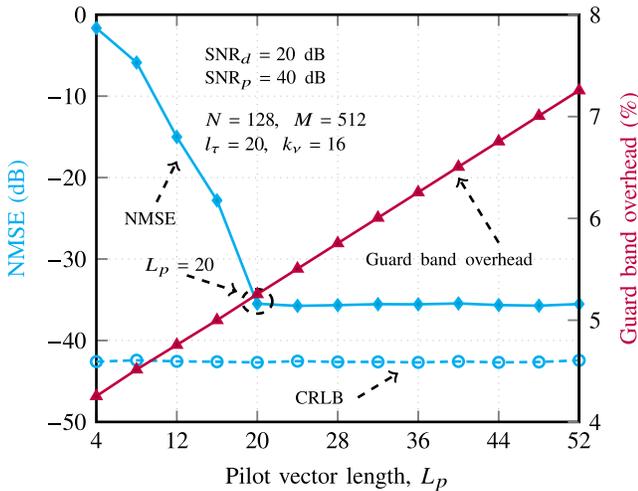
Fig. 11. Optimal pilot vector length of proposed method.

Fig. 12. NMSE for various pilot vectors.
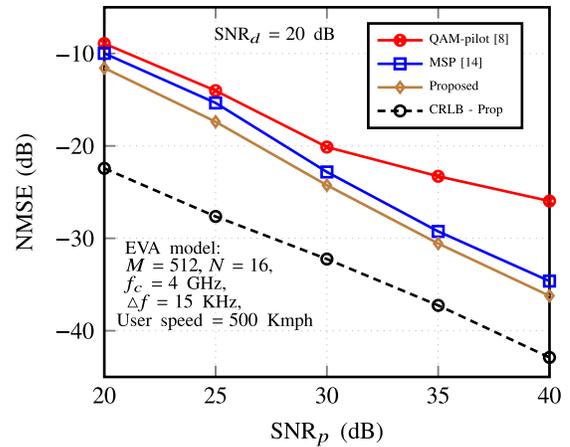
Fig. 13. NMSE of channel estimation in uplink.

specified by CRLB. The simulation results shown in Fig. 10 and Fig. 11 substantiate Lemma 1 and its proof given in APPENDIX A.

2) Pattern Analysis: In Fig. 12, we analyze the NMSE for the pilot-data arrangement shown in Fig. 3, using different types of pilot vectors as discussed in Section IV-B.2. Considering $k_\nu = 16 \ll 128$, we take $L_p = 20$ as per Lemma 1. The differences in NMSE for these various pilot vectors follow the mutual coherence analysis presented in Section IV-B.2. The proposed 'Learned pilot' provides a lower NMSE than 'Gaussian-pilot', as observed in Fig. 12. 'Zadoff-Chu pilot' gives a poor performance compared to 'Learned pilot' and 'Gaussian-pilot'. The NMSE values for 'SCMA cw pilot' are significantly higher than those of other sequences. The 'Random-SP init' curve shows NMSE results when the initialization (line 8 of **Algorithm 1**) is done randomly. By taking the initial estimates sequentially, we can reduce the estimation error. Also, the NMSE performance using non-sparse i.i.d Gaussian random vectors is shown by the 'Non-sparse Gaussian' curve. The slight performance improvement of the sparse pilot vector over the non-sparse one is attributed to minimal multi-user interference in the pilot region.

D. NMSE Comparison With Existing Methods

In Fig. 13, we compare the NMSE of different channel estimation techniques. Considering the parameters as given in the figure, we have $k_\nu \ll 16$ and hence $L_p = 12$ as per Lemma 1. Observe the significant difference between the NMSE values for the sparse signal recovery-based techniques and 'QAM-pilot'. The proposed method and 'MSP' are better than 'QAM-pilot' in terms of NMSE since they are based on estimation of **h** as a vector. Moreover, the proposed method gives a slightly better NMSE than 'MSP'. This improvement may be attributed to the CSC model acting independently on each delay tap and the sequential initial estimate method giving equal preference to all users. For 'QAM-pilot' of [8], an error floor occurs at higher SNR_p. The error floor issue arises as this method detects a path by thresholding the components scalar-wise and then evaluating the corresponding channel coefficient rather than estimating **h** as a vector. The CRLB derived in Section IV-E is also shown in Fig. 13.

E. Complexity Analysis

Fig. 14 highlights the complexity of the proposed CSC modeling in terms of the number of complex multiplications.
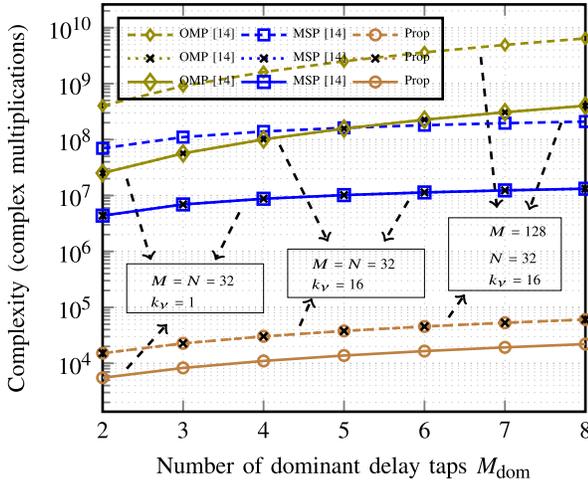
Fig. 14. Complexity analysis of proposed method.
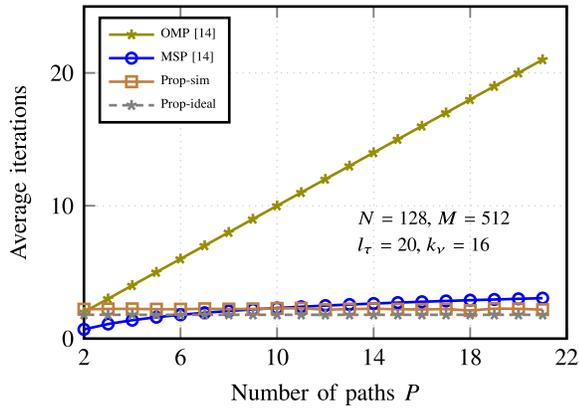
Fig. 15. Average iterations for convergence of **Algorithm 1.**

The legend 'OMP' refers to the direct OMP method discussed in [14]. Observe from TABLE I and Fig. 14 that the proposed method is superior in two aspects of complexity: (i) For a given DD grid of dimension N and M, the proposed method's complexity depends only on the maximum Doppler k_ν of the channel. It is observed that up to $k_\nu + 1 \approx \frac{N}{2}$, the proposed method offers a noticeable complexity reduction; (ii) The complexity of the proposed method is completely independent of the delay dimension M, since each delay path is processed independently. On the other hand, the conventional sparse signal recovery methods always have a constant maximum complexity independent of the Doppler conditions of the channel. Fig. 15 presents the analysis of the average number of iterations required for the convergence of **Algorithm 1** for different values of the number P of multipaths. Observe that the number of iterations required is independent of P. The legends 'Prop-ideal' and 'Prop-sim' show the average number of iterations required for the proposed method as per TABLE I and the simulations respectively. These two curves exhibit high correlation and remain at the same level as P increases. On the other hand, for 'OMP' and 'MSP' algorithms, the number of iterations increases with P.

VI. CONCLUSION

In this paper, we presented a channel estimation technique for OTFS-SCMA based on CSC. The proposed method considers sparse vectors of optimum length as pilots, which follow the same sparsity pattern of the SCMA data codewords of all users. The sparse pilot vectors are designed such that the mutual coherence of the corresponding dictionary is minimized. The non-orthogonal arrangement of the pilot vectors ensures that the guard band overhead is minimal and does not escalate with the number of users. Moreover, the proposed CSC model converts the overall problem to the channel estimation at each delay tap. This reduction of the dimensionality, in turn, lowers the complexity of the sparse recovery algorithm significantly. Finally, impressive BER performances and spectral efficiency are obtained, maintaining a reduced guard band overhead and complexity. These results corroborate the proposed method's suitability for a high Doppler uplink scenario.

APPENDIX A
PROOF OF LEMMA 1

Consider the channel estimation problem defined in (19):

$$\mathbf{y}_l = \mathbf{P}\mathbf{h}_{l-\bar{l}} \quad \text{for } l = \bar{l},\ \bar{l}+1, \ldots, M-1 \quad (39)$$

where $\mathbf{h}_{l-\bar{l}}$ is the sparse signal to be recovered. The minimum value of L_p is derived considering the following points:

1) For **Algorithm 1**, the maximum possible number of columns in $\mathbf{P}_{(\tilde{S}^{(i)})}$ is $2J$. For any matrix $\mathbf{A}$, we have $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ and $\mathbf{A}^\dagger\mathbf{y}$ results in a solution having the number of non-zero elements as rank($\mathbf{A}^\dagger$). Since $\mathbf{P}^\dagger_{(\tilde{S}^{(i)})}\mathbf{y}_l$ (line 13 of **Algorithm 1**) has to recover $2J$ elements in each iteration the following condition must be satisfied:

$$\text{rank}(\mathbf{P}^\dagger_{(\tilde{S}^{(i)})}) \geq 2J \ \Rightarrow L_p \geq 2J. \quad (40)$$

2) Successful sparse signal recovery is guaranteed if a minimum number of observations are available, as discussed in [36] and [40]. In general, for a signal having sparsity of (s/V), for successful recovery, the number U of observations must satisfy $U \geq \lceil cs\log V \rceil$ with $1 < c \leq 2$. Specific to (39), U is limited by the Doppler spread of the channel. Assuming the presence of maximum Doppler paths in the channel, we have $U = L_p + 2k_\nu$ and for successful recovery, the following condition must be satisfied:

$$L_p \geq \lceil cJ \log(J(2k_\nu + 1)) \rceil - 2k_\nu. \quad (41)$$

For the EVA parameters of TABLE III ($k_\nu = 16$) with the minimum $L_p = 12$ given by (40), we have from (41): $c = \frac{L_p + 2k_\nu}{J\log(J(2k_\nu+1))} \approx 1.4$. Thus $L_p = 12$ satisfies the condition in (41). However, if the maximum Doppler paths are present in the channel, then a higher value of L_p is desirable as observed in the simulation results shown in Fig. 10.

3) For the CSC problem, the global sparsity s and the local sparsity s_L are related as $s \approx (s_L/(2p-1))V$, where p

is the length of a dictionary vector and V is the length of the sparse signal to be recovered [33]. Specific to (39), we have a fixed sparsity $s = J$, pilot vector length L_p corresponds to p, and V is $J(2k_\nu + 1)$. Considering the uniform distribution of the Doppler taps, we take $s_L = 1$. The sparsity relation can be expressed as

$$J \approx \frac{1}{2L_p - 1} J(2k_\nu + 1) \Rightarrow L_p \approx k_\nu + 1. \quad (42)$$

For EVA parameters, we get $L_p \approx 17$.

4) Since the pilot vectors follow the same sparsity structure of the codewords with length K, we must have $[L_p]_K = 0$.

To minimize the guard band overhead, L_p is chosen as the nearest integer satisfying the above conditions:

$$L_p \geq \max\{2J, \lceil cJ \log(J(2k_\nu + 1)) \rceil - 2k_\nu, k_\nu + 1\} \quad (43)$$
$$\text{with } [L_p]_K = 0 \text{ and } 1 < c \leq 2$$

REFERENCES

[1] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," 2018, *arXiv:1802.02623*.

[2] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "Multiple access in the delay-Doppler domain using OTFS modulation," 2019, *arXiv:1902.03415*.

[3] V. Khammammetti and S. K. Mohammed, "OTFS-based multiple-access in high Doppler and delay spread wireless channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 528–531, Apr. 2019.

[4] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "OTFS-NOMA: An efficient approach for exploiting heterogenous user mobility profiles," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7950–7965, Nov. 2019.

[5] A. Chatterjee, V. Rangamgari, S. Tiwari, and S. S. Das, "Non orthogonal multiple access with orthogonal time frequency space signal transmission," 2020, *arXiv:2003.06387*.

[6] K. Deka, A. Thomas, and S. Sharma, "OTFS-SCMA: A code-domain NOMA approach for orthogonal time frequency space modulation," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5043–5058, Aug. 2021.

[7] R. Hadani and S. Rakib, "OTFS methods of data channel characterization and uses thereof," U.S. Patent 9 444 514 B2, Sep. 13, 2016.

[8] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in Delay–Doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

[9] L. Zhao, W.-J. Gao, and W. Guo, "Sparse Bayesian learning of delay-Doppler channel for OTFS system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2766–2769, Dec. 2020.

[10] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 206–212.

[11] M. Li, S. Zhang, F. Gao, P. Fan, and O. A. Dobre, "A new path division multiple access for the massive MIMO-OTFS networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 903–918, Apr. 2021.

[12] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, Jr., "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, Aug. 2019.

[13] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1994–2009, Sep. 2020.

[14] O. K. Rasheed, G. D. Surabhi, and A. Chockalingam, "Sparse delay-Doppler channel estimation in rapidly time-varying channels for multi-user OTFS on the uplink," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 4–8.

[15] A. B. Sergienko and V. P. Klimentyev, "Spectral efficiency of uplink SCMA system with CSI estimation," in *Proc. 20th Conf. Open Innov. Assoc. (FRUCT)*, Apr. 2017, pp. 391–397.

[16] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 853–857.

[17] Y. Wang, S. Zhou, L. Xiao, X. Zhang, and J. Lian, "Sparse Bayesian learning based user detection and channel estimation for SCMA uplink systems," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2015, pp. 1–5.

[18] E. Heo, N. Kim, and H. Park, "Sparse structure-based channel estimation for uplink SCMA system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8037–8046, Sep. 2017.

[19] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "OTFS signaling for uplink NOMA of heterogeneous mobility users," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3147–3161, May 2021.

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

[21] K. Price, R. M. Storm, and J. Lampinen, *Differential Evolution*. Berlin, Germany: Springer, 2005.

[22] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

[23] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Tech.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.

[24] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.

[25] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2014, pp. 1–5.

[26] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.

[27] C. Zhang, Y. Luo, and Y. Chen, "A low-complexity SCMA detector based on discretization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2333–2345, Apr. 2018.

[28] 3GPP. (2010). *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception (Release 9)*. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/36_series/36.104/

[29] M. Lewicki and T. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in *Proc. NIPS*. Cambridge, MA, USA: MIT Press, 1998, pp. 730–736.

[30] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 2528–2535.

[31] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 391–398.

[32] B. Wohlberg, "Efficient convolutional sparse coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 7173–7177.

[33] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5687–5701, Nov. 2017.

[34] E. Zisselman, J. Sulam, and M. Elad, "A local block coordinate descent algorithm for the CSC model," in *Proc. IEEE CVPR*, Long Beach, CA, USA, Jun. 2019, pp. 8200–8209.

[35] E. Plaut and R. Giryes, "A greedy approach to $\ell_{0,\infty}$-based convolutional sparse coding," *SIAM J. Imag. Sci.*, vol. 12, no. 1, pp. 186–210, Jan. 2019.

[36] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[37] S. Kay, *Fundamentals of Statistical Processing: Estimation Theory*, vol. 1. London, U.K.: Pearson, 1993.

[38] K. Deka, M. Priyadarsini, S. Sharma, and B. Beferull-Lozano, "Design of SCMA codebooks using differential evolution," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–7.

[39] G. D. Surabhi and A. Chockalingam, "Low-complexity linear equalization for OTFS modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, Feb. 2020.

[40] 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.

Anna Thomas (Student Member, IEEE) received the B.Tech. degree from the Amal Jyothi College of Engineering, Kottayam, India, in 2008, and the M.Tech. degree in signal processing from the Rajagiri School of Engineering, Cochin, India, in 2014. She is currently pursuing the Ph.D. degree with the School of Electrical Sciences, Indian Institute of Technology Goa, India. From 2008 to 2011, she worked as a Software Engineer with QBurst Technologies, Thiruvananthapuram, India. From 2014 to 2018, she worked as an Assistant Professor with Kannur University and Kerala Technological University, Thiruvananthapuram. Her research interests include signal processing for future wireless communications, in the area of orthogonal time frequency space modulation, sparse code multiple access; and coding and information theory applications for the same.
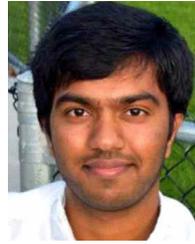
Patchava Raviteja received the M.E. degree in telecommunications from the Indian Institute of Science, Bengaluru, India, in 2014, and the Ph.D. degree from Monash University, Melbourne, Australia, in 2019. From 2014 to 2015, he worked at Qualcomm India Private Ltd., Bengaluru, on WLAN systems design. From 2019 to 2020, he was a Post-Doctoral Research Fellow with Monash University. He is currently working as a Senior Engineer at Qualcomm, Inc., San Diego, CA, USA. His current research interest includes new waveform designs for next generation wireless systems. He was a recipient of Prof. S. V. C. Aiya Medal from the Indian Institute of Science in 2014. He also received the Best Student Paper Award for his publication at the SPAWC 2018 Conference, Greece.

Kuntal Deka received the Ph.D. degree from the Indian Institute of Technology Guwahati, India, in 2016. From 2015 to 2018, he was an Assistant Professor with the Indian Institute of Information Technology Guwahati and from 2018 to 2021, he was an Assistant Professor with the School of Electrical Sciences, Indian Institute of Technology Goa, India. He is currently working as an Assistant Professor at the Department of EEE, Indian Institute of Technology Guwahati. His research interests lie in the areas of error correcting codes, modulation, and multiple access techniques for modern wireless communication systems.

Sanjeev Sharma (Senior Member, IEEE) is currently working as an Assistant Professor at the Electronics Engineering Department, Indian Institute of Technology (IIT) (BHU) Varanasi, India. He has published around 40 journals and conferences in wireless communications. His research interests lie in signal processing for wireless communications and networking, mathematical modeling, simulation, design, and analysis of wireless systems.