OTFS-SCMA: A Code-Domain NOMA Approach for Orthogonal Time Frequency Space Modulation

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

***Abstract*— Orthogonal time frequency space (OTFS) modulation is a two-dimensional (2-D) modulation technique that has the potential to overcome the challenges faced by orthogonal frequency division multiplexing (OFDM) in high Doppler environments. The performance of OTFS in a multi-user scenario with orthogonal multiple access (OMA) techniques has been impressive. Due to the requirement of massive connectivity in 5G and beyond, it is essential to devise and examine the OTFS system with the existing non-orthogonal multiple access (NOMA) techniques. This paper proposes a multi-user OTFS system based on a code-domain NOMA technique called sparse code multiple access (SCMA). This system is referred to as the OTFS-SCMA model. The framework for OTFS-SCMA is designed for both downlink and uplink. First, the sparse SCMA codewords are strategically placed on the delay-Doppler plane. The overall overloading factor of the OTFS-SCMA system is equal to that of the underlying basic SCMA system. The receiver in downlink performs the detection in two sequential phases: first, the conventional OTFS detection using the method of linear minimum mean square error (LMMSE) estimation, and then the SCMA detection. We propose a single-phase detector based on a message-passing algorithm (MPA) to detect multiple users' symbols for the uplink. The expressions for the asymptotic diversity orders of the proposed OTFS-SCMA system are derived for downlink and uplink. OTFS-SCMA provides a significant diversity gain over other multiple access systems for OTFS. Based on the diversity analysis, an algorithm is proposed to devise an optimal codeword allocation scheme. The performance of the proposed OTFS-SCMA system is validated through extensive simulations both in downlink and uplink. We consider delay-Doppler planes of different parameters and various SCMA systems of overloading factor up to 200%. The performance of OTFS-SCMA is compared with those of the existing OTFS-OMA, OFDM-SCMA and OTFS-power-domain (PD)-NOMA techniques. The analysis of OTFS-SCMA with channel estimation is also presented along with the BER performance.**

***Index Terms*— OTFS, SCMA, NOMA, message passing algorithm, delay-Doppler plane.**

I. Introduction

ORTHOGONAL time frequency space (OTFS) has emerged as a promising modulation technique that eliminates the shortcomings of orthogonal frequency division multiplexing (OFDM) in 5G [1], [2]. Many applications facilitated by 5G involve vehicle-to-vehicle and vehicle-to-infrastructure communication scenario with the presence of high-speed vehicles. OFDM is not robust to such high Doppler environments as the sub-carriers do not remain orthogonal. In these scenarios OTFS modulation technique succeeds by operating in the delay-Doppler domain instead of the conventional time-frequency domain. By representing the channel in the delay-Doppler domain, it is possible to convert the time-variant channel response $h(\tau, t)$ to the time-invariant channel response $h(\tau, \nu)$ where t, τ and ν represent time, delay and Doppler shift respectively. Apart from the channel representation, the most distinguishing feature of OTFS modulation is that the information symbols are directly placed on the delay-Doppler grid. The existence of 2-D orthogonal signals in the delay-Doppler domain is the inspiration for developing the OTFS modulation technique, where they behave as localized simultaneously in both time and frequency. An information symbol on a delay-Doppler grid spans the entire allotted time-frequency plane using these orthogonal basis functions. Thus, irrespective of the delay-Doppler grid's positions, all the symbols with the same power experience channels of the same quality. This invariance property provides OTFS the upper hand in high Doppler channel conditions [2].

Non-orthogonal multiple access (NOMA) schemes can be divided into two domains: *power-domain* and *code-domain*. In power-domain (PD) NOMA, different users are identified by using different assigned power levels [3]. In code-domain NOMA, multiple users are distinguished using different codewords assigned to them [4]. Sparse code multiple access (SCMA) is a code-domain NOMA technique where sparse codewords are given to the users' symbols [5], [6]. The sparsity of the codewords facilitates the successful use of the message passing algorithm (MPA) to detect the users' symbols.

Related works: The 5G and beyond communication scenarios are impaired by high Doppler and demand massive connectivity. Therefore it is paramount to study multiple access technologies in the OTFS framework. The authors in [7] proposed an OTFS-OMA scheme such that the multiple users' symbols are spread at an equal interval over the entire delay-Doppler plane. Consequently, these symbols are restricted to non-overlapping contiguous sub-blocks in the time-frequency plane. An OTFS-OMA scheme is presented in [8] where the multiple users' symbols are placed in non-overlapping contiguous sub-blocks of the delay-Doppler plane, multiplexed along either the delay or the Doppler

Manuscript received October 25, 2020; revised January 13, 2021 and February 26, 2021; accepted April 18, 2021. Date of publication April 23, 2021; date of current version August 16, 2021. The associate editor coordinating the review of this article and approving it for publication was F. Zhou. *(Corresponding author: Kuntal Deka.)*

Kuntal Deka and Anna Thomas are with the School of Electrical Sciences, Indian Institute of Technology Goa, Ponda 403401, India (e-mail: kuntal@iitgoa.ac.in; anna183422001@iitgoa.ac.in).

Sanjeev Sharma is with the Department of Electronics Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi 221005, India (e-mail: sanjeev.ece@iitbhu.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3075237.

Digital Object Identifier 10.1109/TCOMM.2021.3075237

axis. The OTFS-OMA method in [9] allocates interleaved symbols of different users in the time-frequency domain. Note that here unlike [7], the allocation is not contiguous in the time-frequency plane. Instead, it is interleaved, i.e., between two symbols of a particular user, symbols from other users can be placed. The authors in [10] considered a heterogeneous PD-NOMA system where the user with the highest velocity operates with OTFS and the remaining users operate with OFDM. In [11], PD-NOMA is considered solely with OTFS. Link level simulations of the coded OTFS-PD-NOMA system are presented.

Contributions: In this paper, we propose an OTFS-NOMA scheme based on a code-domain NOMA technique. The code-domain NOMA approach considered here is SCMA. Although OTFS-NOMA schemes have been studied in the power-domain, this work is the first attempt to build an OTFS-NOMA system in the code-domain, to the best of the authors' knowledge. The next-generation wireless communication systems anticipate two significant challenges: (1) massive connectivity and (2) high Doppler. The objective is to devise a communication scheme to collectively resolve these two issues by considering both the technologies, i.e., OTFS and SCMA. The main contributions are highlighted below.

- We begin with two schemes of allocating the SCMA codewords over the delay-Doppler plane. *Scheme-1* assigns the vector codewords in the vertical sub-columns along the Doppler direction. In *Scheme- 2*, the codewords are placed horizontally in the order of the delay bins. First, we configure the OTFS-SCMA model in the downlink. The downlink detection is a two-phased process: (1) OTFS detection followed by (2) SCMA detection.
- Next, we devise the strategy of OTFS-SCMA in the uplink. As the simple two-stage detection approach adopted in the downlink is not applicable in uplink, we put forward a combined method of OTFS and the SCMA detection for uplink. An MPA-based detector is specifically formulated to detect the multiple users' data at one go.
- The asymptotic diversity orders of OTFS-SCMA are evaluated for both downlink and uplink. Guided by the theoretical diversity analysis, we propose an optimal scheme of allocating the SCMA codewords over the delay-Doppler grid. This scheme can achieve full diversity aided by the knowledge of the multi-paths' delay and Doppler values at the transmitter.
- The BER performances of OTFS-SCMA are evaluated for downlink and uplink. In most cases, the OTFS-SCMA approach is found to provide better results than the existing OTFS-OMA systems. OTFS-SCMA is also found to perform better than the OTFS-PD-NOMA scheme. For the performance comparison with the OTFS-PD-NOMA scheme, we have designed the SCMA codebooks explicitly for an overloading factor of 200%.
- The embedded pilot-based channel estimation (CE) process for the OTFS-SCMA system is designed. Simulation result promises that CE for OTFS-SCMA is efficient with minimal overhead.

Outline: Section II describes the preliminary concepts about OTFS modulation, the SCMA system model and the method of OTFS-PD-NOMA. OTFS-SCMA is proposed in Section III. The simulation results are presented and analyzed in Section IV. Section V concludes the paper.

Notations: $\mathbb{C}$ denotes the set of all complex numbers. $\mathbb{A}$ refers to the underlying alphabet. Boldface upper-case and boldface lower-case letters are used to represent matrices and vectors, respectively. For an $m \times n$ matrix $\mathbf{B}$, vec$(\mathbf{B})$ denotes the $mn \times 1$ column vector, which is obtained by vertical concatenation of the n columns of $\mathbf{B}$. I_N denotes the identity matrix of size $N \times N$. For a vector $\mathbf{h}$, diag$(\mathbf{h})$ is the diagonal matrix, with the main diagonal being $\mathbf{h}$. The complex conjugation of x is denoted by x^*. For a matrix $\mathbf{X}$, $\mathbf{X}^T$ and $\mathbf{X}^\dagger$ represent the transpose of $\mathbf{X}$ without complex conjugation and with complex conjugation, respectively. For any integers k and N, the notation $[k]_N$ refers to $(k \mod N)$. For any real number x, $\lfloor x \rceil$ denotes the integer nearest to x. The notation $|S|$ represents the number of distinct elements in the set S.

II. Preliminaries

A. OTFS

This section briefly explains the representation of the symbols in the delay-Doppler domain and the modulation-demodulation in OTFS. In OTFS, the information symbols (e.g. quadrature amplitude modulation (QAM) symbols) are arranged on a 2-D grid called delay-Doppler plane represented by $\Gamma_{N,M}$, where M and N are the total numbers of the delay and the Doppler bins respectively. $\Gamma_{N,M}$ can accommodate a total of MN QAM symbols. The input signal $x[k,l]$ refers to the symbol placed at the intersection of the k^{th} Doppler bin and the l^{th} delay bin, where $k = 0, 1, \ldots, N-1$ and $l = 0, 1, \ldots, M-1$. Usually, the horizontal and the vertical axis of $\Gamma_{N,M}$ specifies the M delay values and the N Doppler shifts, respectively.

1) Design Parameters of Delay-Doppler Plane: Consider a data transmission frame of duration $T_f = NT$ and bandwidth $B = M\triangle f$ where $\triangle f$ is the subcarrier frequency and T is the symbol duration, such that $T = \frac{1}{\triangle f}$. One must have $\triangle f < \frac{1}{\tau_{\max}}$ and $T < \frac{1}{\nu_{\max}}$ where $\tau_{\max}$ and $\nu_{\max}$ are the delay and the Doppler spread, respectively. $\Gamma_{N,M}$ has a delay interval of $\triangle\tau = \frac{1}{M\triangle f}$ and a Doppler interval of $\triangle\nu = \frac{1}{NT}$. Hence to transmit MN symbols over a frame of duration T_f and bandwidth B, the choice of N and M depends on the delay and Doppler conditions of the channel. A channel with a high Doppler spread $\nu_{\max}$ would require a higher $N\triangle\nu$. It points to a smaller T and a larger $\triangle f$, which directly implies a larger N and a smaller M. Similarly, if the channel has a higher delay spread $\tau_{\max}$, we require a higher $M\triangle\tau$ and $\Gamma_{N,M}$ with a larger M and a smaller N are to be selected [12].

2) Basic Principle of OTFS Modulation-Demodulation: The basic modulation-demodulation involved in OTFS is shown in Fig. 1. It includes two stages of modulation/demodulation, one in the delay-Doppler domain and the other in the time-frequency domain.

3) Operations in Delay-Doppler Domain : Initially, the QAM symbols $(x[k,l]\text{s})$ to be transmitted are arranged on

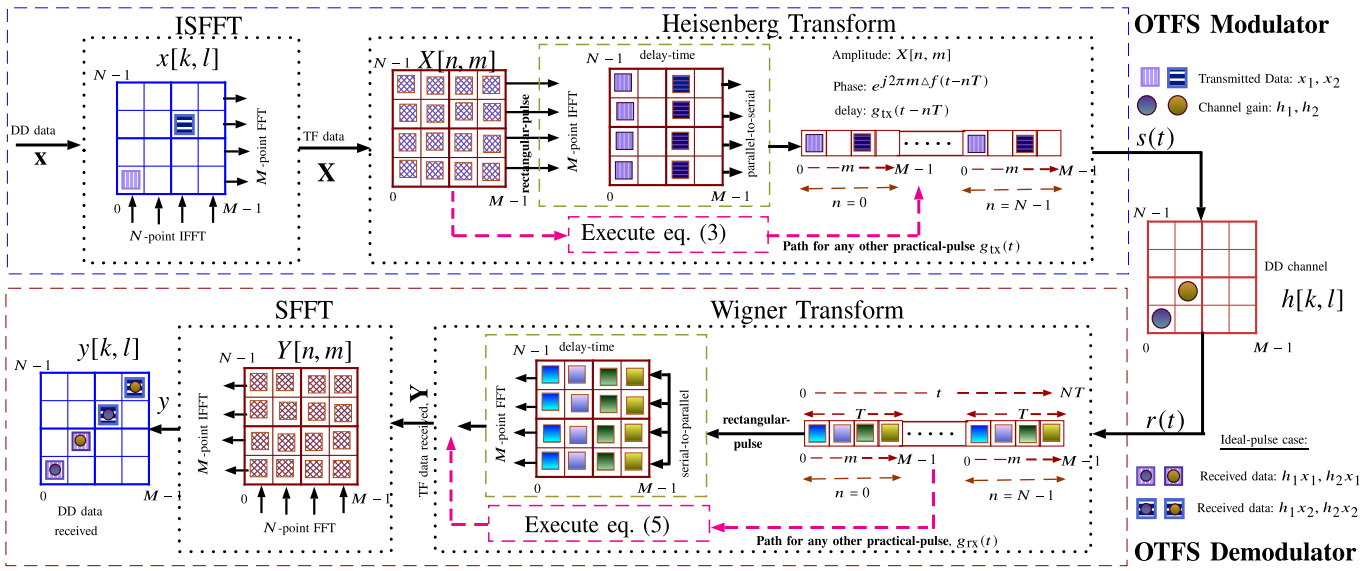


Fig. 1. OTFS Block Diagram (delay-Doppler, frequency-time, and delay-time are plotted as abscissa-ordinate).

$\Gamma_{N,M}$. The symbols in the delay-Doppler domain are transformed into the time-frequency domain using inverse symplectic finite Fourier transform (ISFFT) as:

$$X[n,m] = \frac{1}{MN} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k,l] e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})}. \quad (1)$$

Similarly, at the receiver side, the signals received in the time-frequency domain are finally transformed back to the delay-Doppler domain using symplectic finite Fourier transform (SFFT) as:

$$y[k,l] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y[n,m] e^{-j2\pi(\frac{nk}{N} - \frac{ml}{M})}. \quad (2)$$

Both SFFT and ISFFT can be interpreted as combining M-point IFFTs along each Doppler and N-point FFTs along each delay, and their inverses, respectively. The 2-D grid has to be limited to a dimension of $N \times M$. So a rectangular windowing function is used both at the transmitter and the receiver.

4) Operations in Time-Frequency Domain : Heisenberg transform converts the 2-D modulated signal in the time-frequency domain into the time-domain using any basis pulse $g_{\text{tx}}(t)$ as:

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n,m] e^{j2\pi m \triangle f(t-nT)} g_{\text{tx}}(t - nT). \quad (3)$$

For the particular case of a rectangular pulse, (3) simplifies to a collection of M-point IFFTs along each time, resulting in the delay-time data followed by a parallel-to-serial conversion. Hence for $N = 1$, (3) is equivalent to OFDM modulation. The signal $s(t)$ is transmitted over the wireless channel with impulse response $h(\tau, \nu)$. The received signal $r(t)$ is given by

$$r(t) = \iint h(\tau,\nu) e^{j2\pi\nu(t-\tau)} s(t-\tau)\, d\tau d\nu + z(t) \quad (4)$$

where $z(t)$ is the additive white Gaussian noise (AWGN) signal. The received signal $r(t)$ in the time domain is converted to $Y[n,m]$ in the time-frequency domain by applying Wigner transform:

$$Y[n,m] = \int_{-\infty}^{\infty} g^*_{\text{rx}}(t - nT) r(t) e^{-j2\pi m \triangle f(t-nT)} dt \quad (5)$$

where, $g_{\text{rx}}(t)$ is the basis pulse at the receiver. When $g_{\text{rx}}(t)$ is rectangular, (5) is simply a collection of M-point FFTs along each time, after serial-to-parallel conversion of $r(t)$. Hence for $N = 1$, (5) denotes OFDM demodulation. The selection of $g_{\text{tx}}(t)$ and $g_{\text{rx}}(t)$ is to be done carefully so that ideally, they obey the bi-orthogonality property; otherwise, there will be interference from different delay and Doppler bins. The bi-orthogonality property can be written as:

$$\int e^{-j2\pi m \triangle f(t-nT)} g^*_{\text{rx}}(t - nT) g_{\text{tx}}(t)\, dt \; = \delta(m)\delta(n). \quad (6)$$

The pulses $g_{\text{tx}}(t)$ and $g_{\text{rx}}(t)$, which satisfy the bi-orthogonality property are *ideal* pulses. For the simple *rectangular* pulses, $g_{\text{tx}}(t)$ and $g_{\text{rx}}(t)$ have amplitude $1/\sqrt{T}$ for $t \in [0, T]$ and 0 at all other values. Though the rectangular pulses don't satisfy the bi-orthogonality condition, they are frequently considered in simulations and analysis.

5) Input-Output Relation: The input-output relation refers to the relation between $x[k,l]$s and $y[k,l]$s. The input-output relation helps to design detectors at the receiver side. The channel response in the delay-Doppler domain can be sparsely represented as

$$h(\tau,\nu) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(\nu - \nu_i) \quad (7)$$

where P denotes the number of paths in the channel, $\delta(.)$ represents the Dirac delta function; h_i, τ_i, and ν_i denote the i^{th} path's gain, delay, and Doppler shift respectively. Considering the channel coefficients as i.i.d and a uniform scattering

profile, $h_i \sim \mathcal{CN}(0, \frac{1}{P})$. The delay and the Doppler taps for the i^{th} path can be alternatively represented as [12]

$$\tau_i = \frac{l_{\tau_i}}{M\triangle f}, \quad \nu_i = \frac{k_{\nu_i} + \kappa_{\nu_i}}{NT} \tag{8}$$

with some integers l_{τ_i}, k_{ν_i} and $\kappa_{\nu_i} \in [-0.5, 0.5]$, which represents the delay tap, Doppler tap, and fractional Doppler shift of the channel, respectively, for the desired delay resolution of $\frac{1}{M\triangle f}$ and Doppler resolution of $\frac{1}{NT}$. In wide-band systems, path delays can be approximated to the nearest delay tap using the sampling time of $\frac{1}{M\triangle f}$, which is viewed as the absence of fractional delay. On the other hand, sometimes $\frac{1}{NT}$ may not provide a sufficient Doppler resolution thereby resulting in fractional Doppler shifts. With the above sparse channel representation, the received signal $y[k,l]$s in the delay-Doppler domain for rectangular pulses can be expressed in terms of the input signal $x[k,l]$s as [12]:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{9}$$

where $\mathbf{x} \in \mathbb{A}^{MN\times 1}$, $\mathbf{y} \in \mathbb{C}^{MN\times 1}$, and $\mathbf{z} \in \mathbb{C}^{MN\times 1}$ are the vectorized versions of input ($x[k,l]$s), output ($y[k,l]$s), and AWGN, respectively; $\mathbf{H} \in \mathbb{C}^{MN\times MN}$ is called the coefficient matrix. The procedure of determining $\mathbf{H}$ for any practical waveform is explained in detail in [13] and omitted here. The linear relationship in (9) is valid for both ideal and practical pulse, though the elements of $\mathbf{H}$ change. The coefficient matrix $\mathbf{H}$ has a maximum of $S = \sum_{i=1}^{P}(2N_i+1)$ non-zero elements in each row and in each column, where N_i denotes the number of neighboring Doppler points interfering with the particular point under consideration. In the absence of fractional Doppler shifts (i.e., $N_i = 0$), we have $S = P$. This makes $\mathbf{H}$ a highly sparse matrix. Owing to the sparsity of $\mathbf{H}$, the detection can be carried out with the help of MPA [12], [14]. In place of MPA, we can use other low-complexity linear equalization algorithms like linear minimum mean square error (LMMSE) estimator, by considering the block-circulant property of $\mathbf{H}$ matrix [15], [16].

B. SCMA

SCMA is a sophisticated code-domain NOMA technique. A specific codebook is assigned to each user such that no two users would have to use the same codeword. The basic parameters of SCMA and the other related concepts [5], are briefly explained.

1) SCMA Parameters: An SCMA system is represented as a $J \times K$ model, where J users are sharing K orthogonal resources.[1] As $J > K$, we have an overloaded system with an overloading factor, $\lambda = \frac{J}{K} > 100\%$. As an example, the 6×4 SCMA system shown in Fig. 2 has 6 users sharing 4 resources with an overloading factor of 150%. The system has six distinct codebooks $\{\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_6}\}$, one dedicated for each user [6]. The alphabet size is $A = 4$. Each codebook consists of four different complex codewords of length $K = 4$ representing four different information symbols $\{0, 1, 2, 3\}$. Depending on the input data, one codeword is selected by each user for transmission. The set of the selected codewords is denoted by $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_6\}$ where $\mathbf{x}_j \in \mathbb{C}^{4\times 1}$, $j = 1, \ldots, 6$. The performance of an SCMA system is susceptible to the codebooks [6].

[1] A resource may be time slot/frequency-band/code.

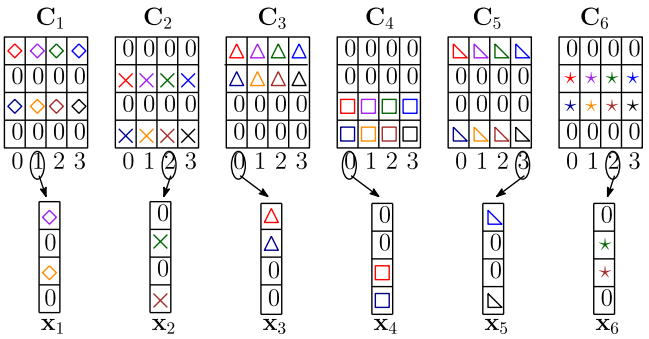


Fig. 2. Basic 6×4 SCMA Model.

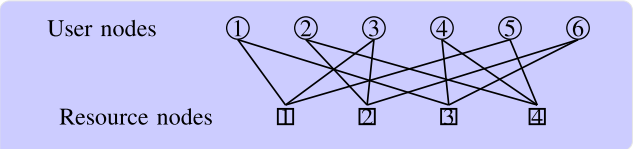


Fig. 3. Factor graph for $J = 6$, $K = 4$ SCMA system.

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{10}$$

2) Factor Graph: Observe from Fig. 2 that the codewords of any particular codebook are sparse and contain 0s in specific locations. For any user, if a codeword has a non-zero element in the k^{th} resource, then it implies that the user is occupying the k^{th} resource. A factor graph can graphically represent the sharing of the resources among multiple users. The factor graph for a $J \times K$ SCMA system contains J user nodes and K resource/factor nodes. An edge is assigned between the j^{th} user node and the k^{th} resource node only if it occupies the k^{th} resource. Fig. 3 shows the factor graph for the SCMA model described in Fig. 2. The degrees of a user node and a resource/factor node are denoted by d_v and d_f respectively. For the factor graph shown in Fig. 3, we have $d_v = 2$ and $d_f = 3$. An alternative representation of the factor graph is the factor matrix, $\mathbf{F}$. Each edge in the factor graph is denoted by a '1' in the factor matrix [4]. For the factor graph shown in Fig. 3, the factor matrix is shown in (10). The sparsity of the factor graph or the matrix indicates that MPA can be used in SCMA detection. It has been extensively reported in the literature that the MPA-based detection provides excellent performance in SCMA [5], [17].

3) Downlink: In the downlink, the base station (BS) first sums up all J users' codewords. This superimposed signal is broadcast to every user. The received signal $\mathbf{y}_i = [y_{i1}, \ldots, y_{iK}]^T$ at the i^{th} user can be expressed as

$$\mathbf{y}_i = \text{diag}\,(\mathbf{h}_i) \sum_{j=1}^{J} \mathbf{x}_j + \mathbf{z}_i \tag{11}$$

where $\mathbf{h}_i = [h_{i1}, h_{i2}, \ldots, h_{iK}]$ is the channel impulse response vector between the BS and the i^{th} user, $\mathbf{z}_i$ is the AWGN at the i^{th} user and is complex Gaussian distributed, i.e., $\mathbf{z}_i \sim \mathcal{CN}(0, N_0\mathbf{I}_K)$.

4) Uplink: In uplink, each user incorporates the respective channel to transmit the information to the BS. The BS receives the signal $\mathbf{y}$ which is given by [18],

$$\mathbf{y} = \sum_{j=1}^{J} \text{diag}(\mathbf{h}_\text{j})\, \mathbf{x}_j + \mathbf{z} \tag{12}$$

where $\mathbf{z}$ is the AWGN at the BS.

C. OTFS-PD-NOMA

Here we briefly explain the method for OTFS-PD-NOMA. A similar kind of OTFS-NOMA is considered in [11]. OTFS-PD-NOMA is presented here as we compare its performance with that of OTFS-SCMA. Consider a downlink situation containing one BS and two users. PD-NOMA is preferable when there is a disparity in the signal-strengths of the users. Suppose the first user is nearer than the second one to the BS. P_1 and P_2 denote the power assigned to the users. As the signal at the first user is stronger than that at the second user, the power levels are selected so that $P_1 < P_2$. The BS superimposes the signals of two users and broadcasts them to both. By considering the input-output relations described in Section II-A.5, we can wirte the vectorized received signal at the first user as:

$$\mathbf{y}_1 = \mathbf{H}_1 \left(P_1\mathbf{x}_1 + P_2\mathbf{x}_2\right) + \mathbf{z}_1 \tag{13}$$

where $\mathbf{H}_1$ represents the input-output relation between the BS and the first user. The successive-interference-cancellation (SIC) is used at the location of the first user. First, the second user's data symbols are estimated from (13) using a conventional MPA-based OTFS detector by assuming the signal from the first user as noise. Suppose, these estimates are given by $\hat{\mathbf{x}}_2^1$. The superscript '1' signifies that these are the estimates at the location of the first user. The effect of the second user is canceled to get the following signal:

$$\mathbf{y}_{1,\text{SIC}} = \mathbf{y}_1 - P_2\mathbf{H}_1\hat{\mathbf{x}}_2^1. \tag{14}$$

The estimates $\hat{\mathbf{x}}_1$ of the data symbols for the first user are obtained from (14) by feeding $\mathbf{y}_{1,\text{SIC}}$ to an MPA detector. The signal received by the second user is given by:

$$\mathbf{y}_2 = \mathbf{H}_2 \left(P_1\mathbf{x}_1 + P_2\mathbf{x}_2\right) + \mathbf{z}_2 \tag{15}$$

where $\mathbf{H}_2$ represents the input-output relation between the BS and the second user. MPA detector is used to find the estimates $\hat{\mathbf{x}}_2$ of the second user from (15).

III. Proposed Method of SCMA-Based OTFS-NOMA

In this section, we propose the system model for OTFS-SCMA in downlink and uplink. First we present two methods of allocating the SCMA codewords over the delay-Doppler grid. The OTFS-SCMA system structure in downlink and uplink and the operations to be carried out at the transmitter and the receiver side are described in detail.

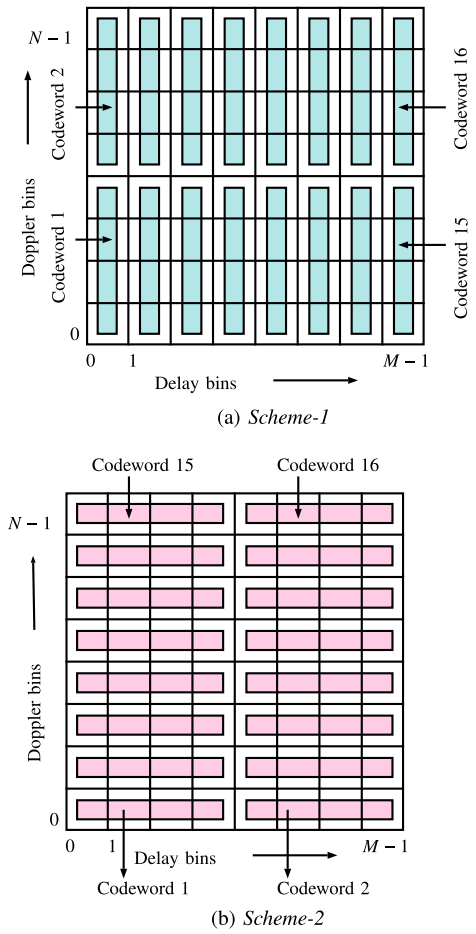


Fig. 4. Allocation of SCMA codewords of length $K = 4$ on delay-Doppler plane $\Gamma_{8,8}$ for a specific user.

Then the diversity analysis is carried out to highlight two aspects: (1) advantage of OTFS-SCMA over OTFS-OMA and (2) optimal scheme of allocating the SCMA codewords over the delay-Doppler grid. Finally, the complexity of the proposed OTFS-SCMA is analyzed for both downlink and uplink.

A. Allocation of SCMA Codewords Over Delay-Doppler Grid

A $J \times K$ SCMA system is considered with an overloading factor $\lambda = \frac{J}{K}$. The length of a complex SCMA codeword is K. We assume that M and N are integer multiples of K, i.e., $[M]_K = [N]_K = 0$. For allocation of the SCMA codewords over the delay-Doppler plane, we consider the following two schemes.

Scheme-1: The SCMA codewords for every j^{th} user, $j = 1, \ldots, J$ are allocated on the delay-Doppler plane $\Gamma_{N,M}$ along the Doppler direction in blocks of size $K \times 1$. This scheme is depicted in Fig. 4(a) for $M = 8, N = 8$ and $K = 4$.

Scheme-2: The SCMA codewords for every j^{th} user, $j = 1, \ldots, J$ are allocated on the delay-Doppler plane $\Gamma_{N,M}$ along the delay direction in blocks of size $1 \times K$. This scheme is depicted in Fig. 4(b) for $M = 8, N = 8$ and $K = 4$.

The $J \times K$ SCMA system is referred to as the *basic SCMA system* as it is repeated over the delay-Doppler grid multiple times. There are MN slots for complex numbers in

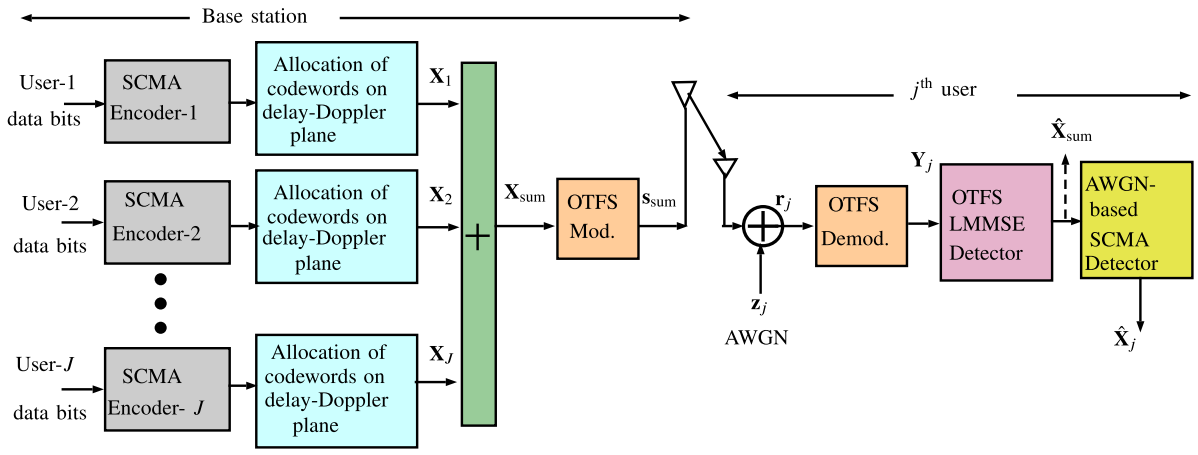


Fig. 5. Block diagram of OTFS-SCMA in downlink.

$\Gamma_{N,M}$. A particular slot or bin is considered as a resource in OTFS-SCMA. The total number of resources in $\Gamma_{N,M}$ is MN. As the length of any SCMA codeword is K, we can accommodate a total of $N_{\text{sym}} = \frac{MN}{K}$ symbols within $\Gamma_{N,M}$ for every user. For J users, JN_{sym} symbols are communicated over MN resources. Thus the overloading factor of the OTFS-SCMA scheme becomes $\lambda = \frac{\text{Total symbols transmitted over resources}}{\text{Total orthogonal resources}} = \frac{JN_{\text{sym}}}{MN} = \frac{J}{K}$. Note that the overloading factor of the OTFS-SCMA system is precisely equal to that of the underlying basic SCMA system.

B. In Downlink

Consider a downlink scenario with J receiving users and one transmitting BS as shown in Fig. 5. Note that only the j^{th} user's receiver is depicted in Fig. 5. In the following, we present the method for *Scheme-1*. The notations and the procedures can be easily extended to *Scheme-2*. Suppose $\mathbf{X}_j \in \mathbb{C}^{N\times M}$ is the delay-Doppler symbol frame for the j^{th} user. The superimposed input signal is given by

$$\mathbf{X}_{\text{sum}} = \sum_{j=1}^{J} \mathbf{X}_j \tag{16}$$

where the summation is done bin-wise. The superimposed signal $\mathbf{X}_{\text{sum}}$ goes through the OTFS modulator to produce the vector $\mathbf{s}_{\text{sum}} \in \mathbb{C}^{MN\times 1}$ as shown in Fig. 5. The BS transmits $\mathbf{s}_{\text{sum}}$ and the signal $\mathbf{r}_j$ received by the j^{th} user is passed through an OTFS demodulator. Suppose the output of the demodulator is $\mathbf{Y}_j \in \mathbb{C}^{N\times M}$. Invoking the input-output relationship between the BS and the j^{th} user, $\mathbf{Y}_j$ can be written as a simple function of the superposition $\mathbf{X}_{\text{sum}}$ of the input SCMA codewords. Suppose $\mathbf{x}_{\text{sum,vec}} = \text{vec}\,(\mathbf{X}_{\text{sum}})$ and $\mathbf{y}_{j,\text{vec}} = \text{vec}\,(\mathbf{Y}_j)$. The effective input-output relation now becomes

$$\mathbf{y}_{j,\text{vec}} = \mathbf{H}_j \mathbf{x}_{\text{sum,vec}} + \mathbf{z}_j \tag{17}$$

where $\mathbf{H}_j$ is the $MN \times MN$ complex coefficient matrix specifying the relationship between the input $\mathbf{x}_{\text{sum,vec}}$ and the output $\mathbf{y}_{j,\text{vec}}$, and $\mathbf{z}_j \in \mathbb{C}^{MN\times 1}$ is the complex AWGN. Observe from (17) that in order to recover $\mathbf{x}_{\text{sum,vec}}$ from $\mathbf{y}_{j,\text{vec}}$, a simple LMMSE-based detector can be used. The estimate of $\mathbf{x}_{\text{sum,vec}}$ is obtained as

$$\hat{\mathbf{x}}_{\text{sum,vec}} = \mathbf{H}_j^{\dagger}\left[\mathbf{H}_j\mathbf{H}_j^{\dagger} + \sigma_n^2\mathbf{I}_{MN}\right]^{-1}\mathbf{y}_{j,\text{vec}}. \tag{18}$$

Note that (18) involves matrix inversion, which increases the computational complexity. To lower the complexity, we may use the method proposed in [15] for OTFS detection. Ideally, the output $\hat{\mathbf{x}}_{\text{sum,vec}}$ of the LMMSE detector is free of delay-Doppler interactions and contains only the multi-user interference. As shown in Fig. 5, $\hat{\mathbf{x}}_{\text{sum,vec}}$ is fed to an MPA-based SCMA detector block to obtain the estimate $\hat{\mathbf{X}}_j$ of the j^{th} user's data $\mathbf{X}_j$. Note that the SCMA-detection block contains N_{sym} basic SCMA detectors. Moreover, as the OTFS detector has already canceled the effects of fading, we consider a simple AWGN channel for the SCMA detector.

C. In Uplink

Consider an uplink scenario with J transmitting users and one receiving BS as shown in Fig. 6. The description of the method pertains to *Scheme-1*. The SCMA codewords are allocated as depicted in Fig. 4(a). There are J different input-output relations for the inputs $\mathbf{X}_j$ of each user's modulator and the corresponding output $\mathbf{y}_{j,\text{vec}}$ of the OTFS demodulator given by $\mathbf{y}_{j,\text{vec}} = \mathbf{H}_j\mathbf{x}_{j,\text{vec}}$, for $j = 1,\ldots,J$. Linear combination of these J relations gives the vectorized version $\mathbf{y}_{\text{vec}}$ of $\mathbf{Y}$, which can be expressed as

$$\mathbf{y}_{\text{vec}} = \sum_{j=1}^{J} \mathbf{H}_j\mathbf{x}_{j,\text{vec}} + \mathbf{z} \tag{19}$$

where $\mathbf{x}_{j,\text{vec}} = \text{vec}\,(\mathbf{X}_j)$ and $\mathbf{z} \in \mathbb{C}^{MN\times 1}$ is the complex AWGN. Alternatively, (19) can be written as

$$\mathbf{y}_{\text{vec}} = \mathbf{H}_{\text{all}}\mathbf{x}_{\text{all}} + \mathbf{z} \tag{20}$$

where

$$\begin{aligned}\mathbf{H}_{\text{all}} &= [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_J]\,\text{and}\\ \mathbf{x}_{\text{all}} &= \left[\mathbf{x}_{1,\text{vec}}^T, \mathbf{x}_{2,\text{vec}}^T, \ldots, \mathbf{x}_{J,\text{vec}}^T\right]^T\end{aligned} \tag{21}$$

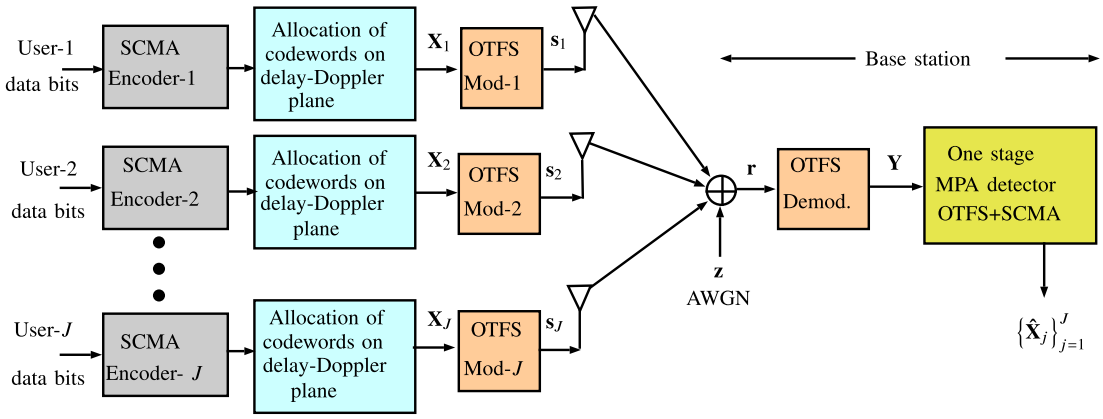


Fig. 6. Block diagram of OTFS-SCMA in uplink.

Unlike downlink, the receiver in uplink involves J channel coefficient matrices, $\{\mathbf{H}_j\}_{j=1}^{J}$. From (21), we can see that has all the J users' channel conditions contribute to $\mathbf{H}_{\text{all}}$ and $\mathbf{x}_{\text{all}}$ is not the superposition of SCMA data from all users, but a vertical concatenation of them. Therefore we cannot separate the delay-Doppler interference of the OTFS modulator and the multi-user interference of SCMA. Therefore, here we cannot adopt the previously-discussed two-stage method of separable detectors. Since $\mathbf{H}_{\text{all}}$ is highly sparse, a combined detector can be designed based on MPA. The resultant detector for uplink has more potential than the two-stage detector in downlink including a linear LMMSE detector.

Observe from (21) that $\mathbf{H}_{\text{all}}$ is an $MN \times JMN$ complex matrix and $\mathbf{x}_{\text{all}}$ is the information vector of length JMN. However, as $\mathbf{x}_{\text{all}}$ contains SCMA codewords, it will be a sparse vector. The number of non-zero elements in $\mathbf{x}_{\text{all}}$ is $\frac{JMNd_v}{K}$, where d_v is the number of non-zero components in a length-K codeword of a $J \times K$ basic SCMA system. As an example, consider an SCMA system with the factor graph shown in Fig. 3. For such a system, $\mathbf{x}_{\text{all}}$ contains $3MN$ non-zero complex numbers. Let $\mathbf{x}_{\text{all,compr}}$ denote the compressed input vector after removing the 0s in $\mathbf{x}_{\text{all}}$. Similarly, let $\mathbf{H}_{\text{all,compr}}$ denote the effective compressed $\mathbf{H}$ matrix after deleting the columns corresponding to the 0s in $\mathbf{x}_{\text{all}}$. Then, the input-output relationship in (20) can be written as

$$\mathbf{y}_{\text{vec}} = \mathbf{H}_{\text{all,compr}}\mathbf{x}_{\text{all,compr}} + \mathbf{z} \tag{22}$$

where $\mathbf{H}_{\text{all,compr}} \in \mathbb{C}^{MN \times \frac{JMNd_v}{K}}$ and $\mathbf{x}_{\text{all,compr}} \in \mathbb{C}^{\frac{JMNd_v}{K} \times 1}$. Note that the consecutive d_v elements of $\mathbf{x}_{\text{all,compr}}$ are the non-zero elements of a particular SCMA codeword. These d_v elements should be considered together as a single entity or variable node as they correspond to a particular information symbol in the input side. Observe from (22), that we have less observations than the number of variables as $J > K$. Therefore, a powerful detector must be employed at the receiver. We consider a single stage of MPA for the detection.

In (22), $\mathbf{x}_{\text{all,compr}} \in \mathbb{C}^{\frac{JMNd_v}{K} \times 1}$ is a collection of $\frac{JMN}{K}$ variable nodes. We call them variable nodes (not user nodes) as they refer to different symbols of the same user. Each variable node comprises d_v complex numbers but they collectively refer to one particular symbol. Similarly, in every row of

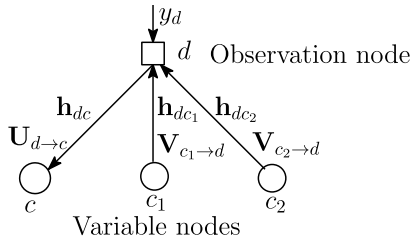


Fig. 7. The update of message from an observation node in the single-stage MPA detection of OTFS-SCMA in uplink.

$\mathbf{H}_{\text{all,compr}}$, the consecutive d_v complex coefficients must be clubbed together. With these new conventions, (22) can be written as

$$\begin{aligned}\mathbf{y}_{\text{vec}} &= \begin{bmatrix} y_1 \\ \vdots \\ y_{MN} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h}_{1,1} & \cdots & \mathbf{h}_{1,\frac{JMN}{K}} \\ \vdots & \cdots & \vdots \\ \mathbf{h}_{MN,1} & \cdots & \mathbf{h}_{MN,\frac{JMN}{K}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{\frac{JMN}{K}} \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_{MN} \end{bmatrix}\end{aligned} \tag{23}$$

where $\mathbf{h}_{dc} \in \mathbb{C}^{1 \times d_v}$ and $\mathbf{x}_c \in \mathbb{C}^{d_v \times 1}$. The corresponding effective factor graph contains MN observation nodes and $\frac{JMN}{K}$ variable nodes. An edge is assigned between the observation node y_d and the variable node $\mathbf{x}_c$ if $\mathbf{h}_{dc} \neq \mathbf{0}_{1\times d_v}$. In MPA, the message update from an observation node is very crucial and this step is further explained with the help of one example furnished below.

Example 1: Consider a $J = 6$ and $K = 4$ SCMA system as shown in Fig. 3 and a delay-Doppler plane with $M = 4$ and $N = 4$. Each user allocates 4 SCMA codewords over one delay-Doppler frame. The total numbers of the observation nodes and the variable nodes are 16 and 24, respectively. Suppose the observation node y_d is connected to the variable nodes $\{c, c_1, c_2\}$ through the coefficient vectors $\{\mathbf{h}_{dc}, \mathbf{h}_{dc_1}, \mathbf{h}_{dc_2}\}$. The scenario is shown in Fig. 7. Here the lengths of each coefficient vector and each symbol vector are equal to $d_v = 2$. Suppose the messages from the variable nodes 'c_1' and 'c_2' towards the observation node 'd' are denoted by

$\mathbf{V}_{c_1\to d}$ and $\mathbf{V}_{c_2\to d}$, respectively. Note that these messages are vectors of length equal to size of the underlying alphabet. The component of $\mathbf{V}_{c_i\to d}$ corresponding to the alphabet member $\mathbf{v}_{c_i}$ is denoted by $V_{c_i\to d}(\mathbf{v}_{c_i})$, $i=1,2$. As per the sum-product rule, the message $\mathbf{U}_{d\to c}$ from the observation node 'd' to the user node 'c' is given by

$$U_{d\to c}(m)=\sum_{(\mathbf{v}_{c_1},\mathbf{v}_{c_2})\in\mathbb{A}_{c_1}\times\mathbb{A}_{c_2}}\frac{1}{\pi N_0}\exp[-\frac{1}{N_0}|y_d-\mathbf{h}_{dc}\mathbf{x}_{cm}-\mathbf{h}_{dc_1}\mathbf{v}_{c_1}-\mathbf{h}_{dc_2}\mathbf{v}_{c_2}|^2]V_{c_1\to d}(\mathbf{v}_{c_1})\,V_{c_2\to d}(\mathbf{v}_{c_2})$$

where N_0 is the one-sided power-spectral density of AWGN. Suppose, the size of each alphabet is $A=4$. Each alphabet contains 4 codewords of length $d_v=2$. The Cartesian product $\mathbb{A}_{c_1}\times\mathbb{A}_{c_2}$ contains $4^2=16$ combinations of $(\mathbf{v}_{c_1},\mathbf{v}_{c_2})$ and the outer summation is done over these 16 combinations. Observe that the product $\mathbf{h}_{dc_i}\mathbf{v}_{c_i}$ is a scalar complex number and thus the argument in the exponential term is a scalar.

The computation of the messages from the variable nodes is relatively simple. The outgoing message from a variable node to an observation node is given by the component-wise product of the incoming messages from the other connected observation nodes. The *a posteriori* probability values of the variable nodes are computed by component-wise multiplication of all incoming messages from the neighboring observation nodes. The estimate of a variable node is considered as that symbol for which the *a posteriori* probability value takes the maximum value.

D. Diversity Analysis

The upper bound on the pairwise error probability (PEP) for maximum-likelihood (ML) detection involves a term of the form $\text{SNR}^{-\rho}$, where ρ is known as the diversity order. As the slope of the BER curve is determined by the diversity order, it has been frequently considered to analyze various communication scenarios' performance. It is a difficult task to derive the closed-form expression of the BER under MPA detection. Therefore, it is customary to consider diversity analysis as the analytical tool for performance evaluation of the OTFS system [19], [20]. Here we analyze the performance gain of the OTFS-SCMA system from the diversity perspective. The input-output relation in (9) can be represented as [19]

$$\mathbf{y}^T=\mathbf{h}'\mathbf{X}^d+\mathbf{z}^T \tag{24}$$

where $\mathbf{y}^T\in\mathbb{C}^{1\times MN}$, $\mathbf{h}'=[h_1,h_2,\ldots,h_P]$ denotes the channel coefficients, $\mathbf{X}^d\in\mathbb{C}^{P\times MN}$ is the restructured input matrix[2] and $\mathbf{z}^T\in\mathbb{C}^{1\times MN}$ is the noise. The c^{th} column of $\mathbf{X}^d$ can be obtained from the input vector $\mathbf{x}$ as

$$\mathbf{X}^d[c]=\begin{bmatrix}x_{[k-k_{\nu_1}]_N+N[l-l_{\tau_1}]_M}\\ x_{[k-k_{\nu_2}]_N+N[l-l_{\tau_2}]_M}\\ \vdots\\ x_{[k-k_{\nu_P}]_N+N[l-l_{\tau_P}]_M}\end{bmatrix} \tag{25}$$

where $c=k+Nl$; l_{τ_i} and k_{ν_i} are the delay tap and Doppler tap, respectively for the i^{th} path, $i=1,2,\ldots,P$. Let $\mathbf{X}^d_m$ and $\mathbf{X}^d_n$ be two distinct data matrices. At high SNR the upper bound on PEP, $P(\mathbf{X}^d_m\to\mathbf{X}^d_n)$ is proportional to $(\text{SNR})^{-r}$, where r is the rank of difference matrix, given by $\mathbf{\Delta}_{mn}=\mathbf{X}^d_m-\mathbf{X}^d_n$. The BER is dominated by the minimum possible value of r of all possible pairs. The asymptotic diversity order of an OTFS system is given by

$$\rho=\min_{m,n\ m\neq n}\text{rank}\,(\mathbf{\Delta}_{mn}). \tag{26}$$

The minimum rank occurs when the same data symbol a_m is assigned to the MN resources of transmitted data vector $\mathbf{x}_m$ and the same data symbol a_n $(a_n\neq a_m)$ is assigned to the MN resources of detected data vector $\mathbf{x}_n$. For an OTFS system with QAM alphabets, the asymptotic diversity order is 1 [19]. Next, we present a theorem for the asymptotic diversity order of the proposed OTFS-SCMA system.

Theorem 1: Consider an OTFS-SCMA system with an $N\times M$ delay-Doppler grid Γ_{NM} and a $J\times K$ SCMA model with N and M being integer multiples of K. Let the wireless channel for the j^{th} user be represented by P multipaths with the integer delay-Doppler tap pairs $\left(l^j_{\tau_i},k^j_{\nu_i}\right)$, $i=1,\cdots,P, j=1,\ldots,J$. Consider the sets $S^j_k=\left\{\left[k^j_{\nu_1}\right]_K,\left[k^j_{\nu_2}\right]_K,\ldots,\left[k^j_{\nu_P}\right]_K\right\}$ and $S^j_l=\left\{\left[l^j_{\tau_1}\right]_K,\left[l^j_{\tau_2}\right]_K,\ldots,\left[l^j_{\tau_P}\right]_K\right\}$. In the downlink, for the j^{th} user, the asymptotic diversity orders for Scheme-1 and Scheme-2 are given by $\left|S^j_k\right|$ and $\left|S^j_l\right|$, respectively. In the uplink, the asymptotic diversity orders for Scheme-1 and Scheme-2 are given by $\min\left\{\left|S^1_k\right|,\left|S^2_k\right|\ldots,\left|S^J_k\right|\right\}$ and $\min\left\{\left|S^1_l\right|,\left|S^2_l\right|\ldots,\left|S^J_l\right|\right\}$, respectively.

Proof: First, we analyze the diversity of *Scheme-1* and *Scheme-2* for single-user, and then we extend the proof to multi-user scenarios. In *Scheme-1*, the SCMA codewords are placed vertically in the Doppler direction. For the input-output relation in (9), the symbol matrix $\mathbf{X}$ is vectorized vertically in a column-wise fashion to obtain $\mathbf{x}$. Thus the consecutive K indices of $\mathbf{x}$ correspond to the consecutive indices of a particular SCMA codeword. By collecting all N columns $c=k+Nl$ $(k=0,\ldots,N-1)$ for a fixed $l\in\{0,1,\ldots,M-1\}$ in a sub-matrix $\mathbf{T}^l\in\mathbb{C}^{P\times N}$, we can write

$$\mathbf{X}^d=\left[\mathbf{T}^0,\mathbf{T}^1,\ldots,\mathbf{T}^{M-1}\right] \tag{27}$$

Observe from (25) that the rows of $\mathbf{T}^l$ are circular shifts of each other depending on Doppler taps' values. The difference matrix for two distinct frames is given by $\mathbf{\Delta}_{mn}=\mathbf{X}^d_m-\mathbf{X}^d_n=\left[\mathbf{\Delta}^0_{mn},\mathbf{\Delta}^1_{mn},\ldots,\mathbf{\Delta}^{M-1}_{mn}\right]$ where $\mathbf{\Delta}^l_{mn}=\mathbf{T}^l_m-\mathbf{T}^l_n$. Observe that the following relation holds

$$\text{rank}\,(\mathbf{\Delta}_{mn})\geq\max_{l:\mathbf{\Delta}^l_{mn}\neq\mathbf{0}_{P\times N}}\text{rank}\left(\mathbf{\Delta}^l_{mn}\right).$$

The minimum rank is obtained when all the non-zero sub-matrices $\mathbf{\Delta}^l_{mn}$s have the same minimum possible rank. As N is an integer multiple of K, $\mathbf{\Delta}^l_{mn}$ can be written as

$$\mathbf{\Delta}^l_{mn}=\begin{bmatrix}\boldsymbol{\delta}_1 & \ldots & \boldsymbol{\delta}_s & \ldots & \boldsymbol{\delta}_{N/K}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ \ldots & \ldots & \ldots & \ldots & \ldots\end{bmatrix} \tag{28}$$

where $\boldsymbol{\delta}_s\in\mathbb{C}^{1\times K}$ is a difference between two SCMA codewords. In (28), only the first row is shown. The remaining

[2]Note that $\mathbf{X}^d\in\mathbb{C}^{P\times MN}$ contains the elements of $\mathbf{X}\in\mathbb{C}^{N\times M}$ in a specific repetitive pattern.

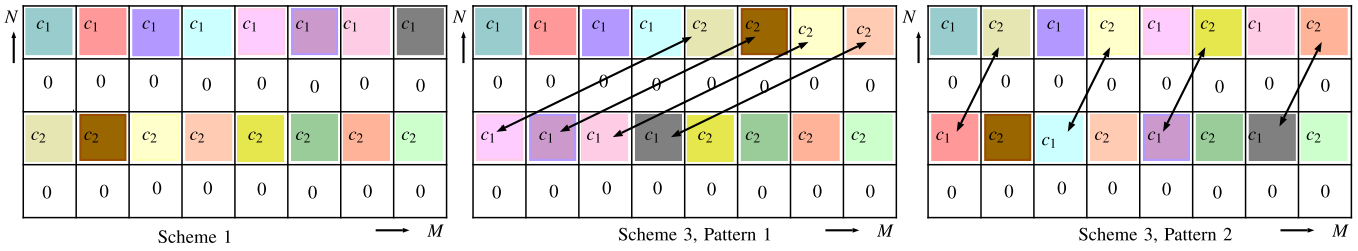


Fig. 8. Two interleaving patterns for *Scheme-3*.

rows are some cyclic shifts of the first row depending on the Doppler tap values. The minimum rank occurs when $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \cdots = \boldsymbol{\delta}_{N/K} = \boldsymbol{\delta}$. In that case, (28) becomes

$$\boldsymbol{\Delta}^l_{mn} = \begin{bmatrix} \boldsymbol{\delta} & \dots & \boldsymbol{\delta} & \dots & \boldsymbol{\delta} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \tag{29}$$

Observe that any circular shift of a row by an integer multiple of K will not result in any change. The amount of the distinguishable circular shift for the i^{th} row is given by $[k_{\nu_i}]_K$. Thus the rank of $\boldsymbol{\Delta}^l_{mn}$ is given by the number of distinct $[k_{\nu_i}]_K$s. This implies that the minimum rank of difference matrix $\boldsymbol{\Delta}_{mn}$ is given by $|S_k|$ where $S_k = \{[k_{\nu_1}]_K, [k_{\nu_2}]_K, \dots, [k_{\nu_P}]_K\}$. For *Scheme-2*, the SCMA codewords are placed horizontally along the direction of delay bins. Note that the row permutations don't change the rank. Any index i in the input $\mathbf{x}$ (or in *Scheme-1*) will refer to the index $i_1 = \lfloor i/N \rfloor + M\,[i]_N$ in *Scheme-2* and vice versa. By doing the permutations between the i^{th} and i_1^{th} rows, the rank condition can be determined as discussed in *Scheme-1*. In that case, the complex codeword element allotted in the index $[k - k_{\nu_i}]_N + N[l - l_{\tau_i}]_M$ for *Scheme-1* will be placed in the index $[l - l_{\tau_i}]_M + M[k - k_{\nu_i}]_N$ for *Scheme-2*. Thus reversing the roles played by the Doppler and the delay terms, N and M, and proceeding similarly, it can be shown that $\text{rank}\,(\boldsymbol{\Delta}_{mn}) = |S_l|$ where $S_l = \{[l_{\tau_1}]_K, \dots, [l_{\tau_P}]_K\}$.

We can extend the above analysis to the multi-user case. First, we consider the output in the downlink scenario:

$$\mathbf{y}_j^T = \mathbf{h}'_j \mathbf{X}^d_{\text{sum}} + \mathbf{z}_j^T \tag{30}$$

where $\mathbf{X}^d_{\text{sum}}$ is the restructured matrix of superimposed data vector $\mathbf{x}_{\text{sum}} = \mathbf{x}_1 + \mathbf{x}_2 + \dots \mathbf{x}_J$. Note that for the SCMA codebook, the sum values of different combinations of J users' codewords are distinct [21]. Thus the superposition of SCMA codewords will not affect $\text{rank}(\boldsymbol{\Delta}_{mn})$. For a specific j^{th} user, we have $\text{rank}(\boldsymbol{\Delta}_{mn}) = \left|S^j_k\right|$ for *Scheme-1* and $\text{rank}(\boldsymbol{\Delta}_{mn}) = \left|S^j_l\right|$ for *Scheme-2* where $S^j_k = \left\{\left[k^j_{\nu_1}\right]_K, \left[k^j_{\nu_2}\right]_K, \dots, \left[k^j_{\nu_P}\right]_K\right\}$ and $S^j_l = \left\{\left[l^j_{\tau_1}\right]_K, \left[l^j_{\tau_2}\right]_K, \dots, \left[l^j_{\tau_P}\right]_K\right\}$.

For the uplink, the input-output relation is given by

$$\mathbf{y}^T = \mathbf{h}'_1 \mathbf{X}^d_1 + \mathbf{h}'_2 \mathbf{X}^d_2 + \dots \mathbf{h}'_J \mathbf{X}^d_J + \mathbf{z}^T. \tag{31}$$

Expressing the above equation in matrix form, we get

$$\begin{aligned} \mathbf{y}^T &= \underbrace{\left[\mathbf{h}'_1\ \mathbf{h}'_2\ \dots\ \mathbf{h}'_J\right]}_{\mathbf{h}'_{\text{eff}}} \underbrace{\begin{bmatrix} \mathbf{X}^d_1 \\ \mathbf{X}^d_2 \\ \vdots \\ \mathbf{X}^d_J \end{bmatrix}}_{\mathbf{X}^d_{\text{eff}}} + \mathbf{z}^T \\ &= \mathbf{h}'_{\text{eff}} \mathbf{X}^d_{\text{eff}} + \mathbf{z}^T \end{aligned} \tag{32}$$

where $\mathbf{h}'_{\text{eff}} \in \mathbb{C}^{1\times JP}$ and $\mathbf{X}^d_{\text{eff}} \in \mathbb{C}^{JP\times MN}$. Proceeding in the same manner as in the single-user case, we can write $\mathbf{X}^d_{\text{eff}} = [\mathbf{X}_{0,\text{eff}}, \mathbf{X}_{0,\text{eff}}, \dots, \mathbf{X}_{0,\text{eff}}]$ where $\mathbf{X}_{0,\text{eff}} \in \mathbb{C}^{PJ\times K}$ for *Scheme-1*. Here the diversity is determined by $\text{rank}\,(\boldsymbol{\Delta}_{mn}) = \text{rank}\left(\mathbf{X}^d_{m_{\text{eff}}} - \mathbf{X}^d_{n_{\text{eff}}}\right) = \text{rank}\left(\mathbf{X}^d_{0,m_{\text{eff}}} - \mathbf{X}^d_{0,n_{\text{eff}}}\right)$. The minimum rank occurs when there is an error in data detection of one user's frame. Let j^{th} user's data detection is an error. Therefore $\text{rank}(\boldsymbol{\Delta}_{mn})$ is equal to $\left|S^j_k\right|$. Because detection error can take place for any user, the minimum rank possible is $\rho = \min\left\{\left|S^1_k\right|, \dots, \left|S^J_k\right|\right\}$ for *Scheme-1*. Similarly, the minimum rank for *Scheme-2* is given by $\rho = \min\left\{\left|S^1_l\right|, \dots, \left|S^J_l\right|\right\}$. ■

Observe from Theorem 1 that OTFS-SCMA can easily achieve higher asymptotic diversity order than that of OTFS-OMA, for which the order is 1. Inspired by the diversity analysis presented above, we propose an interleaved SCMA codeword allocation scheme called *Scheme-3*. In this scheme, the non-zero elements of different codewords are swapped. Scheme 3 is diagrammatically represented for an OTFS frame $\Gamma_{4,8}$ in Fig. 8. Each SCMA codeword has length $K = 4$ and $d_v = 2$ non-zero elements. The codeword $[c_1\ 0\ c_2\ 0]^T$ is repeated throughout the OTFS frame as it determines the asymptotic diversity order. The interleaving pattern can be adopted in multiple ways. Fig. 8 shows only two such patterns obtained from Scheme 1. Observe from Fig. 8 that in Pattern 1, one non-zero element is exchanged diagonally between two halves of the OTFS frame. On the other hand, in Pattern 2, one non-zero element of a codeword is swapped diagonally with that of the adjacent codeword. For analysis, we consider three cases with $P = 2$. In all the cases, we take $(k_{\nu_1}, l_{\tau_1}) = (0, 0)$. The three cases are specified as *Case 1*: $(k_{\nu_2} \neq 0, l_{\tau_2} \neq 0)$, *Case 2*: $(k_{\nu_2} \neq 0, l_{\tau_2} = 0)$ and *Case 3*: $(k_{\nu_2} = 0, l_{\tau_2} \neq 0)$. From [20], the asymptotic diversity order is determined by the minimum possible rank of the matrix $\boldsymbol{\Phi}^\dagger(\boldsymbol{\delta})\boldsymbol{\Phi}(\boldsymbol{\delta})$ (or equivalently $\boldsymbol{\Phi}(\boldsymbol{\delta})$), where

$$\begin{aligned} &\boldsymbol{\Phi}(\boldsymbol{\delta}) \\ &= \left[\boldsymbol{\Xi}_1\boldsymbol{\delta}\ \boldsymbol{\Xi}_2\boldsymbol{\delta} \dots \boldsymbol{\Xi}_P\boldsymbol{\delta}\right], \\ &\boldsymbol{\Xi}_i \end{aligned}$$

$= (\mathbf{I}_M \otimes \mathbf{F}_N)\boldsymbol{\Theta}^{k_{\nu_i}}\boldsymbol{\Pi}^{l_{\tau_i}}(\mathbf{I}_M \otimes \mathbf{F}_N^{\dagger})$
$[\mathbf{F}_N$ is the N-point DFT matrix$]$

$\boldsymbol{\delta}$
$= [\delta_{0,0}\ \delta_{1,0}\ldots\delta_{N-1,0}\ \delta_{0,1}\ldots\delta_{0,M-1}\ \delta_{1,M-1}\delta_{N-1,M-1}]^T$
(error vector), $\delta_{n,m} = \delta_{n+Nm}$
$\boldsymbol{\Theta}^{k_{\nu_i}}$
$= \text{diag}\ [1\ z\ z^2\ldots z^{NM-1}]^{k_{\nu_i}}$ where $z = e^{j2\pi/NM}$ and
$\boldsymbol{\Pi}^{l_{\tau_i}}$
$= \mathbf{I}_{NM}$ forward cyclic shifted by l_{τ_i}.

Considering the ideal pulse shaping, we can see that all columns of $\boldsymbol{\Phi}(\boldsymbol{\delta})$ will be shifted versions of $\boldsymbol{\delta}$. The SCMA codebooks are designed so that any two distinct codewords of a user differ in $d_v = 2$ nonzero elements. Thus one symbol error results in two different non-zero components in $\boldsymbol{\delta}$. Let $\delta^{(1)}_{n_1,m_1}$ and $\delta^{(2)}_{n_2,m_2}$ be the two non-zero components at some arbitrary locations (n_1, m_1) and (n_2, m_2) respectively, and $\delta^{(1)}_{n_1,m_1} \neq \delta^{(2)}_{n_2,m_2}$. We have $\boldsymbol{\Xi}_1 = \mathbf{I}$, which gives $\boldsymbol{\Phi}(\boldsymbol{\delta}) = [\boldsymbol{\delta}\ \ \boldsymbol{\Xi}_2\boldsymbol{\delta}]$. The complete four rows of $\boldsymbol{\Phi}(\boldsymbol{\delta})$ which contain the non-zero error elements can be represented as

$$\begin{bmatrix} \delta^{(1)}_{n_1,m_1} & \delta_{[n_1-k_{\nu_2}],[m_1-l_{\tau_2}]} \\ \delta_{[n_1+k_{\nu_2}],[m_1+l_{\tau_2}]} & \delta^{(1)}_{n_1,m_1} \\ \delta^{(2)}_{n_2,m_2} & \delta_{[n_2-k_{\nu_2}],[m_2-l_{\tau_2}]} \\ \delta_{[n_2+k_{\nu_2}],[m_2+l_{\tau_2}]} & \delta^{(2)}_{n_2,m_2} \end{bmatrix}. \tag{33}$$

$\boldsymbol{\Phi}(\boldsymbol{\delta})$ will have a minimum rank of 2 if none of the following equations is satisfied:

$$\begin{aligned} {\delta^{(1)}_{n_1,m_1}}^2 &= \delta_{[n_1-k_{\nu_2}],[m_1-l_{\tau_2}]}\delta_{[n_1+k_{\nu_2}],[m_1+l_{\tau_2}]} \\ {\delta^{(2)}_{n_2,m_2}}^2 &= \delta_{[n_2-k_{\nu_2}],[m_2-l_{\tau_2}]}\delta_{[n_2+k_{\nu_2}],[m_2+l_{\tau_2}]} \end{aligned} \tag{34}$$

As $[N]_K = 0$, for Pattern 1 and Pattern 2 shown in Fig. 8, we have the following relations:

$$(\delta_{n,m})_{\text{Pattern 1}} = \begin{cases} 0, & \text{if } [n]_2 \neq 0 \\ \delta^{(1)}, & \text{if } [n]_2 = 0\ \&\ 0 \leq m < M/2 \\ \delta^{(2)}, & \text{if } [n]_2 = 0\ \&\ M/2 \leq m < M \end{cases}$$

$$(\delta_{n,m})_{\text{Pattern 2}} = \begin{cases} 0, & \text{if } [n]_2 \neq 0 \\ \delta^{(1)}, & \text{if } [n]_2 = 0\ \&\ [m]_2 = 0 \\ \delta^{(2)}, & \text{if } [n]_2 = 0\ \&\ [m]_2 \neq 0 \end{cases}$$

Pattern 1: Substituting the above conditions in (34), we get $[k_{\nu_2}]_2 = 0$ and $M/2 < l_{\tau_2} < M-1$ and $M/2 < -l_{\tau_2} < M-1$. The last two conditions of the above equation for Pattern 1 are never satisfied, which always gives an asymptotic diversity order of 2 for Case 1. For Case 2, we can see that only k_{ν_2} decides the dependency and which results in the condition of rank 1 as $[k_{\nu_2}]_2 = 0$, which implies that for Case 2, asymptotic diversity order of 2 is not guaranteed. For Case 3, we can see that (34) depends only on l_{τ_2} and it is never satisfied, which gives an asymptotic diversity order of 2 always in Case 3.

Pattern 2: For Case 1, the condition for obtaining a rank of 2 is $[k_{\nu_2}]_2 \neq 0$ or $[l_{\tau_2}]_2 \neq 0$. For Case 2, we can see that dependency of rows banks only on k_{ν_2} and which results in the condition of minimum rank 2 as $[k_{\nu_2}]_2 \neq 0$, which implies that for Case 2, in Pattern 2 also the asymptotic diversity order of 1 can happen. For Case 3, we can see that only l_{τ_2} decides the dependency of rows and an asymptotic diversity order of 2 is achieved only if $[l_{\tau_2}]_2 \neq 0$. The asymptotic diversity orders of the three schemes are summarized in TABLE I. For *Scheme-1* and *Scheme-2*, the results are based on Theorem 1. Removing the condition of $(k_{\nu_1}, l_{\tau_1}) = (0,0)$, we can generalize the diversity order for any two path selections $\{(k_{\nu_1}, l_{\tau_1}), (k_{\nu_2}, l_{\tau_2})\}$ of Pattern 1 and Pattern 2 as given below:

Pattern 1: Diversity order of 2 if $[|k_{\nu_1} - k_{\nu_2}|]_2 \neq 0$ or $l_{\tau_1} \neq l_{\tau_2}$
Pattern 2: Diversity order of 2 if $[|k_{\nu_1} - k_{\nu_2}|]_2 \neq 0$ or $[|l_{\tau_1} - l_{\tau_2}|]_2 \neq 0$

The above analysis for Scheme 3 is done for two fixed interleaving patterns. If the delay and the Doppler taps are known at the transmitter, then the interleaving pattern can be adapted to achieve full diversity. Such an optimal codeword allocation scheme is presented in Algorithm 1. Although there is a scope of achieving full asymptotic diversity by devising an optimal allocation scheme, there are a few challenges: (1) The determination of the optimal scheme requires the knowledge

TABLE I
SUMMARY OF ASYMPTOTIC DIVERSITY ORDERS OF DIFFERENT SCHEMES FOR $P = 2$, GIVEN $(k_{\nu_1}, l_{\tau_1}) = (0,0)$

(k_{ν_2}, l_{τ_2})	Scheme-1	Scheme-2	Scheme-3	
			Pattern 1	Pattern 2
$k_{\nu_2} \neq 0, l_{\tau_2} \neq 0$	2	2	2	1, if $([k_{\nu_2}]_2 = 0)$ AND $([l_{\tau_2}]_2 = 0)$; 2, otherwise.
$k_{\nu_2} \neq 0, l_{\tau_2} = 0$	2	1	1, if $([k_{\nu_2}]_2 = 0)$; 2, otherwise.	1, if $([k_{\nu_2}]_2 = 0)$; 2, otherwise.
$k_{\nu_2} = 0, l_{\tau_2} \neq 0$	1	2	2	1, if $([l_{\tau_2}]_2 = 0)$; 2, otherwise.

Algorithm 1 Optimal Codeword Allocation Scheme

input : OTFS-SCMA Parameters: N, M, K; Channel parameters:
$\mathcal{P} = \{(k_{\nu_1}, l_{\tau_1}), (k_{\nu_2}, l_{\tau_2}), \ldots, (k_{\nu_P}, l_{\tau_P})\}$
output: Optimal scheme

Initialization: $|S_k|$=Number of distinct mod-K Doppler taps; $|S_l|$ = Number of distinct mod-K delay taps;
if $P \leq K$ **then**
  if $|S_k| = P$ **then**
    Scheme-1
  else
    if $|S_l| = P$ **then**
      Scheme-2
    else
      Scheme-3: Using $\mathcal{P}$, design an interleaving pattern such that asymptotic diversity order is P
else
  Scheme-3: Using $\mathcal{P}$, design an interleaving pattern such that asymptotic diversity order is P

of the delay and the Doppler tap values for the multi-paths at the transmitter. It is difficult to estimate these parameters at the transmitter side due to the requirement of the feedback link from the receiver, (2) Due to the interleaving of the SCMA codewords, the multiple users will interfere with each other in a different way than what was envisioned during the codebook design process. The underlying factor graph will also change. For every specific set of delay/Doppler values, an optimal codeword allocation pattern must be identified. By keeping that pattern in mind, the SCMA codebooks must be designed to minimize multi-user interference. It is challenging to realize these operations in a real-time environment and (3) Moreover, due to the changes in the factor graph, the steps of the MPA-based detection will also change. It is not feasible to alter the steps of the variable nodes' updates and check nodes' updates for every possible set of the delay/Doppler values. Due to these challenges, we consider only Scheme 1 and Scheme 2 for simulations in this paper.

E. Complexity Analysis

We present a theorem regarding the degrees of the observation and the variable nodes in the effective factor graph in the uplink. This theorem is crucial for the complexity analysis of the detection process. In the proposed OTFS-SCMA, in addition to the multi-user interference present in the basic SCMA scheme, OTFS modulation will introduce delay-Doppler interference caused by the multi-paths. Hence, there is a difference in the degrees of nodes of the effective factor graph of the OTFS-SCMA system.

Theorem 2: Consider an OTFS-SCMA model with a $J \times K$ basic SCMA system in the uplink. Let the degrees of a resource node and a variable node in the SCMA system be d_v and d_f respectively. Let P denote the number of multipath for the underlying wireless channel for each user. Then for the effective factor graph of the OTFS-SCMA system, the degrees $\{\deg_d\}_{d=1}^{MN}$ of the observation nodes and the degrees $\{\deg_c\}_{c=1}^{\frac{JMN}{K}}$ of the variable nodes satisfy the following:

$$\frac{1}{MN}\sum_{d=1}^{MN}\deg_d \leq Sd_f \tag{35}$$

$$\frac{1}{\left(\frac{JMN}{K}\right)}\sum_{c=1}^{\frac{JMN}{K}}\deg_c \leq Sd_v \tag{36}$$

where $S=\sum_{i=1}^{P}(2N_i+1)$.

The complexity of OTFS-SCMA is analyzed for downlink and uplink in the following. As the sophisticated detector is the main contributor to the overall complexity, we focus only on the detector's complexity.

1) Downlink: The detector in the downlink is a concatenated system of LMMSE and MPA detector. The interference over the delay-Doppler plane is removed by the LMMSE-based OTFS detector. The LMMSE detector involves traditional matrix inversion. Its complexity is $O\left(M^3N^3\right)$. If the low-complexity version of the LMMSE detector in [15] is used, then the complexity would be $O\left(\frac{MN}{2}\log_2 N + 2MNP^2\right)$. The output of the OTFS detector now contains the multi-user interference. The MPA-based SCMA detector removes multi-user interference. The complexity of the SCMA detector is $O\left(\frac{MN}{K}A^{d_f}\right)$ where A is the alphabet size and d_f is the degree of the factor node in the basic SCMA system. For relatively small values of M and N, the complexity of the two-stage detector is approximately $O\left(A^{d_f}\right)$.

2) Uplink: In uplink, the detector consists of only a single MPA. As per Theorem 1, the average degree of an observation node in the effective factor graph is upper bounded by Sd_f. Therefore, the complexity of the detector in uplink is $O\left(MNA^{Sd_f}\right)$ as there are MN observation nodes. In the absence of fractional Doppler shifts, the complexity is $O\left(MNA^{Pd_f}\right)$. Note that the complexity of this detector is high if the number P of multipaths is high. Suppose P is known to be high in a particular situation. In that case, other variants based on expectation propagation [18], Gaussian-approximated MPA [12] etc., may be used to reduce the complexity.

F. Channel Estimation

The embedded pilot-based channel estimation (CE) method [22] is adapted to OTFS-SCMA in the downlink. Although the treatment presented here adheres to *Scheme-1*, it can be easily extended to other schemes. Fig. 9 depicts the entire process of CE and data detection. Using the QAM pilot symbol, CE is performed for the integer Doppler case. Suppose k_ν and l_τ are the maximum Doppler and delay taps respectively. Let $\mathbf{x}_{d_{(i)}}$ denote the superimposed data codewords in the transmitter side and $\mathbf{y}_{d_{(i)}}$ denote the same in the receiver side, $i=1,2,\ldots$. The location $[k_p,l_p]$ of the pilot can be arbitrarily chosen such that $0 \leq l_p - l_\tau \leq l_p \leq l_p + l_\tau \leq M-1$ and $k_p = N-1-2k_\nu$. The pilot, the guard band and the superposed codewords are allocated over the deal-Doppler grid as follows:

$$x[k,l]=\begin{cases} x_p k = k_p\ , & l=l_p \\ 0 & \left\lfloor \frac{k_p-2k_\nu}{K}\right\rfloor K \leq k \leq N-1\ , \\ & 0 \leq l \leq l_p+l_\tau \\ x_{d_{(i)}}[j] & j^{\text{th}} \text{ component of } i^{\text{th}} \text{ superposed} \\ & \text{SCMA codeword} \end{cases}$$

$$\text{where,}\quad i=\begin{cases} l\left\lfloor\frac{N-4k_\nu-1}{K}\right\rfloor+\left\lfloor\frac{k}{K}\right\rfloor+1 & l \leq l_p+l_\tau \\ (l_p+l_\tau+1)\left\lfloor\frac{N-4k_\nu-1}{K}\right\rfloor & \\ \quad +l\frac{N}{K}+\left\lfloor\frac{k}{K}\right\rfloor+1 & l > l_p+l_\tau \end{cases}$$

$$\text{and}\quad j=[k]_K+1.$$

At the receiver, the received symbols $y[k,l], k_p-k_\nu \leq k \leq k_p+k_\nu\ ,\ l_p \leq l \leq l_p+l_\tau$ are used to find the estimate $\hat{\mathbf{H}}$ of the channel, by a thresholding strategy. These symbols are denoted by $y_p[k,l]$s in Fig. 9 and the channel coefficient is estimated if $|y_p[k,l]| \geq \mathcal{T}$, where $\mathcal{T}$ depends on the noise level. The remaining received symbols denoted by $y_{\text{OTFS}}[k,l]$s and $\hat{\mathbf{H}}$ are used for LMMSE-based

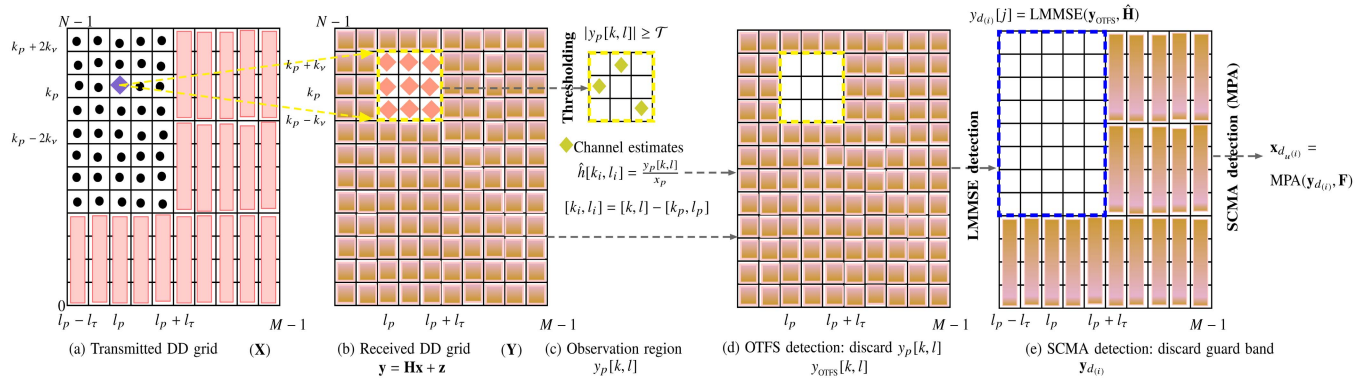


Fig. 9. CE for OTFS-SCMA based on the data-embedded pilot symbol with the parameters $M = 10, N = 12, l_\tau = 2, k_\nu = 1$.

TABLE II
GUARD BAND OVERHEAD

UE speed (Kmph)	30 $k_\nu = 1$	120 $k_\nu = 4$	500 $k_\nu = 16$
Pilot+Guard symbols[1] (N_g)	328	820	2788
Data symbols[2] (N_d)	97812	97074	94122
Guard band overhead $\left(\frac{N_g}{NM}\right)$	.5%	1.25%	4.25%

[1] $N_g = (2l_\tau + 1)(4k_\nu + 4)$

[2] $N_d = J\left[\frac{(M-(2l_\tau+1))N}{K} + \lfloor\frac{N-4k_\nu-1}{K}\rfloor(2l_\tau+1)\right]$

OTFS detection. The relevant outputs $\mathbf{y}_{d_{(i)}}$s of the OTFS detector are identified by taking into account the guard band. Then, $\mathbf{y}_{d_{(i)}}$s and the underlying SCMA factor matrix $\mathbf{F}$ are fed to the MPA-based multi-user detector. The multi-user interference of SCMA is finally canceled to extract the individual user's data. The comparison of the guard band overhead (N_g) and the total number of data symbols (N_d) for different UE speeds are shown in TABLE II with parameters $M = 512$, $N = 128$, $J = 6$, $K = 4$, $f_c = 4$ GHz, $\Delta f = 15$ KHz, $l_\tau = 20$ and EVA propagation model. The overloading factor of 150% is maintained, even in the presence of the guard band. In the uplink scenario, separate pilot symbols have to be used for the J users with a sufficient guard band allotted to each user. It would be an interesting future work to devise an efficient CE technique for OTFS-SCMA in uplink with minimal overhead and detection complexity.

IV. Simulation Results

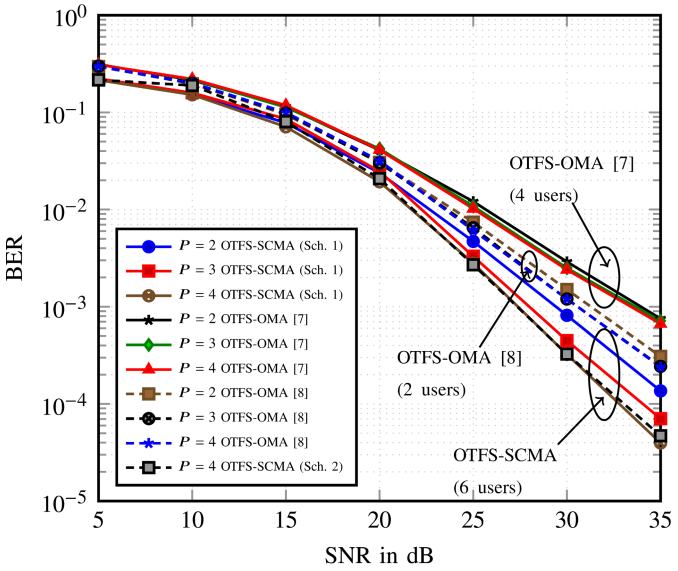
In this section, we present the BER simulation results for OTFS-SCMA systems. The results for both downlink and uplink are presented. In both cases, we consider OTFS with the rectangular pulse. The OTFS system is assumed to be free of fractional Doppler and fractional delay values. It is assumed that the receiver has the perfect knowledge of the channel.

A. Downlink

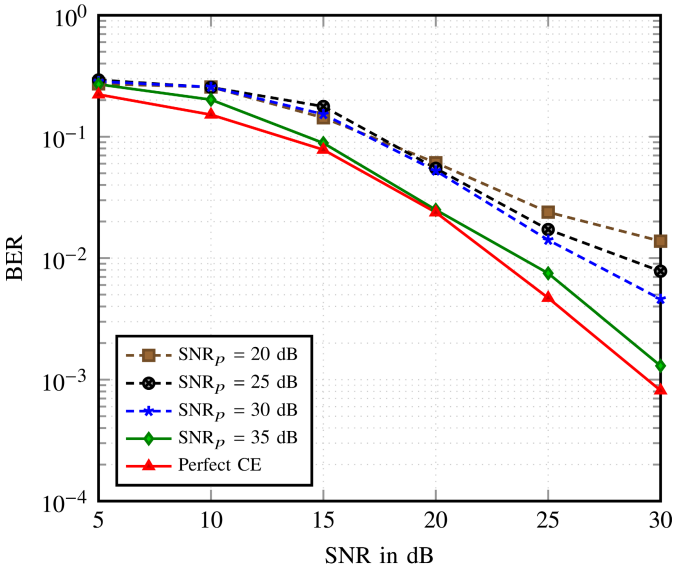
1) Case 1 : $M = 8, N = 8$, 6×4 SCMA System: First we consider a delay-Doppler plane $\Gamma_{8,8}$. An SCMA system with $J = 6$ and $K = 4$ is considered. The SCMA codebooks proposed in [21] are used. The overloading factor of the OTFS-SCMA system is $\lambda = J/K = 150\%$. The BS transmits $MN/K = 16$ symbols for each user. First, as per *Scheme-1*, 16 SCMA codewords of every user are allocated on the respective delay-Doppler plane.

The wireless channel is represented by P propagation paths. The i^{th} path is associated with integer delay (l_{τ_i}) and Doppler (k_{ν_i}) components as explained in Section II-A.4. We consider $\triangle f = 1$ and $T = 1$. The values of l_{τ_i}s and k_{ν_i}s are generated randomly from $\{0, 1, \ldots, P-1\}$. We have considered distinct propagation paths with distinct pairs of (l_{τ_i}, k_{ν_i}) in our simulations. The BER performance of this OTFS-SCMA scheme is presented in Fig. 10(a). The results for $P = 2, 3, 4$ are shown. Observe that as P increases, the BER curve improves due to the increased diversity in the system [19], [20]. Note from Fig. 10(a) that the BER performance of *Scheme-2* is similar to that of *Scheme-1*. Fig. 10(a) also presents the results for OTFS-OMA over $\Gamma_{8,8}$. The OTFS-OMA scheme proposed in [8] is considered first. We consider two users occupying the delay-Doppler plane. The number of symbols each user can transmit is 32. Thus for two users, the BS transmits 64 symbols over 64 slots/resources in the plane. The overloading factor is 100%. We have also considered the OTFS-OMA method proposed in [7] for 4 users. Here also, the overloading factor is 100%. The BER results for these OMA schemes with $P = 2, 3, 4$ are shown in Fig. 10(a). We consider the QAM alphabet as its size $(A = 4)$ is the same as that of the OTFS-SCMA system under consideration. Observe that, even though the OTFS-SCMA system involves 6 users with overloading factor 150%, it performs better than the OTFS-OMA schemes with 2 users and 4 users with an overloading factor of 100%. The performance gain of OTFS-SCMA over OTFS-OMA is attributed to higher diversity as explained in Section III-D. The BER performance with CE using data-embedded pilot method is analyzed in Fig. 10(b). As expected, when SNR_p increases, the BER approaches that of the perfect CE.

2) Case 2 : $M = 12, N = 12$, 12×6 SCMA System: Next, we consider an SCMA system with $J = 12$ and $K = 6$ over $\Gamma_{12,12}$. The overloading factor of this OTFS-SCMA scheme is 200%. The method proposed in [23] is adopted to design the codebooks for the 12×6 SCMA system using the factor matrix



(a) With perfect channel estimation



(b) Embedded pilot CE , P = 2.

Fig. 10. OTFS-SCMA with $N = 8$, $M = 8$ and ($J = 6$, $K = 4$) SCMA system, overloading factor 150%, (*Scheme-1* and *Scheme-2*) in downlink.

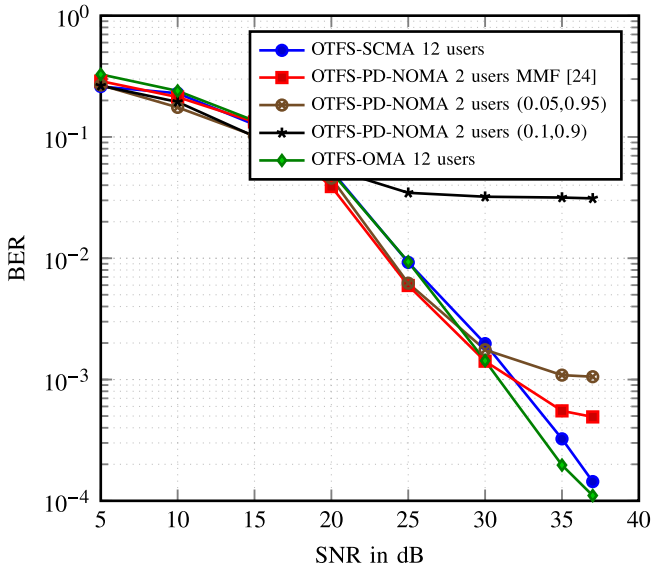


Fig. 11. $N = 12$, $M = 12$ and ($J = 12$, $K = 6$) SCMA system, overloading factor 200% in downlink with $P = 5$.

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{37}$$

$\mathbf{F}$ shown (37). Observe that the factor graph does not contain any 4-cycle and its girth is 6. These codebooks are used in the OTFS-SCMA transmission. Fig. 11 presents the BER performance of the OTFS-SCMA system with *Scheme-1* in the downlink. Although the overloading factor is 200%, the BER performance is impressive. This performance is compared with OTFS-OMA [8] for 12 users and OTFS-PD-NOMA for 2 users, all with $P = 5$. OTFS-PD-NOMA considers 2 users occupying the entire delay-Doppler grid. The QAM alphabet is considered for both the users. At the location of the near user, the principle of SIC is considered for extracting the data of the strong user from the superimposed signal as explained in Section II-C. MPA-based detection [12] is used for both the near and the far users.

The variances of the channel gains for the near user and the far users are related as $\sigma^2_{h_2} = 0.5\sigma^2_{h_1}$ with $\sigma^2_{h_1} = 0$ dB. The optimum power split is obtained by maximizing the worst (i.e. minimum) user rate at each SNR point. This power allocation method in PD-NOMA is known as the maximin fairness (MMF) [24], [25] criterion. As per the MMF criterion, the optimum power splits (P_1, P_2) amongst the near and the far users are obtained as $(0.26, 0.74), (0.2, 0.8), (0.14, 0.86), (0.09, 0.91), (0.05, 0.95), (0.03, 0.97), (0.02, 0.98), (0.01, 0.99)$ at the SNR points considered in Fig. 11. The results for two fixed power-splits of $(0.05, 0.95)$ and $(0.1, 0.9)$ are also shown. Observe that the BER performance of OTFS-SCMA is better than that of OTFS-PD-NOMA. The BER curves for OTFS-PD-NOMA flattens out, resulting in the error floor. Note that the performance of OTFS-SCMA with an overloading factor of 200% is slightly degraded, unlike other cases of 150% compared to OTFS-OMA with 12 users.

3) Case 3 : $M = 512$, $N = 16$, 6×4 SCMA System: Next we consider a more extensive and practical OTFS frame with $M = 512$ and $N = 16$. The simulation parameters are presented in TABLE III. The path values, i.e. l_{τ_i}s and k_{ν_i}s are computed as per the EVA propagation model [26]. The power delay profile is given by [0, −1.5, −1.4, −3.6, −0.6, −9.1, −7.0, −12.0, −16.9] dB against the excess tap delays [0, 30, 150, 310, 370, 710, 1090, 1730, 2510] ns [26]. Only a single path is considered for a particular delay value. The maximum Doppler shift is given by $\nu_{\max} = (f_c v_u)/(3 \times 10^8)$ where f_c is the carrier frequency and v_u is the speed of the user in

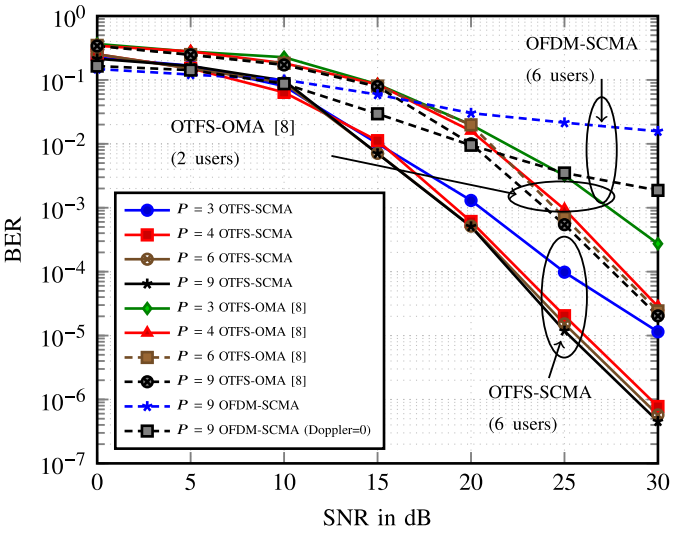


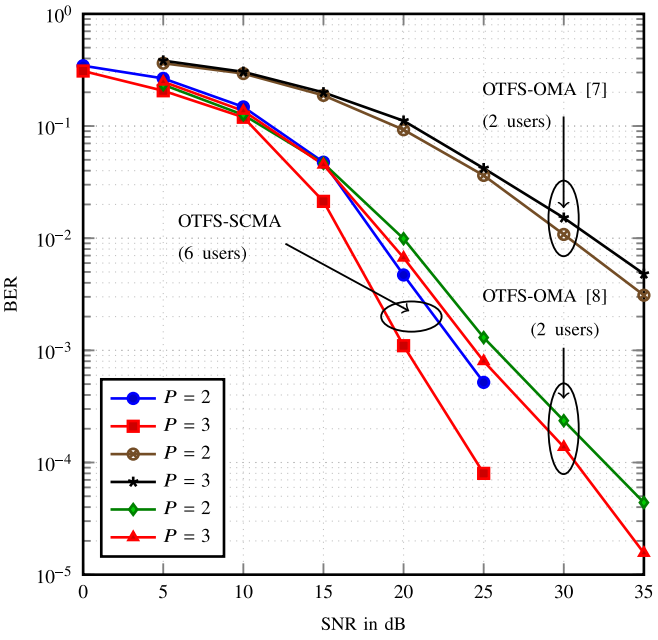
Fig. 12. $N = 16$, $M = 512$ and ($J = 6$, $K = 4$) SCMA system for EVA propagation model [26] in downlink.

TABLE III
SIMULATION PARAMETERS

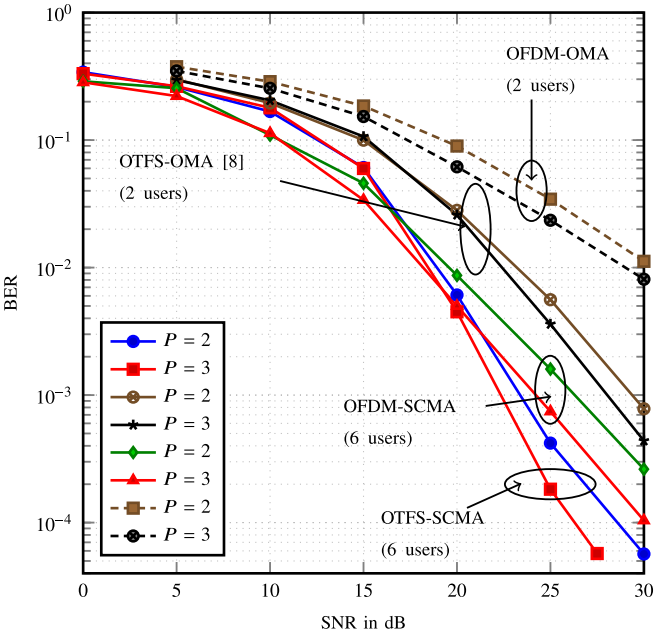
Parameter	Value
Carrier frequency (f_c)	4 GHz
Number of delay bins or sub-carriers (M)	512
Number of Doppler bins (N)	16
Sub-carrier spacing ($\triangle f$)	15 KHz
Mapping or modulation alphabet	6×4 SCMA [21]
User speed (v_u)	500 Kmph
Propagation model	EVA [26]

meter per second. The Doppler shift for the i^{th} path is given by $\nu_i = \nu_{\max} \cos(\theta_i)$ where θ_i has a uniform distribution $\theta_i \sim U[-\pi, \pi]$. From the Doppler shift (ν_i) and the delay (τ_i) as specified in the power delay profile, the integer Doppler (k_{ν_i}) and the integer delay (l_{τ_i}) values can be computed using (8) with $\kappa_{\nu_i} = 0$. We consider the same basic 6×4 SCMA system [21].

The simulation results are shown in Fig. 12. The BER plots of the proposed OTFS-SCMA method are shown for $P = 3, 4, 6$ and 9. Observe that in this case also, the BER decreases as P increases. The reduction in the BER values takes place due to the enhanced diversity. The BER improvement appears to get stagnant for higher values of P, such as $P = 6$ and 9. Fig. 12 also compares the proposed OTFS-SCMA method with the OTFS-OMA method in [8] for the same values of P. We can observe that OTFS-SCMA performs significantly better than OTFS-OMA for any value of P. We have also evaluated the performance of OFDM-SCMA. Fig. 12 shows the performance of OFDM-SCMA for two cases: (1) $P = 9$ with the delay and the Doppler values generated according to the EVA model, (2) $P = 9$ with the delay values generated according to the EVA model and all Doppler shifts set to zero. Observe that OFDM-SCMA performs poorly in both the cases in comparison to OTFS-SCMA and OTFS-OMA.



(a) $M = 8$, $N = 4$



(b) $M = 8$, $N = 8$

Fig. 13. OTFS-SCMA 6×4 SCMA system and OTFS-OMA over QAM in the uplink.

Moreover, OFDM-SCMA without any Doppler shifts performs better than that with non-zero Doppler shifts. Note that the results and the patterns for the EVA propagation model follow exactly those for the simple propagation model explained in the beginning.

B. Uplink

Case 1 ($M = 8, N = 4$, 6×4 SCMA System): First we consider a delay-Doppler grid $\Gamma_{4,8}$ with $J = 6$, $K = 4$ SCMA system. We consider a simple propagation

model for the wireless channel. The integer delay (l_{τ_i}) and the integer Doppler shift (k_{ν_i}) values are generated randomly from $\{0, 1, \ldots, P-1\}$ with the constraint that the pairs (k_{ν_i}, l_{τ_i}), $i = 1, \ldots, P$ are distinct. The effective coefficient matrix $\mathbf{H}_{\text{all,compr}}$ is found out as explained in Section III-C. Using this matrix, the single-stage MPA-based detection is carried out to extract the individual users' data symbols.

Fig. 13(a) presents the BER plots for the proposed OTFS-SCMA system for $P = 2$ and $P = 3$ over $\Gamma_{4,8}$. Observe that as P increases, the BER improves due to the enhancement of the diversity. The results of OTFS-OMA [7], [8] with 2 users are also presented in Fig. 13(a). QAM alphabet is considered for OTFS-OMA. For OTFS-OMA in the uplink, we can carry out the detection with the help of MPA. Note that in the uplink too, OTFS-SCMA performs significantly better than its OMA counterparts. Moreover, the OTFS-OMA in [8] performs better than the OTFS-OMA in [7]. It is noteworthy that the performance of the proposed OTFS-SCMA is better in the uplink than that in downlink. This performance difference may be accredited to the powerful single-stage MPA detector in uplink compared to the downlink's two-stage detection process. *Case 2 ($M = 8, N = 8$, 6×4 SCMA System):* Next we consider the delay-Doppler grid $\Gamma_{8,8}$ along with a 6×4 SCMA system. Observe from Fig. 13(b) that OTFS-SCMA performs better than the OTFS-OMA scheme [8]. Fig. 13(b) also shows the BER plots of OFDM-SCMA and OFDM-OMA for $P = 2$ and $P = 3$. In OFDM-SCMA, the same codebooks considered in OTFS-SCMA are used. For OFDM-OMA, the method presented in [8] is considered for distributing the QAM symbols over the time-frequency grid with 2 users. Observe that OTFS-SCMA and OTFS-OMA perform significantly better than OFDM-SCMA and OFDM-OMA respectively.

V. Conclusion

This paper presented a novel OTFS-NOMA scheme based on SCMA. This method considers pre-designed SCMA codebooks with a particular overloading factor. The sparse SCMA codewords of each user are strategically placed on the delay-Doppler grid. The ratio of the total number of codewords transmitted to the total number of resources in the delay-Doppler plane is precisely equal to the overloading factor of the underlying basic SCMA system. The OTFS-SCMA scheme is devised and analyzed for both downlink and uplink. The receiver in downlink includes a two stage detection process. First, the intermingling of the symbols in different delay-Doppler bins is resolved with the help of an OTFS detector. Subsequently, an SCMA detector eliminates the multi-user interaction. In uplink, the receiver at the BS includes a combined OTFS-SCMA detector based on a single-stage MPA. Both the delay-Doppler and the multi-user interferences are resolved at one shot. The asymptotic diversity orders of the proposed method are derived in the case of downlink and uplink. An algorithm is presented to devise an optimum strategy of allocating the codewords over the delay-Doppler grid such that full diversity is achieved. A data-embedded pilot-aided channel estimation method is developed. Simulation results are presented with different parameters for downlink and uplink with various overloading factors up to 200%. The results showed that the proposed OTFS-SCMA system could yield better BER performances than the conventional OTFS-OMA a and OTFS-PD-NOMA systems.

Acknowledgment

The authors would like to thank the anonymous reviewers and the editor for the constructive suggestions that have significantly improved the paper. The authors are also thankful to Dr. P. Raviteja for multiple discussions at various stages.

References

[1] R. Hadani *et al.*, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–13.

[2] R. Hadani *et al.*, "Orthogonal time frequency space modulation," *CoRR*, vol. abs/1808.00519, pp. 1–13, Aug. 2018. [Online]. Available: http://arxiv.org/abs/1808.00519

[3] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[4] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[5] 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.

[6] 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.

[7] 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.

[8] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "Multiple access in the delay-Doppler domain using OTFS modulation," *CoRR*, vol. abs/1902.03415, pp. 1–9, Feb. 2019. [Online]. Available: http://arxiv.org/abs/1902.03415

[9] R. M. Augustine and A. Chockalingam, "Interleaved time-frequency multiple access using OTFS modulation," in *Proc. IEEE Veh. Technol. Conf.*, no. 1, Sep. 2019, pp. 1–5.

[10] 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.

[11] 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*. [Online]. Available: http://arxiv.org/abs/2003.06387

[12] 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.

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

[14] P. Raviteja, K. T. Phan, Q. Jin, Y. Hong, and E. Viterbo, "Low-complexity iterative detection for orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

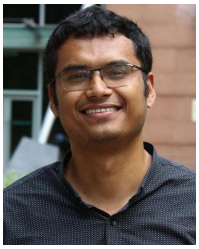
[15] S. Tiwari, S. S. Das, and V. Rangamgari, "Low complexity LMMSE receiver for OTFS," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2205–2209, Dec. 2019.

[16] 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.

[17] J. Dai, K. Niu, C. Dong, and J. Lin, "Improved message passing algorithms for sparse code multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9986–9999, Nov. 2017.

[18] X. Meng, Y. Wu, Y. Chen, and M. Cheng, "Low complexity receiver for uplink SCMA system via expectation propagation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–5.

[19] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3049–3063, Jun. 2019.
[20] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Effective diversity of OTFS modulation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 249–253, Feb. 2020.
[21] S. Zhang *et al.*, "A capacity-based codebook design method for sparse code multiple access systems," in *Proc. WCSP*, Oct. 2016, pp. 1–5.
[22] 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.
[23] 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.
[24] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.
[25] M. Vaezi, Z. Ding, and H. V. Poor, Eds., *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019.
[26] *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission And Reception, 36.104, Release 8*. Accessed: Apr. 24, 2021. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/36_series/36.104/



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. He is currently working as an Assistant Professor with the School of Electrical Sciences, Indian Institute of Technology Goa, India. His research interests lie in the areas of error correcting codes, modulation, and multiple access techniques for modern wireless communication systems.

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 & Technology, 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, Trivandrum, India. From 2014 to 2018, she worked as an Assistant Professor with Kannur University and Kerala Technological University, India. Her research interest includes 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.

Sanjeev Sharma (Member, IEEE) is currently working as an Assistant Professor of electronics engineering with the 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, and simulation, design, and analysis of wireless systems.