Space-Time Shift Keying Aided OTFS Modulation for Orthogonal Multiple Access

Zeping Sui, *Member, IEEE*, Hongming Zhang, *Senior Member, IEEE*, Sumei Sun, *Fellow, IEEE*, Lie-Liang Yang, *Fellow, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Space-time shift keying-aided orthogonal time frequency space modulation-based multiple access (STSK-OTFS-MA) is proposed for reliable uplink transmission in high-Doppler scenarios. As a beneficial feature of our STSK-OTFS-MA system, extra information bits are mapped onto the indices of the active dispersion matrices, which allows the system to enjoy the joint benefits of both STSK and OTFS signalling. Due to the fact that both the time-, space- and DD-domain degrees of freedom are jointly exploited, our STSK-OTFS-MA achieves increased diversity and coding gains. To mitigate the potentially excessive detection complexity, the sparse structure of the equivalent transmitted symbol vector is exploited, resulting in a pair of low-complexity near-maximum likelihood (ML) multiuser detection algorithms. Explicitly, we conceive a progressive residual check-based greedy detector (PRCGD) and an iterative reduced-space check-based detector (IRCD). Then, we derive both the unconditional single-user pairwise error probability (SU-UPEP) and a tight bit error ratio (BER) union-bound for our single-user STSK-OTFS-MA system employing the ML detector. Furthermore, the discrete-input continuous-output memoryless channel (DCMC) capacity of the proposed system is derived. The optimal dispersion matrices (DMs) are designed based on the maximum attainable diversity and coding gain metrics. Finally, it is demonstrated that our STSK-OTFS-MA system achieves both a lower BER and a higher DCMC capacity than its conventional spatial modulation (SM) and its orthogonal frequency-division multiplexing (OFDM) counterparts. As a benefit, the proposed system strikes a compelling BER *vs.* system complexity as well as BER *vs.* detection complexity trade-offs.**

***Index Terms*—Space-time shift keying (STSK), orthogonal time frequency space (OTFS), multiple access, maximum-likelihood detection, low-complexity detection, performance analysis.**

This work was supported by the National Science Foundation of China Grants 62001056 and the China Scholarship Council under Grant 202004910653. This work is also a contribution by Project REASON, a UK Government funded project under the Future Open Networks Research Challenge (FONRC) sponsored by the Department of Science Innovation and Technology (DSIT).

Lajos Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1 and EP/Y026721/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

Zeping Sui is with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, Belfast BT3 9DT, U.K. (e-mail: z.sui@qub.ac.uk).

Hongming Zhang is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (email: zhanghm@bupt.edu.cn).

Sumei Sun is with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: sunsm@i2r.a-star.edu.sg).

Lie-Liang Yang and Lajos Hanzo are with the Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lly@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

I. Introduction

During the last decade, space-time shift keying (STSK) [1]–[4] has been considered a compelling multi-functional multiple-input multiple-output (MIMO) arrangement, where the information bits are jointly mapped onto the conventional amplitude-phase modulated (APM) symbols and the indices of the active dispersion matrices (DMs). To elaborate further, each single APM symbol is dispersed to multiple transmit antennas (TAs) and time-slots by activating one out of Q DMs. Hence, STSK can achieve diversity and multiplexing gains [1]. By contrast, the conventional spatial modulation (SM) activates one TA to transmit a single APM symbol, yielding only receive diversity gain [5].

Orthogonal time frequency space (OTFS) modulation also constitutes a promising candidate for next-generation wireless networks [6]–[8]. Since it is capable of providing reliable transmission in high-mobility scenarios, it has been widely studied in the context of reconfigurable intelligent surfaces (RISs) [9] and low-earth orbit (LEO) satellites [10]. More specifically, in OTFS systems, the information symbols are mapped to the delay-Doppler (DD)-domain, and each symbol is spread across the entire time-frequency (TF)-domain by leveraging the inverse symplectic finite Fourier transform (ISFFT). Therefore, OTFS can attain both time and frequency diversity gains, if channels are time-frequency selective or doubly-selective [11], [12]. In addition, the sparse DD-domain sparse representations of the doubly-selective channels are incorporated into the OTFS theory, and the dimension of the DD-domain channel model is reduced to the number of resolvable paths [6], [7]. Since the communication distance and relative velocity can be approximated as constants within a few milliseconds, the DD-domain channels can be regarded as nearly time-invariant over an entire OTFS frame [13]. Furthermore, the inter-carrier interference (ICI) and inter-symbol interference (ISI) introduced by Doppler and delay spreads remain quasi-orthogonal with the aid of ISFFT, which can hence be processed in the Doppler-domain and delay-domain separately [6]. By contrast, the performance of the conventional orthogonal frequency-division multiplexing (OFDM) suffers severely in the face of high-Doppler doubly-selective channels, since the orthogonality of subcarriers may be destroyed by severe ICI. Therefore, in high-Doppler scenarios, OTFS constitutes a more promising signaling scheme than the conventional OFDM [6], [7].

More recently, OTFS-based non-orthogonal multiple-access (NOMA) communication schemes have been studied [14]–

TABLE I
CONTRASTING OUR CONTRIBUTIONS TO THE LITERATURE

Contributions	**This paper**	[2]	[7]	[8]	[18]	[19]	[20]	[23]
OTFS-OMA	✓				✓	✓		
STSK-aided MA	✓	✓					✓	
BER performance analysis	✓			✓				✓
Coding gain & diversity order analysis	✓			✓			✓	✓
Greedy check detection	✓							
Reduced-space check detection	✓							
Capacity analysis	✓							
Complexity analysis	✓		✓					
Dispersion matrix design	✓							
LDPC coded system	✓	✓		✓				

[17], where different users arranged in the same DD-domain resource blocks (RBs) are distinguished by their unique sparse codewords [14], [15], [17] and power levels [16]. However, the performance of OTFS-NOMA may degrade significantly due to the non-orthogonality-induced interference, and the excessive complexity of the transceiver [18]. As a remedy, OTFS-based orthogonal MA (OTFS-OMA) techniques have been designed in [18], [19], where the RBs of different users are arranged in non-overlapping grids in the TF- and/or DD-domains. In [18], the spectral efficiency (SE) of several OTFS-OMA schemes using rectangular pulse shapes are analyzed. Nevertheless, the above OTFS-OMA schemes only modulate data in the DD-domain. None of them exploits the degrees of freedom in the space-time (ST)-domain, even though it would be beneficial to leverage both transmit and receive diversity gains for further BER performance improvement.

As a parallel development, MA communications schemes have been designed in association with SM/STSK in [2], [20]. Specifically in [2], the STSK-aided OFDM-based multiple access (STSK-OFDM-MA) paradigms have been proposed, where an attractive diversity *vs.* multiplexing gain trade-off was provided. However, these SM/STSK-based schemes are designed based on flat Rayleigh fading or *low-mobility* frequency-selective fading channels, but without considering the ICI imposed by *high-mobility environments*. However, next-generation MA systems aim for providing reliable data transmission under high-mobility scenarios [21], [22]. Therefore, the BER performance of the STSK-OFDM-MA schemes may degrade substantially under doubly-selective channels, yielding a significant system reliability loss. On the other hand, it can be observed from [2] that the STSK-OFDM-MA scheme is capable of providing better BER performance than the conventional solutions, resulting in a more reliable communication system. This observation implies that the reliability of the above-mentioned OTFS-OMA systems can be further enhanced by exploiting the STSK technique to attain both higher diversity and coding gains. Against this backdrop, in this paper we jointly invoke the DD-, space- and time-domain (TD) resources for transmission over doubly-selective channels. Explicitly, by intrinsically amalgamating STSK and OTFS-OMA, we propose space-time shift keying-aided OTFS-based MA (STSK-OTFS-MA) for reliable communications in doubly-selective channels.

The novel contributions of the paper are boldly and explicitly contrasted to the existing literatures in Table I, which are addressed below.

- We propose an STSK-OTFS-MA scheme for supporting the reliable data transmission of multiple users over high-mobility channels, where information is conveyed both by the classic APM symbols and the indices of active DMs. According to the $(N \times M)$ DD-domain grids, we first activate one out of Q DMs for spreading NM APM symbols to both the space and time dimensions, resulting in NM STSK blocks. Then a tailor-made ST mapper is conceived for mapping the elements of the STSK blocks onto the transmitted DD-domain symbol matrices of users, which enables our STSK-OTFS-MA to achieve both transmit and receive diversity gains. Moreover, based on the DD-domain statistics of users, a resource-allocation scheme is introduced for the STSK-OTFS-MA system. Explicitly, the RBs of different users are mapped to the non-overlapping grids along the delay domain to mitigate the multiuser interference (MUI) caused by the Doppler shift. Furthermore, the proposed STSK-OTFS-MA is capable of striking an attractive diversity versus multiplexing gain trade-off. Both the analytical and simulation results illustrated that the STSK-OTFS-MA advocated achieves a better BER performance than its conventional SM-OTFS, single-input-multiple-output (SIMO)-OTFS and STSK-OFDM-MA counterparts. Additionally, the general flexibility of the proposed STSK-OTFS-MA scheme is demonstrated in different-rate low-density parity-check (LDPC)-coded systems. Finally, the BER *vs.* system complexity of the STSK-OTFS-MA and other counterparts are also evaluated.
- A pair of low-complexity near-maximum likelihood detectors (MLDs) are proposed for the STSK-OTFS-MA scheme. Firstly, inspired by the family of greedy algorithms designed for compressed sensing (CS), a progressive residual check-based greedy detector (PRCGD) is conceived, where optimal local choices are obtained at each iteration, yielding a detector approaching the globally optimal performance. Furthermore, commencing with the consideration of detecting the APM and DM-index symbols separately, an iterative reduced-space check-based detector (IRCD) is proposed. Specifically, by sorting the reliabilities of all DM activation patterns (DAPs), a reduced set of DAPs is tested. Finally, the BER performance *vs.* complexity of the MLD, of the PRCGD

and of the IRCD are compared.

- We derive the unconditional single-user pairwise error probability (SU-UPEP) of the STSK-OTFS-MA system. Then by invoking the union-bound technique, we derive the closed-form single-user BER bound of our proposed STSK-OTFS-MA system employing MLD, which is shown to be tight for moderate to high signal-to-noise ratios (SNRs). Then, based on the SU-UPEP, both the diversity order and the ST coding gain achieved by the STSK-OTFS-MA system are determined.
- The single-user discrete-input continuous-output memoryless channel (DCMC) capacity of the STSK-OTFS-MA system is derived, which is demonstrated to outperform its SM counterpart. Additionally, based on the SU-UPEP and the DCMC capacity derived, the design criteria of DMs are proposed, for approaching the maximum attainable diversity order and ST coding gain.

The rest of the paper is structured as follows. In Section II, the proposed STSK-OTFS-MA system model is investigated. Then, low-complexity near-ML multiuser detection algorithms are proposed in Section III. In Section IV, the overall system performance is characterized and the DM design algorithm is detailed. The simulation results are shown in Section V. Finally, the conclusions are offered in Section VI.

Notation: We use the following notation throughout this paper: $\mathbb{C}$ and $\mathbb{R}$ are the ring of complex and real; $\mathbb{B}$ and $\mathbb{Z}_+^M$ represent the real integer set of $\{1,\ldots,M\}$ and the bit set consisting of $\{0,1\}$; $\mathbb{E}[\cdot]$ and $\mathrm{tr}\{\cdot\}$ denote the expectation and trace operator; $x(l)$ and $X(l,k)$ are the lth of vector $\boldsymbol{x}$ and (l,k)th element of matrix $\boldsymbol{X}$, respectively; $\mathrm{vec}(\boldsymbol{A})$ denotes the vector formulated by stacking the columns of $\boldsymbol{A}$ to obtain a single column vector matrix, and $\mathrm{vec}^{-1}(\boldsymbol{a})$ denotes the inverse vectorization operation to form the original matrix; $\otimes$ denotes the Kronecker product of two matrices; $\mathcal{CN}(\boldsymbol{a},\boldsymbol{B})$ is the complex Gaussian distribution having a mean vector $\boldsymbol{a}$ and covariance matrix $\boldsymbol{B}$; $\boldsymbol{A}[:,1:n]$ and $\boldsymbol{A}[1:m,:]$ represent the first n columns and first m rows of a matrix $\boldsymbol{A}$, respectively; $\boldsymbol{I}_N$ and $\boldsymbol{I}_N(l)$ denote an N-dimensional identity matrix and its rows shift by l; The module-N and determinant operations are defined by $[\cdot]_N$ and $\det(\cdot)$; $\delta(\cdot)$ is the delta function; The uniform distribution in the interval $[a,b]$ is denoted by $\mathcal{U}[a,b]$; $\mathrm{sta}\{\boldsymbol{A}_n^{(u)}\}|_{n=0}^{N-1} = [(\boldsymbol{A}_0^{(u)})^T,(\boldsymbol{A}_1^{(u)})^T,\ldots,(\boldsymbol{A}_{N-1}^{(u)})^T]^T$ and $\mathrm{sta}\{\boldsymbol{A}_n^{(u)}\}|_{u=0}^{U-1} = [(\boldsymbol{A}_n^{(0)})^T,(\boldsymbol{A}_n^{(1)})^T,\ldots,(\boldsymbol{A}_n^{(U-1)})^T]^T$ denotes the matrix (vector) formulated by stacking N and U identical-dimensional sub-matrices (sub-vectors) $\boldsymbol{A}_n$ and $\boldsymbol{A}_n^{(u)}$ for $n=0,\ldots,N-1$ and $u=0,\ldots,N-1$, respectively.

II. System Model

A. Transmitter Description

Let us consider a single-cell uplink communication scenario, where the information signals of U users are simultaneously transmitted to a base station (BS). Specifically, we assume that N_t TAs are employed by each user, and the BS uses N_r receive antennas (RAs). Moreover, each TA transmits an OTFS signal having the bandwidth of $B = M\Delta f$ and time-slot duration of $T_f = NT$, where M and N denote the number of subcarriers and time intervals within an OTFS time-slot, while Δf and T represent the subcarrier spacing and symbol duration, respectively. Hence, we have a total of $M_d = NM$ DD-domain RBs and each user occupies $G = M_d/U$ RBs. As shown in Fig. 1, the information bit sequence $\boldsymbol{b}^{(u)} \in \mathbb{B}^{\bar{L}}$ transmitted by the user u is first partitioned into G groups, yielding $\boldsymbol{b}^{(u)} = [\boldsymbol{b}_1^{(u)},\ldots,\boldsymbol{b}_G^{(u)}]$. The gth bit sequence $\boldsymbol{b}_g^{(u)} \in \mathbb{B}^{L_b}$, $g=0,\ldots,G-1$, contains $L_b = \bar{L}/G = L_1+L_2$ bits. Explicitly, the subsequence $\boldsymbol{b}_{1,g}^{(u)} \in \mathbb{B}^{L_1}$ is mapped into an index symbol in $\{1,\ldots,Q\}$ for selecting an active DM in $\mathcal{A} = \{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_Q\}$, where we have $L_1 = \log_2 Q$. The remaining L_2-bit sequences $\boldsymbol{b}_{2,g}^{(u)} \in \mathbb{B}^{L_2}$ are mapped into the normalized quadrature amplitude modulation (QAM)/phase-shift keying (PSK) symbols chosen from the constellation $\mathcal{F} = \{f_1,\ldots,f_V\}$, where $L_2 = \log_2 V$. Hence, in an OTFS frame, the total number of bits transmitted per user can be given by $L = UGL_b = NM\log_2(VQ)$. Assuming that the STSK symbol duration includes T_c OTFS time-slots, the ST codeword matrices of the user u can be expressed as [1]

$$\boldsymbol{S}_{d,g}^{(u)} = f_{l_g}^{(u)}\boldsymbol{A}_{d,g}^{(u)} \in \mathbb{C}^{N_t\times T_c}, \tag{1}$$

where $f_{l_g}^{(u)} \subset \mathcal{F}$ and $\boldsymbol{A}_{d,g}^{(u)} \subset \mathcal{A}$ denote a single QAM/PSK symbol and an active DM, respectively. By introducing $\breve{\boldsymbol{s}}_{d,g}^{(u)} = \mathrm{vec}(\boldsymbol{S}_{d,g}^{(u)})$ and stacking all the G RBs of user u, the uth ST stacked codeword vector is formulated as

$$\breve{\boldsymbol{s}}_d^{(u)} = \mathrm{sta}\{\breve{\boldsymbol{s}}_{d,g}^{(u)}\}|_{g=0}^{G-1}, \quad g=0,\ldots,G-1. \tag{2}$$

Let us parameterize the STSK-OTFS-MA system by the five-tupple (N_t,N_r,T_c,Q,V). Note that the DM set can be generated using diverse design criteria, for example by minimizing the pairwise error probability or by maximizing the DCMC capacity [2], which will be further investigated in Section IV-D. In the STSK pre-processing block, the DMs are assumed to be normalized to maintain the transmitted power, and the constraint is given by [1]

$$\mathrm{tr}(\boldsymbol{A}_q^H\boldsymbol{A}_q) = T_c, \quad q=1,\ldots,Q. \tag{3}$$

As shown in Fig. 2, the ST codewords are fed into the ST mapper, which is detailed in Section II-B, and the uth user's transmitted frame output by the ST mapper $\boldsymbol{s}^{(u)} \in \mathbb{C}^{N_tG\times T_c}$ can be formulated as

$$\boldsymbol{s}^{(u)} = \begin{bmatrix} \boldsymbol{s}_{0,0}^{(u)} & \cdots & \boldsymbol{s}_{0,T_c-1}^{(u)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{s}_{N_t-1,0}^{(u)} & \cdots & \boldsymbol{s}_{N_t-1,T_c-1}^{(u)} \end{bmatrix}, \tag{4}$$

where we have $\boldsymbol{s}_{n_t,t_c}^{(u)} = [S_{d,0}^{(u)}(n_t,t_c),\ldots,S_{d,G-1}^{(u)}(n_t,t_c)]^T \in \mathbb{C}^{G\times 1}$ for $n_t=0,\ldots,N_t-1$ and $t_c=0,\ldots,T_c-1$, which is formulated based on (1). Then the G ST codeword elements of $\boldsymbol{s}_{n_t,t_c}^{(u)}$ are mapped to M_d RBs, yielding

$$\boldsymbol{x}_{n_t,t_c}^{(u)} = \boldsymbol{\mathcal{P}}^{(u)}\boldsymbol{s}_{n_t,t_c}^{(u)} = [x_{n_t,t_c}^{(u)}(0),\ldots,x_{n_t,t_c}^{(u)}(M_d-1)]^T, \tag{5}$$

where $\boldsymbol{\mathcal{P}}^{(u)}$ is the $(M_d\times G)$-element resource allocation matrix. To alleviate the MUI caused by ICI, we introduce our delay-domain index-based RB allocation scheme, i.e., Scheme 1 illustrated in Fig. 3 (a), where each user occupies $J = M/U$

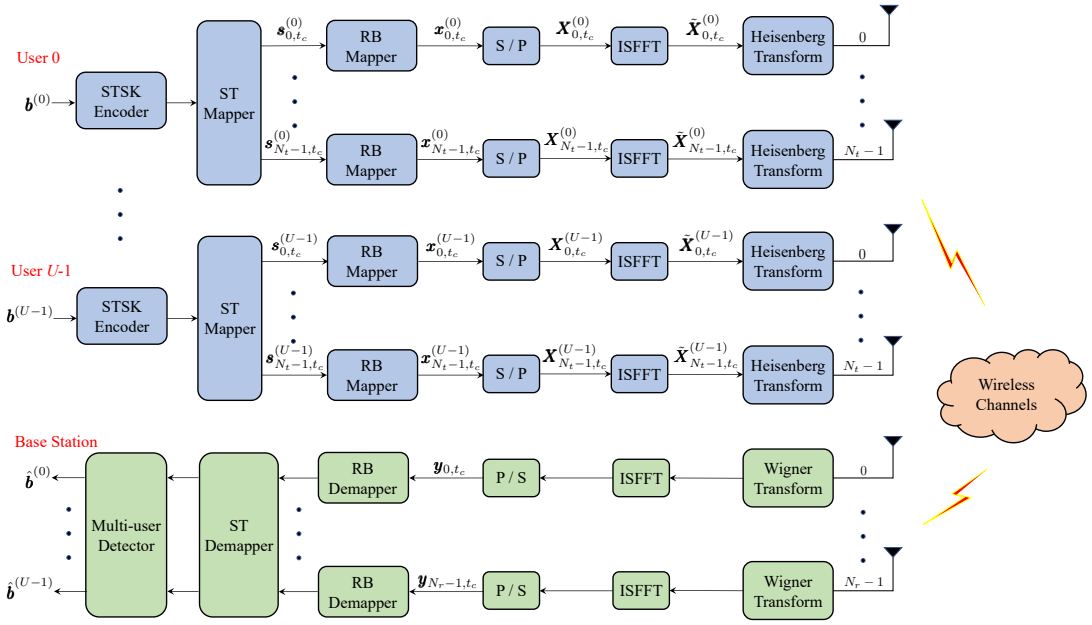

Fig. 1. Illustration of the STSK-OTFS-MA system.

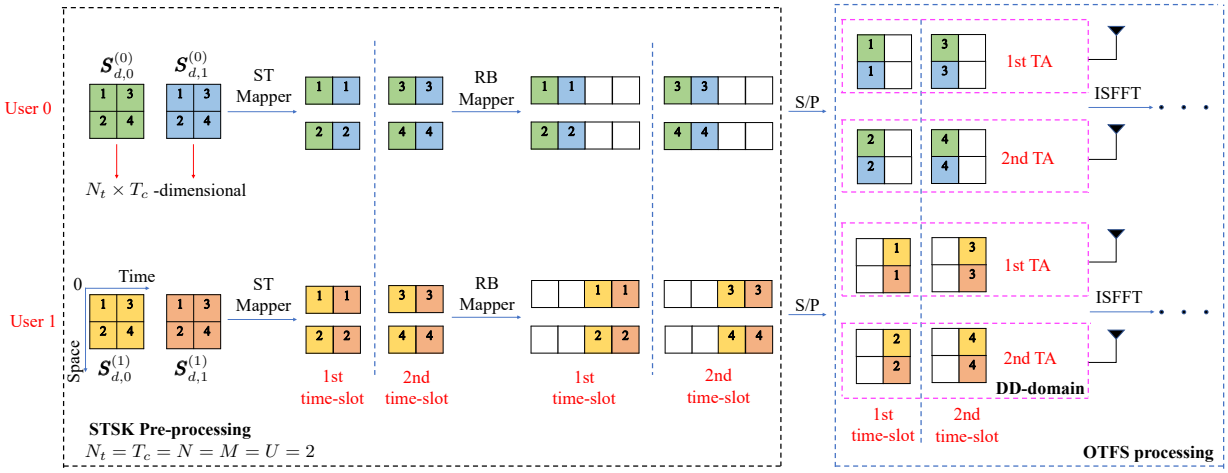

Fig. 2. A toy example of the STSK-OTFS-MA system with $N_t = T_c = N = M = U = 2$.

columns of the DD-domain grids. By contrast, the Doppler-domain index-based resource allocation scheme (Scheme 2) shown in Fig. 3 (b) is invoked as the benchmark. Let us denote the column indices of user u by $\mathcal{L}^{(u)} = \{l_0^{(u)}, \ldots, l_{J-1}^{(u)}\}$. More specifically, if we assume that the ST codewords of user u are assigned to the RBs set $\mathcal{N}^{(u)}$, then the corresponding elements of $\boldsymbol{\mathcal{P}}^{(u)}$ can be expressed as

$$\mathcal{P}^{(u)}(m_d, g) = \begin{cases} 1, & \text{if } m_d \in \mathcal{N}^{(u)} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

for $0 \leq m_d \leq M_d - 1$. The indices in $\mathcal{N}^{(u)}$ are given by $m_d = l_j^{(u)} N + n$ and $g = j^{(u)} N + n$, where $l_j^{(u)} = (J \times u) + j$ for $n = 0, \ldots, N-1$ and $j = 0, \ldots, J-1$.

When considering the OTFS processing block, by defining a DD-domain codeword matrix as $\boldsymbol{X}_{n_t,t_c}^{(u)} = \text{vec}^{-1}(\boldsymbol{x}_{n_t,t_c}^{(u)})$, the TF-domain signal can be formulated using ISFFT as

$$\tilde{X}_{n_t,t_c}^{(u)}(n,m) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \frac{X_{n_t,t_c}^{(u)}(k,l)}{\sqrt{M_d}} e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)}, \tag{7}$$

for $n = 0, \ldots, N-1$ and $m = 0, \ldots, M-1$. The transmitted TD signal is obtained by exploiting the Heisenberg transform, yielding

$$\tilde{s}_{n_t,t_c}^{(u)}(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \tilde{X}_{n_t,t_c}^{(u)}(n,m) g_{\text{tx}}(t - nT) e^{j2\pi m \Delta f (t-nT)}, \tag{8}$$

where $g_{\text{tx}}(t)$ denotes the transmit waveform.

B. Received Signals

Assuming that the synchronization among uplink users is perfect. Let us consider a P-path DD-domain time-varying

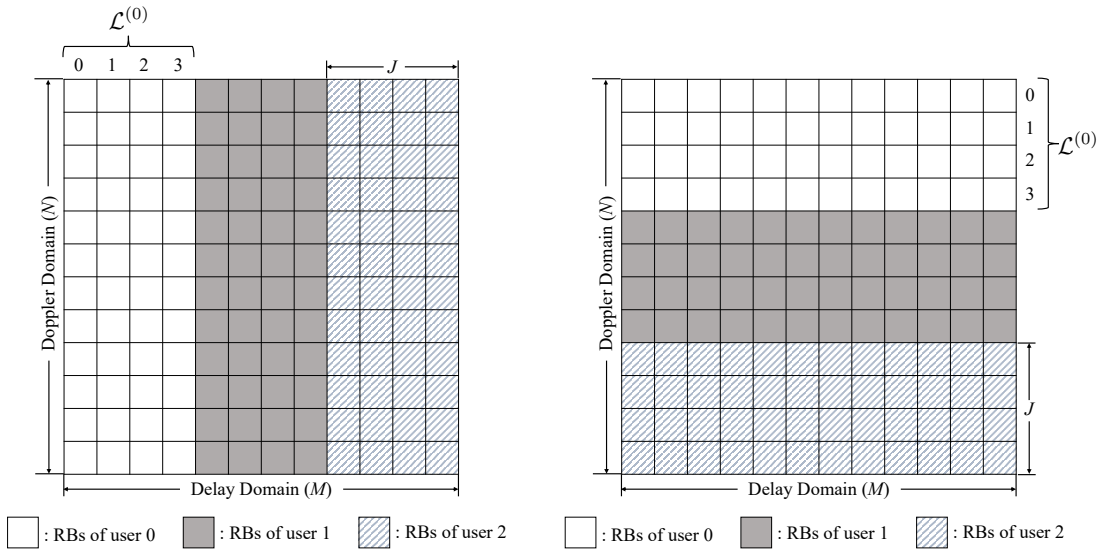

Fig. 3. Illustration of resource allocations (a) Scheme 1 (b) Scheme 2 with $M = 12$, $N = 12$, $U = 3$ and $J = M/U = 4$, where $\mathcal{L}^{(u)}$ represents the column index set of user u for $u = 0, \ldots, U-1$.

multipath channel model between the n_tth TA of user u and the n_rth RA, which can be formulated as $h^{(u)}_{n_r,n_t}(\tau,\nu) = \sum_{i=1}^{P} h^{(u)}_{i,n_r,n_t}\delta(\tau-\tau_i)\delta(\nu-\nu_i)$, where $h^{(u)}_{i,n_r,n_t}$, τ_i and ν_i denote the complex-valued path gain between the n_tth TA and n_rth RA, normalized delay and Doppler shifts introduced by the ith path, respectively [6]. Here we have $h_i \sim \mathcal{CN}(0, 1/P), \forall i$, which is independent of τ_i and ν_i. Therefore, the delay and Doppler shifts corresponding to the ith reflector are given by $\nu_i = \frac{k_i}{NT}, \tau_i = \frac{l_i}{M\Delta f}$, where $l_i = a_i + \alpha_i$ and $k_i = b_i + \beta_i$ represent the normalized delay and Doppler indices associated with the ith path, where a_i and b_i denote the integer delay and Doppler indices, while the fractional components are given by $\alpha_i, \beta_i \in \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$. During the t_cth OTFS time-slot, the TD signal of the n_rth RA received from the n_tth TA of user u can be expressed as [7]

$$r^{(u)}_{n_r,n_t,t_c}(t) = \int\int h^{(u)}_{n_r,n_t}(\tau,\nu)\tilde{s}^{(u)}_{n_t,t_c}(t-\tau)e^{j2\pi\nu(t-\tau)}d\tau d\nu + n^{(u)}_{n_r,n_t,t_c}(t), \tag{9}$$

where $n^{(u)}_{n_r,n_t,t_c}(t)$ is the complex-valued additive white Gaussian noise (AWGN). Based on the Wigner transform, the elements of the corresponding received TF-domain codeword matrix $\tilde{\boldsymbol{Y}}^{(u)}_{n_r,n_t,t_c} \in \mathbb{C}^{N\times M}$ can be obtained as

$$\tilde{Y}^{(u)}_{n_r,n_t,t_c}(n,m) = \int r^{(u)}_{n_r,n_t}(t')g^{*}_{\text{rx}}(t'-nT)e^{-j2\pi m\Delta f(t'-nT)}dt', \tag{10}$$

where $g_{\text{rx}}(t)$ is the receive waveform. Then, upon utilizing the SFFT, the received DD-domain codeword matrix can be formulated as

$$Y^{(u)}_{n_r,n_t,t_c}(k,l) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\frac{\tilde{Y}^{(u)}_{n_r,n_t,t_c}(n,m)}{\sqrt{M_d}}e^{-j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}, \tag{11}$$

for $k = 0, \ldots, N-1$ and $l = 0, \ldots, M-1$. Assuming that both the transmit and receive waveforms satisfy the bi-orthogonal condition, the vector-form input-output relationship for user u can be formulated as [7]

$$\boldsymbol{y}^{(u)}_{n_r,n_t,t_c} = \boldsymbol{H}^{(u)}_{n_r,n_t}\boldsymbol{x}^{(u)}_{n_t,t_c} + \boldsymbol{n}^{(u)}_{n_r,n_t,t_c}, \tag{12}$$

where we have $\boldsymbol{y}^{(u)}_{n_r,n_t,t_c} = \text{vec}\left(\boldsymbol{Y}^{(u)}_{n_r,n_t,t_c}\right)$, and $\boldsymbol{n}^{(u)}_{n_r,n_t,t_c}$ denotes the complex-valued AWGN vector. Moreover, the effective DD-domain channel matrix $\boldsymbol{H}^{(u)}_{n_r,n_t}$ can be expressed as $\boldsymbol{H}^{(u)}_{n_r,n_t} = \sum_{i=1}^{P}\boldsymbol{I}_M(l_i) \otimes \left[\boldsymbol{I}_N(k_i)h^{(u)}_{i,n_r,n_t}e^{-j2\pi\frac{l_ik_i}{M_d}}\right]$ [24]. Let $\boldsymbol{H}^{(u)}_{n_r} = \left[\boldsymbol{H}^{(u)}_{0,n_r}, \ldots, \boldsymbol{H}^{(u)}_{N_t-1,n_r}\right]$ and $\boldsymbol{x}^{(u)}_{t_c} = \text{sta}\{\boldsymbol{x}^{(u)}_{n_t,t_c}\}|^{N_t-1}_{n_t=0}$. Then the DD-domain codeword vector received by the n_rth RA within the t_cth time-slot can be expressed as

$$\begin{aligned}\boldsymbol{y}_{n_r,t_c} &= \sum_{u=0}^{U-1}\boldsymbol{y}^{(u)}_{n_r,t_c} + \boldsymbol{n}_{n_r,t_c} = \sum_{u=0}^{U-1}\sum_{n_t=0}^{N_t-1}\boldsymbol{H}^{(u)}_{n_r,n_t}\boldsymbol{x}^{(u)}_{n_t,t_c} + \boldsymbol{n}_{n_r,t_c} \\ &= \sum_{u=0}^{U-1}\boldsymbol{H}^{(u)}_{n_r}\boldsymbol{x}^{(u)}_{t_c} + \boldsymbol{n}_{n_r,t_c},\end{aligned} \tag{13}$$

where $\boldsymbol{n}_{n_r,t_c}$ is the corresponding complex AWGN vector with a zero mean and a covariance matrix of $N_0\boldsymbol{I}_{M_d}$, expressed as $\mathcal{CN}(0, N_0\boldsymbol{I}_{M_d})$. Hence, the average SNR per RA is given by $\gamma = 1/N_0$. By invoking all the signals of N_r RAs and let $\bar{\boldsymbol{y}}_{t_c} = \text{sta}\{\boldsymbol{y}_{n_r,t_c}\}^{N_r-1}_{n_r=0}$, then it can be shown that the end-to-end DD-domain input-output relationship for the time-slot t_c is given as

$$\bar{\boldsymbol{y}}_{t_c} = \sum_{u=0}^{U-1}\bar{\boldsymbol{H}}^{(u)}\boldsymbol{x}^{(u)}_{t_c} + \tilde{\boldsymbol{n}}_{t_c}, \tag{14}$$

for $t_c = 0, \ldots, T_c-1$, where $\tilde{\boldsymbol{n}}_{t_c} = \text{sta}\{\boldsymbol{n}_{n_r,t_c}\}|^{N_r-1}_{n_r=0}$ denotes the stacked noise vector at the BS side. The DD-domain MIMO channel matrix of the uth user can be expressed as $\bar{\boldsymbol{H}}^{(u)} = \text{sta}\{\boldsymbol{H}^{(u)}_{n_r}\}|^{N_r-1}_{n_r=0}$. Moreover, we invoke the relationship between the ST codewords and the stacked codeword vector shown in (5), the transmitted codeword vector $\boldsymbol{x}^{(u)}_{t_c}$ can

be formulated as

$$\begin{aligned}\boldsymbol{x}_{t_c}^{(u)} &= \left[\left(\boldsymbol{\mathcal{P}}^{(u)}\boldsymbol{s}_{0,t_c}^{(u)}\right)^T,\ldots,\left(\boldsymbol{\mathcal{P}}^{(u)}\boldsymbol{s}_{N_t-1,t_c}^{(u)}\right)^T\right]^T \\ &= \left(\boldsymbol{I}_{N_t}\otimes\boldsymbol{\mathcal{P}}^{(u)}\right)\left[\left(\boldsymbol{s}_{0,t_c}^{(u)}\right)^T,\ldots,\left(\boldsymbol{s}_{N_t-1,t_c}^{(u)}\right)^T\right]^T \\ &= \bar{\boldsymbol{\mathcal{P}}}^{(u)}\boldsymbol{s}_{t_c}^{(u)}, \end{aligned} \tag{15}$$

where $\bar{\boldsymbol{\mathcal{P}}}^{(u)}$ is the equivalent resource allocation matrix of user u. Then, upon applying (15) to (14), $\bar{\boldsymbol{y}}_{t_c}$ can be rewritten as

$$\begin{aligned}\bar{\boldsymbol{y}}_{t_c} &= \sum_{u=0}^{U-1}\bar{\boldsymbol{H}}^{(u)}\bar{\boldsymbol{\mathcal{P}}}^{(u)}\boldsymbol{s}_{t_c}^{(u)}+\tilde{\boldsymbol{n}}_{t_c} \\ &= \sum_{u=0}^{U-1}\boldsymbol{\Omega}^{(u)}\boldsymbol{s}_{t_c}^{(u)}+\tilde{\boldsymbol{n}}_{t_c}, \quad t_c=0,\ldots,T_c-1, \end{aligned} \tag{16}$$

where $\boldsymbol{\Omega}^{(u)} = \bar{\boldsymbol{H}}^{(u)}\bar{\boldsymbol{\mathcal{P}}}^{(u)}$ is a $(M_dN_r \times GN_t)$-dimensional matrix.

By considering all the T_c time-slots, let us define the overall received symbol vector and the transmitted STSK symbol vector of user u as $\tilde{\boldsymbol{y}} = \text{sta}\{\bar{\boldsymbol{y}}_{t_c}\}|_{t_c=0}^{T_c-1}$ and $\bar{\boldsymbol{s}}^{(u)} = \text{sta}\{\boldsymbol{s}_{t_c}^{(u)}\}|_{t_c=0}^{T_c-1}$, respectively. Consequently, $\tilde{\boldsymbol{y}}$ can be obtained as

$$\begin{aligned}\tilde{\boldsymbol{y}} &= \sum_{u=0}^{U-1}\left(\boldsymbol{I}_{T_c}\otimes\boldsymbol{\Omega}^{(u)}\right)\bar{\boldsymbol{s}}^{(u)}+\tilde{\boldsymbol{n}} \\ &= \sum_{u=0}^{U-1}\tilde{\boldsymbol{\Omega}}^{(u)}\bar{\boldsymbol{s}}^{(u)}+\tilde{\boldsymbol{n}} = \tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{s}}+\tilde{\boldsymbol{n}}, \end{aligned} \tag{17}$$

where $\tilde{\boldsymbol{\Omega}} = \left[\tilde{\boldsymbol{\Omega}}^{(0)},\ldots,\tilde{\boldsymbol{\Omega}}^{(U-1)}\right] \in \mathbb{C}^{M_dN_rT_c\times GN_tT_cU}$, $\tilde{\boldsymbol{s}} = \text{sta}\{\bar{\boldsymbol{s}}^{(u)}\}|_{u=0}^{U-1} \in \mathbb{C}^{GN_tT_cU\times 1}$ and $\tilde{\boldsymbol{n}} = \text{sta}\{\tilde{\boldsymbol{n}}_{t_c}\}|_{t_c=0}^{T_c-1}$.

Now we further detail the ST mapper shown in Fig. 1. For deriving the associated input-output relationship, we stack the transmitted ST codewords of U users shown in (2), yielding $\tilde{\boldsymbol{s}}_d = \text{sta}\{\check{\boldsymbol{s}}_d^{(u)}\}|_{u=0}^{U-1}$. As illustrated in Fig. 1 and Fig. 2 of Section II-A, the symbol vector $\tilde{\boldsymbol{s}}_d$ is essentially obtained by rearranging all the GN_tT_cU ST codeword elements $S_{d,g}^{(u)}(n_t,t_c)$ based on the following order 1 of *1) TA index 2) time-slot index 3) RB index 4) user index*. However, according to the STSK and multiuser system design principles [1], [19], [20], [25] and to the transceiver structure of Section II-A and Section II-B, the above-mentioned ST codeword elements should be placed in the following order 2 of *1) RB index 2) TA index 3) user index 4) time-slot index*, which can be further verified based on our derivation of (12)-(17). Therefore, we have $\tilde{\boldsymbol{s}} = \boldsymbol{\Upsilon}\tilde{\boldsymbol{s}}_d$, where $\boldsymbol{\Upsilon}$ is the $(GN_tT_cU\times GN_tT_cU)$-element ST mapping matrix, whose elements are defined as

$$\Upsilon(d_x,d_y) = \begin{cases}1, & \text{if } d_x = g+n_tG+uN_tG+t_cGUN_t \\ & \text{and } d_y = n_t+t_cN_t+gN_tT_c+uGN_tT_c \\ 0, & \text{otherwise,}\end{cases} \tag{18}$$

for $0\le d_x,d_y\le GN_tUT_c-1$. It should be noted that d_x and d_y are formulated based on order 2 and order 1, respectively. Finally, based on (17) and (18), the end-to-end input-output relationship for the OTFS frame can be expressed as

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{\Omega}}\boldsymbol{\Upsilon}\tilde{\boldsymbol{s}}_d+\tilde{\boldsymbol{n}} = \tilde{\boldsymbol{\Omega}}\boldsymbol{\Upsilon}\bar{\boldsymbol{\chi}}\boldsymbol{K}+\tilde{\boldsymbol{n}} = \boldsymbol{C}\boldsymbol{K}+\tilde{\boldsymbol{n}}, \tag{19}$$

where $\bar{\boldsymbol{\chi}} = \boldsymbol{I}_{UG}\otimes\boldsymbol{\chi}$ with $\boldsymbol{\chi} = [\text{vec}(\boldsymbol{A}_1),\ldots,\text{vec}(\boldsymbol{A}_Q)] \in \mathbb{C}^{N_tT_c\times Q}$. Moreover, the equivalent transmitted symbol vector can be defined as $\boldsymbol{K} = \text{sta}\{\boldsymbol{K}^{(u)}\}_{u=0}^{U-1}$, in which the sub-vectors can be formulated as $\boldsymbol{K}^{(u)} = \text{sta}\{\bar{\boldsymbol{K}}_{\mathcal{I}_g^{(u)}}\}|_{g=0}^{G-1}$, and the index sets are given by $\mathcal{I}_g^{(u)} = \{q_g^{(u)}, l_g^{(u)}\}$, where $q_g^{(u)} \in \{1,\ldots,Q\}$ and $l_g^{(u)} \in \{1,\ldots,V\}$ for $u=0,\ldots,U-1$ and $g=0,\ldots,G-1$. Moreover, the gth equivalent transmitted symbol vector of the uth user can be formulated as

$$\bar{\boldsymbol{K}}_{\mathcal{I}_g^{(u)}} = [\underbrace{0,\cdots,0}_{q_g^{(u)}-1}, f_{l_g^{(u)}}, \underbrace{0,\cdots,0}_{Q-q_g^{(u)}}]^T \in \mathbb{C}^{Q\times 1}, \tag{20}$$

where the $l_g^{(u)}$th QAM/PSK symbol $f_{l_g^{(u)}}$ is located in the $q_g^{(u)}$th element, while the active DM in the gth STSK block of user u is denoted as $\boldsymbol{A}_{q_g^{(u)}}$. According to (19)-(20), it can be observed that a total of M_d ST codewords are transmitted in the entire system, hence only M_d elements of $\bar{\boldsymbol{K}}$ have a non-zero value. Therefore, the index candidate sets of non-zero valued elements are represented as $\mathcal{Q} = \{\mathcal{Q}_1,\ldots,\mathcal{Q}_C\}$, which has $C = 2^{M_dL_1} = Q^{M_d}$ index candidate subsets; The cth subset can be expressed as $\mathcal{Q}_c = \{\mathcal{Q}_c(0),\ldots,\mathcal{Q}_c(M_d-1)\} \subset \mathcal{Q}$, whose elements obey $\mathcal{Q}_c(m_d) \in \mathbb{Z}_+^{QM_d}$ for $m_d = 0,\ldots,M_d-1$ and $c=1,\ldots,C$. For a given $\boldsymbol{K}$, the index candidate subset is denoted as $\mathcal{I}$, where we have $\mathcal{I} = \mathcal{Q}_c \subset \mathcal{Q}$, and the APM symbols can be expressed as $\boldsymbol{K}_d = [K_d(0),\ldots,K_d(M_d-1)]^T \in \mathcal{F}^{M_d\times 1}$. For the sake of demonstration, all the candidates of the ST codeword vectors can be expressed as a specifically designed codebook, yielding

$$\mathcal{B} = \left\{\boldsymbol{B}_1,\ldots,\boldsymbol{B}_{2^L} : \boldsymbol{B}_i \in \mathbb{C}^{QM_d}, i=1,\ldots,2^L\right\}, \tag{21}$$

which will be discussed in Section IV, while $\boldsymbol{K}$ is a vector having elements selected from $\mathcal{B}$. Based on (19), the received symbol vector $\tilde{\boldsymbol{y}}$ for a given $\boldsymbol{K}$ follows the Gaussian probability density function (PDF) of

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{K}) = \frac{1}{(\pi N_0)^{M_dN_rT_c}}\exp\left(-\frac{||\tilde{\boldsymbol{y}}-\boldsymbol{C}\boldsymbol{K}||^2}{N_0}\right). \tag{22}$$

Finally, the attainable rate can be formulated as

$$R = \frac{L}{NMT_c} = \frac{\log_2(VQ)}{T_c} \text{ bits/s/Hz}. \tag{23}$$

III. Multiuser Detection in STSK-OTFS-MA Systems

In this section, we first introduce the optimal maximum *a posteriori* detector (MAPD) or MLD designed for our STSK-OTFS-MA system. Typically, the complexity of the optimal MLD becomes excessive even for moderate constellations. To mitigate this problem, a pair of detectors are proposed. Specifically, we commence by deriving a low-complexity PRCGD, and then we detail the IRCD. Finally, the complexity analysis of the proposed detectors is provided.

A. Maximum A Posteriori Detector (MAPD)

Given the conditional PDF in (22), the optimum MAPD maximizes the *a posteriori* probability of the equivalent transmitted vector $\boldsymbol{K}$, yielding $\boldsymbol{K}^{\mathrm{MAP}} = \arg\max_{\boldsymbol{B}_i \in \mathcal{B}} \{p(\boldsymbol{B}_i|\tilde{\boldsymbol{y}})\}$. Assuming that the mapping process of different candidates in $\mathcal{B}$ is independent and equiprobable, the MAPD is equivalent to the MLD, which can be written as

$$\boldsymbol{K}^{\mathrm{ML}} = \arg\min_{\boldsymbol{B}_i \in \mathcal{B}} \left\{ ||\tilde{\boldsymbol{y}} - \boldsymbol{C}\boldsymbol{B}_i||^2 \right\}. \tag{24}$$

B. Progressive Residual Check Greedy Detector (PRCGD)

We firstly rewrite the input-output relationship of (19) as

$$\tilde{\boldsymbol{y}} = \boldsymbol{C}\boldsymbol{K} + \tilde{\boldsymbol{n}} = \boldsymbol{C}\boldsymbol{\Upsilon}_{\mathcal{I}}\boldsymbol{K}_d + \tilde{\boldsymbol{n}} = \boldsymbol{C}_{\mathcal{I}}\boldsymbol{K}_d + \tilde{\boldsymbol{n}}, \tag{25}$$

where $\boldsymbol{C}_{\mathcal{I}} = \boldsymbol{C}\boldsymbol{\Upsilon}_{\mathcal{I}} \in \mathbb{C}^{M_d N_r T_c \times M_d}$ and $\boldsymbol{\Upsilon}_{\mathcal{I}}$ is a $(QM_d \times M_d)$-element mapping matrix associated with $\mathcal{I}$. Assuming that the candidates in the index candidate sets $\mathcal{Q}$ and in the constellation set $\mathcal{F}$ are independent and equiprobable, similar to (24), the corresponding joint MLD can be formulated as $\left(\mathcal{I}^{\mathrm{ML}}, \boldsymbol{K}_d^{\mathrm{ML}}\right) = \arg\min_{\mathcal{Q}_c \subset \mathcal{Q}, \boldsymbol{f} \in \mathcal{F}^{M_d}} \left\{ \|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{Q}_c}\boldsymbol{f}\|^2 \right\}$. Since the complexity of the MLD is excessive, we propose the PRCGD. Based on the sparse structure of the transmitted symbol vector $\boldsymbol{K}$, our objective is to harness the philosophy of greedy algorithms [26], which are often employed for sparse recovery in low-complexity multiuser detection (MUD). Specifically, the proposed PRCGD can provide locally optimal detection results based on the elements of $\boldsymbol{K}$. Moreover, our PRCGD includes the reliability sorting and progressive detection stages detailed below.

At the reliability sorting stage, the received symbol vector $\tilde{\boldsymbol{y}}$ is first processed by linear minimum mean square error (LMMSE) estimation to obtain the soft estimates of $\boldsymbol{K}$ as $\tilde{\boldsymbol{K}} = \left(\boldsymbol{C}^H\boldsymbol{C} + \frac{1}{\gamma_s}\boldsymbol{I}_{QM_d}\right)^{-1}\boldsymbol{C}^H\tilde{\boldsymbol{y}}$, where $\tilde{\boldsymbol{K}} = [\tilde{K}(0), \ldots, \tilde{K}(QM_d - 1)]^T \in \mathbb{C}^{QM_d \times 1}$ and $\gamma_s = \gamma/Q$ is the generalized average SNR per symbol. The elements in $\tilde{\boldsymbol{K}}$ having relatively high magnitudes should also have relatively high probabilities of being active in $\boldsymbol{K}$, which becomes more pronounced in the high-SNR region. Hence, inspired by [26], [27], we can order the magnitudes of the elements in $\tilde{\boldsymbol{K}}$ in descending order to reflect the reliability of the index symbols, yielding

$$\mathcal{J} = \{j_1, \ldots, j_{QM_d}\} \quad \text{subject to} \quad \left|\tilde{K}(j_1)\right|^2 \geqslant \ldots \geqslant \left|\tilde{K}(j_{QM_d})\right|^2, \tag{26}$$

where we have $j_l \in \{1, \ldots, QM_d\}$ for $l = 1, \ldots, QM_d$ and $j_l \neq j_q, \forall l \neq q$. Then, based on the reliability set $\mathcal{J}$, the PRCGD enters the progressive detection stage, which is detailed below.

During the progressive detection stage, the PRCGD carries out the index symbol detection and the symbol-wise APM symbol detection separately. To elaborate further, we first select j_t from the reliability reliability set $\mathcal{J}$ and exploit C_t DAPs in the tth iteration, yielding $\mathcal{Q}^t = \left\{\mathcal{Q}_1^t, \ldots, \mathcal{Q}_{C_t}^t\right\} \subset \mathcal{Q}$, where $\mathcal{Q}_{c_t}^t(m_d) \in \mathbb{Z}_+^{QM_d}$ for $m_d = 0, \ldots, M_d - 1$ and $c_t = 1, \ldots, C_t$. Based on the reliability set $\mathcal{J}$, the DAPs are chosen under the constraint that j_t is a common value in all selected subsets, i.e., we have $\bigcap_{c_t=1}^{C_t} \mathcal{Q}_{c_t}^t = j_t$. Upon invoking the DAPs $\mathcal{Q}_{c_t}^t$ as the *a priori* information, the ensuing APM symbol estimation can be formulated as the following optimization problem of $\hat{\boldsymbol{K}}_{c_t,d} = \arg\min_{\boldsymbol{a} \in \mathbb{C}^{M_d \times 1}} ||\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}\boldsymbol{a}||^2$. Then the corresponding least square solution can be formulated as

$$\begin{aligned}\hat{\boldsymbol{K}}_{c_t,d} &= \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}^{\dagger}\tilde{\boldsymbol{y}} = \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}^{\dagger}\boldsymbol{C}_{\mathcal{I}}\boldsymbol{K}_d + \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}^{\dagger}\tilde{\boldsymbol{n}} \\ &= \boldsymbol{K}_d + \boldsymbol{r}_{\mathcal{Q}_{c_t}^t,\mathcal{I}} + \bar{\boldsymbol{n}},\end{aligned} \tag{27}$$

where we have the residual interference $\boldsymbol{r}_{\mathcal{Q}_{c_t}^t,\mathcal{I}} = \boldsymbol{0}$ in the case that all the index symbols are detected correctly, i.e., $\mathcal{Q}_{c_t}^t = \mathcal{I}$, and $\bar{\boldsymbol{n}} = \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}^{\dagger}\tilde{\boldsymbol{n}}$ is the corresponding AWGN vector. Based on the DAPs $\mathcal{Q}_{c_t}^t$, the estimates of the APM symbols $\boldsymbol{f}_{c_t}^t = [f_{c_t}^t(0), \ldots, f_{c_t}^t(M_d - 1)]^T$ can be obtained by symbol-wise ML detection, yielding,

$$f_{c_t}^t(m_d) = \arg\min_{f_v \in \mathcal{F}} \left|\hat{K}_{c_t,d}(j) - f_v\right|^2, \tag{28}$$

for $m_d = 0, \ldots, M_d - 1$ and $v = 1, \ldots, V$. Hence, after testing all the C_t DAPs, the PRCGD delivers the corresponding APM candidate sets $\hat{\mathcal{F}}^t = \{\boldsymbol{f}_1^t, \ldots, \boldsymbol{f}_{C_t}^t\}$.

Algorithm 1 Progressive Residual Check Greedy Detector

Require: $\tilde{\boldsymbol{y}}$, $\boldsymbol{C}$, $\mathcal{Q}$, γ_s and ϵ_0.
Preparation: Set the maximum number of iteration T_1, $\epsilon_\infty = \infty$, $\mathcal{I}^{\mathrm{PRCGD}} = \emptyset$ and $\boldsymbol{K}_d^{\mathrm{PRCGD}} = \emptyset$.
//Reliability Sorting:
Employ LMMSE detection according as
$\tilde{\boldsymbol{K}} = \left(\boldsymbol{C}^H\boldsymbol{C} + \frac{1}{\gamma_s}\boldsymbol{I}_{QM_d}\right)^{-1}\boldsymbol{C}^H\tilde{\boldsymbol{y}}$.
Obtain the measurements of the index reliabilities as
$\mathcal{J} = \{j_1, \ldots, j_{QM_d}\}$ subject to $\left|\tilde{K}(j_1)\right|^2 \geqslant \ldots \geqslant \left|\tilde{K}(j_{QM_d})\right|^2$.
//Progressive Detection:
for $t = 1$ to T_1 **do**
  Collect the DAPs $\mathcal{Q}^t = \left\{\mathcal{Q}_1^t, \ldots, \mathcal{Q}_{C_t}^t\right\}$, where $\bigcap_{c_t=1}^{C_t} \mathcal{Q}_{c_t}^t = j_t$.
  if $\epsilon^t < \epsilon_0$ **then**
    break
  else
    for $c_t = 1$ to C_t **do**
      Compute the least square solution as:
      $\hat{\boldsymbol{K}}_{c_t,d} = \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}^{\dagger}\tilde{\boldsymbol{y}}$
      Carry out APM symbol estimation as
      $f_{c_t}^t(j) = \arg\min_{f_v \in \mathcal{F}} \left|\hat{K}_{c_t,d}(j) - f_v\right|^2$, for $m_d = 0, \ldots, M_d - 1$.
    end for
    Obtain the local optimal set as
    $(\mathcal{I}^t, \boldsymbol{K}_d^t) = \arg\min_{\mathcal{Q}_{c_t}^t \subset \mathcal{Q}^t, \boldsymbol{f}_{c_t}^t \in \hat{\mathcal{F}}^t} \left\{ \left\|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{Q}_{c_t}^t}\boldsymbol{f}_{c_t}^t\right\|^2 \right\}$.
    Compute the residual error as
    $\epsilon^t = \left\|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{I}^t}\boldsymbol{K}_d^t\right\|^2$.
    if $\epsilon^t < \epsilon_0$ **then**
      $\mathcal{I}^{\mathrm{PRCGD}} = \mathcal{I}^t$ and $\boldsymbol{K}_d^{\mathrm{PRCGD}} = \boldsymbol{K}_d^t$
      break
    else if $\epsilon^t < \epsilon_\infty$ **then**
      $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \mathcal{Q}^t$, $\epsilon_\infty = \epsilon^t$, $\mathcal{I}^{\mathrm{PRCGD}} = \mathcal{I}^t$ and $\boldsymbol{K}_d^{\mathrm{PRCGD}} = \boldsymbol{K}_d^t$
    else
    end if
  end if
end for
Output $\mathcal{I}^{\mathrm{PRCGD}} = \mathcal{I}^t$ and $\boldsymbol{K}_d^{\mathrm{PRCGD}} = \boldsymbol{K}_d^t$.

Now we have obtained the estimates of the APM symbols and the DAPs grouped as $\{\boldsymbol{f}_{c_t}^t, \mathcal{Q}_{c_t}^t\}_{c_t=1}^{C_t}$. Therefore, the locally optimal set can be formulated as $(\mathcal{I}^t, \boldsymbol{K}_d^t) = \underset{\mathcal{Q}_{c_t}^t \subset \mathcal{Q}^t, \boldsymbol{f}_{c_t}^t \in \hat{\mathcal{F}}^t}{\arg\min} \left\{ \left\| \tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{Q}_{c_t}^t} \boldsymbol{f}_{c_t}^t \right\|^2 \right\}$, where the residual error can be expressed as $\epsilon^t = \|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{I}^t} \boldsymbol{K}_d^t\|^2$. Typically, the progressive detection terminates in the case of $\epsilon^t < \epsilon_0$, where ϵ_0 is the predefined termination parameter. However, if this condition cannot be satisfied after testing all DAPs, the PRCGD returns the corresponding set with minimum residual error. The proposed PRCGD is summarized in Algorithm 1.

C. Iterative Reduced-Space Check Detector (IRCD)

The main philosophy of IRCD is to carry out the detection of DAPs and APM symbols separately. Then near-ML detection performance can be obtained by only testing a reduced set of the entire DAP space $\mathcal{Q}$. In contrast to the PRCGD, the index reliabilities of all elements in $\tilde{\boldsymbol{K}}$ are invoked in the IRCD. Explicitly, the IRCD first derives the index reliability metric based on each DAP $\mathcal{Q}_c$ after employing the LMMSE detection, yielding $\rho_c = \sum_{m_d=0}^{M_d-1} \left| \tilde{K}(i_{m_d}^c) \right|^2$ for $i_{m_d}^c = \mathcal{Q}_c(m_d) \in \{0, \ldots, QM_d - 1\}$ and $c = 1, \ldots, C$. Then the reliability metrics of all $C = Q^{M_d}$ DAPs are sorted in descending order, which can be formulated as

$$\mathcal{R} = \{i_1, \ldots, i_C\} \quad \text{subject to } \rho_{i_1} \geqslant \ldots \geqslant \rho_{i_C}, \tag{29}$$

where we have $i_c \in \{1, \ldots, C\}$ for $c = 1, \ldots, C$ and $\rho_{i_p} \neq \rho_{i_q}, \forall p \neq q$. Similar to the PRCGD, the DAP associated with a higher value ρ_{i_c} in $\mathcal{R}$ can be regarded as the correct result with a higher probability, especially in the high-SNR scenarios. It should be noted that our IRCD first tests the DAPs corresponding to ρ_{i_1} with the highest priority in the following stage, where we carry out the detection of the APM symbols, as discussed below.

In the second stage, the proposed IRCD first selects the DAP $\mathcal{I}^t$ based on the reliability metric ρ_{i_t} in the tth iteration. Consequently, the detected APM symbols $\boldsymbol{K}_d^t \in \mathcal{F}^{M_d}$ can be obtained based on the least square approach and the symbol-wise ML detection of (27) and (28), respectively. Hence, the detected set and APM symbols in the tth iteration can be grouped as $\mathcal{G}^t = \{\mathcal{I}^t, \boldsymbol{K}_d^t\}$. Assume that there are T_2 DAPs to be tested during the second stage. Hence, the final detected DAP and the APM symbols can be obtained as $(\mathcal{I}^{\text{IRCD}}, \boldsymbol{K}_d^{\text{IRCD}}) = \underset{(\mathcal{I}^t, \boldsymbol{K}_d^t) \subset \mathcal{G}}{\arg\min} \|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{I}^t} \boldsymbol{K}_d^t\|^2$, where $\mathcal{G} = \{\mathcal{G}^1 \cup \mathcal{G}^2 \ldots \cup \mathcal{G}^{T_2}\}$. Our proposed IRCD is summarized in Algorithm 2.

D. Detection Complexity Analysis

It can be concluded from (19) that the MLD evaluates all the M_d STSK blocks. Hence the complexity of the optimum MLD is on the order of $\mathcal{O}[(VQ)^{M_d}]$, which is excessive for high values of M_d.

Based on our analysis in Section III-B, it can be readily shown that the complexity of the PRCGD relies on the number of iterations and the value of C_t. In more detail, we have the best scenario, if the PRCGD terminates after the first iteration and only a single DAP is considered. Therefore, the corresponding complexity is given by $\mathcal{O}(M_d V)$, since only the symbol-wise detection of (28) is employed. By contrast, the worst-case scenario is when all the DAPs in $\mathcal{Q}$ are tested. Since there are $C = Q^{M_d}$ DAPs, the overall complexity is on the order of $\mathcal{O}(Q^{M_d} M_d V)$. In more general cases, our PRCGD only has to consider a subset of $\mathcal{Q}$, having $C_1 < C$ DAPs. Under this condition, the complexity of the PRCGD is on the order of $\mathcal{O}(C_1 M_d V)$.

Based on Section III-C, the complexity of each IRCD iteration is on the order of $\mathcal{O}(M_d V)$. Hence, the overall complexity of IRCD is given by $\mathcal{O}(T_2 M_d V)$, and the worst scenario happens when the IRCD terminates after $T_2 = Q^{M_d}$ iterations, i.e., all the DAPs are tested. However, since the IRCD measures the index reliabilities of all DAPs, near-ML detection performance can be achieved by only testing a small subset of the entire DAP set, which is shown in Section V. Therefore, it can be readily demonstrated that the complexity of our IRCD can be significantly lower than that of the MLD.

Algorithm 2 Iterative Reduced-space Check Detector

Require: $\tilde{\boldsymbol{y}}$, $\boldsymbol{C}$, $\mathcal{Q}$ and γ_s.
Preparation: Set the maximum number of iteration T_2, and $\mathcal{G} = \emptyset$.
//Reliability Sorting:
Employ LMMSE detection as
$\tilde{\boldsymbol{K}} = \left(\boldsymbol{C}^H \boldsymbol{C} + \frac{1}{\gamma_s} \boldsymbol{I}_{QM_d}\right)^{-1} \boldsymbol{C}^H \tilde{\boldsymbol{y}}$.
Compute the index reliability metrics based on DAPs $\mathcal{Q}_c$ as
$\rho_c = \sum_{m_d=0}^{M_d-1} \left| \tilde{K}(i_{m_d}^c) \right|^2$ for $i_{m_d}^c \in \{0, \ldots, QM_d - 1\}$ and $c = 1, \ldots, C$.
Obtain the measurements of the index reliabilities as
$\mathcal{R} = \{i_1, \ldots, i_C\}$ subject to $\rho_{i_1} \geqslant \ldots \geqslant \rho_{i_C}$.
//Reduced-space Check Detection:
for $t = 1$ to T_2 **do**
    Collect the DAP $\mathcal{L}^t$ according to ρ_{i_t}.
    Carry out APM symbol estimation $\boldsymbol{K}_d^t$ based on (27) and (28).
    Obtain the detected results as $\mathcal{G}^t = \{\mathcal{I}^t, \boldsymbol{K}_d^t\}$.
    Calculate the residual error as
    $\epsilon^t = \left\|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{I}^t} \boldsymbol{K}_d^t\right\|^2$.
    Group the detection candidate sets as $\mathcal{G} = \mathcal{G} \cup \mathcal{G}^t$.
end for
Compute the final detection results as
$(\mathcal{I}^{\text{IRCD}}, \boldsymbol{K}_d^{\text{IRCD}}) = \underset{(\mathcal{I}^t, \boldsymbol{K}_d^t) \subset \mathcal{G}}{\arg\min} \left\|\tilde{\boldsymbol{y}} - \boldsymbol{C}_{\mathcal{I}^t} \boldsymbol{K}_d^t\right\|^2$.
return $\mathcal{I}^{\text{IRCD}}$ and $\boldsymbol{K}_d^{\text{IRCD}}$.

E. System Complexity Analysis

The system complexity of the SM/STSK-based schemes is identical to the corresponding MLD complexity [1], hence we provide our system complexity analysis in this section. We commence by specifying the SIMO-OTFS, SM-OTFS and STSK-OFDM-MA systems with the aid of their parameters as (N_r, V), (N_t, N_r, V) and (N_t, N_r, T_c, Q, V), respectively. Since all the combinations of the symbols have to be evaluated in the MLD, the complexity of SIMO-OTFS (N_r, V) and SM-OTFS (N_t, N_r, V) is on the order of $\mathcal{O}(V^{MN})$ and $\mathcal{O}[(N_t V)^{MN}]$ [25], respectively.

The STSK-aided OFDM-MA (STSK-OFDM-MA) (N_t, N_r, T_c, Q, V) system can be viewed as a special case of STSK-OTFS-MA (N_t, N_r, T_c, Q, V) associated with $N = 1$. Therefore, the corresponding system complexity can be expressed as $\mathcal{O}[(QV)^M]$.

IV. Performance Analysis of the Single-User System and Dispersion Matrix Design

In this section, we commence with the BER analysis of the single-user STSK-OTFS-MA system and derive its performance upper-bound, referred to as the SU-UPEP. Then the diversity order and ST coding gain are derived. Moreover, the single-user DCMC capacity of the STSK-OTFS-MA system is discussed. Finally, based on the SU-UPEP and the DCMC capacity, we gain deeper insights into the criteria of DM design in Section IV-D, and the algorithm for optimizing the DMs is derived.

A. Analysis of Single-User Bit Error Ratio Performance

The vector-form input-output relationship in (12) can be expressed as $\left(\boldsymbol{y}^{(u)}_{n_r,n_t,t_c}\right)^T = \breve{\boldsymbol{h}}^{(u)}_{n_r,n_t}\breve{\boldsymbol{X}}^{(u)}_{n_t,t_c} + \left(\boldsymbol{n}^{(u)}_{n_r,n_t,t_c}\right)^T$, where $\breve{\boldsymbol{h}}^{(u)}_{n_r,n_t} = \left[\breve{h}^{(u)}_{n_r,n_t}(1),\ldots,\breve{h}^{(u)}_{n_r,n_t}(P)\right]$ with $\breve{h}^{(u)}_{n_r,n_t}(i) = h^{(u)}_{i,n_r,n_t}e^{-j2\pi\nu_i\tau_i}, \forall i$, and the m_dth column of $\breve{\boldsymbol{X}}^{(u)}_{n_t,t_c} \in \mathbb{C}^{P\times M_d}$ can be obtained as

$$\breve{\boldsymbol{X}}^{(u)}_{n_t,t_c}[:,m_d] = \begin{bmatrix} x^{(u)}_{n_t,t_c}([k-k_1]_N + N[l-l_1]_M) \\ \vdots \\ x^{(u)}_{n_t,t_c}([k-k_P]_N + N[l-l_P]_M) \end{bmatrix}, \quad (30)$$

where we have $m_d = k + Nl$ for $k = 0,\ldots,N-1$ and $l = 0,\ldots,M-1$. Similar to (13), the n_rth received codeword vector within the t_cth OTFS time-slot can be formulated as

$$\boldsymbol{y}^T_{n_r,t_c} = \sum_{u=0}^{U-1}\sum_{n_t=0}^{N_t-1}\breve{\boldsymbol{h}}^{(u)}_{n_r,n_t}\breve{\boldsymbol{X}}^{(u)}_{n_t,t_c} + \boldsymbol{n}^T_{n_r,t_c} = \sum_{u=0}^{U-1}\breve{\boldsymbol{h}}^{(u)}_{n_r}\breve{\boldsymbol{X}}^{(u)}_{t_c} + \boldsymbol{n}^T_{n_r,t_c}, \quad (31)$$

where we have $\breve{\boldsymbol{h}}^{(u)}_{n_r} = \left[\breve{\boldsymbol{h}}^{(u)}_{0,n_r},\ldots,\breve{\boldsymbol{h}}^{(u)}_{N_t-1,n_r}\right]$ and $\breve{\boldsymbol{X}}^{(u)}_{t_c} = \text{sta}\{\breve{\boldsymbol{X}}^{(u)}_{n_t,t_c}\}|^{N_t-1}_{n_t=0}$. Therefore, the end-to-end input-output relationship of the t_cth time-slot is given by $\breve{\boldsymbol{Y}}_{t_c} = \sum_{u=0}^{U-1}\breve{\boldsymbol{H}}^{(u)}\breve{\boldsymbol{X}}^{(u)}_{t_c} + \breve{\boldsymbol{n}}_{t_c}$ for $t_c = 0,\ldots,T_c-1$, where $\breve{\boldsymbol{Y}}_{t_c} = [\boldsymbol{y}_{0,t_c},\ldots,\boldsymbol{y}_{N_r-1,t_c}]^T$ is the matrix of received signal, $\breve{\boldsymbol{n}}_{t_c} = [\boldsymbol{n}_{0,t_c},\ldots,\boldsymbol{n}_{N_r-1,t_c}]^T$ denotes the noise matrix, and the channel matrix can be expressed as $\breve{\boldsymbol{H}}^{(u)} = \text{sta}\{\breve{\boldsymbol{h}}^{(u)}_{n_r}\}|^{N_r-1}_{n_r}$. Finally, by defining $\boldsymbol{Y} = \left[\breve{\boldsymbol{Y}}_0,\ldots,\breve{\boldsymbol{Y}}_{T_c-1}\right]$, $\breve{\boldsymbol{X}}^{(u)} = \left[\breve{\boldsymbol{X}}^{(u)}_0,\ldots,\breve{\boldsymbol{X}}^{(u)}_{T_c-1}\right]$ and $\breve{\boldsymbol{n}} = [\breve{\boldsymbol{n}}_0,\ldots,\breve{\boldsymbol{n}}_{T_c-1}]$, the input-output relationship for the entire transmitted frame can be formulated as

$$\boldsymbol{Y} = \sum_{u=0}^{U-1}\breve{\boldsymbol{H}}^{(u)}\breve{\boldsymbol{X}}^{(u)} + \breve{\boldsymbol{n}} = \boldsymbol{H}\boldsymbol{X} + \breve{\boldsymbol{n}}, \quad (32)$$

where we have $\boldsymbol{X} = \text{sta}\{\breve{\boldsymbol{X}}^{(u)}\}|^{U-1}_{u=0}$ and $\boldsymbol{H} = \left[\breve{\boldsymbol{H}}^{(u)},\ldots,\breve{\boldsymbol{H}}^{(U-1)}\right]$. In a single-user scenario, we have $\boldsymbol{H} \in \mathbb{C}^{N_r\times PN_t}$ and $\boldsymbol{X} \in \mathbb{C}^{PN_t\times M_dT_c}$. For notational simplicity, the index (u) is omitted in the rest of this section, since only a single user is considered. Therefore, the MLD associated with the input-output relationship shown in (32) can be formulated as $\boldsymbol{X}^{\text{ML}} = \arg\min_{\boldsymbol{D}_i\in\mathcal{D}}\left\{||\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{D}_i||^2\right\}$, where the equivalent candidate space $\mathcal{D}$ can be formulated based on $\mathcal{B}$ in (21) and the mapping relationship in (30), yielding $\mathcal{D} = \left\{\boldsymbol{D}_1,\ldots,\boldsymbol{D}_{2^L} : \boldsymbol{D}_i \in \mathbb{C}^{PN_t\times M_dT_c}, i = 1,\ldots,2^L\right\}$.

Let us consider the pairwise error event $\{\boldsymbol{X}^c \to \boldsymbol{X}^e\}$, where $\boldsymbol{X}^c = \boldsymbol{D}_i$ denotes the transmitted codeword matrix, while $\boldsymbol{X}^e = \boldsymbol{D}_j$, $\forall i \neq j$, represents the erroneous detection results of the MLD, i.e., we have $\boldsymbol{D}_i \neq \boldsymbol{D}_j$ for $\boldsymbol{D}_i, \boldsymbol{D}_j \in \mathcal{D}$. Let us define furthermore the error matrix space $\mathcal{E} = \{\boldsymbol{E} = \boldsymbol{D}_i - \boldsymbol{D}_j, \forall \boldsymbol{D}_i, \boldsymbol{D}_j \in \mathcal{D}, \forall i \neq j\}$. Then, the conditional PEP for a given channel matrix $\boldsymbol{H}$ is obtained as $P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H}) = P\left(||\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}^c||^2 \geqslant ||\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}^e||^2\right)$. Let us denote the elements of the corresponding matrices as $H(a,b)$, $X(b,c)$, $n(a,c)$ and $Y(a,c)$ for $a = 0,\ldots,N_r-1$, $b = 0,\ldots,M_dN_t-1$ and $c = 0,\ldots,M_dT_c-1$, respectively. Then, after some further algebraic simplifications, it can be shown that $P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H})$ may be written as

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H}) = P\left(\sum_{c=0}^{M_dT_c-1}\sum_{a=0}^{N_r-1}\Re\left\{n^*(a,c)\sum_{b=0}^{PN_t-1}z\right\} \geqslant \frac{1}{2}\sum_{c=0}^{M_dT_c-1}\sum_{a=0}^{N_r-1}\left|\sum_{b=0}^{PN_t-1}z\right|^2\right), \quad (33)$$

where $z = H(a,b)\left[X^e(b,c) - X^c(b,c)\right]$. Consequently, when defining the modified Euclidean distance between two codeword matrices $\boldsymbol{X}^e$ and $\boldsymbol{X}^c$ as $\delta(\boldsymbol{X}^c,\boldsymbol{X}^e) = \sum_{c=0}^{M_dT_c-1}\sum_{a=0}^{N_r-1}\left|\sum_{b=0}^{PN_t-1}z\right|^2 = ||\boldsymbol{H}(\boldsymbol{X}^e-\boldsymbol{X}^c)||^2$, and considering that $\sum_{c=0}^{M_dT_c-1}\sum_{a=0}^{N_r-1}\Re\left\{n^*(a,c)\sum_{b=0}^{PN_t-1}z\right\}$ in (33) is a complex-valued Gaussian random variable with zero mean and a variance of $||\boldsymbol{H}(\boldsymbol{X}^e-\boldsymbol{X}^c)||^2/2\gamma$, (33) can be further expressed as

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H}) = Q\left(\sqrt{\frac{\gamma\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}{2}}\right), \quad (34)$$

where $Q(x) = \frac{1}{\pi}\int_0^{\pi/2}\exp\left(-\frac{x^2}{2\sin^2\theta}d\theta\right)$ is the Gaussian Q-function with $x > 0$. Note that (34) can be alternatively represented as $P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H}) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\exp\left(-\frac{\gamma\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}{4\sin^2\theta}\right)d\theta$ [28]. By averaging the integration of $P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e|\boldsymbol{H})$ with respect to the distribution of $\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)$, the SU-UPEP can be formulated as [28]

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e) = \mathbb{E}_{\boldsymbol{H}}\left[\frac{1}{\pi}\int_0^{\frac{\pi}{2}}\exp\left(-\frac{\gamma\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}{4\sin^2\theta}\right)d\theta\right] = \frac{1}{\pi}\int_0^{\frac{\pi}{2}}\Gamma_{\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}\left(-\frac{\gamma}{4\sin^2\theta}\right)d\theta, \quad (35)$$

where $\Gamma_{\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}(t)$ is the moment generating function (MGF) of $\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)$. Moreover, based on the analysis in [29], $\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)$ can be formulated as $\delta(\boldsymbol{X}^c,\boldsymbol{X}^e) = \sum_{i=0}^{M_dT_c-1}||\boldsymbol{H}\left(\boldsymbol{X}^c[:,i] - \boldsymbol{X}^e[:,i]\right)||^2$
$= \tilde{\boldsymbol{h}}[\boldsymbol{I}_{N_r}\otimes\boldsymbol{R}]\tilde{\boldsymbol{h}}^H$, where the codeword difference matrix is defined as $\boldsymbol{R} = (\boldsymbol{X}^e-\boldsymbol{X}^c)(\boldsymbol{X}^e-\boldsymbol{X}^c)^H$ and $\tilde{\boldsymbol{h}} = (\text{vec}(\boldsymbol{H}^T))^T \in \mathbb{C}^{1\times PN_rN_t}$. Since the non-zero elements of $\tilde{\boldsymbol{h}}$ obeys zero mean and variance of $1/2P$ per real dimension Gaussian distribution

[23], we can derive the MGF $\Gamma_{\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}(t)$ based on the approach of [29], yielding

$$\Gamma_{\delta(\boldsymbol{X}^c,\boldsymbol{X}^e)}(t) = \det[\boldsymbol{I}_{PN_rN_t} - t(\boldsymbol{I}_{N_r} \otimes \boldsymbol{R})/P]^{-1}. \quad (36)$$

Upon substituting (36) into (35), the SU-UPEP can now be expressed as

$$\begin{aligned} &P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e) \\ &= \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\det\left(\boldsymbol{I}_{N_rPN_t} + \frac{\gamma}{4P\sin^2\theta}(\boldsymbol{I}_{N_r}\otimes\boldsymbol{R})\right)\right]^{-1} d\theta \\ &= \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\det\left(\boldsymbol{I}_{PN_t} + \frac{\gamma}{4P\sin^2\theta}\boldsymbol{R}\right)\right]^{-N_r} d\theta. \end{aligned} \quad (37)$$

Let us define $r = \text{rank}(\boldsymbol{R})$ and the nonzero eigenvalues of $\boldsymbol{R}$ as $\{\lambda_1, \ldots, \lambda_r\}$. Then (37) can be expressed as

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \left[\prod_{j=1}^{r}\left(1 + \frac{\lambda_j\gamma}{4P\sin^2\theta}\right)\right]^{-N_r} d\theta. \quad (38)$$

Finally, by leveraging the union bound technique, the average bit error ratio (ABER) of the single-user STSK-OTFS-MA system can be approximated as

$$P_e \approx \frac{1}{2^L L}\sum_{\boldsymbol{b}^c}\sum_{\boldsymbol{b}^e} D_b(\boldsymbol{b}^c,\boldsymbol{b}^e)P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e), \quad (39)$$

where $D_b(\cdot,\cdot)$ denotes the Hamming distance function between two bit sequences, while $\boldsymbol{b}^c$ and $\boldsymbol{b}^e$ are the corresponding binary representations of $\boldsymbol{X}^c$ and $\boldsymbol{X}^e$.

B. Diversity Order and Coding Gain

In (38) we have $\lambda_j\gamma/(4P\sin^2\theta) \geqslant \lambda_j\gamma/4P$. Hence, the upper-bound of $P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e)$ can be formulated as

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e) \leqslant \frac{1}{2}\left[\prod_{j=1}^{r}\left(1 + \frac{\lambda_j\gamma}{4P}\right)\right]^{-N_r}. \quad (40)$$

Moreover, for high SNRs ($\gamma \gg 1$), (40) can be formulated as

$$P_E(\boldsymbol{X}^c \to \boldsymbol{X}^e) \leqslant \frac{1}{2}\left[\left(\prod_{j=1}^{r}\lambda_j\right)^{1/r}\left(\frac{\gamma}{4P}\right)\right]^{-rN_r}, \quad (41)$$

where the exponent of the SNR is often referred to as the **diversity order** obtained by the MLD, which is

$$G_D = \min_{\forall \boldsymbol{E}\in\mathcal{E}} rN_r = \min_{\forall \boldsymbol{E}\in\mathcal{E}} \text{rank}(\boldsymbol{R})\cdot N_r, \quad (42)$$

and the maximum achievable diversity order is $G_{\text{D,max}} = \min\{PN_t, M_dT_c\}\cdot N_r$. Given the values of P and M_d, it can be observed that the maximum achievable diversity order depends on the settings of N_t and T_c. Although the diversity order can be increased by increasing the value of T_c in the case of $(PN_t > M_dT_c)$, the transmit diversity order cannot be further improved, if we increase M_dT_c beyond PN_t. In more detail, the system with a lower value of T_c may result in a higher transmission rate in (21) as well as a low computation complexity, since the dimension of DMs is lower.

The coding gain of the STSK-OTFS-MA system can be expressed as $G_C = \min_{\forall \boldsymbol{E}\in\mathcal{E}}\left(\prod_{j=1}^{r}\lambda_j\right)^{1/r}$. It can be inferred from (41) that the diversity order G_D dominates the decay-rate of the SU-UPEP as the value of SNR increases. Furthermore, the coding gains G_C determines the horizontal shift of the STSK-aided SU-UPEP curve from the benchmark SU-UPEP curve of $\frac{1}{2}(\gamma/4P)^{-G_D}$.

C. DCMC Capacity

Now we derive the DCMC capacity of the single-user STSK-OTFS-MA scheme. Based on (19) and (21), the single-user DCMC capacity can be formulated as [30]

$$\begin{aligned} C_{\text{DCMC}} = \frac{1}{M_dT_c}\max_{p(\boldsymbol{B}_1)\ldots p(\boldsymbol{B}_{2^L})}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_i)p(\boldsymbol{B}_i) \\ \times\log_2\left[\frac{p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_i)}{\sum_{j=1}^{2^L}p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_j)p(\boldsymbol{B}_j)}\right]d\tilde{\boldsymbol{y}}. \end{aligned} \quad (43)$$

where $p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_i)$ is given by (22), when assuming that $\boldsymbol{B}_i$ is transmitted. It should be noted that (43) is maximized under the condition that all candidate matrices in the space $\mathcal{B}$ are independent and equiprobable, i.e., $p(\boldsymbol{B}_i) = 1/2^L, \forall i$. Hence we have

$$\log_2\left[\frac{p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_i)}{\sum_{j=1}^{2^L}p(\tilde{\boldsymbol{y}}|\boldsymbol{B}_j)p(\boldsymbol{B}_j)}\right] = \log_2\left(2^L\right) - \log_2\sum_{j=1}^{2^L}\exp(\Psi_{i,j}), \quad (44)$$

where $\Psi_{i,j} = \gamma\left[-||\boldsymbol{C}(\boldsymbol{B}_i - \boldsymbol{B}_j) + \tilde{\boldsymbol{n}}||^2 + ||\tilde{\boldsymbol{n}}||^2\right]$ can be formulated by substituting (22) into (44). Therefore, based on the assumption that the codeword matrix candidates are transmitted at the same probabilities, the DCMC capacity of the single-user STSK-OTFS-MA system can be expressed as

$$C_{\text{DCMC}} = \frac{1}{M_dT_c}\left\{L - \frac{1}{2^L}\sum_{i=1}^{2^L}\mathbb{E}_{\boldsymbol{C}}\left[\log_2\sum_{j=1}^{2^L}\exp\left(\Psi_{i,j}\right)\right]\right\}. \quad (45)$$

D. Design of Dispersion Matrices

Based on the above analysis, let us now discuss the design of DMs. To obtain the best system performance, the DMs can be designed based either on the SU-UPEP of (35) or on the DCMC capacity of (45).

The asymptotic diversity order of the OTFS system is one at an infinite SNR, and higher diversity can be attained in finite-SNR scenarios [23]. However, similar to OFDM, OTFS also requires transmit precoding schemes to attain full diversity [23]. Therefore, the DMs of both STSK-OTFS-MA and STSK-OFDM-MA systems can be designed with the aid of our proposed Algorithm 3.

It can be shown that the above two methods lead to the same design, detailed as follows.

Proposition 1: The design of DMs aiming for minimizing the SU-UPEP in the high-SNR region of (41) and that aiming for maximizing the DCMC capacity of (45) lead to the same

results, which can be formulated as **maximizing** the following metrics:

$$\Lambda_{\mathrm{D}} = \min_{\forall \boldsymbol{E} \in \mathcal{E}} \mathrm{rank}(\boldsymbol{R}), \quad \Lambda_{\mathrm{C}} = \min_{\forall \boldsymbol{E} \in \mathcal{E}} \prod_{j=1}^{r} \lambda_j. \tag{46}$$

Proof: It can be readily shown that the elements of $\{\boldsymbol{Y}, \check{\boldsymbol{n}}\}$ in (32) are the interleaved versions of the elements of $\{\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{n}}\}$ in (19). Therefore, given a pairwise error event $\{\boldsymbol{X}^c \to \boldsymbol{X}^e\}$ and a SNR γ, we have $||\boldsymbol{C}(\boldsymbol{B}_i - \boldsymbol{B}_j)||^2 = ||\boldsymbol{H}(\boldsymbol{D}_i - \boldsymbol{D}_j)||^2$. Furthermore, similar to the derivation shown in Section IV-A, it can be readily shown that based on (45) and the MGF technique, maximizing the DCMC capacity is equivalent to minimizing

$$\mathbb{E}_{\boldsymbol{H}}\left[-\gamma\delta(\boldsymbol{X}^c, \boldsymbol{X}^e)\right] \leqslant \frac{1}{2}\left[\prod_{j=1}^{r}\left(1 + \frac{\lambda_j \gamma}{P}\right)\right]^{-N_r}, \tag{47}$$

which can be further upper-bounded by $\frac{1}{2}\left[\left(\prod_{j=1}^{r}\lambda_j\right)^{1/r}\left(\frac{\gamma}{P}\right)\right]^{-rN_r}$ under the condition of $\gamma \gg 1$. It may now be observed that (41) and (47) are in similar forms. Note that it is more important to maximize the diversity order G_D in (42) than to maximize the coding gain G_C, since it is the diversity order that dominates the slope of the ABER curve [26]. Since the value of N_r is fixed, we arrive at the criteria Λ_{D} in (46) by searching for the minimum rank value within the corresponding error matrix space $\mathcal{E}$. Furthermore, the coding gain is given by $\min_{\forall \boldsymbol{E} \in \mathcal{E}}\left(\prod_{j=1}^{r}\lambda_j\right)^{\frac{1}{\Lambda_{\mathrm{D,max}}}}$. Since $\Lambda_{\mathrm{D,max}}$ is a constant for $\forall \boldsymbol{E} \in \mathcal{E}$, the corresponding criteria Λ_{C} can be derived in (46). ■

Based on the above analysis, let us now delve into the details of designing the DM set. Let us assume that $\tilde{N}$ consecutive Monte Carlo simulations are employed. Let us define the DM set, design criteria, codeword difference matrix, error matrix, and error matrix space for the $\tilde{n}$th experiment as $\mathcal{A}_{\tilde{n}}$, $\Lambda_{\mathrm{D},\tilde{n}}$, $\Lambda_{\mathrm{C},\tilde{n}}$, $\boldsymbol{R}_{\tilde{n}}$, $\boldsymbol{E}_{\tilde{n}}$ and $\mathcal{E}_{\tilde{n}}$ for $\tilde{n} = 1, \ldots, \tilde{N}$, respectively. For the $\tilde{n}$th experiment, we first randomly generate the Q full-rank $(\bar{T} \times \bar{T})$-dimensional unitary matrices $\bar{\boldsymbol{A}}_{q,\tilde{n}}$ for $q = 1, \ldots, Q$, where $\bar{T} = \max\{N_t, T_c\}$. Then the DMs can be formulated as

$$\boldsymbol{A}_{q,\tilde{n}} = \begin{cases} \tilde{\boldsymbol{A}}_{q,\tilde{n}}[:, 1:T_c], & \text{if } N_t > T_c \\ \sqrt{\frac{T_c}{N_t}}\tilde{\boldsymbol{A}}_{q,\tilde{n}}[1:N_t, :], & \text{if } T_c > N_t, \end{cases} \tag{48}$$

where the constant $\sqrt{\frac{T_c}{N_t}}$ is used for satisfying the power constraint in (3), and we have $\mathcal{A}_{\tilde{n}} = \{\boldsymbol{A}_{1,\tilde{n}}, \ldots, \boldsymbol{A}_{Q,\tilde{n}}\}$. Moreover, the DM set of all $\tilde{N}$ simulations is denoted by $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_{\tilde{N}}\}$. Assuming that there are $\check{N}$ out of $\tilde{N}$ DM sets that maximize the diversity order, it can be readily shown from (46) that the candidates of the optimal DM set are obtained as $\check{\mathcal{A}} = \arg\max_{\mathcal{A}_{\tilde{n}} \subset \mathcal{A}} \Lambda_{\mathrm{D},\tilde{n}} = \arg\max_{\mathcal{A}_{\tilde{n}} \subset \mathcal{A}} \left\{\min_{\forall \boldsymbol{E}_{\tilde{n}} \in \mathcal{E}_{\tilde{n}}} \{\mathrm{rank}(\boldsymbol{R}_{\tilde{n}})\}\right\}$, where $\check{\mathcal{A}} = \left\{\check{\mathcal{A}}_1, \ldots, \check{\mathcal{A}}_{\check{N}}\right\}, \check{\mathcal{A}}_{\check{n}} \subset \mathcal{A}, \forall \check{n}$. Finally, the optimal DM set can be obtained as $\mathcal{A}^{\mathrm{opt}} = \arg\max_{\check{\mathcal{A}}_{\check{n}} \subset \check{\mathcal{A}}} \Lambda_{\mathrm{C},\check{n}} = \arg\max_{\check{\mathcal{A}}_{\check{n}} \subset \check{\mathcal{A}}} \left\{\min_{\forall \boldsymbol{E}_{\check{n}} \in \mathcal{E}_{\check{n}}} \prod_{j=1}^{r} \lambda_j\right\}$. The proposed DM design method is summarized in Algorithm 3.

Algorithm 3 Dispersion Matrix Design

Require: The values of Q, T_c, N_t and $\bar{T} = \max\{N_t, T_c\}$.
1: **Preparation**: Set $\tilde{N}$ as the number of Monte Carlo simulations.
2: **for** $\tilde{n} = 1$ to $\tilde{N}$ **do**
3: Randomly generate Q full-rank $(\bar{T} \times \bar{T})$-dimensional unitary matrices $\bar{\mathcal{A}}_{\tilde{n}} = \{\bar{\boldsymbol{A}}_{1,\tilde{n}}, \ldots, \bar{\boldsymbol{A}}_{Q,\tilde{n}}\}$.
4: Generate the DM set $\mathcal{A}_{\tilde{n}} = \{\boldsymbol{A}_{1,\tilde{n}}, \ldots, \boldsymbol{A}_{Q,\tilde{n}}\}$ based on (48) as
5: $\boldsymbol{A}_{q,\tilde{n}} = \begin{cases} \tilde{\boldsymbol{A}}_{q,\tilde{n}}[:, 1:T_c], & \text{if } N_t > T_c \\ \sqrt{\frac{T_c}{N_t}}\tilde{\boldsymbol{A}}_{q,\tilde{n}}[1:N_t, :], & \text{if } T_c > N_t, \end{cases}$
6: **end for**
7: Collect all DM matrices $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_{\tilde{N}}\}$.
8: Obtain $\check{N}$ out of $\tilde{N}$ DM sets that achieve the maximum diversity order as
9: $\check{\mathcal{A}} = \arg\max_{\mathcal{A}_{\tilde{n}} \subset \mathcal{A}} \Lambda_{\mathrm{D},\tilde{n}} = \arg\max_{\mathcal{A}_{\tilde{n}} \subset \mathcal{A}} \left\{\min_{\forall \boldsymbol{E}_{\tilde{n}} \in \mathcal{E}_{\tilde{n}}} \{\mathrm{rank}(\boldsymbol{R}_{\tilde{n}})\}\right\}$, where $\check{\mathcal{A}} = \left\{\check{\mathcal{A}}_1, \ldots, \check{\mathcal{A}}_{\check{N}}\right\}, \check{\mathcal{A}}_{\check{n}} \subset \mathcal{A}, \forall \check{n}$.
10: Generate the optimal DM set as
11: $\mathcal{A}^{\mathrm{opt}} = \arg\max_{\check{\mathcal{A}}_{\check{n}} \subset \check{\mathcal{A}}} \Lambda_{\mathrm{C},\check{n}} = \arg\max_{\check{\mathcal{A}}_{\check{n}} \subset \check{\mathcal{A}}} \left\{\min_{\forall \boldsymbol{E}_{\check{n}} \in \mathcal{E}_{\check{n}}} \prod_{j=1}^{r} \lambda_j\right\}$.
12: **return** $\mathcal{A}^{\mathrm{opt}}$.

V. Performance Results

In this section, we provide the simulation results for characterizing the overall performance of STSK-OTFS-MA systems. Unless specifically defined, $N = 4$ time intervals are considered for an OTFS subframe and the entire OTFS frame has $T_c = 2$ subframes while resource allocation scheme 1 is invoked. The DMs are obtained from Algorithm 3. The subcarrier spacing and carrier frequency are $\Delta f = 15$ kHz and $f_c = 4$ GHz. The normalized maximum Doppler and delay shifts are set to $k_{\max} = M - 1$ and $l_{\max} = N - 1$ [23], respectively. The normalized indices of delay and Doppler associated with the ith path are given by $l_i \in \mathcal{U}[0, l_{\max}]$ and $k_i \in \mathcal{U}[-k_{\max}, k_{\max}]$, respectively. Furthermore, the SIMO-OTFS and SM-OTFS systems are specified as (N_r, V) and (N_t, N_r, V), respectively. More explicitly, the SM-OTFS can also be viewed as a single-user STSK-OTFS-MA $(N_t, N_r, 1, V, Q = N_t)$ system, where $\boldsymbol{A}_q = \boldsymbol{I}_{N_t}[:, q]$ for $q = 1, \ldots, Q$ [1].

In Fig. 4, the single-user BER performance of STSK-OTFS-MA $(2, N_r, 2, 2, 2)$ systems detected by the MLD is compared to the corresponding BER upper-bound shown in (39), where $M = 8$ subcarriers are employed. Explicitly, the BER performance is evaluated with different numbers of RAs and channel paths. Based on the simulation results of Fig. 4, we have the following observations. Firstly, as the SNR γ increases, the analytical BER upper-bound approaches with the simulated BER. Secondly, given $P = 4$, the union-bound converges to the BER of MLD in the case of $\gamma > 18$ dB for $N_r = 1$, and $\gamma > 8$ dB for $N_r = 2$ in general. And again, the upper-bounds approach the simulated BER values in the moderate- and high-SNR scenarios for other values of P. Furthermore, it can be observed that all the upper-bounds are tight below 10^{-3} BER in general. Finally, the higher the values of P and N_r, the higher the BER performance gain achieved by our STSK-OTFS-MA system.

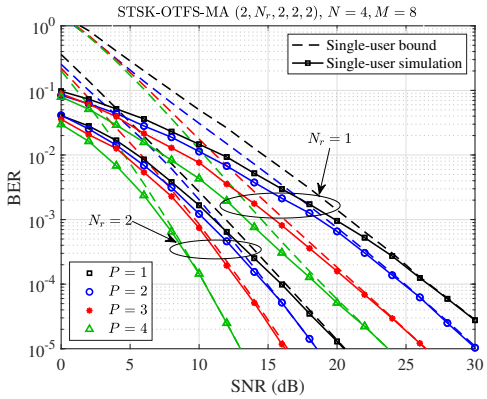

Fig. 4. Single-user BER performance using MLD and upper-bounds for STSK-OTFS-MA $(2, N_r, 2, 2, 2)$ systems with $N = 4$, $M = 8$ and $N_r = \{1, 2\}$, communicating over doubly-selective channels having the different number of paths at the same transmission rate of 1 bits/s/Hz. The upper-bounds are calculated based on (39).

This is because both frequency diversity and space diversity can be achieved by the proposed STSK-OTFS-MA system, and a higher diversity order can be attained as P or N_r increases, which are consistent with our analytical results in (42).

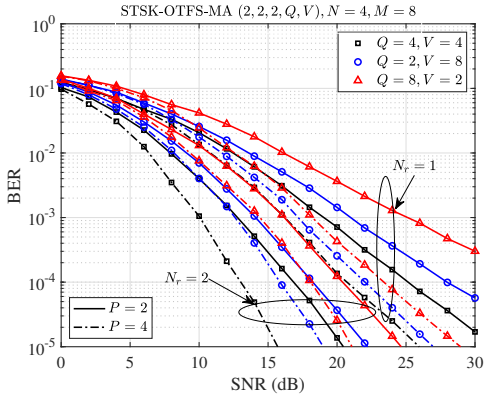

Fig. 5. Single-user BER performance employing MLD for STSK-OTFS-MA $(2, 2, 2, Q, V)$ systems with $P = \{2, 4\}$ and $N_r = \{1, 2\}$ but invoking a different number of DMs and modulation orders at the same transmission rate of 2 bits/s/Hz.

Fig. 5 depicts the single-user BER performance of STSK-OTFS-MA $(2, 2, 2, Q, V)$ systems for $P = \{2, 4\}$ and $N_r = \{1, 2\}$, where $M = 8$ subcarriers are employed and the effect of different combinations of $\{Q, V\}$ are investigated under the constraint of $R = 2$ bits/s/Hz. As shown in Fig. 5, for all BER curves specified by $P = \{2, 4\}$ and $N_r = \{1, 2\}$, the system exploiting $Q = 4$ and QPSK modulation is capable of achieving the best BER performance among the three sets of parameters $\{Q, V\}$ considered. This observation implies that for a given transmission rate R, an optimum combination of the parameters $\{Q, V\}$ can be found, which leads to the best BER performance. Finally, given the same other parameters used, we observe based on Fig. 4 and Fig. 5 that our STSK-OTFS-MA system is capable of attaining a better BER performance for relatively lower values of the parameters $\{Q, V\}$.

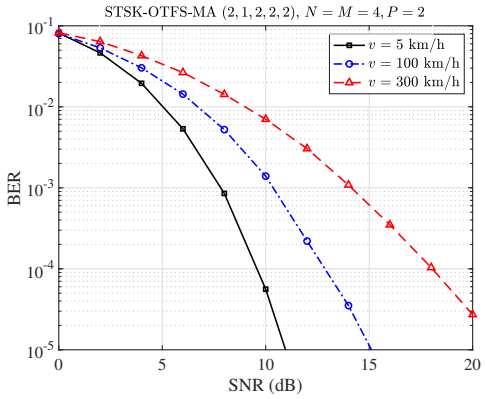

Fig. 6. BER performance of two-user STSK-OTFS-MA $(2, 1, 2, 2, 2)$ system using MLD with fractional delay and Doppler shifts under different mobile velocities.

In Fig. 6, the BER performance of the two-user STSK-OTFS-MA $(2, 1, 2, 2, 2)$ system using MLD with fractional delay and Doppler shifts are investigated in the case of different mobile speeds, where we have $N = 4$ and $M = 4$. It is observed that the BER performance degrades as a higher speed is encountered, which is owing to the higher Doppler frequency. Explicitly, at a BER of 10^{-4}, the $v = 5$ km/h scenario yields about 2.5 dB and 8 dB SNR gain compared to the $v = 100$ km/h and $v = 300$ km/h cases.

In Fig. 7, we investigate the multiuser BER performance of our STSK-OTFS-MA $(2, N_r, 2, 2, 2)$ system employing a different number of RAs. Observe from Fig. 7 that the BER performance degrades when the system supports more users. This is because a higher MUI is experienced as U increases. More specifically, it can be observed that the four-user system equipped with $N_r = 1$ RA suffers from about a 2 dB performance loss compared to the single-user system at a BER of 10^{-4}. The corresponding performance erosion is reduced to about 1 dB for $N_r = 2$. This trend is reminiscent of the channel hardening phenomenon of the classic massive MIMO uplink detection [31].

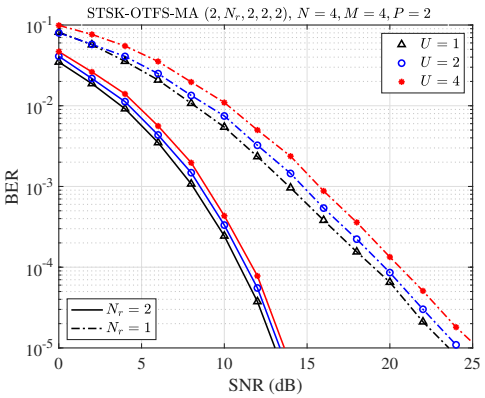

Fig. 7. BER performance utilizing MLD for STSK-OTFS-MA $(2, N_r, 2, 2, 2)$ systems with $N = 4$, $M = 4$ and $N_r = \{1, 2\}$ supporting a different number of users at the same transmission rate of $R = 1$ bits/s/Hz.

Fig. 8 compares the BER performance of conventional SIMO-OTFS, of SM-OTFS as well as of the single-user

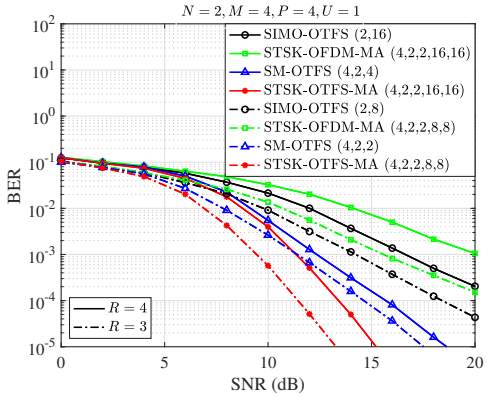

Fig. 8. BER performance of the conventional SIMO-OTFS scheme, the SM-OTFS scheme, the STSK-OFDM-MA scheme, and our STSK-OTFS-MA scheme both invoking MLD for the cases of the same transmission rate of $R = 3$ and $R = 4$ bits/s/Hz.

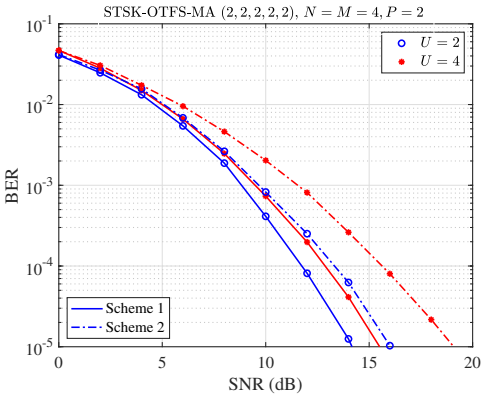

Fig. 9. BER performance of the STSK-OTFS-MA $(2, 2, 2, 2, 2)$ systems with $N = M = 4$ and supporting a different number of users by invoking proposed resource allocation schemes as shown in Fig. 3.

STSK-OFDM-MA and the proposed single-user STSK-OTFS-MA systems, where $R = 3$ and $R = 4$ bits/s/Hz are considered, and the same values of N_r are employed under the same constraint of R. Based on Fig. 8, we have the following observations. Firstly, for a given rate, the BER performance of SM-OTFS is better than that of the conventional SIMO-OTFS. This is because SM-OTFS may rely on a lower-order modulation scheme than SIMO-OTFS. Moreover, spatial diversity can be attained by SM-OTFS. Secondly, it is found that the proposed STSK-OTFS-MA schemes are capable of achieving better performance than the SM-OTFS systems in all the scenarios considered, resulting in about 4 dB gain at a BER of 10^{-5} for the rate of $R = 4$ bits/s/Hz, and about a 5 dB gain at a BER of 10^{-5} in the $R = 3$ bits/s/Hz scenario. This observation can be explained by our analytical results of Section IV-B. As an ST coding scheme, the proposed STSK-OTFS-MA system can also achieve time diversity in addition to space diversity and frequency diversity, yielding a higher diversity order than the other schemes. Additionally, the maximum coding gain can be achieved by taking full advantage of Algorithm 3. Therefore, our OTFS-STSK-MA scheme outperforms the SIMO-OTFS and SM-OTFS schemes. Furthermore, STSK-OFDM-MA attains the worst BER performance at a given rate, since the ICI introduced by high-mobility channels is ignored in STSK-OFDM-MA. Specifically, at a BER of 10^{-3}, our proposed STSK-OTFS-MA is capable of attaining about 6 dB and 9 dB SNR gains in the $R = 3$ and $R = 4$ bits/s/Hz scenarios, respectively. Finally, the BER performances of all systems improve, as the rate is reduced due to the lower values of Q and V used. This observation is consistent with the conclusions in [1].

Fig. 9 evaluates the BER performance of STSK-OTFS-MA $(2, 2, 2, 2, 2)$ systems supporting $U = 2$ and $U = 4$ users by exploiting the proposed resource allocation schemes shown in Fig. 3. Observe from Fig. 9 that the Scheme 1-based system is capable of attaining better BER performance than the system utilising Scheme 2. Moreover, the performance gap becomes wider when our STSK-OTFS-MA supports more users. Explicitly, at a BER of 10^{-5}, the two-user Scheme 1-based system attains about 2 dB SNR gain over its Scheme 2 counterpart, while the gain escalates to 4 dB, when supporting $U = 4$ users. This is because the MUI becomes higher as the number of supported users increases. Additionally, the efficiency of our MUI mitigation and resource allocation Scheme 1 is boldly illustrated.

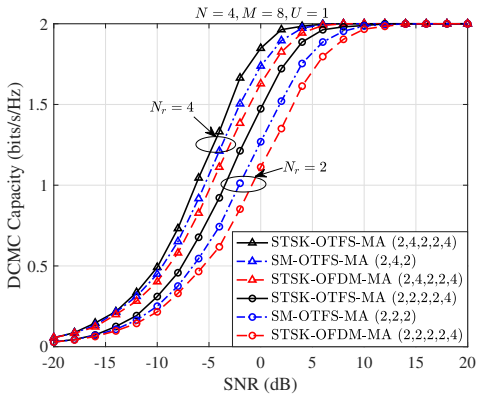

Fig. 10. The single-user DCMC capacity of the SM-OTFS $(2, N_r, 2)$, STSK-OFDM-MA $(2, N_r, 2, 2, 4)$ and our STSK-OTFS-MA $(2, N_r, 2, 2, 4)$ systems with different number of RAs.

In Fig. 10, the single-user DCMC capacities are investigated for both the SM-OTFS $(2, N_r, 2)$, for the STSK-OFDM-MA $(2, N_r, 2, 2, 4)$ and for our STSK-OTFS-MA $(2, N_r, 2, 2, 4)$ schemes at a given rate of $R = 2$ bits/s/Hz, where $N_r = 2$ and $N_r = 4$ are considered respectively. Based on Fig. 10, we have the following observations. Firstly, the asymptotic capacities of both the SM-OTFS, STSK-OFDM-MA and of the STSK-OTFS-MA systems are $R = 2$ bits/s/Hz, which is independent of the number of RAs, since the rate was limited to 2 bits/s/Hz. Furthermore, given a value of N_r, it is shown in Fig. 10 that the proposed STSK-OTFS-MA system always outperforms the SM-OTFS and STSK-OFDM-MA schemes. This observation can also be inferred from Fig. 8 and [1], since our STSK-OTFS-MA is capable of attaining extra time diversity and ST coding gains over the SM-OTFS scheme, while the capacity of STSK-OFDM-MA erodes due to the ICI.

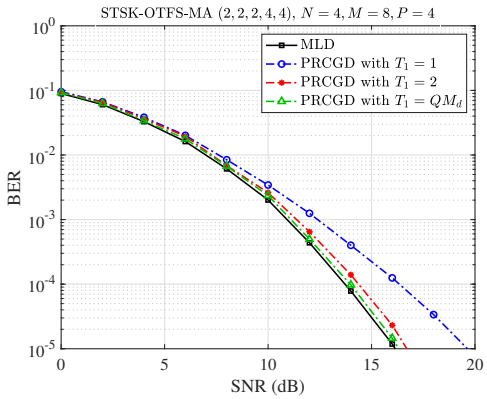

Fig. 11. Two-user BER performance of the STSK-OTFS-MA $(2,2,2,4,4)$ system using MLD and the proposed PRCGD with the different number of iterations operating at $R=2$ bits/s/Hz.

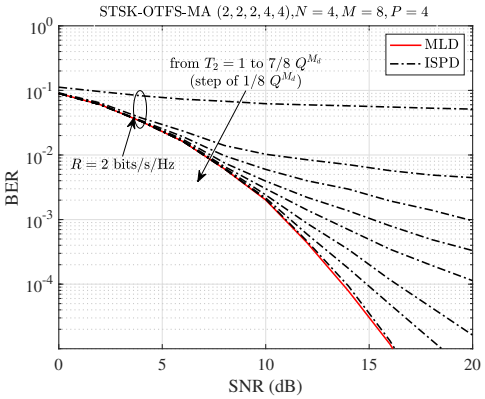

Fig. 12. Two-user BER performance of the STSK-OTFS-MA $(2,2,2,4,4)$ systems employing MLD and the proposed IRCD with the different number of iterations operating at $R=2$ bits/s/Hz.

Fig. 11 characterizes the BER performance of both the MLD and PRCGD conceived for the STSK-OTFS-MA $(2,2,2,4,4)$ system operating at $R=2$ bits/s/Hz. We observe from the results of Fig. 11 that the proposed PRCGD relying on two iterations is capable of attaining a BER performance close to that in the $T_1=Q^{M_d}$ case. Moreover, as depicted in Fig. 11, a near-ML BER performance is attainable when the PRCGD invokes as few as two iterations. It should be noted that T_1 denotes the upper-bound of the actual number of iterations in Algorithm 1.

The BER performance of our IRCD designed for STSK-OTFS-MA is characterized in Fig. 12, where the BER curve of MLD is depicted as the benchmark, while the other parameters are the same as those for Fig. 11. It should be emphasized that the IRCD improves a relatively low complexity by invoking an adequate number of T_2 than the MLD for the same system. As shown in Fig. 12, the higher the value of T_2, the better BER performance our STSK-OTFS-MA system becomes. Explicitly, in the cases of $T_2>5/8Q^{M_d}$, the IRCD is capable of achieving a better performance than the PRCGD with $T_1=1$, as shown in Fig. 11. Moreover, we can observe from Fig. 12 that at a BER of 10^{-4}, the IRCD with $T_2=6/8Q^{M_d}$ attains a gain of about 1.5 dB over the case of using $T_2=5/8Q^{M_d}$, while the IRCD associated with $T_2=7/8Q^{M_d}$ iterations can also achieve a gain of about 1.5 dB over the IRCD with $T_2=5/8Q^{M_d}$. Furthermore, Fig. 12 clearly shows that the IRCD with $T_2=7/8Q^{M_d}$ is capable of attaining nearly the same performance as the MLD. Therefore, from the above observations we conclude that the IRCD with $T_2=5/8Q^{M_d}$ to $T_2=7/8Q^{M_d}$ can be implemented to achieve a desirable BER performance in contrast to the PRCGD associated with $T_1=1$ and $T_1=2$, as shown in Fig. 11, while improving a considerably lower complexity than the MLD.

To further compare IRCD and PRCGD, in Fig. 13 we characterize the BER performance of these two detectors in three-user STSK-OTFS-MA $(2,2,2,2,4)$ systems, yielding a rate of $R=1.5$ bits/s/Hz. Specifically, the number of iterations is set to $T_1=2$ and $T_2=5/8Q^{M_d}$ for PRCGD and IRCD, respectively. Observe from Fig. 13 that the BER

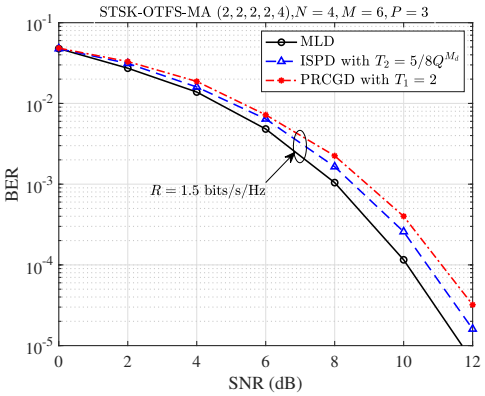

Fig. 13. Three-user BER performance of the STSK-OTFS-MA $(2,2,2,2,4)$ systems using MLD, our IRCD with $T_1=2$, and the proposed PRCGD with $T_2=5/8Q^{M_d}$ operating at $R=1.5$ bits/s/Hz.

performance of the PRCGD with $T_1=2$ iterations is about 0.5 dB and 1 dB worse than that of the IRCD with $T_2=5/8Q^{M_d}$ and the MLD, respectively. To elaborate further, the PRCGD needs 0.5 dB higher SNR than the IRCD to achieve the BER of 10^{-5}. Furthermore, it can be concluded that the ISPD associated with $T_2=5/8Q^{M_d}$ iterations obtain a good BER performance compared to the MLD, despite its lower complexity.

To illustrate the general flexibility of our STSK-OTFS-MA scheme, the BER performance of both uncoded, $1/2$-rate and $2/3$-rate LDPC coded STSK-OTFS-MA $(2,2,2,2,4)$ systems using MLD are evaluated in Fig. 14. All the remaining parameters are consistent with those in Fig. 13. In this context, the sum-product decoding algorithm is harnessed [32]. As observed in Fig. 14, the LDPC-coded system is capable of attaining an a substantial performance improvement compared to the conventional uncoded system. Moreover, at a BER of 10^{-5}, the $1/2$-rate LDPC coded system attains about 2 dB and 7 dB SNR gain compared to the $2/3$-rate LDPC coded and uncoded systems, respectively.

Fig. 15 portrays the corresponding computational complexity of the MLD, IRCD, and PRCGD employed in Fig. 13. We have the following observations based on Fig. 15. Firstly, the

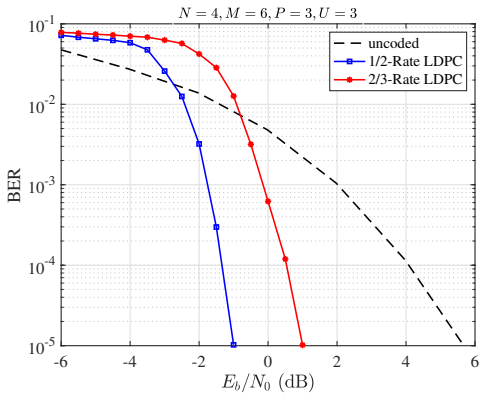

Fig. 14. BER performance of both the uncoded, rate-$1/2$ and rate-$2/3$ LDPC coded multiuser STSK-OTFS-MA $(2, 2, 2, 2, 4)$ systems invoking MLD.

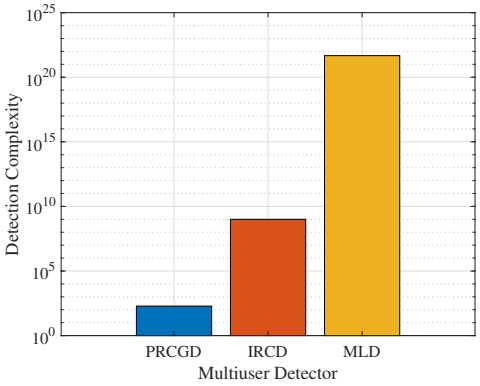

Fig. 15. Multiuser detection complexity of the STSK-OTFS-MA $(2, 2, 2, 2, 4)$ systems invoking MLD, our IRCD with $T_1 = 2$, and the proposed PRCGD with $T_2 = 5/8Q^{M_d}$ operating at $R = 1.5$ bits/s/Hz.

complexity of PRCGD with $T_1 = 2$ is much lower than that of IRCD and of MLD. This can be explained by the fact that our PRCGD tests each DAP uniquely, and repeated searches can be avoided, as illustrated in Algorithm 1. Moreover, the PRCGD employs simple symbol-based detection for APM symbols, whereas the MLD detects all the APM symbols jointly, which can be seen by comparing (24) and (28). Secondly, the IRCD with $T_2 = 5/8Q^{M_d}$ can provide about 12 orders of magnitude complexity reduction over MLD. Since the reliability sorting of all the DAPs is exploited, the full-research process of MLD can be avoided in the proposed IRCD, yielding a near-ML performance at a significantly lower complexity than the MLD. Finally, based on Fig. 13 and Fig. 15, it can be concluded that satismetricy BER performances can be attained by invoking both the PRCGD with $T_1 = 2$ and IRCD with $T_2 = 5/8Q^{M_d}$, which impose much lower complexity than that of the MLD. Furthermore, the PRCGD with $T_1 = 2$ attains a more attractive BER vs. complexity trade-off than the IRCD.

In Fig. 16, the system complexity of the conventional STSK-OFDM-MA, SIMO-OTFS, SM-OTFS and the proposed STSK-OTFS-MA employed in Fig. 8 is investigated. It is observed that STSK-OFDM-MA exhibits the lowest system complexity at a given rate among the OTFS-based systems, followed by the SIMO-OTFS and SM-OTFS paradigms. This

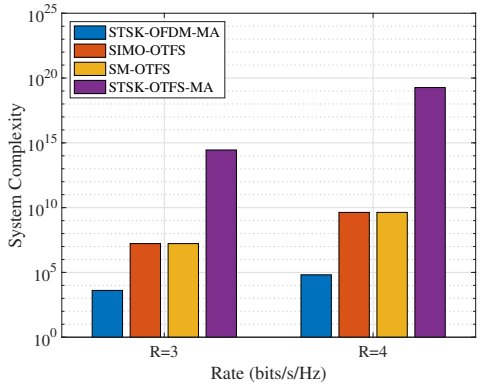

Fig. 16. System complexity of the conventional SIMO-OTFS scheme, the SM-OTFS scheme, the STSK-OFDM-MA scheme, and our STSK-OTFS-MA scheme invoked in Fig. 8 for a transmission rate of $R = 3$ and $R = 4$ bits/s/Hz.

is because $N = 1$ is invoked in the STSK-OFDM-MA scheme, which significantly reduces the system complexity. However, as shown in Fig. 8, the BER performance of STSK-OFDM-MA is the worst. The best-performing STSK-OTFS-MA imposes the highest system complexity at a given rate as seen in Fig. 8 and Fig. 10. Hence, it is demonstrated that our proposed STSK-OTFS-MA strikes a beneficial performance *vs.* system complexity trade-off.

VI. Summary and Conclusions

An STSK-OTFS-MA system has been proposed, where each DD-domain APM symbol is spread over both the space and time dimensions by invoking DMs. Our theoretical derivations illustrated that the proposed STSK-OTFS-MA scheme takes full advantage of both time, frequency, space diversity and also attains ST coding gains. Then, a DD-domain RB allocation scheme has been conceived to mitigate the MUI. Moreover, a pair of low-complexity detectors have been proposed for STSK-OTFS-MA based on greedy algorithms and a codebook of DAPs. Furthermore, based on the MGF technique, the asymptotical BER upper-bound of single-user STSK-OTFS-MA has been derived. Our simulation results have shown that the upper-bound becomes tight at high SNRs. Additionally, the DCMC capacity of our STSK-OTFS-MA scheme has been quantified. Finally, by jointly leveraging the DCMC capacity and the BER union-bound, attractive DM design criteria have been proposed for attaining the maximum attainable diversity and coding gains. Both the analytical and simulation results have demonstrated the superiority of our STSK-OTFS-MA system in terms of both its BER and DCMC capacity. We also demonstrated that there exists an optimal combination of the DM sets and the modulation order. Finally, our simulation results demonstrated that both the proposed PRCGD and IRCD are capable of achieving near-ML BER performances at reduced complexity, while the proposed STSK-OTFS-MA scheme is capable of attaining better BER performance at an acceptive system complexity compared to other counterparts.

REFERENCES

[1] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3219–3230, 2010.

[2] M. I. Kadir, S. Sugiura, J. Zhang, S. Chen, and L. Hanzo, "OFDMA/SC-FDMA aided space–time shift keying for dispersive multiuser scenarios," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 408–414, 2013.

[3] H. Elfadil, M. Maleki, N. Behdad, and H. R. Bahrami, "Trellis-coded space-time shift keying," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 5888–5901, 2018.

[4] S. Lu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Compressed-sensing-aided space-time frequency index modulation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6259–6271, 2018.

[5] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[6] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

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

[8] S. Li, J. Yuan, W. Yuan, Z. Wei, B. Bai, and D. W. K. Ng, "Performance analysis of coded OTFS systems over high-mobility channels," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 6033–6048, 2021.

[9] M. Li, S. Zhang, Y. Ge, F. Gao, and P. Fan, "Joint channel estimation and data detection for hybrid RIS aided millimeter wave OTFS systems," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6832–6848, 2022.

[10] J. Shi, J. Hu, Y. Yue, X. Xue, W. Liang, and Z. Li, "Outage probability for OTFS based downlink LEO satellite communication," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 3355–3360, 2022.

[11] Z. Wei, W. Yuan, S. Li, J. Yuan, G. Bharatula, R. Hadani, and L. Hanzo, "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 136–144, 2021.

[12] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7976–7980, 2020.

[13] Z. Yuan, F. Liu, W. Yuan, Q. Guo, Z. Wang, and J. Yuan, "Iterative detection for orthogonal time frequency space modulation with unitary approximate message passing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 714–725, 2022.

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

[21] Y. Ma, G. Ma, N. Wang, Z. Zhong, and B. Ai, "OTFS-TSMA for massive internet of things in high-speed railway," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 519–531, 2022.

[15] A. Thomas, K. Deka, P. Raviteja, and S. Sharma, "Convolutional sparse coding based channel estimation for OTFS-SCMA in uplink," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5241–5257, 2022.

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

[17] H. Wen, W. Yuan, Z. Liu, and S. Li, "OTFS-SCMA: A downlink NOMA scheme for massive connectivity in high mobility channels," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[18] V. Khammammetti and S. K. Mohammed, "Spectral efficiency of OTFS based orthogonal multiple access with rectangular pulses," *IEEE Transactions on Vehicular Technology*, pp. 1–16, 2022.

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

[20] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Layered multi-group steered space-time shift-keying for millimeter-wave communications," *IEEE Access*, vol. 4, pp. 3708–3718, 2016.

[22] J. Shi, Z. Li, J. Hu, Z. Tie, S. Li, W. Liang, and Z. Ding, "OTFS enabled LEO satellite communications: A promising solution to severe doppler effects," *IEEE Network*, pp. 1–7, 2023.

[23] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3049–3063, 2019.

[24] F. Liu, Z. Yuan, Q. Guo, Z. Wang, and P. Sun, "Message passing-based structured sparse signal recovery for estimation of OTFS channels with fractional doppler shifts," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7773–7785, 2021.

[25] Z. Sui, H. Zhang, Y. Xin, T. Bao, L.-L. Yang, and L. Hanzo, "Low complexity detection of spatial modulation aided OTFS in doubly-selective channels," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2023.

[26] H. Zhang, C. Jiang, L.-L. Yang, E. Basar, and L. Hanzo, "Linear precoded index modulation," *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 350–363, 2019.

[27] Z. Sui, S. Yan, H. Zhang, L.-L. Yang, and L. Hanzo, "Approximate message passing algorithms for low complexity OFDM-IM detection," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9607–9612, 2021.

[28] L.-L. Yang, *Multicarrier communications*. John Wiley & Sons, 2009.

[29] G. Taricco and E. Biglieri, "Exact pairwise error probability of space-time codes," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 510–513, 2002.

[30] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multi-dimensional signal sets," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 528–536, 2006.

[31] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5205–5219, 2018.

[32] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes," *Electronics letters*, vol. 33, no. 6, pp. 457–458, 1997.