

Enhanced Index Modulation Aided Orthogonal Time Frequency Space with Variable Active Grids and Multiple Constellations

Kai Zhang, Xue-Qin Jiang, *Senior Member, IEEE*, Han Hai, *Member, IEEE*,
Runhe Qiu, and Shahid Mumtaz, *Senior Member, IEEE*

Abstract—Present orthogonal time frequency space modulation with index modulation (OTFS-IM) techniques activate a given number of grids in the delay-Doppler (DD) domain and employ fixed constellation as changing them would result in ambiguous detection. Motivated by this, we propose a novel enhanced OTFS with index modulation (E-OTFS-IM) scheme which permits changing the number of active grids conveying modulation symbols from any constellation. The proposed E-OTFS-IM scheme alleviates ambiguous detection by jointly mapping information to grid activation pattern in DD domain and 2D modulation symbols. We also propose an efficient encoding method for E-OTFS-IM. In addition, an updated message passing (MP) detection suitable for E-OTFS-IM as well as its low-complexity improvement are developed. We derive a average bit error probability (ABEP) upper bound for our proposed E-OTFS-IM scheme employing the maximum likelihood detector. Additionally, we also propose an enhanced scheme named E-OTFS-IM with in-phase/quadrature (E-OTFS-IM-IQ) to further improve the spectral efficiency and reliability. In E-OTFS-IM-IQ, the principle of E-OTFS-IM is extended to in-phase and quadrature branches independently. The efficient encoding method and MP detector for this modified scheme is also introduced. Simulation results confirm that the derived ABEP upper bound is tight in the high signal-to-noise ratio region, and show that our proposed E-OTFS-IM improves the performance of conventional OTFS and OTFS-IM. The results also indicate the proposed E-OTFS-IM-IQ achieves better bit error rate performance than E-OTFS-IM.

Index Terms—Orthogonal time frequency space modulation, index modulation, spectral efficiency, in-phase/quadrature.

I. INTRODUCTION

ORTHOGONAL frequency division multiplexing (OFDM) modulation techniques achieve good robustness in

Manuscript received June 11, 2023; revised January 04, 2024; accepted March 2, 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61801106, in part by the 6G-SENSES project from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Unions Horizon Europe research and innovation programme under Grant Agreement No. 101139282, and in part by the Fundamental Research Funds for the Central Universities and Graduate Student Innovation Fund of Donghua University under Grant CUSF-DH-D-2021053. The associate editor coordinating the review of this paper and approving it for publication was A. Al-Dweik. (*Corresponding author: X.-Q. Jiang.*)

K. Zhang, H. Hai, X.-Q. Jiang, and R. Qiu, are with the School of Information Science and Technology, Donghua University, Shanghai, 201620, China (e-mail: 2171294@mail.dhu.edu.cn, hhai@dhu.edu.cn, xqjiang@dhu.edu.cn, qiurh@dhu.edu.cn).

S. Mumtaz is with Department of Applied Informatics, Silesian University of Technology Akademicka 16 44-100 Gliwice, Poland and also with the Department of Computer Sciences, Nottingham Trent University, NG1 4FQ Nottingham, U.K. (e-mail: dr.shahid.mumtaz@ieee.org).

time-invariant channels and have been widely used in mobile communication systems. However, for the high mobility scenarios in the 5G and future mobile communication prospects [1], where the channel is typically time-varying, OFDM suffers from severe inter-carrier interference caused by high Doppler spread, resulting in degraded performance. Therefore, future wireless communication systems require new multi-carrier modulation techniques for reliable transmission in high mobility scenarios.

Recently, orthogonal time frequency space (OTFS) modulation [2]-[4] was proposed that is robust to Doppler shifts and effectively resistant to the effects of fast time-varying channels in high-mobility scenarios. Information is transmitted via a matrix in the delay-Doppler (DD) domain in OTFS modulation. This ensures that each symbol transmitted within a frame experiences a nearly identical DD domain channel. The modulation is then extended to the time-frequency domain, occupying all time-frequency resources. So even if the channel is time-frequency doubly-selective, OTFS still has the potential for achieving full diversity [5]-[7]. Due to the block-circulant, sparse and constant characteristics of the channel in the DD domain [8], many low-complexity detector designs were proposed [9]-[13]. Besides, the estimation difficulty of time-varying channel in DD domain was proven to be much lower than that in the time-frequency domain [14], [15]. These advantages make OTFS modulation a promising choice for dealing with wireless communications in high mobility scenarios.

In addition to reliability, users are demanding higher spectrum efficiency (SE) and energy efficiency as the requirements of various applications continue to rise. Driven by this purpose, the concept of index modulation (IM) was proposed and applied in various communication systems [16]-[18]. Several IM-based OFDM schemes have been proposed in [19]-[22], where the active subcarrier patterns convey information beyond the modulation symbols to enhance SE without extra energy consumption. In [23]-[25], the number of active subcarriers is variable for achieving a higher throughput than OFDM with IM (OFDM-IM). Specially, it is for the first time proposed in [25] that the active subcarrier number can be combined with subcarrier activation states to form high-dimension signal constellations, achieving significant performance gains compared with existing OFDM-IM schemes.

For improving reliability under high mobility, OTFS with IM (OTFS-IM) was recently proposed in [26]. Concretely, the

OTFS-IM scheme transmits additional information via grid activation pattern (GAP) in the DD domain apart from the conventional modulation symbols. The silence grids in the DD domain can be regarded as guard symbols with improved tolerability against high Doppler effect relative to conventional OTFS. Notice that OTFS-IM also exhibits overwhelming performance advantages over OFDM-IM, since the OFDM waveform suffers from severe inter-carrier interference due to the time variation of the high-mobility channel. To achieve a better trade-off between reliability and SE, OTFS with dual-mode IM (OTFS-DM-IM) [27] was proposed, where two separate sets of selected and unselected grids are modulated by two distinguishable constellation alphabets. By extending IM independently to the I- and Q- domains, OTFS with in-phase and quadrature IM (OTFS-I/Q-IM) was proposed in [28]. OTFS with improved IM (OTFS-IIM) was proposed to enhance the bit error rate (BER) performance by transmitting each index twice [29]. Considering the Doppler spread caused by the off-grid Doppler shifts, Doppler-IM with OTFS (DoIM-OTFS) that activates a block of Doppler resource bins was proposed in [30]. The diversity and peak-to-average power ratio of OTFS-IM were investigated in [32]. IM is also available for OTFS based uplink multiple access (MA) in [33], aiming to provide OTFS-MA with more flexible resource utilization and serve more users. OTFS-IMMA activates a subset of available DD resources through IM to allow multiple users to share the same DD resources.

Due to the more complex signal structure of OTFS-IM, detection design for OTFS-IM also deserves attention, as detections for OTFS cannot be employed directly. Minimum mean squared error (MMSE) equalizer was adopted to detect the symbol bits and index bits in [26]-[29], while the high complexity of MMSE due to matrix inverse operator is intractable in practice. In order to achieve a compromise between reliability and complexity, the customized message-passing (MP) detector for DoIM-OTFS was proposed in [30] as well. A multi-layer MP detection for DoIM-OTFS was also proposed in [31], which with faster convergence and higher accuracy by introducing additional pattern nodes. A symbol-by-symbol aided expectation propagation (SS-EP) detector was designed for OTFS-IM in [34] as well as its low complexity improvement. Performing MP detection by treating each IM symbol as a whole vector, the vector-by-vector-aided MP (VV-MP) detector was proposed in [35].

However, the existing OTFS-IM schemes use separate mapping, which divides the information bits into two parts that are mapped to an GAP and conventional modulation symbols, respectively. Because of this, they must fix the number of active grids in DD domain and the cardinality of the constellation. Changing the number of active grids and the cardinality of the constellation causes a variable length of information bits, resulting in ambiguous detection at the receiver and making resource scheduling and allocation difficult. The proposed scheme solves the above problems and without imposing any artificial constraints. The contributions of this paper are summarized as follows:

- We propose a novel transmission scheme termed enhanced OTFS with index modulation (E-OTFS-IM) where

the number of active grids transmitting modulation symbols from any cardinality of constellations is allowed to be changed. The proposed E-OTFS-IM scheme avoids the ambiguous detection by transmitting a constant length of information bits through joint-mapping. We design a computationally efficient encoding method for this joint-mapping suited for the practical case of larger parameters. We also modified the MP detector to adapt it to the proposed E-OTFS-IM. And a max-sum (MS) based MP detector is proposed to further reduce the detection complexity. By deriving an upper bound on the asymptotic average bit error probability (ABEP) with the optimal maximum likelihood (ML) detector, we determine the diversity and coding gain of the E-OTFS-IM.

- We further propose E-OTFS-IM with in-phase/quadrature (E-OTFS-IM-IQ) scheme with better reliability and higher transmission rate than E-OTFS-IM. By extending the principles of E-OTFS-IM to I- and Q- branches independently, another degree of freedom can be added to the system for improving SE and resource utilisation in the DD domain. We also propose the corresponding computationally efficient encoding method and the updated MP detector for E-OTFS-IM-IQ.
- Simulation results confirm that the derived upper bound for ABEP is tight in the high signal-to-noise ratio (SNR) region, and show that E-OTFS-IM improves the BER performance of conventional OTFS and OTFS-IM. Furthermore, E-OTFS-IM-IQ achieves better BER performance than E-OTFS-IM at the same SE.

The rest of this paper is organized as follows. The system model of conventional OTFS-IM is reviewed in Section II. Ambiguous detection caused by variable active grids is expounded in Section III. In Section IV, we propose the E-OTFS-IM scheme and its low-complexity detection. The analysis of BER performance is presented in Section V. In Section VI, the E-OTFS-IM-IQ scheme and its low-complexity detectors are proposed. The simulation results and comparisons of BER performance are presented in Section VII. This paper is ended with conclusions in Section VIII.

Notations: Matrices and vectors are denoted by upper- and lower- case boldface letters, respectively. $\binom{n}{k}$ represents the number of combinations to take k elements from n different elements. \otimes is the Kronecker product. $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote ceil function and floor function operations, respectively. $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and Hermitian transpose, respectively. The probability of an event is denoted by $\Pr(\cdot)$. The operator $\mathbb{E}[\cdot]$ represents expectation. The Euclidean norm is denoted by $\|\cdot\|$ and the 0-norm is denoted by $\|\cdot\|_0$. $Q(\cdot)$ represents the Gaussian Q -function. \mathbf{F}_N stands for N -point discrete Fourier transform (DFT) matrix.

II. REVIEW OF CONVENTIONAL OTFS-IM SCHEME

Consider an OTFS-IM system with M subcarriers and N symbols, where the subcarrier spacing and symbol duration are Δf and T , respectively. A 2D phase shift keying (PSK)/quadrature amplitude modulation (QAM) constellation \mathcal{S} is adopted. A total of B incoming bits are divided into G groups,

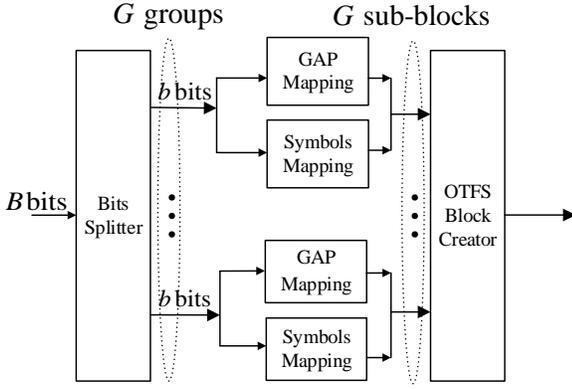


Fig. 1. Block diagram of OTFS-IM by separate mapping.

each of which contains $b = B/G$ bits. Each group of b information bits is mapped to an OTFS sub-block of size $n = MN/G$ in the DD domain. In each sub-block, there are M' grids along the delay dimension and N' grids along the Doppler dimension, so the size n of each sub-block can also be calculated as $n = M'N'$. Each sub-block consists of an $M' \times N'$ signal matrix. Let k be the fixed number of active grids in each sub-block. Thus, there are $\binom{n}{k}$ GAPs. k modulation symbols from the constellation \mathcal{S} are carried by the active grid indices of a chosen GAP, and “0”s are carried by the inactive grid indices. The conventional information bits to a sub-block realization mapping is referred to as separate mapping. Information bits in the OTFS-IM scheme mapped to one sub-block are divided into two parts as shown in Fig. 1. The first $\lfloor \log_2 \binom{n}{k} \rfloor$ bits are mapped to GAPs, and the second $k \log_2(|\mathcal{S}|)$ bits are mapped to modulation symbols from a 2D constellation \mathcal{S} . The SE of OTFS-IM is

$$\eta_{\text{IM}} = \frac{Gb}{MN} = \frac{1}{n} \left(\lfloor \log_2 \binom{n}{k} \rfloor + k \log_2(|\mathcal{S}|) \right). \quad (1)$$

The fixed number of active grids limits the potential of OTFS-IM. We will expound the phenomenon of ambiguous detection that occurs as a result of conventional separate mapping with variable active grids in a sub-block in the next section.

III. AMBIGUOUS DETECTION IN OTFS-IM SYSTEM WITH VARIABLE ACTIVE GRIDS

Let k_1, \dots, k_R ($0 < k_1 < \dots < k_R \leq n$) denote the numbers of active grids allowed in any OTFS sub-block and R is the total number of cases of the variable active grids. For example, in the case of $n = 4$, $R = 2$, $k_1 = 1$ and $k_2 = 3$, the number of activated grids in a sub-block of size four can be one or three. i.e., the activation ratio of grids in the sub-block can be varied. This suggests that different grid activation ratios can be applied to different sub-blocks. Let us consider the case of separate mapping wherein $b = b_1 + b_2$ bits are mapped to a sub-block. Here, b_1 and b_2 denote the number of bits mapped to the GAP and the symbols from constellation \mathcal{S} , respectively. The total number of GAPs is $\sum_{i=1}^R \binom{n}{k_i}$, thus $b_1 = \lfloor \log_2 \sum_{i=1}^R \binom{n}{k_i} \rfloor$, which is constant for choosing a single pattern for transmission

from the possible GAPs. The number of bits mapped to the symbols from constellation \mathcal{S} is $b_2 = k_i \log_2(|\mathcal{S}|)$, where k_i , $i = 1, \dots, R$, is the number of active grids in any sub-block. Now, k_i varies for each sub-block, which also varies the total number of b for each group. This creates ambiguous detection at the receiver, as the receiver might detect wrong number of bits.

Consider an example with $n = 4$, $R = 2$, $k_1 = 1$ and $k_2 = 3$. All the GAPs are shown as

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

In each GAP, “1” means that the grid is active and “0” means that the grid is inactive. The total number of GAPs is $\binom{4}{1} + \binom{4}{3} = 8$, and $\log_2(8) = 3$ bits are mapped to the GAPs. However, the number of symbols from constellation \mathcal{S} in each sub-block changes with different number of active grids, resulting in one or three symbols. Because the receiver lacks this knowledge, the incorrect number of symbols might be detected. As a result, the total amount of bits in a frame might be incorrectly detected.

Ambiguous detection can be avoided by imposing artificial restriction, such as fixing the total number of active grids in an OTFS transmission block and therefore the number of transmitted symbols in a transmission block. Ambiguity in detection can be resolved by detecting a fixed number of transmitted symbols in the entire OTFS transmission block. Considering that this approach requires detection based on the entire OTFS transmission block, the inability to detect symbols and indices for each sub-block individually may lead to a high detection complexity. This is the biggest drawback of this approach.

It is also possible to use constellations with different cardinalities for variable active grids resulting in a constant b_2 , similar to the approach in [36]. However, the cardinalities of these constellations must meet certain conditions, which makes certain constellations not applicable for all k_i s. Still, it is possible that constellations with huge cardinalities have to be used, which severely degrades the BER of OTFS-IM. For example, for constellations \mathcal{S}_1 , \mathcal{S}_2 and \mathcal{S}_3 corresponding to $k_1 = 1$, $k_2 = 3$ and $k_3 = 5$, respectively, their cardinalities must satisfy $|\mathcal{S}_1| = 2^{15z}$, $|\mathcal{S}_2| = 2^{5z}$ and $|\mathcal{S}_3| = 2^{3z}$ for fixed b_2 , where $z \in \mathbb{Z}^+$.

IV. THE PROPOSED E-OTFS-IM SCHEME

We now propose the E-OTFS-IM scheme to enhance the OTFS-IM by varying the number of active grids in each OTFS sub-block avoiding ambiguous detection without imposing any artificial constraints. Furthermore, these active grids can carry modulation symbols from any constellation without restriction.

A. Implementation of E-OTFS-IM

In E-OTFS-IM, the allowed numbers of active grids in a sub-block with n grids are k_1, k_2, \dots, k_R ($0 < k_1 < k_2 < \dots < k_R \leq n$). The 2D constellation \mathcal{S}_i is used for the GAPs

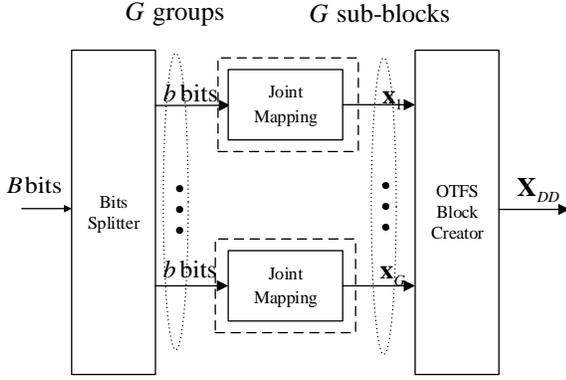


Fig. 2. Block diagram of E-OTFS-IM by joint mapping.

with k_i active grids for $i = 1, \dots, R$. Denote the GAP with k_i active grids as $\mathbf{a} = [a_1, a_2, \dots, a_n]^T \in \{0, 1\}^{n \times 1}$, $\|\mathbf{a}\|_0 = k_i$, where $a_j = 1$ if the j -th grid is active, and $a_j = 0$, otherwise. Furthermore, the vector form of the g -th sub-block in the DD domain by combining the chosen GAP and symbols can be denoted as

$$\mathbf{x}^g = [x_1^g, x_2^g, \dots, x_n^g]^T, \quad (2)$$

where $x_j^g \in \{0, \mathcal{S}_i\}$ for $g = 1, \dots, G$ and $j = 1, \dots, n$. The GAP for \mathbf{x}^g can be presented by $\mathbf{a}^g = [a_1^g, a_2^g, \dots, a_n^g]^T$ with

$$a_j^g = \begin{cases} 1, & \text{if } x_j^g \in \mathcal{S}_i \\ 0, & \text{if } x_j^g = 0. \end{cases} \quad (3)$$

Let \mathcal{X} be the set of possible sub-block realizations. The total number of possible GAPs is $\sum_{i=1}^R \binom{n}{k_i}$. The number of sub-block realizations associated with one GAP with k_i active grids is $|\mathcal{S}_i|^{k_i}$. Thus, the total number of all possible sub-block realizations is

$$|\mathcal{X}| = \sum_{i=1}^R \binom{n}{k_i} |\mathcal{S}_i|^{k_i}. \quad (4)$$

Even with variable active grids and modulation symbols from any constellation, the total number of possible sub-block realizations $|\mathcal{X}|$ is a constant. Therefore, we can jointly map information bits of constant length b to a sub-block such that

$$b = \lceil \log_2 |\mathcal{X}| \rceil = \left\lceil \log_2 \left[\sum_{i=1}^R \binom{n}{k_i} |\mathcal{S}_i|^{k_i} \right] \right\rceil, \quad (5)$$

resulting in an SE of

$$\eta_E = \frac{Gb}{MN} = \frac{1}{n} \lceil \log_2 |\mathcal{X}| \rceil. \quad (6)$$

The required sub-block is chosen from a mapping table. The mapping table activates the grids, and chooses the symbols carried by the activated grids based on the information bits of length b , denoted \mathbf{b} , transmitted by the sub-block as shown in Fig. 2. This alleviates the problem of ambiguous detection, as E-OTFS-IM achieves bijective mapping between constant-length information bits and sub-block realizations. We provide an example of E-OTFS-IM for illustration with $n = 8$, $k_1 = 2$,

TABLE I
MAPPING TABLE OF E-OTFS-IM WITH $n = 8$, $k_1 = 2$,
 $k_2 = 3$, 4QAM FOR \mathcal{S}_1 AND BPSK FOR \mathcal{S}_2

Z	Information Bits	Sub-block
0	00000000	$[-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, 0, 0, 0, 0, 0, 0]^T$
1	00000001	$[-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i, 0, 0, 0, 0, 0, 0]^T$
2	00000010	$[-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, 0, 0, 0, 0, 0, 0]^T$
\vdots	\vdots	\vdots
447	11011111	$[0, 0, 0, 0, 0, \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i, \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i]^T$
448	11100000	$[-1, -1, -1, 0, 0, 0, 0, 0]^T$
449	11100001	$[-1, -1, 1, 0, 0, 0, 0, 0]^T$
450	11100010	$[-1, 1, -1, 0, 0, 0, 0, 0]^T$
\vdots	\vdots	\vdots
510	11111110	$[1, 0, 0, 1, -1, 0, 0, 0]^T$
511	11111111	$[1, 0, 0, 1, 1, 0, 0, 0]^T$

$k_2 = 3$, 4QAM for \mathcal{S}_1 and BPSK for \mathcal{S}_2 . This means that for a sub-block consisting of 8 grids, two of the eight grids may carry 4QAM symbols or three of the eight grids may carry BPSK symbols. In this case, there are totally $|\mathcal{X}| = 896$ possible sub-block realizations from (4), leading to totally 9 bits transmitted through a sub-block ($b = 9$) and 512 allowed sub-block realizations for transmission. The corresponding mapping table is given in TABLE I. From TABLE I, the sub-block can be directly obtained by certain information bits.

By mapping all G sub-blocks, the OTFS block creator generates a DD domain symbol \mathbf{X}_{DD} whose vectorized form is

$$\mathbf{x} = \text{vec}(\mathbf{X}_{DD}) = [x_1, x_2, \dots, x_{MN}]^T. \quad (7)$$

The transmitted symbols in the DD domain are then converted to the time-frequency domain by the Inverse Symplectic Fourier transform (ISFFT) operation. Then through the Heisenberg transform and the pulse-shaping waveform $g_{\text{tx}}(t)$, the transmit signal in the time domain can be presented by

$$\mathbf{X}_T = \mathbf{G}_{\text{tx}} \mathbf{F}_M^H (\mathbf{F}_M \mathbf{X}_{DD} \mathbf{F}_N^H) = \mathbf{G}_{\text{tx}} \mathbf{X}_{DD} \mathbf{F}_N^H. \quad (8)$$

We assume that the sampling interval is T/M and $\mathbf{G}_{\text{tx}} = \text{diag}[g_{\text{tx}}(0), g_{\text{tx}}(T/M), \dots, g_{\text{tx}}((M-1)T/M)] \in \mathbb{C}^{M \times M}$ represents the filter operating at the transmitter. Vectorising \mathbf{X}_T in column-wise yields a one-dimensional vector $\mathbf{x}_T \in \mathbb{C}^{MN \times 1}$, which is transmitted over a time-varying channel after the addition of a cyclic prefix (CP).

After removing the CP, the received vector in the time domain can be presented by

$$\mathbf{r} = \mathbf{H}_T \mathbf{x}_T + \mathbf{n}, \quad (9)$$

where $\mathbf{n} \sim \mathcal{CN}(0, N_0)$ is the additive Gaussian noise. The

channel matrix in the time domain is given by [8]

$$\mathbf{H}_T = \sum_{i=1}^P h_i \mathbf{\Pi}^{\nu_i} \mathbf{\Delta}^{\kappa_i + \varepsilon_i}, \quad (10)$$

where the forward cyclic-shift matrix $\mathbf{\Pi} = \text{circ} \{ [0, 1, 0, \dots, 0]_{MN \times 1}^T \} \in \mathbb{C}^{MN \times MN}$, $\mathbf{\Delta} = \text{diag} [1, \exp(-2\pi j/MN), \dots, \exp(-2\pi j(MN-1)/MN)]$ and P is the number of propagation paths. The channel gain, delay and Doppler taps of the i -th path are defined as h_i , $\tau_i = \frac{\nu_i}{M\Delta f}$ and $\nu_i = \frac{\kappa_i + \varepsilon_i}{NT}$, respectively, where ν_i and κ_i represent integer delay and Doppler taps and ε_i denotes the fractional Doppler. The received vector $\mathbf{r} \in \mathbb{C}^{MN \times 1}$ is rearranged into $\mathbf{R} \in \mathbb{C}^{M \times N}$. The Wigner transform with the pulse-shaping waveform $g_{rx}(\tau)$ is then applied to \mathbf{R} , followed by the Symplectic Fourier transform (SFFT). The received signal in the DD domain can be presented by

$$\mathbf{Y} = \mathbf{F}_M^H (\mathbf{F}_M \mathbf{G}_{rx} \mathbf{R}) \mathbf{F}_N = \mathbf{G}_{rx} \mathbf{R} \mathbf{F}_N. \quad (11)$$

Here, $\mathbf{G}_{rx} = \text{diag} [g_{rx}(0), g_{rx}(T/M), \dots, g_{rx}((M-1)T/M)] \in \mathbb{C}^{M \times M}$ represents the filter operating at the receiver. By vectorising the \mathbf{Y} in column-wise, organising (8), (9), and (11) yields

$$\mathbf{y} = (\mathbf{F}_N \otimes \mathbf{G}_{rx}) \mathbf{H}_T (\mathbf{F}_N^H \otimes \mathbf{G}_{tx}) \mathbf{x} + (\mathbf{F}_N \otimes \mathbf{G}_{rx}) \mathbf{n} = \mathbf{H} \mathbf{x} + \mathbf{w}. \quad (12)$$

The joint maximum a posteriori detection rule for estimating the transmitted signals is given by

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \{0 \cup \dot{\mathcal{S}}\}^{MN \times 1}} \Pr(\mathbf{x} | \mathbf{y}, \mathbf{H}), \quad (13)$$

where $\dot{\mathcal{S}} = \bigcup_{i=1}^R \mathcal{S}_i$.

B. Efficient Encoding Method

Explicitly, utilizing a mapping table is quite straightforward, and the ambiguity in detection can be avoided due to constant-length information bits by jointly considering GAPs and 2D modulation symbols. For large values of n , R , k_i s and $|\mathcal{S}_i|$ s, the size of mapping table will become prohibitively large. Implementation of an encoding map by a mapping table with very large size is impractical. To address this issue, we propose a more efficient E-OTFS-IM mapping method.

The following encoding procedure maps the information bits \mathbf{b} with constant length b to sub-block.

- 1) Convert \mathbf{b} to a decimal number Z , $Z \in [0, 2^b - 1]$. Determine the indication parameter r according to

$$r = \begin{cases} 1, & Z \in [0, \binom{n}{k_1} |\mathcal{S}_1|^{k_1} - 1] \\ 2, & Z \in [\binom{n}{k_1} |\mathcal{S}_1|^{k_1}, \binom{n}{k_2} |\mathcal{S}_2|^{k_2} - 1] \\ \vdots \\ R, & Z \in [\binom{n}{k_{R-1}} |\mathcal{S}_{R-1}|^{k_{R-1}}, 2^b - 1] \end{cases}. \quad (14)$$

Determine the number of symbol bits $b_2 = k_r \log_2 |\mathcal{S}_r|$ and the number of GAP bits is $b_1 = b - b_2$. Denote the first b_1 bits and the remainder b_2 bits of \mathbf{b} as \mathbf{b}_1 and \mathbf{b}_2 , respectively.

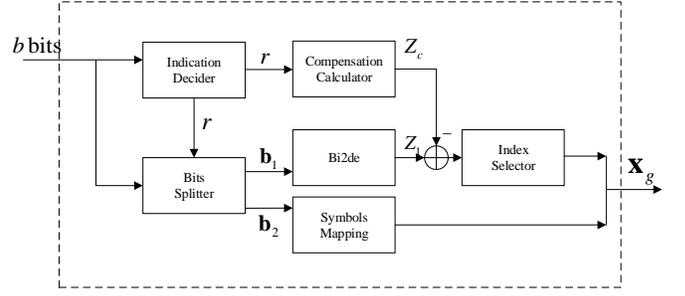


Fig. 3. Block diagram of the proposed efficient encoding method for the g -th sub-block.

- 2) For $r = 1$, set the compensation value $Z_c = 0$. For $r \neq 1$, convert $\binom{n}{k_{r-1}} |\mathcal{S}_{r-1}|^{k_{r-1}}$ to the binary sequence \mathbf{b}' of length b and convert the first b_1 bits of \mathbf{b}' to the decimal number Z_c .
- 3) Convert \mathbf{b}_1 to the decimal number Z_1 and subtract Z_c from Z_1 . The difference $Z_1 - Z_c$ is fed into the index selector to obtain the indices of active grids through the combinatorial method in [19]. The GAP is then obtained from the indices of active grids.
- 4) Select k_r constellation points independently from the constellation \mathcal{S}_r according to \mathbf{b}_2 .
- 5) Construct the sub-block $\mathbf{x}^g = [x_1^g, \dots, x_n^g]^T$ by combining the GAP and symbols.

The above procedure is illustrated by a block diagram in Fig. 3. The module in the dashed line in Fig. 2 can be directly replaced by the module in Fig. 3.

Remark: The index selector requires the input to be continuous integers starting from zero. This method is implemented by mapping an all-zero sequence representing 0 to an all-one sequence representing $2^b - 1$ into sub-block realizations in the order of active grid numbers k_1 to k_R . As a result, $r = 1$ must meet the input requirement of index selector. For $i = 1, 2, \dots, R-1$, the sub-block realizations with k_{i+1} active grids begins to be mapped only when all $\binom{n}{k_i} |\mathcal{S}_i|^{k_i}$ sub-block realizations with k_i active grids have been mapped, which leads to the input of index selector for k_{i+1} active grids to start at Z_c . This corresponds to the cases of $r = 2, 3, \dots, R$. Therefore, Z_c needs to be subtracted from Z_1 in order to meet the input requirement of the index selector.

We also take $n = 8$, $k_1 = 2$ and $k_2 = 3$ with 4QAM for \mathcal{S}_1 and BPSK for \mathcal{S}_2 to demonstrate how this method works. There are 512 allowed sub-block realizations for transmission, thus $b = 9$ bits can be transmitted by a sub-block. Without losing generality, we take g -th sub-block to illustrate.

Convert \mathbf{b} to decimal number Z . According to (14), the interval $[0, 511]$ where Z is located is divided by $\binom{n}{k_1} |\mathcal{S}_1|^{k_1} = \binom{8}{2} 4^2 = 448$ into two sub-intervals $[0, 447]$ and $[448, 511]$.

If $Z \in [0, 447]$, this means that two 4QAM symbols are selected and two of the eight grids are activated to carry these two 4QAM symbols. In this case, $r = 1$, $b_2 = 2 \log_2(4) = 4$ and $b_1 = 9 - 4 = 5$. The first 5 bits and the remainder 4 bits of \mathbf{b} are considered as \mathbf{b}_1 and \mathbf{b}_2 , respectively. The compensation

value is $Z_c = 0$. Convert \mathbf{b}_1 to Z_1 and feed $Z_1 - Z_c = Z_1$ into the index selector to obtain the the indices of active grids and the GAP \mathbf{a}^g . The two selected 4QAM symbols $\mathbf{s}^g = [s_1, s_2]^T$, $s_1, s_2 \in \mathcal{S}_1$ are obtained according to \mathbf{b}_2 . Then, construct the sub-block \mathbf{x}^g by combining \mathbf{a}^g and \mathbf{s}^g .

If $Z \in [448, 511]$, this means that three BPSK symbols are selected and three grids are activated from the eight grids to carry these three BPSK symbols. In this case, $r = 2$, $b_2 = 3\log_2(2) = 3$ and $b_1 = 9 - 3 = 6$. The first 6 bits and the remainder 3 bits of \mathbf{b} are considered as \mathbf{b}_1 and \mathbf{b}_2 , respectively. Convert $\binom{n}{k_{r-1}}|\mathcal{S}_{r-1}|^{k_{r-1}} = \binom{n}{k_1}|\mathcal{S}_1|^{k_1} = 448$ to the 9-bit binary sequence $\mathbf{b}' = [111000000]$. Then convert the first six bits of \mathbf{b}' , $[111000]$, to decimal as $Z_c = 56$. This means that the input of index selector for $k_2 = 3$ starts from 56, which does not satisfy the input requirements of the index selector. So in this case, after \mathbf{b}_1 is converted into decimal Z_1 , $Z_c = 56$ needs to be subtracted from Z_1 . The indices of active grids is obtain from $Z_1 - Z_c$ via the index selector and the GAP \mathbf{a}^g is also obtained. The three BPSK symbols $\mathbf{s}^g = [s_1, s_2, s_3]^T$, $s_1, s_2, s_3 \in \mathcal{S}_2$ are obtained from \mathbf{b}_2 . Then, construct the sub-block \mathbf{x}^g by combining \mathbf{a}^g and \mathbf{s}^g .

Here is a concrete instance. Assuming the incoming information bits $\mathbf{b} = [111111110]$ with $Z = 510$, we have $r = 2$, $b_2 = 3$ and $b_1 = 6$. The compensation value is $Z_c = 56$. Covert the first six bits $\mathbf{b}_1 = [111111]$ to $Z_1 = 63$ and input $Z_1 - Z_c = 63 - 56 = 7$ into the index selector to obtain the indices of active grids $\{1, 3, 4\}$. According to $\{1, 3, 4\}$, the GAP $\mathbf{a}^g = [1, 0, 0, 1, 1, 0, 0, 0]^T$ is obtained. Map the remaining three bits $\mathbf{b}_2 = [110]$ to obtain three BPSK symbols $\mathbf{s}^g = [1, 1, -1]^T$. Activate the first, fourth and fifth grids of the g -th OTFS sub-block to carry 1, 1, and -1 respectively, and construct the sub-block $\mathbf{x}^g = [1, 0, 0, 1, -1, 0, 0, 0]^T$. ■

C. Updated MP Detection

MP algorithm is a practical choice to trade off between the reliability and detection complexity. In this subsection, the updated MP algorithm for E-OTFS-IM is developed.

\mathbf{H} is a sparse matrix and only L elements out of MN are non-zero in each row and column of \mathbf{H} . Let $\mathcal{I}(d)$ and $\mathcal{J}(c)$ denote the sets of indexes with non-zero elements in the d -th row and c -th column, respectively. The system can be modeled as a sparsely-connected factor graph. In this factor graph, each observation node $y[d]$ is connected to the set of L variable nodes $\{x[c], c \in \mathcal{I}(d)\}$. Similarly, each variable node $x[c]$ is connected to the set of L observation nodes $\{y[d], d \in \mathcal{J}(c)\}$. For the updated MP algorithm, its steps in the t -th iteration are described as follows.

- 1) The mean $\mu_{d,c}^{(t)}$ and variance $(\sigma_{d,c}^{(t)})^2$ of the interference term $\xi_{d,c}^{(t)}$ are passed as messages from $y[d]$ to $\{x[c], c \in \mathcal{I}(d)\}$. $\xi_{d,c}^{(t)}$ can be approximated as a Gaussian random variable, which is given by

$$\xi_{d,c}^{(t)} = \sum_{e \in \mathcal{I}(c), e \neq c} x[e]H[d, e] + w[d]. \quad (15)$$

The mean and variance of $\xi_{d,c}^{(t)}$ are given by

$$\mu_{d,c}^{(t)} = \sum_{e \in \mathcal{I}(c), e \neq c} \sum_{x \in \{\check{\mathcal{S}} \cup 0\}} p_{e,d}^{(t-1)}(x)H[d, e]x, \quad (16)$$

and

$$(\sigma_{d,c}^{(t)})^2 = \sum_{e \in \mathcal{I}(d), e \neq c} \left(\sum_{x \in \{\check{\mathcal{S}} \cup 0\}} p_{e,d}^{(t-1)}(x) |H[d, e]x|^2 - \left| \sum_{x \in \{\check{\mathcal{S}} \cup 0\}} p_{e,d}^{(t-1)}(x)H[d, e]x \right|^2 \right) + \sigma^2. \quad (17)$$

The variable node $x[c]$ calculates the probability according to

$$\zeta_{d,c}^{(t)}(x) \propto \exp \left(-\frac{|y[d] - \mu_{d,c}^{(t)} - H[d, c]x|^2}{(\sigma_{d,c}^{(t)})^2} \right). \quad (18)$$

- 2) The message from variable nodes $x[c]$ to observation node $\{y[d], d \in \mathcal{J}(c)\}$ can be calculated by

$$\begin{aligned} \tilde{p}_{c,d}^{(t)}(x) &\propto \prod_{e \in \mathcal{J}(c), e \neq d} \Pr(y[e] | x[c] = x, \mathbf{H}) \\ &= \prod_{e \in \mathcal{J}(c), e \neq d} \frac{\zeta_{e,c}^{(t)}(x)}{\sum_{x \in \{\check{\mathcal{S}} \cup 0\}} \zeta_{e,c}^{(t)}(x)}. \end{aligned} \quad (19)$$

With a damping factor $\Delta \in (0, 1]$, the practical message from $x[c]$ to $y[d]$ is

$$p_{c,d}^{(t)}(x) = \Delta \cdot \tilde{p}_{c,d}^{(t)}(x) + (1 - \Delta) \cdot p_{c,d}^{(t-1)}(x). \quad (20)$$

- 3) Calculate the convergence indicator $\gamma^{(t)}$ as

$$\gamma^{(t)} = \frac{1}{MN} \sum_{c=1}^{MN} \mathbb{1} \left(\max_{x \in \{\check{\mathcal{S}} \cup 0\}} p_c^{(t)}(x) \geq 1 - \epsilon \right), \quad (21)$$

where $\mathbb{1}(\cdot)$ is indicator function, and

$$p_c^{(t)}(x) = \prod_{e \in \mathcal{J}(c)} \frac{\zeta_{e,c}^{(t)}(x)}{\sum_{x \in \{\check{\mathcal{S}} \cup 0\}} \zeta_{e,c}^{(t)}(x)}. \quad (22)$$

If $\gamma^{(t)} > \gamma^{(t-1)}$, then update the probability vector of symbols as $\mathbf{p}_c = \mathbf{p}_c^{(t)}$ for $c = 1, 2, \dots, MN$.

- 4) Repeat the above steps till $\gamma^{(t)} = 1$ or the maximum number of iterations, t_{max} , is reached.

Once the stopping criteria are satisfied, we next calculate the log likelihood ratio (LLR) of each grid according to

$$\lambda(c) = \ln \frac{\sum_{x \in \check{\mathcal{S}}} p_c(x)}{p_c(0)}, \quad c = 1, \dots, MN. \quad (23)$$

The processed LLR values are calculated as

$$\begin{aligned} \bar{\boldsymbol{\lambda}} &= [H_0(\lambda(1)), H_0(\lambda(2)), \dots, H_0(\lambda(MN))]^T \\ &= [\bar{\boldsymbol{\lambda}}_1^T, \bar{\boldsymbol{\lambda}}_2^T, \dots, \bar{\boldsymbol{\lambda}}_G^T]^T, \end{aligned} \quad (24)$$

where $\bar{\boldsymbol{\lambda}}_g = [H_0(\lambda_g(1)), H_0(\lambda_g(2)), \dots, H_0(\lambda_g(n))]^T$ is used for GAP estimation for g -th sub-block. $H_0(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$

is the Heaviside step function. Determine the Hamming distance between $\bar{\lambda}_g$ and each legitimate GAP, and find the set of possible legitimate GAPs associated with the smallest Hamming distance d_{\min} , denoted \mathcal{A} . If $d_{\min} = 0$, $\bar{\lambda}_g$ is the estimated legitimate GAP. If $d_{\min} > 1$, reverse the element with the smallest LLR magnitude in $\bar{\lambda}_g$ to reduce the minimum Hamming distance until $d_{\min} = 1$. If $d_{\min} = 1$, find the positions in $\bar{\lambda}_g$ which is different from GAPs in \mathcal{A} and reverse the element with the smallest LLR magnitude to get the estimation of GAP. Finally, determine the modulation symbols over these active grids as

$$\hat{x}[c] = \arg \max_{x \in \mathbb{A}} p_c(x), \quad c \in \mathbb{A}, \quad (25)$$

where \mathbb{A} is the set of active grids. Thereafter, the reverse procedure of the encoding procedure mentioned in the last subsection is implemented sub-block by sub-block to obtain an estimate of the transmitted information bits.

The computational complexity of the U-MP detection for E-OTFS-IM scheme in each iteration is dominated by (16), (17), (20), (21) and (25). To sum up, the U-MP detection has a complexity order $\mathcal{O}(t_{max} MNL (|\mathcal{S}| + 1))$.

D. Max-sum Based Updated MP Detection

Obviously, a large number of division operations are required to compute $p_{c,d}$ in each iteration, which introduces a significant complexity. In this subsection, MS based updated MP (MS-U-MP) detection is proposed to achieve a reduction in division operations by MS approximation strategy [37], which alleviates the complexity of the implementation, but with some performance loss.

We calculate the MAP rule (13) by the following approximation:

$$\begin{aligned} \hat{x}[c] &= \arg \max_{x \in \{\mathcal{S} \cup 0\}} \Pr(x[c] = x | \mathbf{y}, \mathbf{H}) \\ &= \arg \max_{x \in \{\mathcal{S} \cup 0\}} \Pr(x) \Pr(\mathbf{y} | x[c] = x, \mathbf{H}) \\ &\approx \arg \max_{x \in \{\mathcal{S} \cup 0\}} p_c(x) \prod_{d \in \mathcal{J}(c)} \Pr(y[d] | x[c] = x, \mathbf{H}). \end{aligned} \quad (26)$$

The interference term is still approximated by a Gaussian random variable, the same as the expression in (15). The mean and variance of the interference in the t -th iteration are given in (16) and (17), respectively.

The variable node $x[c]$ first computes the LLR as

$$\begin{aligned} \tilde{\alpha}_{c,d}^{(t)}(x) &= \log \frac{p_c^{(t-1)}(x) \prod_{e \in \mathcal{J}(c), e \neq d} \Pr(y[e] | x[c] = x, \mathbf{H})}{p_c^{(t-1)}(s) \prod_{e \in \mathcal{J}(c), e \neq d} \Pr(y[e] | x[c] = s, \mathbf{H})} \\ &= \log \frac{p_c^{(t-1)}(x)}{p_c^{(t-1)}(s)} + \sum_{e \in \mathcal{J}(c), e \neq d} \log \frac{\Pr(y[e] | x[c] = x, \mathbf{H})}{\Pr(y[e] | x[c] = s, \mathbf{H})} \\ &= \log \frac{p_c^{(t-1)}(x)}{p_c^{(t-1)}(s)} + \sum_{e \in \mathcal{J}(c), e \neq d} \log \frac{\zeta_{e,c}^{(t)}(x)}{\zeta_{e,c}^{(t)}(s)}. \end{aligned} \quad (27)$$

Here, $s \in \{\mathcal{S} \cup 0\}$. Then, the message from variable nodes $x[c]$ to observation node $\{y[d], d \in \mathcal{J}(c)\}$ can be calculated

by

$$\tilde{p}_{c,d}^{(t)}(x) = \frac{\exp(\tilde{\alpha}_{c,d}^{(t)}(x))}{\sum_{x \in \{\mathcal{S} \cup 0\}} \exp(\tilde{\alpha}_{c,d}^{(t)}(x))}. \quad (28)$$

By taking logarithms on both sides of (28) and replacing $\sum_{x \in \{\mathcal{S} \cup 0\}} \exp(\tilde{\alpha}_{c,d}^{(t)}(x))$ with the dominant term $\exp\left(\max_{x \in \mathcal{S} \cup 0} (\tilde{\alpha}_{c,d}^{(t)}(x))\right)$, we can obtain

$$\tilde{p}_{c,d}^{(t)}(x) \approx \exp\left(\tilde{\alpha}_{c,d}^{(t)}(x) - \max_{x \in \mathcal{S} \cup 0} (\tilde{\alpha}_{c,d}^{(t)}(x))\right). \quad (29)$$

To compensate for the loss due to approximation while maintaining a similar computational complexity, we use here the normalised MS, which gives a better approximation by multiplying (29) by a positive scaling factor ρ less than 1. In addition to extensive simulation, ρ can be selected by interpolation method [37] or deep neural network method [38]. With a damping factor $\Delta \in (0, 1]$, the practical message from $x[c]$ to $y[d]$ is

$$p_{c,d}^{(t)}(x) = \Delta \cdot \rho^{(t)} \cdot \tilde{p}_{c,d}^{(t)}(x) + (1 - \Delta) \cdot p_{c,d}^{(t-1)}(x). \quad (30)$$

The posterior LLR for each transmitting symbol is given as

$$\tilde{\alpha}_c^{(t)}(x) = \log \frac{p_c^{(t-1)}(x) \prod_{e \in \mathcal{J}(c)} \Pr(y[e] | x[c] = x, \mathbf{H})}{p_c^{(t-1)}(s) \prod_{e \in \mathcal{J}(c)} \Pr(y[e] | x[c] = s, \mathbf{H})}, \quad (31)$$

and the posterior probability is given as

$$p_c^{(t)}(x) = \frac{\exp(\tilde{\alpha}_c^{(t)}(x))}{\sum_{x \in \{\mathcal{S} \cup 0\}} \exp(\tilde{\alpha}_c^{(t)}(x))}. \quad (32)$$

The convergence indicator is also calculated by (21). The criterion for convergence and the subsequent process are the same as those in Section IV-C.

The U-MP detection calculates the probability according to (19) and (22), which require MNL division operations per iteration. The proposed MS-U-MP detection achieves an effective reduction in computational complexity by eliminating a large number of division operations. Compared to the original U-MP detection, MS-U-MP detection can achieve less performance loss by adding a small number of multiplications, which puts only a small pressure on the complexity.

V. PERFORMANCE ANALYSIS

In this section, we derive an ABEP upper bound of E-OTFS-IM for the ML detector. Substitute the channel representation in (10) into the input-output relationship in (12) and rewrite the input-output relationship as [6]

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^P h_i (\mathbf{F}_N \otimes \mathbf{G}_{\text{rx}}) \mathbf{\Pi}^{\mathbf{L}_i} \mathbf{\Delta}^{\kappa_i + \varepsilon_i} (\mathbf{F}_N^H \otimes \mathbf{G}_{\text{tx}}) \mathbf{x} + \mathbf{w} \\ &= \mathbf{\Phi}(\mathbf{x}) \mathbf{h} + \mathbf{w}, \end{aligned} \quad (33)$$

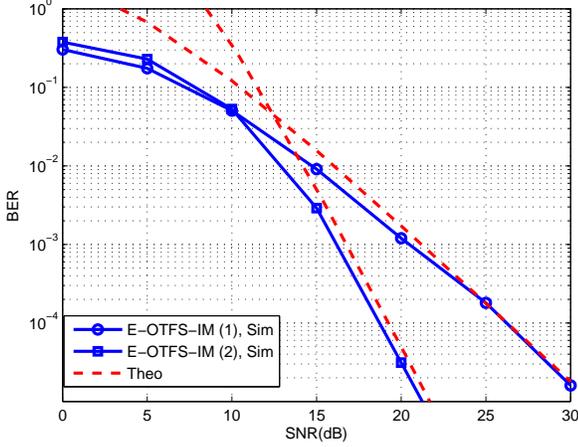


Fig. 4. Comparison of ABEP and simulation results of E-OTFS-IM.

where $\mathbf{h} = [h_1, \dots, h_P]^T$, and Φ is expressed as

$$\Phi(\mathbf{x}) = \begin{bmatrix} \Xi_1 & | & \dots & | & \Xi_P \\ \hline MN \times MN & & & & MN \times MN \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \vdots \\ \mathbf{x} \end{bmatrix}_{PMN \times P} \quad (34)$$

with $\Xi_i = (\mathbf{F}_N \otimes \mathbf{G}_{rx}) \mathbf{\Pi}^{L_i} \mathbf{\Delta}^{K_i + \epsilon_i} (\mathbf{F}_N^H \otimes \mathbf{G}_{tx})$. For the above system, the conditional pairwise error probability (PEP) for misjudging the transmitted \mathbf{x} as $\hat{\mathbf{x}}$ can be expressed as

$$\begin{aligned} \Pr(\mathbf{x} \rightarrow \hat{\mathbf{x}} | \mathbf{h}) &= Q \left(\sqrt{\frac{\|\Phi(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}\|^2}{2N_0}} \right) \\ &= Q \left(\sqrt{\frac{\mathbf{h}^H \Psi(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}}{2N_0}} \right), \end{aligned} \quad (35)$$

where $\Psi(\mathbf{x} - \hat{\mathbf{x}}) = \Phi^H(\mathbf{x} - \hat{\mathbf{x}})\Phi(\mathbf{x} - \hat{\mathbf{x}})$. Denote $\chi = \frac{\mathbf{h}^H \Psi(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}}{2N_0}$, according to the exponential approximation $Q(x) \approx \frac{1}{12}e^{-\frac{x^2}{2}} + \frac{1}{4}e^{-\frac{2}{3}x^2}$ [39], the PEP can be derived as

$$\Pr(\mathbf{x} \rightarrow \hat{\mathbf{x}}) \approx \mathbb{E}_\chi \left(\frac{1}{12}e^{-\frac{\chi}{2}} + \frac{1}{4}e^{-\frac{2}{3}\chi} \right). \quad (36)$$

Since $\Psi(\mathbf{x} - \hat{\mathbf{x}})$ is a non-negative definite Hermitian matrix, there exists a unitary matrix \mathbf{U} and an eigenvalue matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_P)$ satisfying $\Psi(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$. Thus, χ can be rewritten as $\chi = \frac{1}{2N_0} \sum_{q=1}^r \lambda_q |\tilde{h}_q|^2$, where $r = \text{rank}(\Psi(\mathbf{x} - \hat{\mathbf{x}}))$, λ_q s are the non-zero eigenvalues of $\Psi(\mathbf{x} - \hat{\mathbf{x}})$, and \tilde{h}_q is the q -th element in $\tilde{\mathbf{h}} = \mathbf{U}\mathbf{h}$ with $\tilde{h}_q \sim \mathcal{CN}(0, 1/P)$. The moment generating function of χ is

$$M_\chi(s) = \prod_{q=1}^r \left(1 - \frac{\lambda_q s}{2N_0 P} \right)^{-1}. \quad (37)$$

By substituting (37) into (29), we can obtain the expression of PEP as

$$\Pr(\mathbf{x} \rightarrow \hat{\mathbf{x}}) \approx \frac{1}{12} \prod_{q=1}^r \left(1 + \frac{\lambda_q}{4N_0 P} \right)^{-1} + \frac{1}{4} \prod_{q=1}^r \left(1 + \frac{\lambda_q}{3N_0 P} \right)^{-1} \quad (38)$$

According to the union boundary technique, the ABEP can be upper bounded by

$$P_e \leq \frac{1}{B2^B} \sum_{\mathbf{x}} \sum_{\hat{\mathbf{x}}} e(\mathbf{x}, \hat{\mathbf{x}}) \Pr(\mathbf{x} \rightarrow \hat{\mathbf{x}}), \quad (39)$$

where B indicates the total number of bits transmitted in an OTFS frame, and $e(\mathbf{x}, \hat{\mathbf{x}})$ is the Hamming distance between binary sequences associated with \mathbf{x} and $\hat{\mathbf{x}}$.

Furthermore, for high SNRs, (38) can be expressed as

$$\begin{aligned} \Pr(\mathbf{x} \rightarrow \hat{\mathbf{x}}) &\approx \frac{1}{12} \prod_{q=1}^r \left(\frac{\lambda_q}{4N_0 P} \right)^{-1} + \frac{1}{4} \prod_{q=1}^r \left(\frac{\lambda_q}{3N_0 P} \right)^{-1} \\ &= \frac{1}{12} \left[\left(\prod_{q=1}^r \lambda_q \right)^{1/r} \frac{1}{4N_0 P} \right]^{-r} \\ &\quad + \frac{1}{4} \left[\left(\prod_{q=1}^r \lambda_q \right)^{1/r} \frac{1}{3N_0 P} \right]^{-r}. \end{aligned} \quad (40)$$

We observe that the order of diversity of the E-OTFS-IM system is determined by r , which is $G_d = \min_{\forall(\mathbf{x}-\hat{\mathbf{x}})} r$. The coding gain of the E-OTFS-IM system is $G_c = \min_{\forall(\mathbf{x}-\hat{\mathbf{x}})} \left(\prod_{q=1}^r \lambda_q \right)^{1/r}$.

In Fig.4, we compare the BER curves detected by the ML detector and the curves for the derived ABEP upper bound. The parameters are $M = 4$, $N = 4$ and subcarrier spacing $\Delta f = 3.75$ kHz. The velocity of mobile user is set to 253 km/h. The parameters of the considered two systems are (1) $P = 2$, $G = 4$, $k_1 = 1$, $k_2 = 3$, 4QAM and BPSK for \mathcal{S}_1 and \mathcal{S}_2 , respectively; and (2) $P = 4$, $G = 2$, $k_1 = 2$, $k_2 = 4$, BPSK for \mathcal{S}_1 and \mathcal{S}_2 . The two considered systems have the same SE of 1.25 bps/Hz. It can be seen that the theoretical bounds get closer to the simulation results as the SNR increases, and match the simulation results well at high SNR. This verifies the validity of the derived upper bound. It can also be observed that larger diversity can be used to obtain better BER performance with the increase of the number of independent multipath P .

VI. THE PROPOSED E-OTFS-IM-IQ SCHEME

In E-OTFS-IM, the 2D modulation symbols in each sub-block is transmitted jointly via I- and Q-branches. Alternatively, decoupling the I- and Q-branches can add another degree of freedom to the system and thus make a faster increase in SE. In this section, we extend the principles of E-OTFS-IM to I- and Q-branches independently and propose the E-OTFS-IM-IQ scheme.

A. Implementation of E-OTFS-IM-IQ

For convenience of illustration, we follow a similar notation to Section IV. We apply two pulse amplitude modulation (PAM) constellations $\tilde{\mathcal{S}}_i$ and $\hat{\mathcal{S}}_i$ standing for the real and imaginary parts of an amplitude-phase constellation \mathcal{S}_i , respectively. The allowed numbers of active grids in a sub-block with n grids for I/Q branches are k_1, k_2, \dots, k_R ($0 < k_1 < k_2 < \dots < k_R \leq n$). Denote the GAP with k_i active grids for the I-branch

as $\bar{\mathbf{a}} = [\bar{a}_1, \dots, \bar{a}_n]^T \in \{0, 1\}^{n \times 1}$, $\|\bar{\mathbf{a}}\|_0 = k_i$, where $\bar{a}_j = 1$ if the j -th grid is active for the I-branch, and $\bar{a}_j = 0$, otherwise. Furthermore, the vector form of the g -th sub-block in the DD domain for the I-branch can be denoted as

$$\bar{\mathbf{x}}^g = [\bar{x}_1^g, \bar{x}_2^g, \dots, \bar{x}_n^g]^T, \quad (41)$$

where $\bar{x}_j^g \in \{0, \bar{\mathcal{S}}_i\}$ for $g = 1, \dots, G$ and $j = 1, \dots, n$. The GAP for $\bar{\mathbf{x}}^g$ can be presented by $\bar{\mathbf{a}}^g = [\bar{a}_1^g, \bar{a}_2^g, \dots, \bar{a}_n^g]^T$ with

$$\bar{a}_j^g = \begin{cases} 1, & \text{if } \bar{x}_{I,j}^g \in \bar{\mathcal{S}}_i \\ 0, & \text{if } \bar{x}_{I,j}^g = 0. \end{cases} \quad (42)$$

The Q-branch is modulated in a similar manner. The vector form of the g -th sub-block in the DD domain for the Q-branch is denoted as

$$\tilde{\mathbf{x}}^g = [\tilde{x}_1^g, \tilde{x}_2^g, \dots, \tilde{x}_n^g]^T, \quad (43)$$

where $\tilde{x}_j^g \in \{0, \tilde{\mathcal{S}}_i\}$ for $g = 1, \dots, G$ and $j = 1, \dots, n$. Afterwards, the two sub-blocks for the I- and Q- branches are combined to obtain the vector form of the g -th sub-block $\mathbf{x}^g \triangleq [x_1^g, \dots, x_n^g]^T = \bar{\mathbf{x}}^g + j\tilde{\mathbf{x}}^g$. Let \mathcal{X}_I and \mathcal{X}_Q be the sets of possible sub-block realizations for I- and Q- branches, respectively. The total number of possible GAPs for I/Q branch is $\sum_{i=1}^R \binom{n}{k_i}$. Thus, the total number of all possible sub-block realizations for the I-branch is

$$|\mathcal{X}_I| = \sum_{i=1}^R \binom{n}{k_i} |\bar{\mathcal{S}}_i|^{k_i}. \quad (44)$$

Similarly, the total number of all possible sub-block realizations for the Q-branch is

$$|\mathcal{X}_Q| = \sum_{i=1}^R \binom{n}{k_i} |\tilde{\mathcal{S}}_i|^{k_i}. \quad (45)$$

After obtaining the sub-block realizations for both of the I- and Q- branches, the total number of possible sub-block realizations for transmission can be calculated as $|\mathcal{X}_I| \cdot |\mathcal{X}_Q|$. Therefore, we can jointly map information bits of constant length $\lceil \log_2 (|\mathcal{X}_I| \cdot |\mathcal{X}_Q|) \rceil$ to a sub-block, and the SE of E-OTFS-IM-IQ is thus

$$\eta_{\text{E-OTFS-IM-IQ}} = \frac{1}{n} \lceil \log_2 (|\mathcal{X}_I| \cdot |\mathcal{X}_Q|) \rceil. \quad (46)$$

For clarity, we provide an example of E-OTFS-IM-IQ with $R = 2$, $k_1 = 1$, $k_2 = 2$ and $n = 4, 8$ in Table II. It can be observed that for the same modulation order, the SE of E-OTFS-IM-IQ can even reach twice the SE of E-OTFS-IM. It is clear that E-OTFS-IM-IQ with lower order constellations is able to achieve the same SE as E-OTFS-IM with higher order constellations. This suggests that E-OTFS-IM-IQ may achieve better BER performance than E-OTFS-IM.

B. Efficient Encoding Method

The mapping of the b bits to a sub-block can also be implemented by either a mapping table or a designed efficient encoding method for E-OTFS-IM-IQ. In this subsection, we propose a computationally effective E-OTFS-IM-IQ encoding method, which maps the information bits \mathbf{b} of constant length b to sub-block, as described below:

TABLE II
SE COMPARISON FOR E-OTFS-IM-IQ AND E-OTFS-IM
WITH $R = 2$, $k_1 = 1$, $k_2 = 2$ and $n = 4, 8$

n	Constellations	SE
4	I:[2PAM, 2PAM] Q:[2PAM, 2PAM]	2.5
4	[BPSK, BPSK]	1.25
4	[16QAM, 16QAM]	2.5
8	I:[2PAM, 2PAM] Q:[2PAM, 2PAM]	1.75
8	[BPSK, BPSK]	0.875
8	[32QAM, 32QAM]	1.75

- 1) Convert \mathbf{b} to a decimal number Z , $Z \in [0, 2^b - 1]$.
- 2) Calculate the parameters according to the following formula

$$Z_I = \left\lfloor \frac{Z}{|\mathcal{X}_Q|} \right\rfloor, \quad Z_Q = \text{mod}(Z, |\mathcal{X}_Q|). \quad (47)$$

- 3) Convert Z_I to a binary sequence $\bar{\mathbf{b}}$ of length $\lceil \log_2 (|\mathcal{X}_I|) \rceil$, and convert Z_Q to a binary sequence $\tilde{\mathbf{b}}$ of length $\lceil \log_2 (|\mathcal{X}_Q|) \rceil$.
- 4) Map $\bar{\mathbf{b}}$ and $\tilde{\mathbf{b}}$ according to the method in Section IV-B to construct $\bar{\mathbf{x}}$ and $\tilde{\mathbf{x}}$, respectively. The complex sub-block for transmission is $\mathbf{x} = \bar{\mathbf{x}} + j\tilde{\mathbf{x}}$.

We take $n = 4$, $k_1 = 1$, $k_2 = 3$, the 2PAM constellation $\{1, -1\}$ for $\bar{\mathcal{S}}_1, \bar{\mathcal{S}}_2, \tilde{\mathcal{S}}_1$ and $\tilde{\mathcal{S}}_2$ to demonstrate the mapping method. We assume the information bits as $\mathbf{b} = [1010101101]$. Without losing generality, we take g -th sub-block to illustrate. We have $|\mathcal{X}_I| = 40$, $|\mathcal{X}_Q| = 40$, $\lceil \log_2 (|\mathcal{X}_I|) \rceil = 6$, and $\lceil \log_2 (|\mathcal{X}_Q|) \rceil = 6$.

First of all, \mathbf{b} is converted to a decimal number $Z = 685$. The decimal number representing the sub-block information for the I-branch is calculated as $Z_I = \lfloor Z/|\mathcal{X}_Q| \rfloor = \lfloor 685/40 \rfloor = 17$, and the decimal number representing the sub-block information of the Q-component is calculated as $Z_Q = \text{mod}(Z, |\mathcal{X}_Q|) = \text{mod}(685, 40) = 5$.

Next, $Z_I = 17$ and $Z_Q = 5$ are converted to the 6-bit binary sequences $\bar{\mathbf{b}} = [010001]$ and $\tilde{\mathbf{b}} = [000101]$, respectively.

Then, $\bar{\mathbf{b}}$ is used to construct $\bar{\mathbf{x}}$. Since $Z_I = 17$, we have $r = 2$ and the compensation value $Z_c = 1$. Convert the first three bits $[010]$ to $Z_1 = 2$ and use $Z_1 - Z_c = 1$ to obtain the GAP for I-branch $\bar{\mathbf{a}}^g = [1, 1, 0, 1]^T$. The last three bits $[001]$ are modulated to three 2PAM symbols $\bar{\mathbf{s}}^g = [-1, -1, 1]^T$. The constructed sub-block for I-branch is $\bar{\mathbf{x}}^g = [-1, -1, 0, 1]^T$. Similarly, $\tilde{\mathbf{b}}$ is used to construct $\tilde{\mathbf{x}}$. Since $Z_Q = 5$, we have $r = 1$ and the compensation value $Z_c = 0$. The first five bits $[00010]$ are used to obtain the GAP for Q-branch $\tilde{\mathbf{a}} = [0, 0, 1, 0]^T$ and the last bit $[1]$ is modulated to $\tilde{\mathbf{s}}^g = [1]^T$. The constructed sub-block for Q-branch is $\tilde{\mathbf{x}}^g = [0, 0, 1, 0]^T$. The complex sub-block for transmission is $\mathbf{x}^g = \bar{\mathbf{x}}^g + j\tilde{\mathbf{x}}^g = [-1, -1, j, 1]^T$. ■

TABLE III
SIMULATION PARAMETERS

Parameters	Value
Carrier frequency (GHz)	4
Subcarrier spacing (kHz)	15
Number of paths (P)	4
Size of a frame (M, N)	(32,32)
User equipment speed (km/h)	506

C. Updated MP Detection

The updated MP algorithm for E-OTFS-IM-IQ can be performed in a similar manner to that for E-OTFS-IM. The received signal in (12) can be rewritten in the real valued field as

$$\underbrace{\begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix}}_{\mathbf{y}'} = \underbrace{\begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}}_{\mathbf{H}'} \underbrace{\begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}}_{\mathbf{x}'} + \underbrace{\begin{bmatrix} \Re(\mathbf{w}) \\ \Im(\mathbf{w}) \end{bmatrix}}_{\mathbf{w}'}, \quad (48)$$

where $\mathbf{y}' \in \mathbb{R}^{2MN \times 1}$, $\mathbf{H}' \in \mathbb{R}^{2MN \times 2MN}$, $\mathbf{x}' \in \mathbb{R}^{2MN \times 1}$ and $\mathbf{w}' \in \mathbb{R}^{2MN \times 1}$. Subsequently, the procedures in Section IV-C can be implemented similarly. We omit the process for the sake of brevity. After obtaining the estimates of the real and imaginary parts of the received signal, the reverse procedure of the encoding procedure is implemented sub-block by sub-block. \hat{Z}_I and \hat{Z}_Q for each sub-block can be obtained, and \hat{Z} is calculated from $\hat{Z} = \hat{Z}_I \cdot |\mathcal{X}_Q| + \hat{Z}_Q$. Finally, convert \hat{Z} to binary to get an estimate of the transmitted information bits.

VII. SIMULATION RESULTS

In this section, computer simulations of the proposed E-OTFS-IM and E-OTFS-IM-IQ are carried out under the assumption of perfect channel estimation. The simulation parameters for are shown in Table III. The Doppler parameters are given as $\nu_i = \nu_{max} \cos(\theta_i)$, where ν_{max} is the maximum Doppler shift corresponding to the user equipment speed and $\theta_i \sim \mathcal{U}[-\pi, \pi]$ [10]. The maximum delay is $\tau_{max} = \frac{3}{M\Delta f}$. We set $\Delta = 0.4$, $\epsilon = 0.01$, and the maximum iteration number $t_{max} = 20$.

The abbreviation “E-OTFS-IM ($n, \{k_1, \dots, k_R\}, [\mathcal{S}_1, \dots, \mathcal{S}_R]$)” denotes the E-OTFS-IM scheme with sub-block length n , the R allowed numbers of active grids and the R adopted constellations. “E-OTFS-IM-IQ ($n, \{k_1, \dots, k_R\}, [\tilde{\mathcal{S}}_1, \dots, \tilde{\mathcal{S}}_R], [\hat{\mathcal{S}}_1, \dots, \hat{\mathcal{S}}_R]$)” denotes the E-OTFS-IM-IQ scheme with sub-block length n , the R allowed numbers of active grids and the PAM constellations for I- and Q- branches. “OTFS-IM (n, k, \mathcal{S})” denotes the OTFS-IM scheme with sub-block length n , k active grids and the adopted constellation.

From Fig. 5, it can be observed that “OTFS-IM (4, 3, BPSK)” can transmit 5 bits per sub-block with the SE of 1.25 bps/Hz. E-OTFS-IM can flexibly increase the bits transmitted per sub-block by adjusting active grids and adopted constellations. It can also be observed that “E-OTFS-IM (4, {1, 2, 3}, [BPSK BPSK, BPSK])” can transmit 6 bits per sub-block for the SE of 1.5 bps/Hz, achieving a 20% improvement in SE compared with “OTFS-IM (4, 3, BPSK)”. In the low SNR region, due to the protection of more silent grids, the BER

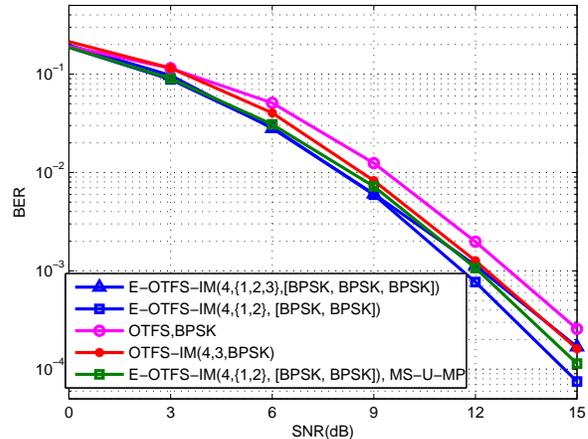


Fig. 5. BER performance comparison of E-OTFS-IM, OTFS-IM and conventional OTFS with $n = 4$ and BPSK.

performance of “E-OTFS-IM (4, {1, 2, 3}, [BPSK BPSK, BPSK])” is better. In high SNR region, since the systems are almost immune to GAP detection errors, their performance is almost the same. Compared to “OTFS, BPSK”, “E-OTFS-IM (4, {1, 2, 3}, [BPSK, BPSK, BPSK])” achieves a 50% improvement in SE and a 0.75 dB performance gain at the BER of 10^{-3} . In addition, “E-OTFS-IM (4, {1, 2}, [BPSK, BPSK])” achieves a 0.7 dB performance gain compared with “OTFS-IM (4, 3, BPSK)” at the BER of 10^{-3} . Since the average number of active grids per sub-block of “E-OTFS-IM (4, {1, 2}, [BPSK, BPSK])” is 1.75 less than 3, it is more resilient against the Doppler effect. Compared to “OTFS, BPSK”, “E-OTFS-IM (4, {1, 2}, [BPSK, BPSK])” achieves a 1.5 dB performance gain at the BER of 10^{-3} and a 25% improvement in SE as well. It can also be seen that the proposed MS-U-MP detection approaches the U-MP detection with an acceptable loss of performance. Specifically, the BER performance gap between U-MP and MS-U-MP is about 0.5 dB at a BER of 10^{-3} .

We can observe from Fig.(6a) that, for the same SE of 2 bps/Hz, “E-OTFS-IM (4, {1, 3}, [4QAM, 4QAM])” achieves 0.6dB and 2.2dB performance gain over “OTFS-IM (4, 3, 4QAM)” and “OTFS, 4QAM” at the BER of 10^{-2} , respectively. The inactive grids in DD domain can be regarded as guard symbols resulting in better performance, and the increasing number of inactive grids in DD domain can improve the tolerability against Doppler effect. The average number of active grids per sub-block of “E-OTFS-IM (4, {1, 3}, [4QAM, 4QAM])” is less than 3, which makes the “E-OTFS-IM (4, {1, 3}, [4QAM, 4QAM])” have better performance. We can observe from Fig.6b that “E-OTFS-IM (8, {1, 3}, [4QAM, 4QAM])” with the SE of 1.375 bps/Hz outperforms “OTFS-IM (8, 5, BPSK)” with the SE of 1.25 bps/Hz. The average number of active grids per sub-block is 2.96875, which is much smaller than 5, leading to more inactive grids in DD domain. Even if 4QAM is adopted, the BER performance is better. “E-OTFS-IM (8, {3, 5}, [4QAM, BPSK])” with the SE of 1.5bps/Hz outperforms “OTFS-IM (8, 5, BPSK)”, and a noticeable performance gain appears in the region of

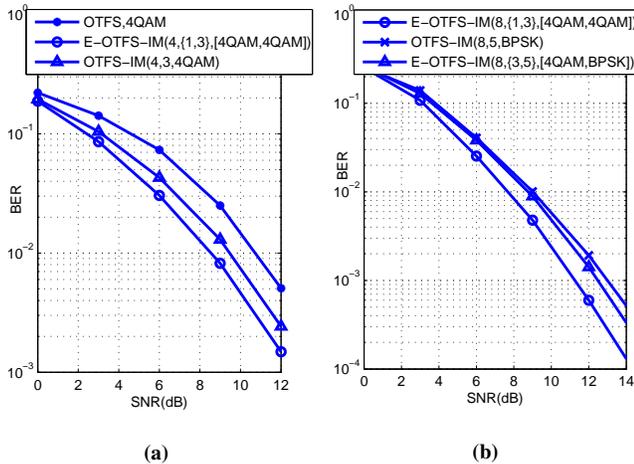


Fig. 6. BER performance comparison of E-OTFS-IM, OTFS-IM and conventional OTFS with (a) $n = 4$; (b) $n = 8$.

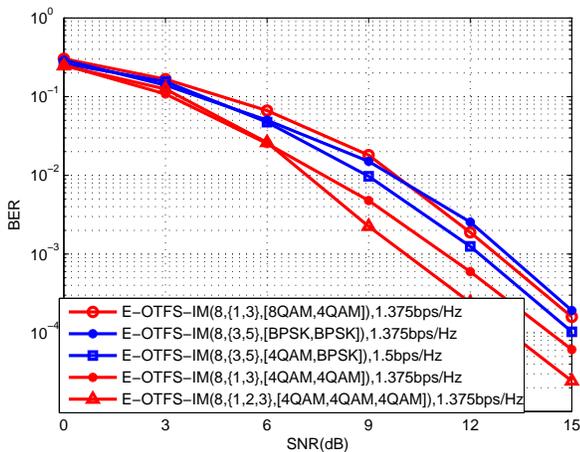


Fig. 7. BER performance comparison for E-OTFS-IM with different configurations.

SNR > 6 dB. The sub-blocks with three active grids and 4QAM provide better BER performance. The variable number of active grids in the E-OTFS-IM leads to a larger number of sub-block realizations with a large Euclidean distance. From the above observations, it can be concluded that by subtly adjusting the active grids and adopted constellations, both SE and BER performance can be improved.

To further observe the effect of variable active grids and different constellations on performance, in Fig.7, we compare the BER performance of E-OTFS-IM with $n = 8$. It can be observed that as SNR increases, “E-OTFS-IM (8, {3, 5}, [BPSK, BPSK])” exhibits the worst performance of all systems. This can be explained that the sub-blocks with five active grids and BPSK are predominant for “E-OTFS-IM (8, {3, 5}, [BPSK, BPSK])” (1600 of total 2048 possible sun-block realizations), while the sub-blocks with three active grids and 4QAM are predominant for the other four systems. Since the sub-blocks with three active grids and 4QAM outperform significantly the sub-blocks with five active grids and BPSK, this leads to “E-OTFS-IM(8, {3, 5}, [BPSK, BPSK])” exhibits the worst

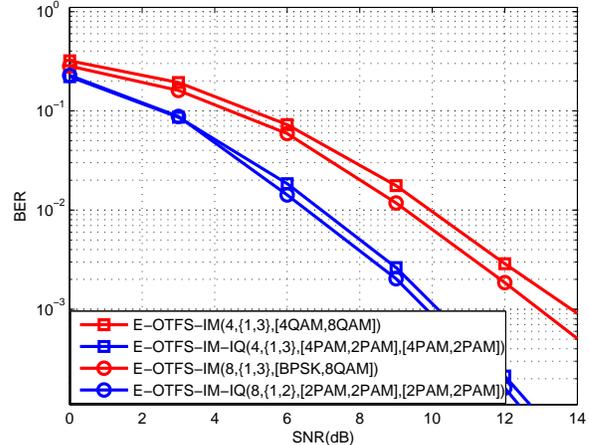


Fig. 8. BER performance comparison of E-OTFS-IM and E-OTFS-IM-IQ with different configurations.

performance. “E-OTFS-IM, (8, {1, 3}, [8QAM, 4QAM])” with the SE of 1.375bps/Hz and “E-OTFS-IM (8, {3, 5}, [4QAM, BPSK])” with the SE of 1.5bps/Hz achieve a close performance. And in the high SNR region, the performance of “E-OTFS-IM (8, {1, 3}, [8QAM, 4QAM])” is slightly inferior to “E-OTFS-IM (8, {3, 5}, [4QAM, BPSK])”. This can be explained that although the sub-blocks with three active grids and 4QAM is predominant, 8QAM causes the performance loss compared to BPSK and 4QAM. Not only that, “E-OTFS-IM (8, {3, 5}, [4QAM, BPSK])” has a higher SE than “E-OTFS-IM (8, {1, 3}, [8QAM, 4QAM])” and “E-OTFS-IM (8, {3, 5}, [BPSK, BPSK])”. From the above observations, it can be concluded that not only SE can be improved by adjusting k_i s and different constellations, but also BER performance can be improved. “E-OTFS-IM (8, {1, 3}, [4QAM, 4QAM])” achieves better performance than “E-OTFS-IM (8, {3, 5}, [4QAM, BPSK])” because the sub-blocks with one active grid and 4QAM outperform the sub-blocks with five active grids and BPSK. And “E-OTFS-IM (8, {1, 2, 3}, [4QAM, 4QAM, 4QAM])” performs better than “E-OTFS-IM (8, {1, 3}, [4QAM, 4QAM])” due to the less average number of active grids (2.375 versus 2.5).

In Fig.8, we compare the BER performance of E-OTFS-IM and E-OTFS-IM-IQ with different configurations. For the SE of 2.75 bps/Hz, “E-OTFS-IM-IQ (4, {1, 3}, [4PAM, 2PAM], [4PAM, 2PAM])” achieves a 4 dB performance gain over “E-OTFS-IM (4, {1, 3}, [4QAM, 8QAM])” at the BER of 10^{-3} . This is because lower-order modulations are employed in E-OTFS-IM-IQ. In E-OTFS-IM, however, higher-order modulations must be adopted to achieve the same SE. For the SE of 1.75 bps/Hz, “E-OTFS-IM-IQ (8, {1, 2}, [2PAM, 2PAM], [2PAM, 2PAM])” achieves a 3 dB performance gain over “E-OTFS-IM (8, {1, 3}, [BPSK, 8QAM])” at the BER of 10^{-3} . This can be interpreted as that E-OTFS-IM-IQ not only employs lower-order modulations, but also activates fewer grids than E-OTFS-IM, which improves the BER performance. The results in Fig.5 show that E-OTFS-IM-IQ achieves a significant BER performance improvement compared to E-

OTFS-IM and E-OTFS-IM-IQ can employ lower-order signal constellations and/or activate fewer grids to achieve the same SE as E-OTFS-IM. Due to the presence of a large number of sub-blocks with low-order constellations, E-OTFS-IM-IQ achieves a larger Euclidean distance than E-OTFS-IM, resulting in a better performance.

VIII. CONCLUSION

In this paper, we proposed the E-OTFS-IM scheme by permitting the number of active grids conveying modulation symbols from any constellation to be changed. The proposed E-OTFS-IM resolves the ambiguous detection by transmitting a constant length of information bits via joint mapping. We also proposed an efficient encoding method for E-OTFS-IM. In addition, we updated the MP detector with LLR criterion for our proposed E-OTFS-IM, and proposed its low-complexity improvement with MS approximation strategy for E-OTFS-IM. We derived an ABEP upper bound for the E-OTFS-IM system, which is verified to be tight in high SNR region by simulation. We further proposed E-OTFS-IM-IQ to improve the SE and the error performance of E-OTFS-IM, where E-OTFS-IM-IQ applies the principle of E-OTFS-IM independently to the I- and Q-branches for more possible sub-block realizations. Likewise, we provided the corresponding efficient encoding method and the updated MP detector for E-OTFS-IM-IQ. Simulation results showed that E-OTFS-IM outperforms conventional OTFS and OTFS-IM under the same transmission rate, and E-OTFS-IM-IQ achieves better BER performance than E-OTFS-IM.

In the future, first, variational Bayes algorithm and expectation propagation algorithm for the proposed schemes can be utilized to further reduce complexity and improve detection accuracy. Second, bit-wise detection for IM aided OTFS deserves further research.

REFERENCES

- [1] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359-383, Jan. 2022.
- [2] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *Proc. 2017 IEEE Wireless Commun. Netw. Conf.*, San Francisco, CA, USA, Mar. 2017, pp. 1-7.
- [3] 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 Commun.*, vol. 28, no. 4, pp. 136-144, Aug. 2021.
- [4] L. Xiao, S. Li, Y. Qian, and D. Chen, and T. Jiang, "An Overview of OTFS for Internet of Things: Concepts, Benefits, and Challenges," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7596-7618, May. 2022.
- [5] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3049-3063, Jun. 2019.
- [6] E. Biglieri, P. Raviteja, and Y. Hong, "Error performance of orthogonal time frequency space (OTFS) modulation," in *Proc. IEEE Int. Conf. Commun. Workshops*, Shanghai, China, May. 2019, pp. 1-6.
- [7] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Effective diversity of OTFS modulation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 249-253, Feb. 2020.
- [8] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 957-961, Jan. 2019.
- [9] G. D. Surabhi, and A. Chockalingam, "Low-complexity linear equalization for otfs modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330-334, Feb. 2020.
- [10] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501-6515, Oct. 2018.
- [11] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational Bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7976-7980, Apr. 2020.
- [12] Z. Yuan, F. Liu, W. Yuan, Q. Guo, Z. Wang, and J. Yuan, "Iterative detection for orthogonal time frequency space modulation using approximate message passing with unitary transformation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 714-725, Jul. 2021.
- [13] H. Li, Y. Dong, C. Gong, Z. Zhang, X. Wang, and X. Dai, "Low complexity receiver via expectation propagation for OTFS modulation," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3180-3184, Oct. 2021.
- [14] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906-4917, May. 2019.
- [15] 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 Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7773-7785, Dec. 2021.
- [16] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168-175, Jun. 2016.
- [17] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693-16746, 2017.
- [18] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 315-348, 1st Quart., 2019.
- [19] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536-5549, Nov. 2013.
- [20] R. Fan, Y. J. Yu, and Y. L. Guan, "Generalization of orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5350-5359, Oct. 2015.
- [21] B. Zheng, F. Chen, M. Wen, F. Ji, H. Yu, and Y. Liu, "Low-complexity ML detector and performance analysis for OFDM with in-phase/quadrature index modulation," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1893-1896, Nov. 2015.
- [22] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4786-4801, Jul. 2017.
- [23] A. M. Jaradat, J. M. Hamamreh, and H. Arslan, "OFDM with subcarrier number modulation," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 914-917, Dec. 2018.
- [24] S. Dang, G. Ma, B. Shihada, and M.-S. Alouini, "Enhanced orthogonal frequency-division multiplexing with subcarrier number modulation," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7907-7920, Oct. 2019.
- [25] M. Wen, J. Li, S. Dang, Q. Li, S. Mumtaz, and H. Arslan, "Joint-mapping orthogonal frequency division multiplexing with subcarrier number modulation" *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4306-4318, Jul. 2021.
- [26] Y. Liang, L. Li, P. Fan and Y. Guan, "Doppler resilient orthogonal time frequency space (OTFS) systems based on index modulation," in *Proc. 2020 IEEE 91st Veh. Technol. Conf. (VTC2020-Spring)*, Antwerp, Belgium, May. 2020, pp. 1-5.
- [27] J. Fu, C. Hou, W. Xiang, L. Yan, and Y. Hou, "Orthogonal time frequency space (OTFS) with dual-mode index modulation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 991-995, May. 2021.
- [28] D. Feng, J. Zheng, B. Bai, J. Jiang, and L. Zheng, "In-phase and quadrature index modulation aided OTFS transmission," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1318-1322, Jun. 2022.
- [29] H. Ren, W. Xu, and L. Wang, "Orthogonal time-frequency space with improved index modulation," in *Proc. 2021 15th Int. Conf. Signal Process. and Commun. Syst. (ICSPCS)*, Sydney, Australia, Dec. 2021, pp. 1-6.
- [30] M. Qian, Y. Ge, M. Wen, and F. Ji, "A novel block-wise index modulation scheme for high-mobility OTFS communications," in *Proc. 2022 IEEE/CIC Int. Conf. Commun. in China*, Sanshui, Foshan, China, Aug. 2022, pp. 1-5.
- [31] M. Qian, F. Ji, Y. Ge, M. Wen, X. Cheng, and H. V. Poor, "Block-Wise Index Modulation and Receiver Design for High-Mobility OTFS Communications," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5726-5739, Oct. 2023.
- [32] J. K. Francis, R. M. Augustine, and A. Chockalingam, "Diversity and PAPR enhancement in OTFS using indexing," in *Proc. 2021 IEEE 93rd Veh. Tech. Conf. (VTC2021-Spring)*, Helsinki, Finland, Apr. 2021, pp. 1-6.

- [33] S. Dogan-Tusha, A. Tusha, S. Althunibat, and K. Qaraqe, "Orthogonal Time Frequency Space Multiple Access Using Index Modulation," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 15858-15866, Dec. 2023.
- [34] L. Li, L. Xiao, X. Zhang, G. Liu, S. Li, and T. Jiang, "An efficient symbol-by-symbol aided expectation propagation detector for OTFS with index modulation," *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2046-2050, Jul. 2022.
- [35] Y. Wu, L. Xiao, Y. Xie, G. Liu, and T. Jiang, "Efficient signal detector design for OTFS with index modulation," *Digit. Commun. Netw.*, <https://doi.org/10.1016/j.dcan.2022.10.026>.
- [36] C.-C. Cheng, H. Sari, S. Sezginer, and Y. T. Su, "Enhanced spatial modulation with multiple signal constellations," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2237-2248, Jun. 2015.
- [37] Y. Zhang, L. Ge, X. You, and C. Zhang, "Belief propagation detection based on max-sum algorithm for massive MIMO systems," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Nanjing, China, Dec. 2017, pp. 1-6.
- [38] X. Tan, W. Xu, K. Sun, Y. Xu, Y. Beery, X. You and C. Zhang, "Improving massive MIMO message passing detectors with deep neural network," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1267-1280, Feb. 2020.
- [39] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840-845, Jul. 2003.



Kai Zhang received the B.S. degree in electronic and information engineering from Wenhua College of Huazhong University of Science and Technology, Wuhan, China. He is currently pursuing the Ph.D. degree in information and communication intelligence system with Donghua University, Shanghai, China. His main research interests include wireless communications, index modulation and MIMO.



Xue-Qin Jiang (Senior Member, IEEE) received the B.S. degree in computer science from the Nanjing Institute of Technology, Nanjing, China. He received the M.S and Ph.D. degrees in electronics engineering from Chonbuk National University, Jeonju, South Korea. He is now a Full Professor with the School of Information Science and Technology, Donghua University, Shanghai, China. He has authored or coauthored more than 60 SCI papers. His main research interests include wireless communications, quantum key distribution, and channel coding. He is currently serving as an Editor for IEEE Communications Letters and IEEE Access.



Han Hai (Member, IEEE) received his B.S. degree in information and computing sciences from South Central University for Nationalities, P.R. China, and the M.S. and the Ph.D. degrees in electronics engineering from Chonbuk National University, Korea, in 2013 and 2018, respectively. He is currently an assistant professor at College of Information Science and Technology, Donghua University, Shanghai, China. His research interests include wireless communications, MIMO and spatial modulation.



Runhe Qiu received the B.S. degree from Shanghai University of Science and Technology, Shanghai, in 1983 and the M.S. degree from China University of Mining and Technology, Xuzhou, in 1995, both in electrical engineering. He received Ph.D. degree in physical radio engineering from East China Normal University, in 2007. As a professor, he currently works in College of Information Science and Technology, Donghua University. He is the author/co-author of more than 70 technical papers. His research interests include cognitive radio technology, cooperative communication, wireless broad-band communication and application.



Shahid Mumtaz (Senior Member, IEEE) is an IET Fellow, IEEE ComSoc and ACM Distinguished Speaker, recipient of IEEE ComSoc Young Researcher Award (2020), founder and EiC of the IET Journal of Quantum Communication, Vice-Chair of the Europe/Africa Region IEEE Com Soc Green Communications Computing Society, and Vice-Chair of IEEE Standard P1932.1: Standard for Licensed/Unlicensed Spectrum Interoperability in Wireless Mobile Networks. He is the author of 4 technical books, 12 book chapters, 300+ technical papers (200+ IEEE journals/transactions, 100+ conference proceedings), and received 2 IEEE best paper awards in the area of mobile communications. Most of his publication is in the field of wireless communication. He is serving as Scientific Expert and Evaluator for various research funding agencies. He was awarded an Alain Bensoussan Fellowship in 2012. He was the recipient of the NSFC Researcher Fund for Young Scientist in 2017 from China.