Online Bayesian Learning Aided Sparse CSI Estimation in OTFS Modulated MIMO Systems for Ultra-High-Doppler Scenarios

Anand Mehrotra *Graduate Student Member, IEEE*, Suraj Srivastava *Member, IEEE*, Shaik Asifa, Aditya K. Jagannatham *Senior Member, IEEE*, Lajos Hanzo *Life Fellow, IEEE*

***Abstract*—Online Bayesian learning-assisted channel state information (CSI) estimation schemes are conceived for single input single output (SISO) and multiple input multiple output (MIMO) orthogonal time frequency space (OTFS) modulated systems. To begin with, an end-to-end system model is derived in the delay-Doppler (DD)-domain, followed by an online CSI estimation (CE) framework for SISO-OTFS systems. Next, the sequential minimum mean square error (MMSE) estimator is derived for this model which utilizes expectation maximization (EM) based sparse Bayesian learning (SBL) for initialization of the online estimation procedure. Additionally, a low-complexity detection technique is developed for the system under consideration, which is accomplished via an analogous time-frequency (TF)-domain system model that leads to a block-diagonal TF-domain channel matrix. The paradigm designed for online CE is subsequently extended to MIMO-OTFS systems. The corresponding DD-domain CSI is shown to be simultaneously row and group sparse. Hence a novel EM-based row and group sparse Bayesian learning scheme is developed for determining the initialization parameters for the above online algorithm. As a further continuation, a low-complexity detector is also proposed for MIMO-OTFS systems based on an iterative block matrix inversion technique. Furthermore, time-recursive Bayesian Cramer-Rao lower bounds (BCRLBs) are derived to benchmark the MSE performance of the proposed schemes for both the systems. Finally, simulation results are presented to demonstrate the efficiency of the proposed online estimation techniques.**

***Index Terms*—OTFS, DD-domain, sparse CE, high-mobility, Bayesian learning, low-complexity detector.**

I. Introduction

Ultra-high Doppler arises when the transmitter and the receiver are moving relative to one another at high velocities, such as in high-speed trains [1], [2] with velocities in the range 400-500Km/hr or for air-plane users [3] travelling at speeds in the range of 800-1000 km/hr. In such a scenario, the current generation of multi-carrier modulation schemes, such as orthogonal frequency division multiplexing (OFDM), suffer from a high level of inter-carrier interference (ICI) [4]–[7]. This arises due to the significant Doppler shift that perturbs the orthogonality of the subcarriers in an OFDM system. As a result, the performance of the system is considerably degraded. The ICI analysis of OFDM systems under ultra-high Doppler is presented by the authors of [8], together with a time-domain filtering technique for combating ICI. Along similar lines, the performance degradation of these systems due to Doppler spreading is investigated by the authors of [9]. Furthermore, several channel estimation approaches, such as Doppler-aided channel estimation with parallel ICI cancellation [10] and interpolation-based channel estimation [11], have been presented to improve the quality of channel estimates in such high Doppler systems. In addition, the rapidly varying channel in such scenarios also necessitates its frequent estimation, which leads to a significant increase in the pilot overhead.

To address this issue and enhance system performance in high-Doppler scenarios, the novel orthogonal time frequency and space (OTFS) modulation scheme has been introduced [5], [12], [13] precisely addresses this challenge. The fact that OTFS operates in the DD-domain, where the wireless channel is time-invariant over a significantly longer observation window, is one of its key characteristics. This mitigates the degradation caused by high Doppler, allowing for the high-integrity demodulation of the information symbols in high-speed scenarios. However, the accuracy of the CSI estimate (CE) obtained is of critical importance in OTFS systems, since it affects the overall performance of the system. Hence, we proceed with a brief overview of the research contributions relevant to OTFS channel estimation in the open literature.

A. Literature review

The conventional CE scheme includes the transmission of pilot impulses in a single OTFS frame [5], [14], [15]. The impulse-based technique has subsequently been extended to channel estimation in MIMO-OTFS systems [16]. While this approach has significant appeal due to its low complexity, a noteworthy disadvantage is the associated spectral inefficiency, which arises due to the fact that the entire frame is comprised of pilot symbols. To address this issue, [15], [17] introduced an embedded pilot-based CSI estimator, where the data and pilot symbols are sent in the same OTFS frame with sufficient guards to eliminate interference. In an environment characterized by an ultra-high Doppler, the maximum Doppler ν_{max}

L. 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)
The work of Aditya K. Jagannatham was supported in part by the Qualcomm Innovation Fellowship; in part by the Qualcomm 6G UR Gift; and in part by the Arun Kumar Chair Professorship.
A. Mehrotra, S. Srivastava, and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, UP 208016, India (e-mail: anandme@iitk.ac.in, ssrivast@iitk.ac.in, adityaj@iitk.ac.in), S. Asifa is with Mediatek Pvt. Ltd. India (e-mail: asifashaik20@iitk.ac.in), L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

is very high. This directly impacts traditional CSI estimation methods discussed above [18]. Due to the high maximum Doppler ν_{max}, it becomes necessary to allocate additional guard symbols around the pilot symbols along the Doppler axis, albeit this is only possible at the cost of reduced spectral efficiency. The problem is even more severe in the MIMO OTFS scenario, wherein pilot impulses are transmitted in the DD-domain with suitable guard intervals between the pilots of different transmit antennas to avoid overlap. Moreover, in such a scenario the interference also increases both with the delay τ and Doppler ν of the channel. In particular, they are affected by the value of the cross-ambiguity function [19].

Additionally, the DD-domain wireless channel response [4], [19] typically has a small number of dominant reflectors, which leads to a sparse channel. Consequently, exploiting this characteristic feature, one may estimate it at a much lower pilot overhead than is possible in a standard TF-domain OFDM system [17]. By reformulating the task of CE as a sparse estimation problem, researchers have been able to take advantage of the sparse nature of DD-domain wireless channel [21]–[25], thereby achieving a significant improvement in the estimation performance over the simplistic impulse-based pilot schemes. The seminal research paper by Shen *et al.* [21] provides a methodology for sparse downlink CE in a massive MIMO-OTFS system. This approach takes into account the sparsity of the channel and employs the orthogonal matching pursuit (OMP) for sparse signal recovery. The authors of [23] extend this framework to an uplink OTFS-based (MA) scenario and thereafter use an OMP and subspace pursuit-based algorithm. Zhao *et al.* in [22] proposed a novel pilot pattern that does not require a DD-domain-guard band. This special frame is then exploited to develop the pertinent sparse channel estimation problem, but exclusively for SISO-OTFS systems. In another recent addition, the authors of [25] exploited the sparse signal recovery paradigm for the estimation of the DD domain channel by employing the powerful Bayesian learning framework in order to determine the estimate. However, the CE scheme therein is limited to SISO-OTFS systems and it is difficult to extend to MIMO systems. To decrease the pilot overhead, the authors of [24], [27] proposed a scheme that transmits the TF-domain pilots on a common TF resource block for all the TAs.

To address the challenge posed by ultra-high Doppler effects, an intriguing alternative involves leveraging online estimation techniques, as previously explored by the authors [28], [29] within the context of OFDM systems. This prior research has inspired us to investigate the applicability of such an approach to OTFS-based systems. Nevertheless, it is important to note that a direct adaptation of these OFDM-based techniques to OTFS systems is not feasible, primarily because the former relies on exploiting angular domain sparsity, while the latter exhibits sparsity in the DD-domain.

Therefore, our proposed estimation approach relies on developing a sequential estimation framework. This framework adds a cyclic prefix to each pilot vector in the time domain. Following this they are transmitted sequentially. This approach offers a notable advantage, eliminating the need for guard bands, which depend on the Doppler parameter. Furthermore, the proposed scheme demonstrates consistent performance across a wide range of frequencies, as long as the parameters are thoughtfully selected to meet the condition $\nu_{max} < 1/T$ and $\tau_{max} < 1/\Delta f$ [19]. This adaptability sets our scheme apart from other conventional methods, making it an attractive solution for scenarios characterized by high Doppler shifts in both SISO and MIMO OTFS systems.

In these systems, it is also common practice to use the non-linear recursive message passing (MP) detector [19] or variational Bayes (VB) detector [30]. These schemes, however, assume that perfect CSI information is available at the receiver. Furthermore, imperfect CSI scenarios present many challenges, such as miss-detection or false detection, which occur when the number of multipath components detected is less or more, respectively. Furthermore, the existing methods operate iteratively, potentially encountering convergence challenges. Similarly, the authors of [31], [26], [32] explored low-complexity detectors in the DD-domain. By utilizing circulant and quasi-banded structures. By contrast, the author of [33] proposes a modified MRC detector for the DD-domain and its low-complexity delay-time domain variant is also implemented in this treatise. Another appealing alternative is to develop low-complexity TF-domain linear detectors based on the zero-forcing (ZF) or minimal mean squared error (MMSE) principles, which can be easily implemented on OFDM-based OTFS design using pre- and post-processing blocks.

The problem statement formulated in this paper proposes a novel system model for sequential estimation of the DD-domain CSI for SISO systems. Furthermore, it also extends the work to the MIMO-OTFS scenario, where it also exploits the inherent row and group sparsity along with sequential estimation. The other existing BL-based OTFS CE schemes [5], [14], [15], [16], [17] use batch processing for the estimation of the channel coefficients. Briefly, in batch processing, the entire block of pilot symbols of size $M \times N_p$ is processed at a single instant, which increases the computational cost and processing delay. By contrast, in our case, we have used BL for sequential estimation, which is based on individually processing each $M \times 1$ length pilot vector that helps significantly reduce the processing delay of parameters estimation. The novel contributions of this paper are described next and are also boldly compared to the existing literature in Table-I.

B. Contributions of the paper

1) The end-to-end system model is derived for a SISO-OTFS system using arbitrary Tx-Rx pulse shapes in the DD-domain. At the receiver, a procedure is conceived for pilot extraction followed by the formulation of the CE paradigm as a sparse estimation problem.
2) A novel online estimation framework is proposed, which utilizes the expectation maximization (EM) based sparse Bayesian learning (SBL) technique for determining the initialization parameters for channel estimation in SISO-OTFS systems.
3) A low complexity data detector is developed for the data processing unit (DPU) that collects outputs corresponding to the transmitted data signal. An equivalent TF-domain

TABLE I: Contrasting the proposed solution to the existing literature on OTFS CE

	[20]	[16]	[21]	[17]	[22]	[23]	[24]	[25]	[26]	Proposed
MIMO system		✓		✓						✓
Sparse DD-domain channel			✓		✓	✓	✓	✓		✓
Row-group sparsity										✓
Flexible pilot overhead	✓						✓			✓
Online CE										✓
Practical pulse shape			✓	✓	✓		✓			✓
Low complexity detector									✓	✓
DD-domain guard band		✓	✓	✓				✓		

system model is derived to achieve this, wherein the channel matrix constructed in the TF-domain is explicitly shown to be of block diagonal nature.

4) Furthermore, the above formulations and algorithms are also extended to MIMO-OTFS systems after the development of the corresponding end-to-end DD-domain system model. Interestingly, the CSI for this scenario is shown to be both row and group sparse in nature. To suitably exploit this important property, an online row and group sparse Bayesian learning (RGBL) technique is proposed for CE in these systems.
5) A novel iterative block matrix inversion approach based on the Schur complement is also derived for symbol detection in MIMO-OTFS systems which is proven to have significantly lower complexity than the conventional LMMMSE detector.
6) Additionally, the time-recursive Bayesian Cramer-Rao lower bounds (BCRLBs) have been derived for both SISO and MIMO-OTFS systems to benchmark the CE performance.

C. Organization of the paper

The rest of this work is organized as follows. The DD-domain SISO-OTFS system model is derived in Section-II. Section-III expresses the CSI estimation problem as a sparse estimation problem and develops a novel online estimation framework for the same. In addition, a low complexity detector is proposed data detection, followed by Section-IV that derives the DD-domain MIMO-OTFS system model. Next, in section V develops the online estimation framework and a low complexity detector in these systems. Section-VI presents the time-recursive Bayesian Cramer-Rao Lower Bounds (BCRLB) that serve as benchmarks to characterize the MSE performance of the proposed algorithms. Section-VII gives a complexity analysis of the proposed algorithm. This is followed by our simulation results in Section-VIII and conclusion in Section-IX. Also, the recursive procedure for finding block inversion is derived in Appendix A.

D. Notation

The following notation is used throughout this paper. Uppercase boldface letters $(\mathbf{A})$ represent matrices, while lowercase boldface letters $(\mathbf{a})$ denote vectors. The operation $\mathrm{vec}(\mathbf{A})$ converts a matrix to its corresponding vector via stacking its columns, while $\mathrm{vec}^{-1}(\mathbf{a})$ denotes vector to matrix conversion. The property $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\mathrm{vec}\left(\mathbf{B}\right)$ is used frequently in the paper, where $\otimes$ denotes the Kronecker product of two matrices. The notation $[\cdot]_M$ represents the modulo-M operation. The operator $\mathrm{diag}\left(\cdot\right)$, $\mathrm{blkdiag}\left(\cdot\right)$ arranges the elements in diagonal, and block-diagonal matrices, respectively, and $\mathbf{I}_k$ and $\mathbf{0}_k$ represent the identity and zero matrices, respectively, of size $k \times k$. The discrete Fourier transform (DFT) matrix of order K is represented by $\mathbf{F}_K$, where its (i,j)-th element is given as $\mathbf{F}_K(i,j) = \frac{1}{\sqrt{K}} e^{-j2\pi\frac{ij}{K}}$.

II. SISO-OTFS system Model

Consider a SISO-OTFS system wherein the frame duration T_{fd} is given as $T_{fd} = NT$ and the bandwidth by $BW = M\Delta f$. The quantities T (seconds) and Δf (Hz) represent the symbol duration and sub-carrier spacing, respectively, such that $T\Delta f = 1$. The parameters N and M denote the number of symbols along the time and frequency axes, respectively. The information symbols are placed in the delay-Dopler (DD) domain, where the delay and Doppler axes are sampled at integer multiples of $\Delta\tau = \frac{1}{BW}$ and $\Delta\nu = \frac{1}{\Delta f}$.

A. OTFS modulation

Let the 2D-array of information symbols represented in the DD-domain be given by the matrix $\mathbf{X}_{\text{DD}} \in \mathbb{C}^{M\times N}$, where each symbol is selected from a suitable constellation with symbol power σ_d^2. At the transmitter, the DD-domain symbol matrix $\mathbf{X}_{\text{DD}}$ is converted to its TF-domain equivalent $\mathbf{X}_{\text{TF}} \in \mathbb{C}^{M\times N}$ by using the inverse symplectic finite Fourier transform (ISFFT). Furthermore, $\mathbf{X}_{\text{TF}}$ is a linear transformation of $\mathbf{X}_{\text{DD}}$ given by $\mathbf{X}_{\text{TF}} = \mathbf{F}_M\mathbf{X}_{\text{DD}}\mathbf{F}_N^H$, where $\mathbf{F}_M$ and $\mathbf{F}_N$ are DFT matrices of sizes M and N, respectively. The equivalent time-domain signal is obtained from the TF-domain signal employing the Heisenberg transform, which is further sampled at the rate of $\frac{M}{T}$. The resultant transmit signal matrix $\mathbf{S} \in \mathbb{C}^{M\times N}$ can be modeled as

$$\mathbf{S} = \boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H\mathbf{X}_{\text{TF}} = \boldsymbol{\zeta}_{\text{tx}}\mathbf{X}_{\text{DD}}\mathbf{F}_N^H, \tag{1}$$

where $\boldsymbol{\zeta}_{\text{tx}} = \mathrm{diag}\left\{\boldsymbol{\xi}_{\text{tx}}\left(\frac{mT}{M}\right)\right\}_{m=0}^{M-1} \in \mathbb{C}^{M\times M}$ is obtained from the samples of the transmit pulse $\boldsymbol{\xi}_{\text{tx}}(t)$. In a manner analogous to OFDM, a cyclic prefix (CP) of length L is appended to each column of $\mathbf{S}$, denoted by $\mathbf{s}_i$, where $i \in [1, N]$, prior to transmission, to eliminate inter-frame interference (IFI). The various steps described above for OTFS modulation are schematically depicted in Fig. 1.

B. DD-domain wireless channel model

Let $h(\tau,\nu)$ represent the DD-domain wireless channel, where τ and ν are the delay and Doppler variables, respectively. The typical DD-domain channel is composed of only a few dominant reflectors (ρ), with the i-th reflector introducing a delay-shift τ_i and Doppler-shift ν_i, $1 \leq i \leq \rho$, in the transmitted signal. Thus, the DD-domain wireless channel $h(\tau,\nu)$ can be modeled as

$$h(\tau,\nu) = \sum_{i=1}^{\rho} h_i \delta(\tau-\tau_i)\delta(\nu-\nu_i), \tag{2}$$

where h_i is the complex gain of the i-th path and $\delta(\cdot)$ denotes the Dirac-delta function. For a typical wideband system associated with $M = 32$ and $\Delta f = 15KHz$, the delay resolution obtained is $\Delta\tau = \frac{1}{M\Delta f} = 2.08\mu sec$, which is low. Hence, one can safely assume that the delays of the multipath components are integer multiples of the delay resolution [4], [19], [34]. The delay and Doppler shifts associated with the i-th multi-path component are defined as $\tau_i = \frac{l_i}{M\Delta f}$, $\nu_i = \frac{k_i}{NT}$, respectively, where $k_i = \text{round}(k_i) + \kappa_{\nu_i}$, with $|\kappa_{\nu_i}| < \frac{1}{2}$ [17], [19], [35]. For an under-spread wireless channel, we have $\max(l_i) << M$ and $\max(k_i) << N$ [17], [19], [36]. At the receiver, the output signal $r_n(t)$ is sampled at the Nyquist rate. The m-th sample $r_n(m) = r_n(t)|_{t=\frac{mT}{M}}$ of the n-th received vector $\mathbf{r}_n$ after removing CP can be expressed as

$$r_n(m) = \sum_{i=1}^{\rho} h_i e^{j2\pi\frac{k_i(m+n(M+L))-l_i)}{MN}} \mathbf{s}\big([m+nM-l_i]_{MN}\big) + w_n(m), \tag{3}$$

where $w_n(m)$ is the m-th sample of the noise vector $\mathbf{w}_n$. Stacking all the samples $r_n(m)$ for $0 \leq m \leq M-1$, one obtains the vector $\mathbf{r}_n = [r_n(0), r_n(1), \cdots, r_n(M-1)]^T$. Furthermore, the received column vector $\mathbf{r}_n$ can be expressed as $\mathbf{r}_n = \mathbf{H}_n\mathbf{s}_n + \mathbf{w}_n$, where the equivalent time-domain channel matrix $\mathbf{H}_n \in \mathbb{C}^{M\times M}$ obeys

$$\mathbf{H}_n = \sum_{i=1}^{\rho} h_i \left(\bar{\mathbf{P}}\right)^{l_i} \left(\bar{\mathbf{D}}_{i,n}\right).$$

The quantity $\bar{\mathbf{P}} \in \mathbb{C}^{M\times M}$ denote permutation (forward cyclic shift) and $\bar{\mathbf{D}}_{i,n} \in \mathbb{C}^{M\times M}$ is diagonal matrices defined as

$$\bar{\mathbf{P}} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}_{M\times M} \tag{4}$$

$$\bar{\mathbf{D}}_{i,n} = \begin{cases} \text{diag}\Big\{\omega_i^{n(M+L)}, \omega_i^{n(M+L)+1} \cdots \omega_i^{n(M+L)+M-1-l_i}, \\ \omega_i^{n(M+L)-l_i} \cdots \omega_i^{n(M+L)-1}\Big\}, & \text{if } l_i \neq 0, \\ \text{diag}\Big\{\omega_i^{n(M+L)}, \omega_i^{n(M+L)+1} \cdots \omega_i^{n(M+L)+M-1}\Big\}, \\ & \text{if } l_i = 0, \end{cases} \tag{5}$$

where $\omega_i = e^{\frac{j2\pi k_i}{MN}}$. The received symbol vector $\mathbf{r} \in \mathbb{C}^{MN\times 1}$, obtained by stacking all the vectors $\mathbf{r}_n$, can be modeled as

$$\mathbf{r} = \left[\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_N^T\right]^T = \widetilde{\mathbf{H}}\mathbf{s} + \mathbf{w}, \tag{6}$$

where the effective channel matrix $\widetilde{\mathbf{H}} \in \mathbb{C}^{MN\times MN}$ and the noise vector $\mathbf{w} \in \mathbb{C}^{MN\times 1}$, which is comprised of independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) samples, and has the multivariate distribution $\mathcal{N}(0,\sigma^2\mathbf{I}_{MN})$ are, respectively, defined as

$$\widetilde{\mathbf{H}} = \text{blkdiag}(\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_N),$$
$$\mathbf{w} = \left[\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_N^T\right]^T,$$

while the transmitted symbol vector $\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_N^T\right]^T \in \mathbb{C}^{MN\times 1}$ is given by

$$\mathbf{s} = \text{vec}(\mathbf{S}) = \left(\mathbf{F}_N^H \otimes \boldsymbol{\zeta}_{\text{tx}}\right)\mathbf{x}_{\text{DD}}, \tag{7}$$

where $\mathbf{x}_{\text{DD}} = \text{vec}\left(\mathbf{X}_{\text{DD}}\right)$.

C. OTFS demodulator

At the receiver, the received symbol matrix is constructed as $\mathbf{R} = \text{vec}^{-1}(\mathbf{r})$. The demodulated signal $\mathbf{Y}_{\text{TF}} \in \mathbb{C}^{M\times N}$ in the TF-domain is subsequently obtained by employing the discrete Wigner transform to $\mathbf{R}$ [19], which is mathematically expressed as

$$\mathbf{Y}_{\text{TF}} = \mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}}\mathbf{R}, \tag{8}$$

where $\boldsymbol{\zeta}_{\text{rx}} = \text{diag}\left\{\xi_{\text{rx}}\left(\frac{mT}{M}\right)\right\}_{m=0}^{M-1} \in \mathbb{C}^{M\times M}$ is obtained from the samples of the receive pulse $\xi_{\text{rx}}(t)$. Subsequently, the demodulated DD-domain symbol matrix $\mathbf{Y}_{\text{DD}} \in \mathbb{C}^{M\times N}$ is derived by applying the simplectic finite Fourier transform (SFFT) to the TF-domain demodulated symbol matrix $\mathbf{Y}_{\text{TF}}$ as follows $\mathbf{Y}_{\text{DD}} = \mathbf{F}_M^H\mathbf{Y}_{\text{TF}}\mathbf{F}_N = \boldsymbol{\zeta}_{\text{rx}}\mathbf{R}\mathbf{F}_N$. This equation can be reduced to the equivalent form

$$\mathbf{y}_{\text{DD}} = \text{vec}\left(\mathbf{Y}_{\text{DD}}\right) = \left(\mathbf{F}_N \otimes \boldsymbol{\zeta}_{\text{rx}}\right)\mathbf{r}, \tag{9}$$

where $\mathbf{y}_{\text{DD}} \in \mathbb{C}^{MN\times 1}$. Substituting $\mathbf{r}$, $\mathbf{s}$ from (6), (7), respectively, into (9), the end-to-end DD-domain relationship can now be obtained by

$$\mathbf{y}_{\text{DD}} = \left(\mathbf{F}_N \otimes \boldsymbol{\zeta}_{\text{rx}}\right)\left(\widetilde{\mathbf{H}}\mathbf{s} + \mathbf{w}\right),$$

which yields

$$\begin{aligned}\mathbf{y}_{\text{DD}} &= \left(\mathbf{F}_N \otimes \boldsymbol{\zeta}_{\text{rx}}\right)\widetilde{\mathbf{H}}\left(\mathbf{F}_N^H \otimes \boldsymbol{\zeta}_{\text{tx}}\right)\mathbf{x}_{\text{DD}} + \left(\mathbf{F}_N \otimes \boldsymbol{\zeta}_{\text{rx}}\right)\mathbf{w} \\ &= \mathbf{H}_{\text{DD}}\mathbf{x}_{\text{DD}} + \mathbf{w}_{\text{DD}}. \end{aligned} \tag{10}$$

Data detection can now be performed using the MMSE-based linear detector formulated as

$$\widehat{\mathbf{x}}_{\text{DD}}^{\text{MMSE}} = \left(\mathbf{H}_{\text{DD}}^H\mathbf{R}_{w,\text{DD}}^{-1}\mathbf{H}_{\text{DD}} + \mathbf{I}_{MN}\right)^{-1}\mathbf{H}_{\text{DD}}^H\mathbf{R}_{w,\text{DD}}^{-1}\mathbf{y}_{\text{DD}}, \tag{11}$$

where the covariance matrix $\mathbf{R}_{w,\text{DD}} = \mathbb{E}\left[\mathbf{w}_{\text{DD}}\mathbf{w}_{\text{DD}}^H\right] \in \mathbb{C}^{MN\times MN}$ of the noise obeys $\mathbf{R}_{w,\text{DD}} = \sigma^2\left[\mathbf{I}_N \otimes \left(\boldsymbol{\zeta}_{\text{rx}}\boldsymbol{\zeta}_{\text{rx}}^H\right)\right]$.

The CE model of SISO-OTFS systems is detailed in the following section.

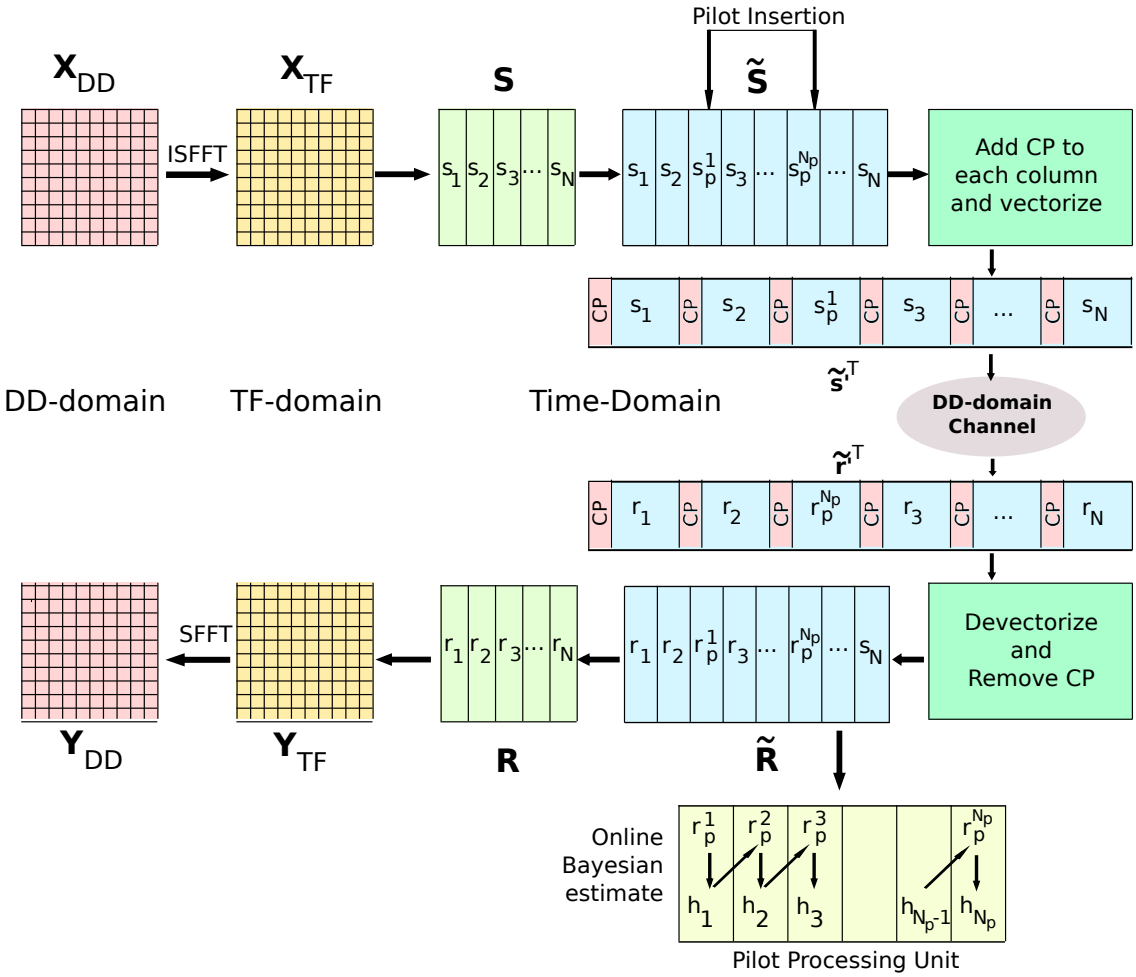


Fig. 1: Block diagram of online estimation scheme in CP-aided OTFS systems

III. Online estimation model for SISO-OTFS systems

Consider an under-spread wireless channel that obeys $\max(l_i) \leq M_\tau \leq\leq M$, $\max(k_i) \leq N_\nu \leq\leq N$, where the maximum delay and Doppler spreads are given by M_τ, N_ν, respectively. For introducing fractional Doppler, consider a grid of size $M_\tau \times G_\nu$, so that $G_\nu >> N_\nu$, i.e., each grid interval corresponding to an integer Doppler shift is divided into multiple intervals. The jth Doppler-grid point, $-G_\nu/2 \leq j \leq G_\nu/2 - 1$, corresponds to a Doppler-shift of $\nu_j = \frac{jN_\nu}{G_\nu NT}$ Hz. Thus, the DD-domain channel $h(\tau, \nu)$ can be equivalently expressed as

$$h(\tau,\nu) = \sum_{i=0}^{M_\tau - 1} \sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j}\delta(\tau - \tau_i)\delta(\nu - \nu_j), \qquad (11)$$

where the (i, j)-th point on the delay-Doppler grid corresponds to the i-th delay tap τ_i and j-th Doppler tap ν_j, with $h_{i,j}$ being the associated path gain. It should be noted that only a small number of coefficients, $h_{i,j}$, viz $\rho\ (<< M_\tau G_\nu)$, of the whole set of $M_\tau G_\nu$ elements are non-zero due to the existence of only a few dominant reflectors in the wireless channel.

Furthermore, prior to transmission, the pilot vectors $\{\mathbf{s}_p^1, \mathbf{s}_p^2, \ldots, \mathbf{s}_p^{\mathrm{N_P}}\}$ are inserted at equi-spaced intervals between the symbol vectors $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$. The modified symbol matrix $\tilde{\mathbf{S}} \in \mathbb{C}^{M\times(N+N_p)}$ comprising both the data and pilot symbol vectors is given by

$$\tilde{\mathbf{S}} = \Big[\mathbf{s}_1,\ \mathbf{s}_2,\ \mathbf{s}_p^1,\ \mathbf{s}_3,\ \ldots,\ \mathbf{s}_p^{\mathrm{N_P}},\ \ldots,\ \mathbf{s}_N\Big].$$

A representation of the frame structure, including the data and pilot symbols, is shown in Fig. 1. The locations of the transmitted pilot vectors are given by the set $\mathbb{L} = \{l_1, l_2, \cdots, l_{N_p}\}$ and are known at the receiver. Subsequently, a CP of length L is added to each column of $\tilde{\mathbf{S}}$ before transmitting it over the DD-domain channel. At the receiver, the CP is removed from each column of the received vector to obtain the matrix $\widetilde{\mathbf{R}} \in \mathbb{C}^{M\times(N+N_p)}$. The received column vector corresponding to the data or pilot output is given by

$$\mathbf{r}_{n'} = \mathbf{H}_{n'}\tilde{\mathbf{s}}_{n'} + \mathbf{w}_{n'}, \qquad (12)$$

where $1 \leq n' \leq (N + N_p)$. The matrix $\mathbf{H}_{n'} \in \mathbb{C}^{M\times M}$ is modeled as

$$\mathbf{H}_{n'} = \sum_{i=0}^{M_\tau - 1} \sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j}\left(\bar{\mathbf{P}}\right)^i \bar{\mathbf{D}}_{i,n'}, \qquad (13)$$

where the quantity $\bar{\mathbf{D}}_{i,n'} \in \mathbb{C}^{M\times M}$ is modified as

$$\bar{\mathbf{D}}_{i,n'} = \begin{cases} \text{diag}\Big\{\omega_j^{n'(M+L)}, \omega_j^{n'(M+L)+1}\cdots\omega_j^{n'(M+L)+M-1-i}, \\ \omega_j^{n'(M+L)-i}\cdots\omega_j^{n'(M+L)-1}\Big\}, & \text{if } i\neq 0, \\ \text{diag}\Big\{\omega_j^{n'(M+L)}, \omega_j^{n'(M+L)+1}\cdot\cdot\,\omega_j^{n'(M+L)+M-1}\Big\}, \\ & \text{if } i=0, \end{cases}$$

where $\omega_k = e^{j2\pi\frac{kN_\nu}{G_\nu MN}}$. Note that equation (13) assumes the number of multipath components to be unknown at the receiver. Also, the exact locations of the delay and Doppler indices are unknown and assumed to be present in the grid of size (M_τ, G_ν) for the underspread channel. Moreover, this is a sparse representation of the channel where $\mathbf{h}$ is the sparse vector whose i,jth element are given by $h_{i,j}$. Only ρ components are non-zero out of a total of $M_\tau G_\nu$ coefficients. Furthermore, the output vectors corresponding to the data symbols are arranged as the vector $\mathbf{r} = \left[\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_N^T\right]^T$. The pilot outputs are employed for sequential CE in the pilot processing unit (PPU). The output corresponding to the k-th pilot vector at location l_k is given by the equation

$$\begin{aligned}\mathbf{r}_p^k &= \sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j}\left(\bar{\mathbf{P}}\right)^i \bar{\mathbf{D}}_{i,k}\mathbf{s}_p^k + \mathbf{w}_k \\ &= \sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j}\boldsymbol{\psi}_{i,j,k} + \mathbf{w}_k, \end{aligned} \tag{14}$$

where $\boldsymbol{\psi}_{i,j,k} \in \mathbb{C}^{M\times 1}$. The above expression can be recast as

$$\mathbf{r}_p^k = \boldsymbol{\Psi}_k\mathbf{h} + \mathbf{w}_k, \tag{15}$$

where the dictionary matrix $\boldsymbol{\Psi}_k \in \mathbb{C}^{M\times M_\tau G_\nu}$ and the channel coefficient vector $\mathbf{h} \in \mathbb{C}^{M_\tau G_\nu\times 1}$ are expressed as

$$\begin{aligned}\boldsymbol{\Psi}_k =& \big[\boldsymbol{\psi}_{0,-G_\nu/2,k}, \cdots, \boldsymbol{\psi}_{0,G_\nu/2-1,k}, \cdots, \boldsymbol{\psi}_{M_\tau-1,-G_\nu/2,k}, \\ & \cdots, \boldsymbol{\psi}_{M_\tau-1,G_\nu/2-1,k}\big] \end{aligned} \tag{16}$$

$$\begin{aligned}\mathbf{h} =& \big[h_{0,-G_\nu/2}, \cdots, h_{0,G_\nu/2-1}, \cdots, h_{M_\tau-1,-G_\nu/2}, \cdots,, \\ & h_{M_\tau-1,G_\nu/2-1}\big]^T. \end{aligned} \tag{17}$$

Moreover, $\mathbf{w}_k$ is the additive Gaussian noise vector having the covariance matrix $\mathbf{R}_{w,k} = \mathbb{E}\left[\mathbf{w}_k\mathbf{w}_k^H\right] = \sigma^2\mathbf{I}_M$. The CE problem in (15) reduces to a sparse signal recovery problem owing to sparse nature of $\mathbf{h}$. Note also that since the sparse CSI is estimated sequentially from the output vectors $\mathbf{r}_p^k$, the proposed algorithm is online in nature, which is described next.

A. Online Bayesian learning for sparse CE in SISO-OTFS systems

In order to estimate the sparse CSI, consider the first pilot vector at location l_1. Then the corresponding output is given as

$$\mathbf{r}_p^1 = \boldsymbol{\Psi}_1\mathbf{h} + \mathbf{w}_1, \tag{18}$$

where the channel coefficients $\mathbf{h}$ are initially assigned the Gaussian prior given by

$$f(\mathbf{h};\boldsymbol{\Lambda}) = \prod_{i=0}^{M_\tau G_\nu-1}\frac{1}{(\pi\lambda_i)}\exp\left(-\frac{|\mathbf{h}(i)|^2}{\lambda_i}\right). \tag{19}$$

Here, λ_i represents the unknown hyperparameter associated with the i-th component of the vector $\mathbf{h}$ and $\boldsymbol{\Lambda} = \text{diag}\left(\{\lambda_i\}_{i=0}^{M_\tau G_\nu-1}\right) \in \mathbb{R}^{+M_\tau G_\nu\times M_\tau G_\nu}$ denotes the diagonal hyperparameter matrix. Upon employing the maximum likelihood framework to estimate the hyperparameter matrix $\boldsymbol{\Lambda}$, one obtains

$$\begin{aligned}\hat{\boldsymbol{\Lambda}} &= \arg\max_{\boldsymbol{\Lambda}}\log p(\mathbf{r}_p^1;\boldsymbol{\Lambda},\sigma^2) \\ &= \arg\max_{\boldsymbol{\Lambda}}\log\left(\frac{1}{(2\pi)^M|\boldsymbol{\Sigma}_{\mathbf{r}_p^1}|}\exp\left(-\frac{(\mathbf{r}_p^1)^H(\boldsymbol{\Sigma}_{\mathbf{r}_p^1})^{-1}(\mathbf{r}_p^1)}{2}\right)\right) \\ &= \arg\max_{\boldsymbol{\Lambda}}\left(-\frac{(\mathbf{r}_p^1)^H(\boldsymbol{\Sigma}_{\mathbf{r}_p^1})^{-1}(\mathbf{r}_p^1)}{2} - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{r}_p^1}| - \frac{M}{2}\log 2\pi\right), \end{aligned} \tag{20}$$

where $\boldsymbol{\Sigma}_{\mathbf{r}_p^1} = \boldsymbol{\Psi}_1\boldsymbol{\Lambda}\boldsymbol{\Psi}_1^H + \sigma^2\mathbf{I}_M$. The above problem can be expressed as

$$\hat{\boldsymbol{\Lambda}} = \arg\min_{\boldsymbol{\Lambda}}\left((\mathbf{r}_p^1)^H(\boldsymbol{\Sigma}_{\mathbf{r}_p^1})^{-1}(\mathbf{r}_p^1) + \log|\boldsymbol{\Sigma}_{\mathbf{r}_p^1}|\right). \tag{21}$$

The second term $\log|\boldsymbol{\Sigma}_{\mathbf{r}_p^1}|$ above is a non-convex function in $\boldsymbol{\Lambda}$, which makes the hyperparamter estimation problem intractable. Thus, the likelihood maximization above can be achieved via the iterative EM algorithm. Let the complete information set for this procedure be constructed as $\{\mathbf{r}_p^1, \mathbf{h}\}$, where $\mathbf{h}$ is the hidden variable, and $\mathbf{r}_p^1$ is the observation variable. The E-step of the EM algorithm evaluates the conditional expectation of the log-likelihood function of the complete information set $\{\mathbf{r}_p^1, \mathbf{h}\}$ as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Lambda}|\widehat{\boldsymbol{\Lambda}}^{(m-1)}) &= \mathbb{E}_{\mathbf{h}|\mathbf{r}_p^1;\widehat{\boldsymbol{\Lambda}}^{(m-1)}}\{\log\left[f(\mathbf{r}_p^1,\mathbf{h};\boldsymbol{\Lambda})\right]\} \\ &= \mathbb{E}\left\{\log\left[p\left(\mathbf{r}_p^1 \mid \mathbf{h}\right)\right]\right\} + \mathbb{E}\left\{\log\left[p\left(\mathbf{h};\boldsymbol{\Lambda}\right)\right]\right\}, \end{aligned} \tag{22}$$

where the quantity $\widehat{\boldsymbol{\Lambda}}^{(m-1)}$ represents the estimate of $\boldsymbol{\Lambda}$ in iteration $(m-1)$, which is defined as $\widehat{\boldsymbol{\Lambda}}^{(m-1)} = \text{diag}\{\widehat{\lambda}_i^{(m-1)}\}_{i=0}^{M_\tau G_\nu-1} \in \mathbb{R}^{M_\tau G_\nu\times M_\tau G_\nu}$. In the subsequent M-step, the log-likelihood above is maximized with respect to Λ to obtain the hyperparameter estimates in iteration m, which can be expressed as

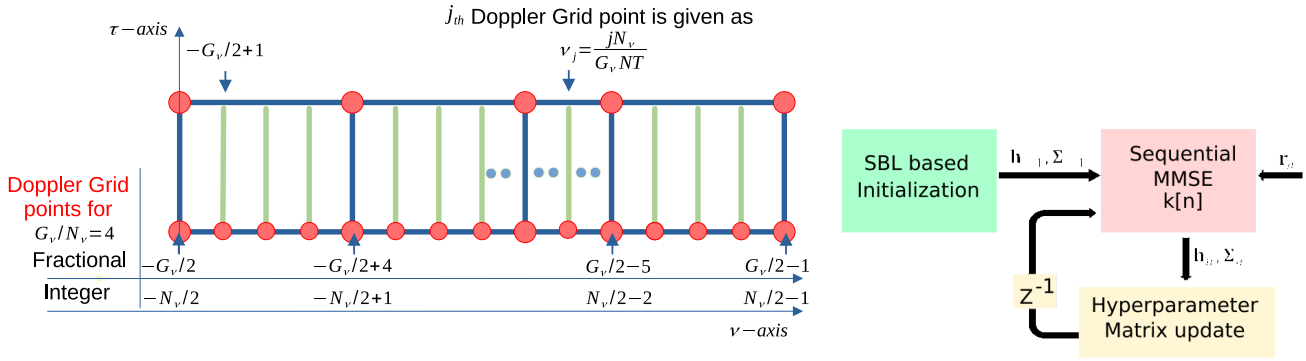
$$\begin{aligned}\widehat{\boldsymbol{\Lambda}}^{(m)} &= \arg\max_{\boldsymbol{\Lambda}}\mathcal{L}\left(\boldsymbol{\Lambda}|\widehat{\boldsymbol{\Lambda}}^{(m-1)}\right) \\ &= \arg\max_{\boldsymbol{\Lambda}}\mathbb{E}\left\{\log\left[p\left(\mathbf{h};\boldsymbol{\Lambda}\right)\right]\right\}. \end{aligned} \tag{23}$$

Upon substituting (19) into (23), followed by further simplifications, one obtains

$$= \arg\max_{\boldsymbol{\Lambda}}\sum_{i=0}^{M_\tau G_\nu-1}\left[-\log\lambda_i - \frac{1}{\lambda_i}\mathbb{E}\{|\mathbf{h}(\mathbf{i})|^2\}\right]. \tag{24}$$

The above expression of (24) can be simplified by decoupling $\boldsymbol{\Lambda}$ in terms of λ_i, yielding

$$\lambda_i^{(m)} = \mathbb{E}_{\mathbf{h}|\mathbf{r}_p^1;\widehat{\boldsymbol{\Lambda}}^{(m-1)}}\{|\mathbf{h}(\mathbf{i})|^2\}. \tag{25}$$



(a) (b)

Fig. 2: (a) Grid layout for fractional Doppler (b) BL-based sequential estimate

The conditional PDF is given by $f(\mathbf{h}|\mathbf{r}_p^1;\widehat{\mathbf{\Lambda}}^{m-1}) = \mathcal{CN}(\widehat{\boldsymbol{\mu}}^{(m)},\mathbf{\Sigma}^{(m)})$, with the *aposteriori* mean vector $\widehat{\boldsymbol{\mu}}^{(m)} \in \mathbb{C}^{M_\tau G_\nu \times 1}$ and associated covariance matrix $\mathbf{\Sigma}^{(m)} \in \mathbb{C}^{M_\tau G_\nu \times M_\tau G_\nu}$ formulated as

$$\widehat{\boldsymbol{\mu}}^{(m)} = \mathbf{\Sigma}^{(m)} \left(\mathbf{\Psi_1}^{(m-1)}\right)^H \mathbf{R}_{w,1}^{-1}\mathbf{r}_p^1,$$

$$\mathbf{\Sigma}^{(m)} = \left[\left(\mathbf{\Psi_1}^{(m-1)}\right)^H \mathbf{R}_{w,1}^{-1}\mathbf{\Psi_1}^{(m-1)} + \left(\widehat{\mathbf{\Lambda}}^{(m-1)}\right)^{-1}\right]^{-1}. \tag{26}$$

Based on (25), it can be shown that the hyperparameter estimates are given by

$$\widehat{\lambda}_i^{(m)} = \mathbf{\Sigma}^{(m)}(i,i) + |\widehat{\boldsymbol{\mu}}^{(m)}(i)|^2, \tag{27}$$

The converged hyperparameter estimate $\widehat{\mathbf{\Lambda}}_{SBL} = \widehat{\mathbf{\Lambda}}^{(m)}$ is employed to initialize the sequential MMSE estimator of Fig.2(b) with the parameter values $\mathbf{\Sigma}_0 = \widehat{\mathbf{\Lambda}}^{(m)}$, $\widehat{\mathbf{h}}_0 = 0_{M_\tau G_\nu \times 1}$. Let $\widehat{\mathbf{h}}_{k-1}$ represent the LMMSE estimate of $\mathbf{h}$ obtained from the $(k-1)$-st transmitted pilot vector at location l_{k-1}, and the associated error covariance matrix be denoted by $\mathbf{\Sigma}_{k-1}$. Upon employing the sequential LMMSE estimation procedure, the estimate $\widehat{\mathbf{h}}_k$ and its error covariance $\mathbf{\Sigma}_k$ can be recursively updated as

$$\begin{aligned}\mathbf{\Sigma}_k &= \left(\mathbf{I}_{M_\tau G_\nu} - \mathbf{K}_k\mathbf{\Psi}_k\right)\mathbf{\Sigma}_{k-1}, \widehat{\mathbf{h}}_k \\ &= \widehat{\mathbf{h}}_{k-1} + \mathbf{K}_k\left(\mathbf{r}_p^k - \mathbf{\Psi}_k\widehat{\mathbf{h}}_{k-1}\right),\end{aligned} \tag{28}$$

where the gain $\mathbf{K}_k \in \mathbb{C}^{M_\tau G_\nu \times M}$ is given by

$$\mathbf{K}_k = \mathbf{\Sigma}_{k-1}\mathbf{\Psi}_k^H\left(\mathbf{\Psi}_k\mathbf{\Sigma}_{k-1}\mathbf{\Psi}_k^H + \mathbf{R}_{w,k}\right)^{-1}. \tag{29}$$

The estimated CSI $\widehat{\mathbf{h}}_{N_p}$ corresponding to all the N_p pilot vectors can be expressed by the matrix $\widehat{\mathbf{H}}_{N_p} = \text{vec}^{-1}(\widehat{\mathbf{h}}_{N_p}) \in \mathbb{C}^{M_\tau \times G_\nu}$. Furthermore, as defined before, due to the sparse nature of the channel, only a few coefficients out of the total $M_\tau \times G_\nu$ coefficients are non-zero. Hence, the delay and Doppler can be computed from the estimated channel coefficient matrix $\widehat{\mathbf{H}}_{N_p}$ by determining the set of K largest values and saving the respective Doppler indices in $\mathcal{S}$. The index i ranges from 0 to $M_\tau - 1$, while j ranges from $-G_\nu/2$ to $G_\nu/2-1$. K is an integer value close to the number of taps. Moreover, the corresponding Doppler can be determined by multiplying the effective Doppler resolution $\Delta\widetilde{\nu}$, defined as $\Delta\widetilde{\nu} = \frac{N_\nu}{G_\nu}\Delta\nu$, with the set of indices in $\mathcal{S}$. Note that the effective Doppler resolution is considerably reduced due to the subdivision of integer Doppler grids. This enables the efficient handling of fractional Doppler shifts. Therefore, the channel matrix estimate $\widehat{\mathbf{H}}_n$ corresponding to the n-th transmitted column vector is given by

$$\widehat{\mathbf{H}}_n = \sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2-1}^{G_\nu/2-1} \widehat{\mathbf{H}}_{N_p,i,j}\left(\bar{\mathbf{P}}\right)^i\bar{\mathbf{D}}_{i,n}. \tag{30}$$

Finally, the overall channel estimate $\widehat{\widetilde{\mathbf{H}}}$ of $\widetilde{\mathbf{H}}$ is obtained as $\widehat{\widetilde{\mathbf{H}}} = \text{blkdiag}(\widehat{\mathbf{H}}_1, \cdots, \widehat{\mathbf{H}}_N)$. The next section presents a low-complexity decoder designed for recovering of the data vectors.

Algorithm 1: OSBL for SISO-OTFS systems

Input: Received vector $\mathbf{r}_p^k$, pilot dictionary matrix $\mathbf{\Psi}_k$, noise covariance matrix $\mathbf{R}_{w,k}$, threshold ϵ and maximum iterations $N_{\max}$.

Initialization: $\widehat{\lambda}_i^{(0)} = 1, \forall\ 0 \le i \le M_\tau G_\nu - 1, \widehat{\mathbf{\Lambda}}^{(0)} = \mathbf{I}_{M_\tau G_\nu}$, $\widehat{\mathbf{\Lambda}}^{(-1)} = \mathbf{0}$, $m = 0$.

Output: $\widehat{\mathbf{h}}^{\text{OSBL}} = \widehat{\mathbf{h}}_{N_p}$.

```
for k ← 1 to N_p do
    if k ← 1 do
        while ||Λ̂^(m) − Λ̂^(m−1)||_F^2 ≥ ε and m < N_max do
            m ← m + 1
            E-step: Compute the mean and aposteriori covariance given by (26).
            M-step: Compute the hyperparameter estimates
            for i ← 0 to M_τ G_ν − 1 do
                λ̂_i^(m) = Σ^(m)(i,i) + |μ̂^(m)(i)|^2
    else: Compute Gain K_k using (29), Σ_k and ĥ_k using (28)
```

B. Low-complexity data detection

Consider the end to end system model of (10). A significant disadvantage of the DD-domain linear MMSE detector described in (11) is that it requires the computation of the inverse of a matrix of size $MN \times MN$. As a result, its complexity scales as $\mathcal{O}(M^3N^3)$, which renders it challenging to implement in practice. Thus, the central motivation of this section is to develop a low-complexity TF-domain data detection scheme.

The time domain output symbol matrix $\widetilde{\mathbf{R}}$, obtained after the removal of the CP from each column, is comprised of contributions from both the data and pilots. In order to perform data detection, the output vectors corresponding to the data symbols are separated and arranged in the form of a vector $\mathbf{r} = \left[\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_N^T\right]^T$. The corresponding matrix can be constructed as $\mathbf{R} = \text{vec}^{-1}(\mathbf{r}) \in \mathbb{C}^{M\times N}$. The TF-domain received symbol matrix $\mathbf{Y}_{\text{TF}}$ is computed by substituting $\mathbf{R}$ into equation (8), and subsequently, its vectorized version $\mathbf{y}_{\text{TF}}$ is determined as

$$\begin{aligned}\mathbf{y}_{\text{TF}} = \text{vec}(\mathbf{Y}_{\text{TF}}) &= \text{vec}(\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}}\mathbf{R}) \\ &= [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{r}. \end{aligned} \tag{31}$$

Upon substituting the expression of $\mathbf{r}$ from (6) into (31), one obtains the relationship

$$\mathbf{y}_{\text{TF}} = [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\widetilde{\mathbf{H}}\mathbf{s} + [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{w}. \tag{32}$$

From the relationship between the matrices $\mathbf{S}$ and $\mathbf{X}_{\text{TF}}$ shown in (1), the vector $\mathbf{s}$, obtained by vectorizing $\mathbf{S}$, can be modeled as $\mathbf{s} = \text{vec}(\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H\mathbf{X}_{\text{TF}}) = [\mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]\mathbf{x}_{\text{TF}}$, where $\mathbf{x}_{\text{TF}} = \text{vec}(\mathbf{X}_{\text{TF}})$. By substituting the above equation for $\mathbf{s}$ into (32), $\mathbf{y}_{\text{TF}}$ can be equivalently written as

$$\begin{aligned}\mathbf{y}_{\text{TF}} =&[\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\widetilde{\mathbf{H}}[\mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]\mathbf{x}_{\text{TF}} \\ &+ [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{w}. \end{aligned} \tag{33}$$

Thus, the end-to-end system model in the TF-domain can be obtained by simplifying the above relationship as

$$\mathbf{y}_{\text{TF}} = \mathbf{H}_{\text{TF}}\mathbf{x}_{\text{TF}} + \mathbf{w}_{\text{TF}}, \tag{34}$$

The above end-to-end system model in (34) is in the TF-domain. This is in contrast to other works such as [36], [18], where the DD-domain relation is derived. The proposed methodology yields a solution that is not optimal, yet achieves a performance close to that of the conventional DD-domain detector at a significantly lower complexity. In the above expression the TF-domain channel $\mathbf{H}_{\text{TF}} \in \mathbb{C}^{MN\times MN}$ is given by

$$\mathbf{H}_{\text{TF}} = [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\widetilde{\mathbf{H}}[\mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]. \tag{35}$$

Since $\widetilde{\mathbf{H}} = \text{blkdiag}(\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_N)$, which is block diagonal, it can be readily deduced that the TF-domain channel matrix $\mathbf{H}_{\text{TF}}$ above is block circulant [37]. The nth block of that matrix $\mathbf{H}_{\text{TF}}$ is given as $\mathbf{H}_{\text{TF},n} = (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})\mathbf{H}_n(\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)$. When rectangular pulse shaping filters are used at the receiver and transmitter, i.e., $\boldsymbol{\zeta}_{\text{rx}} = \boldsymbol{\zeta}_{\text{tx}} = \mathbf{I}_M$, it follows that $\mathbf{H}_{\text{TF}} = [\mathbf{I}_N \otimes \mathbf{F}_M]\widetilde{\mathbf{H}}[\mathbf{I}_N \otimes \mathbf{F}_M^H]$ and $\mathbf{w}_{\text{TF}} = (\mathbf{I}_N \otimes \mathbf{F}_M)\mathbf{w}$. Let $\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} \in \mathbb{C}^{MN\times 1}$ denote the MMSE estimate of $\mathbf{x}_{\text{TF}}$ formulated as

$$\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} = \left(\mathbf{H}_{\text{TF}}^H\mathbf{R}_{w,\text{TF}}^{-1}\mathbf{H}_{\text{TF}} + \mathbf{I}_{MN}\right)^{-1}\mathbf{H}_{\text{TF}}^H\mathbf{R}_{w,\text{TF}}^{-1}\mathbf{y}_{\text{TF}}, \tag{36}$$

where the covariance matrix $\mathbf{R}_{w,\text{TF}} \in \mathbb{C}^{MN\times MN}$ of the noise is $\mathbf{R}_{w,\text{TF}} = \mathbb{E}\left[\mathbf{w}_{\text{TF}}\mathbf{w}_{\text{TF}}^H\right] = \sigma^2[\mathbf{I}_N \otimes \left(\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}}\boldsymbol{\zeta}_{\text{rx}}^H\mathbf{F}_M^H\right)]$. The above expression can be further simplified as

$$\begin{aligned}&\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} \\ &= (\mathbf{I}_N \otimes \mathbf{F}_M)\left(\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}} + \sigma^2\mathbf{I}_{MN}\right)^{-1}\widetilde{\mathbf{H}}^H\left(\mathbf{I}_N \otimes \mathbf{F}_M^H\right)\mathbf{y}_{\text{TF}}. \end{aligned} \tag{37}$$

An important observation is that the TF-domain channel matrix $\mathbf{H}_{\text{TF}}$ and the noise covariance matrix $\mathbf{R}_{w,\text{TF}}$ are block-diagonal. It can be inferred that the inverse term from (36) is also block diagonal. Thus the computational complexity of (37) is significantly lower. The MMSE estimate can be determined in a block-wise fashion, with the MMSE estimate for the n-th block, having a dimension of $\widehat{\mathbf{x}}_{\text{TF},n}^{\text{MMSE}} \in \mathbb{C}^{M\times 1}$, and given as

$$\widehat{\mathbf{x}}_{\text{TF},n}^{\text{MMSE}} = \mathbf{F}_M\left(\mathbf{H}_n^H\mathbf{H}_n + \sigma^2\mathbf{I}_M\right)^{-1}\mathbf{H}_n^H\mathbf{F}_M^H\mathbf{y}_{\text{TF},n}, \tag{38}$$

where $\mathbf{y}_{\text{TF},n} \in \mathbb{C}^{M\times 1}$ is TF-domain response corresponding to the n-th symbol vector $\mathbf{x}_{\text{TF},n}$. Moreover, as $\mathbf{H}_n^H\mathbf{H}_n$ is a Hermitian matrix, $\mathbf{H}_n^H\mathbf{H}_n + \sigma^2\mathbf{I}_M$ is a positive definite matrix, since $\sigma^2 > 0$. Thus, the matrix $\mathbf{H}_n^H\mathbf{H}_n + \sigma^2\mathbf{I}_M$ is always invertible [32]. The computational complexity of (38) is of the order $\mathcal{O}(M^3)$ for each of the N blocks. Thus, it can be deduced that the overall inversion complexity of the TF-domain detector (36) is of the order $\mathcal{O}(M^3N)$. By contrast the DD-domain linear MMSE detector of (11) has complexity order of $\mathcal{O}(M^3N^3)$.

The TF-domain estimate $\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}}$ obtained in (36), is now converted to the DD-domain estimate $\widehat{\mathbf{x}}_{\text{DD}} \in \mathbb{C}^{MN\times 1}$ using the relationship

$$\begin{aligned}\widehat{\mathbf{x}}_{\text{DD}} = \text{vec}(\widehat{\mathbf{X}}_{\text{DD}}) &= \text{vec}(\mathbf{F}_M^H\widehat{\mathbf{X}}_{\text{TF}}^{\text{MMSE}}\mathbf{F}_N) \\ &= \left(\mathbf{F}_N \otimes \mathbf{F}_M^H\right)\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}}. \end{aligned} \tag{39}$$

The online estimation framework is extended next for CSI acquisition in MIMO-OTFS systems.

IV. MIMO-OTFS SYSTEM MODEL

Consider an OTFS modulated MIMO system equipped with N_t transmit antennas (TAs) and N_r receive antennas (RAs). For the r-th RA and t-th TA, the associated DD-domain wireless channel $h_{r,t}(\tau, \nu)$ is formulated as

$$h_{r,t}(\tau, \nu) = \sum_{i=1}^{\rho} h_{i,r,t}\delta(\tau - \tau_i)\delta(\nu - \nu_i), \tag{40}$$

where ρ is the number of multipath between receiver r-th RA and t-th TA [16], [38] for $1 \le r \le N_r$, $1 \le t \le N_t$, and $h_{i,r,t}$ represents the wireless channel coefficient corresponding to the i-th multi-path component. Let $\mathbf{X}_{\text{DD},t} \in \mathbb{C}^{M\times N}$ denote the DD-domain transmit symbol matrix for the t-th TA. The

received symbol matrix $\mathbf{Y}_{\text{DD},r} \in \mathbb{C}^{M\times N}$ corresponding to the r-th RA is expressed as

$$\mathbf{y}_{\text{DD},r} = \sum_{t=1}^{N_t} \mathbf{H}_{\text{DD},r,t}\mathbf{x}_{\text{DD},t} + \mathbf{w}_{\text{DD},r}, \tag{41}$$

where $\mathbf{x}_{\text{DD},t} = \text{vec}(\mathbf{X}_{\text{DD},t}) \in \mathbb{C}^{MN\times 1}$, $\mathbf{H}_{\text{DD},r,t} \in \mathbb{C}^{MN\times MN}$ is the DD-domain channel matrix between the r-th RA and t-th TA that is expressed as

$$\mathbf{H}_{\text{DD},r,t} = (\mathbf{F}_N \otimes \boldsymbol{\zeta}_{\text{rx}})\mathbf{H}_{r,t}(\mathbf{F}_N^H \otimes \boldsymbol{\zeta}_{\text{tx}}). \tag{42}$$

The matrix $\mathbf{H}_{r,t} \in \mathbb{C}^{MN\times MN}$ obeys the relationship $\mathbf{H}_{r,t} = \text{blkdiag}(\mathbf{H}_1, \cdots, \mathbf{H}_N)$, where the matrix $\mathbf{H}_n$ is formulated as follows:

$$\mathbf{H}_n = \sum_{i=1}^{L_p} h_{i,r,t}\left(\bar{\mathbf{P}}\right)^{\text{l}_i}\bar{\mathbf{D}}_{i,n}, \tag{43}$$

where $\bar{\mathbf{P}}$ and $\bar{\mathbf{D}}$ are as described for a SISO-OTFS system in Section II. Now, stacking all the received output vectors $\mathbf{y}_{\text{DD},r}$ for $1 \le r \le N_r$, as $\bar{\mathbf{y}}_{\text{DD}} = \left[\mathbf{y}_{\text{DD},1}^T, \mathbf{y}_{\text{DD},2}^T, \cdots, \mathbf{y}_{\text{DD},N_r}^T\right]^T \in \mathbb{C}^{MNN_r\times 1}$, the end-to-end relationship for the MIMO-OTFS system is expressed as follows

$$\bar{\mathbf{y}}_{\text{DD}} = \bar{\mathbf{H}}_{\text{DD}}\bar{\mathbf{x}}_{\text{DD}} + \bar{\mathbf{w}}_{\text{DD}}, \tag{44}$$

where $\bar{\mathbf{x}}_{\text{DD}} \in \mathbb{C}^{MNN_t\times 1}$ and $\bar{\mathbf{w}}_{\text{DD}} \in \mathbb{C}^{MNN_r\times 1}$ are obtained by stacking the transmit DD-domain symbols for all the TAs and noise vectors at the RAs, respectively. These are defined as

$$\begin{aligned}\bar{\mathbf{x}}_{\text{DD}} &= \left[\mathbf{x}_{\text{DD},1}^T, \mathbf{x}_{\text{DD},2}^T, \cdots, \mathbf{x}_{\text{DD},N_t}^T\right]^T,\\ \bar{\mathbf{w}}_{\text{DD}} &= \left[\mathbf{w}_{\text{DD},1}^T, \mathbf{w}_{\text{DD},2}^T, \cdots, \mathbf{w}_{\text{DD},N_r}^T\right]^T.\end{aligned} \tag{45}$$

The DD-domain channel matrix $\bar{\mathbf{H}}_{\text{DD}} \in \mathbb{C}^{MNN_r\times MNN_t}$ is given by $\bar{\mathbf{H}}_{\text{DD}} = \{\mathbf{H}_{\text{DD},r,t}\}_{r=1:N_r}^{t=1:N_t}$ Furthermore, upon using (44), the MMSE-based linear detector can be formulated as

$$\bar{\mathbf{x}}_{\text{DD}}^{\text{MMSE}} = \left(\bar{\mathbf{H}}_{\text{DD}}^H\bar{\mathbf{R}}_{\text{w,DD}}^{-1}\bar{\mathbf{H}}_{\text{DD}} + \mathbf{I}_{MNN_t}\right)^{-1}\bar{\mathbf{H}}_{\text{DD}}^H\bar{\mathbf{R}}_{\text{w,DD}}^{-1}\bar{\mathbf{y}}_{\text{DD}}, \tag{46}$$

where $\bar{\mathbf{R}}_{\text{w,DD}} = \mathbb{E}\left[\bar{\mathbf{w}}_{\text{DD}}\bar{\mathbf{w}}_{\text{DD}}^H\right] = \left(\mathbf{I}_{N_r} \otimes \bar{\mathbf{R}}_{\text{w,DD},r}\right) \in \mathbb{C}^{MNN_r\times MNN_r}$ denotes the covariance matrix of the noise vector $\bar{\mathbf{w}}_{\text{DD}}$ and $\bar{\mathbf{R}}_{w,\text{DD},r} = \sigma^2\left[\mathbf{I}_N \otimes \left(\boldsymbol{\zeta}_{\text{rx}}\boldsymbol{\zeta}_{\text{rx}}^H\right)\right]$.

V. Online estimation model for MIMO-OTFS systems

Note that the suggested channel estimation algorithm does not assume knowledge of the number of multipath components ρ at the receiver. The robustness of the algorithm arises due to the fact that it considers the dominant components to be present in a grid of size $M_\tau \times G_\nu$ and subsequently determines the number of multipath components along with the corresponding channel coefficients.

Thus, the DD-domain channel $h_{r,t}(\tau,\nu)$ can be equivalently expressed as

$$h_{r,t}(\tau,\nu) = \sum_{i=0}^{M_\tau}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j,r,t}\delta(\tau-\tau_i)\delta(\nu-\nu_j),$$

where $h_{i,j,r,t}$ is the channel coefficient corresponding to the ith delay index, jth Doppler index and (r,t) RA/TA pair. At the t-th TA, applying the ISFFT and Heisenberg transforms to the transmitted symbol matrix $\mathbf{X}_{\text{DD},t}$, one obtains the symbol matrix $\mathbf{S}_t \in \mathbb{C}^{M\times N}$. Furthermore, pilot vectors $\{\mathbf{s}_{p,t}^1, \mathbf{s}_{p,t}^2, \ldots, \mathbf{s}_{p,t}^{\text{N}_\text{p}}\}$ are inserted at equi-spaced intervals in between the symbol vectors $\{\mathbf{s}_{1,t}, \mathbf{s}_{2,t}, \ldots, \mathbf{s}_{N,t}\}$. The modified symbol matrix $\tilde{\mathbf{S}}_t \in \mathbb{C}^{M\times(N+N_p)}$ contains both data and pilot symbol vectors $\tilde{\mathbf{S}}_t = [\mathbf{s}_{1,t}, \mathbf{s}_{2,t}, \mathbf{s}_{p,t}^1, \mathbf{s}_{3,t}, \ldots, \mathbf{s}_{p,t}^{\text{N}_\text{p}}, \ldots, \mathbf{s}_{N,t}]$, while the pilot locations $\mathbb{L} = \{l_1, l_2, \cdots, l_{N_p}\}$ are the same for all the N_t TAs. Subsequently, a CP of length L is added to each column vector of the resultant matrix $\widetilde{\mathbf{S}}_t \in \mathbb{C}^{M\times(N+N_p)}$ prior to transmission. At the r-th RA, the contribution from all the TAs corresponding to the n'-th column vector is received followed by CP removal, which is expressed as

$$\mathbf{r}_{n',r} = \sum_{t=1}^{N_t}\mathbf{H}_{n',r,t}\mathbf{s}_{n',t} + \mathbf{w}_{n',r}, \tag{47}$$

where $1 \le n' \le N + N_p$ and $\mathbf{H}_{n',r,t} \in \mathbb{C}^{M\times M}$ is the channel corresponding to the r-th RA, t-th TA and n'-th column vector, which is defined as

$$\mathbf{H}_{\mathbf{n',r,t}} = \sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j,r,t}\left(\bar{\mathbf{P}}\right)^i\bar{\mathbf{D}}_{i,n'}. \tag{48}$$

Note that all the outputs corresponding to the data are processed by the DPU, whereas those corresponding to the pilots, when encountered, are processed by the PPU. The latter are used for the ensuing CE. Consider now the output $\mathbf{r}_{p,r}^k$ corresponding to the k-th pilot at location l_k. Upon substituting $\mathbf{H}_{\mathbf{k,r,t}}$ from (48) into (47), $\mathbf{r}_{p,r}^k$ can be expressed as

$$\begin{aligned}\mathbf{r}_{p,r}^k &= \sum_{t=1}^{N_t}\sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j,r,t}\left(\bar{\mathbf{P}}\right)^i\bar{\mathbf{D}}_{i,k}\mathbf{s}_{p,t}^k + \mathbf{w}_{p,r}^k\\ &= \sum_{t=1}^{N_t}\sum_{i=0}^{M_\tau-1}\sum_{j=-G_\nu/2}^{G_\nu/2-1} h_{i,j,r,t}\bar{\boldsymbol{\psi}}_{i,j,k,t} + \mathbf{w}_{p,r}^k,\end{aligned} \tag{49}$$

where we have $\bar{\boldsymbol{\psi}}_{i,j,k,t} = \left(\bar{\mathbf{P}}\right)^i\bar{\mathbf{D}}_{i,k}\,\mathbf{s}_{p,t}^k \in \mathbb{C}^{M\times 1}$. This model can be further simplified as

$$\mathbf{r}_{p,r}^k = \sum_{t=1}^{N_t}\boldsymbol{\Psi}_{k,t}\mathbf{h}_{r,t} + \mathbf{w}_{p,r}^k, \tag{50}$$

where $\boldsymbol{\Psi}_{k,t} \in \mathbb{C}^{M\times M_\tau G_\nu}$ is the dictionary matrix defined as $\boldsymbol{\Psi}_{k,t} = \left[\bar{\psi}_{0,-G_\nu/2,k,t}, \cdots, \bar{\psi}_{0,G_\nu/2-1,k,t}, \cdots, \bar{\psi}_{M_\tau-1,-G_\nu/2,k,t}, \cdots, \bar{\psi}_{M_\tau-1,G_\nu/2-1,k,t}\right]$ and the channel coefficient vector corresponding to the r-th RA and t-th TA is defined as $\mathbf{h}_{r,t} = \left[h_{0,-G_\nu/2,r,t}\cdots h_{0,G_\nu/2-1,r,t}\cdots h_{M_\tau-1,-G_\nu/2,r,t}\cdots h_{M_\tau-1,G_\nu/2-1,r,t}\right]^T$. The expression in (50) can be written in the compact form

$$\mathbf{r}_{p,r}^k = \widetilde{\boldsymbol{\Psi}}_k\mathbf{h}_r + \mathbf{w}_{p,r}^k, \tag{51}$$

where $\widetilde{\boldsymbol{\Psi}}_k \in \mathbb{C}^{M\times M_\tau G_\nu N_t}$ is defined as $\widetilde{\boldsymbol{\Psi}}_k = \left[\boldsymbol{\Psi}_{k,1}, \boldsymbol{\Psi}_{k,2}, \cdots, \boldsymbol{\Psi}_{k,N_t}\right]$ and $\mathbf{h}_r \in \mathbb{C}^{M_\tau G_\nu N_t\times 1}$ is formed as $\left[\mathbf{h}_{r,1}^T, \cdots, \mathbf{h}_{r,N_t}^T\right]^T$. By concatenating the received vectors

across all the RAs, the received matrix $\mathbf{R}_p^k \in \mathbb{C}^{M\times N_r}$ can be expressed as

$$\mathbf{R}_p^k = \widetilde{\mathbf{\Psi}}_k \mathbf{H} + \mathbf{W}_p^k, \tag{52}$$

where we have $\mathbf{R}_p^k = [\mathbf{r}_{p,1}^k, \mathbf{r}_{p,2}^k, \cdots, \mathbf{r}_{p,N_r}^k]$, $\mathbf{H} \in \mathbb{C}^{M_\tau G_\nu N_t \times N_r}$ is defined as

$$\mathbf{H} = \big[\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{N_r}\big]. \tag{53}$$

In the MIMO-OTFS system, upon concatenating the pilot outputs, the resultant multiple measurement vector (MMV) $\mathbf{R}_p^k$ is sparse and the resultant problem can be formulated as a sparse estimation problem as shown in (52). For the MIMO-OTFS channel, the delay and Doppler shifts corresponding to the multipath components are identical for all the TA and RA pairs. Thus, the sparsity profile of the vectors $\mathbf{h}_{r,t}$, is identical, and the vector $\mathbf{h}_r$, which is obtained via stacking $\mathbf{h}_{r,t} \forall t$, exhibits a group-sparse structure [34], i.e., the i-th group comprising the coefficients $i \in M_\tau G_\nu$, $\mathbf{h}_r[(t-1)M_\tau G_\nu + i]_1^{N_t}$, are either all zero or non-zero. Furthermore, the ith group of rows denoted by the row indices $[(t-1)M_\tau G_\nu + i]_1^{N_t}$, become either simultaneously zero or non-zero. Thus, the channel matrix $\mathbf{H}$ is row-group sparse in nature. The row-group sparse CE problem defined in (52) can be simplified using our proposed model discussed next.

A. Online Bayesian learning aided sparse CE for MIMO-OTFS systems

The online estimation framework assigns the parameterized Gaussian prior of

$$f(\mathbf{h}_{r,t}; \mathbf{\Lambda}) = \prod_{i=0}^{M_\tau G_\nu - 1} \frac{1}{(\pi\lambda_i)} \exp\left(-\frac{|\mathbf{h}_{r,t}(i)|^2}{\lambda_i}\right), \tag{54}$$

to the channel vector $\mathbf{h}_{r,t}$. Furthermore, since the sparsity profile of $\mathbf{h}_{r,t}$ is identical for each RA and TA pair, the prior corresponding to $\widetilde{\mathbf{H}}$ is given as

$$f\left(\widetilde{\mathbf{H}}; \mathbf{\Lambda}\right) = \prod_{r=1}^{N_r} \prod_{t=1}^{N_t} \prod_{i=0}^{M_\tau G_\nu - 1} \frac{1}{(\pi\lambda_i)} \exp\left(-\frac{|\mathbf{h}_{r,t}(i)|^2}{\lambda_i}\right). \tag{55}$$

For the first transmitted pilot vector at index l_1, the output response at the receiver is given by

$$\mathbf{R}_p^1 = \widetilde{\mathbf{\Psi}}_1 \mathbf{H} + \mathbf{W}_p^1. \tag{56}$$

The initial estimate of the hyperparameter matrix $\mathbf{\Lambda}$ is obtained iteratively using the EM algorithm. In the E-step, for the m-th iteration, the update equations for the mean $\widehat{\boldsymbol{\mathcal{M}}}^{(m)} \in \mathbb{C}^{M_\tau G_\nu N_t \times N_r}$ and *aposteriori* covariance $\mathbf{\Sigma}^{(m)} \in \mathbb{C}^{M_\tau G_\nu N_t \times M_\tau G_\nu N_t}$ are given as

$$\widehat{\boldsymbol{\mathcal{M}}}^{(m)} = \mathbf{\Sigma}^{(m)} (\widetilde{\mathbf{\Psi}}_{\mathbf{1}}^{(m-1)})^H \mathbf{R}_{w_1}^{-1} \mathbf{R}_p^1, \tag{57}$$

$$\mathbf{\Sigma}^{(m)} = [(\widetilde{\mathbf{\Psi}}_{\mathbf{1}}^{(m-1)})^H \mathbf{R}_{w_1}^{-1} \widetilde{\mathbf{\Psi}}_{\mathbf{1}}^{(m-1)} + (\mathbf{I}_{N_t} \otimes (\widehat{\mathbf{\Lambda}}^{(m-1)})^{-1})]^{-1}. \tag{58}$$

Algorithm 2: ORGBL for MIMO-OTFS systems

Input: Received matrix $\mathbf{R}_p^k$, pilot dictionary matrix $\mathbf{\Psi}_k$, noise covariance matrix $\mathbf{R}_{w,k}$, threshold ϵ and maximum iterations $N_{\max}$.

Initialization: $\widehat{\lambda}_i^{(0)} = 1, \forall\ 0 \le i \le M_\tau G_\nu - 1, \widehat{\mathbf{\Lambda}}^{(0)} = \mathbf{I}_{M_\tau G_\nu}$, $\widehat{\mathbf{\Lambda}}^{(-1)} = \mathbf{0}$, $m = 0$.

Output: $\widehat{\mathbf{H}}^{\text{OSBL}} = \widehat{\mathbf{H}}_{N_p}$.

```
for k ← 1 to N_p do
    if k ← 1 do
        while ||Λ̂^(m) − Λ̂^(m−1)||_F^2 ≥ ε and m < N_max do
            m ← m + 1
            E-step: Compute the mean and aposteriori covariance.
            The mean and aposteriori covariance are given by (57), (58), respectively.
            M-step: Compute the hyperparameter estimates
            for i ← 0 to M_τ G_ν − 1 Compute λ̂_i^(m) using (59).
    else: Compute Σ_k, Gain matrix K_k, Ĥ_k using (60), (61).
```

In the subsequent M-step, the hyperparameter update $\widehat{\lambda}_i^{(m)}$ is given as

$$\widehat{\lambda}_i^{(m)} = \frac{1}{N_r N_t} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} \left| \widehat{\boldsymbol{\mathcal{M}}}^{(m)} \big[(t-1)N_t + i, r-1\big] \right|^2 + \frac{1}{N_t} \sum_{t=1}^{N_t} \mathbf{\Sigma}^{(m)} \big[(t-1)N_t + i, (t-1)N_t + i\big]. \tag{59}$$

The estimated hyperparameter matrix update $\widehat{\mathbf{\Lambda}}^{(m)} \in \mathbb{R}^{M_\tau G_\nu \times M_\tau G_\nu}$ is employed for initializing the sequential MMSE estimator of Fig. 2, with $\mathbf{\Sigma}_0 = \mathbf{I}_{N_t} \otimes \widehat{\mathbf{\Lambda}}^{(m)}$, $\widehat{\mathbf{H}}_0 = 0_{M_\tau G_\nu N_t \times N_r}$. As per the sequential MMSE paradigm, the parameters can be updated as

$$\widehat{\mathbf{H}}_k = \widehat{\mathbf{H}}_{k-1} + \mathbf{K}_k (\mathbf{R}_p^k - \widetilde{\mathbf{\Psi}}_k \widehat{\mathbf{H}}_{k-1}), \tag{60}$$

where the gain matrix $\mathbf{K}_k$ and $\mathbf{\Sigma}_k$ is given by

$$\mathbf{K}_k = \mathbf{\Sigma}_{k-1} \widetilde{\mathbf{\Psi}}_k^H \left(\widetilde{\mathbf{\Psi}}_k \mathbf{\Sigma}_{k-1} \widetilde{\mathbf{\Psi}}_k^H + \widetilde{\mathbf{R}}_{\widetilde{\mathbf{w}}_p^k}\right)^{-1},$$
$$\mathbf{\Sigma}_k = \left(\mathbf{I}_{M_\tau N_\nu N_t} - \mathbf{K}_k \widetilde{\mathbf{\Psi}}_k\right) \mathbf{\Sigma}_{k-1}. \tag{61}$$

The resultant channel estimate obtained after training with N_p pilot vectors is denoted by $\widehat{\mathbf{H}}_{N_p}$ and the associated error covariance matrix is $\mathbf{\Sigma}_{N_p}$. The estimate of the channel matrix $\widehat{\mathbf{H}}_{n,r,t}$ corresponding to the r-th RA, t-th TA, n-th transmitted column vector is

$$\widehat{\mathbf{H}}_{n,r,t} = \sum_{i=0}^{M_\tau - 1} \sum_{j=0}^{G_\nu - 1} \widehat{h}_{i,j,r,t} \left(\bar{\mathbf{P}}\right)^i \bar{\mathbf{D}}_{i,n}, \tag{62}$$

where we have

$$\widehat{h}_{i,j,r,t} = \widehat{\mathbf{H}}_{N_p}\left[i(G_\nu + 1) + j + (t-1)M_\tau G_\nu, r-1\right].$$

Algorithm 3: Iterative block matrix inversion

Input: Block matrix $\mathbf{G}$, Number of TAs N_t, $\mathbf{G}_{(0)} = \mathbf{G}_{11}$, $\mathbf{G}_{(0)}^{-1} = \mathbf{G}_{11}^{-1}$.

for $t = 1 : N_t - 1$ **do**

Initialization: $\mathbf{A} = \mathbf{B} = \mathbf{C} = \mathbf{D} = \mathbf{Z} = [\,]$

$$\mathbf{A} \leftarrow \mathbf{G}(1:MNt, 1:MNt), \mathbf{B} \leftarrow \mathbf{G}(1:MNt, MNt+1:MN(t+1))$$
$$\mathbf{C} \leftarrow \mathbf{G}(MNt+1:MN(t+1), 1:MNt), \mathbf{D} \leftarrow \mathbf{G}(MNt+1:MN(t+1), MNt+1:MN(t+1))$$

$$\mathbf{G}_{(t)} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

Compute $(\mathbf{G}_{(t)}/\mathbf{A}) = \mathbf{D} - \mathbf{C}\mathbf{G}_{(t-1)}^{-1}\mathbf{B}$ and $(\mathbf{G}_{(t)}/\mathbf{A})^{-1}$

$$\mathbf{G}_{(t)}^{-1} = \begin{bmatrix} \mathbf{G}_{(t-1)}^{-1} + \mathbf{G}_{(t-1)}^{-1}\mathbf{B}(\mathbf{G}_{(t)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(t-1)}^{-1} & -\mathbf{G}_{(t-1)}^{-1}\mathbf{B}(\mathbf{G}_{(t)}/\mathbf{A})^{-1} \\ -(\mathbf{G}_{(t)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(t-1)}^{-1} & (\mathbf{G}_{(t)}/\mathbf{A})^{-1} \end{bmatrix}.$$

Output: $\mathbf{G}^{-1} = \mathbf{G}_{(N_t-1)}^{-1}$

Thus, the estimate of the channel matrix $\widehat{\mathbf{H}}_{r,t}$ is formulated as $\widehat{\mathbf{H}}_{r,t} = \text{blkdiag}(\widehat{\mathbf{H}}_{1,r,t}, \cdots, \widehat{\mathbf{H}}_{N,r,t})$ corresponding to the r-th RA, and the t-th TA. The detailed description of online estimation process for a MIMO-OTFS system is given by Algorithm 2. The next sub-section describes a low complexity detector for MIMO-OTFS systems.

B. Low complexity data detection for MIMO-OTFS

Consider the end-to-end system model of (44). The resultant DD-domain linear MMSE detector of (46) has a computational complexity order of $\mathcal{O}(M^3N^3N_t^3)$. Therefore, we have conceive a low complexity TF-domain detector, as described next. The TF-domian output signal $\mathbf{Y}_{\text{TF},r}$ is given by $\mathbf{Y}_{\text{TF},r} = \mathbf{F}_{\text{M}}\boldsymbol{\zeta}_{\text{rx}}\mathbf{R}_r$. Vectorizing $\mathbf{Y}_{\text{TF},r}$, one obtains

$$\mathbf{y}_{\text{TF},r} = \text{vec}(\mathbf{Y}_{\text{TF},r}) = [\mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{r}_r,$$

where

$$\mathbf{r}_r = \sum_{t=1}^{N_t} \mathbf{H}_{r,t}\mathbf{s}_t + \mathbf{w}_r.$$

Upon vertically concatenating the received vectors $\mathbf{y}_{\text{TF},r}$ at each RA, the resultant model is $\bar{\mathbf{y}}_{\text{TF}} = [\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{r}$, where $\mathbf{y}_{\text{TF}} \in \mathbb{C}^{MNN_r\times 1}$ is $[\mathbf{y}_{\text{TF},1}^T, \cdots \mathbf{y}_{\text{TF},N_r}^T]^T$, $\mathbf{r} \in \mathbb{C}^{MNN_r\times 1}$ is $[\mathbf{r}_1^T, \mathbf{r}_2^T, \cdots, \mathbf{r}_{N_r}^T]^T$. Thus, $\bar{\mathbf{y}}_{\text{TF}}$ can be expressed as

$$\bar{\mathbf{y}}_{\text{TF}} = [\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})](\bar{\mathbf{H}}\mathbf{s} + \mathbf{w}), \tag{63}$$

where the channel matrix is given by $\bar{\mathbf{H}} = \text{blkmtx}\{\mathbf{H}_{r,t}\}_{r=1:N_r}^{t=1:N_t}$ The quantity $\mathbf{s} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_{N_t}^T]^T$, where $\mathbf{s}_t$ is defined as

$$\begin{aligned}\mathbf{s}_t = \text{vec}(\mathbf{S}_t) &= \zeta_{\text{tx}}\mathbf{F}_M^H\mathbf{X}_{\text{TF},t} \\ &= [\mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]\mathbf{x}_{\text{TF},t}.\end{aligned}$$

Thus $\mathbf{s}$ can be expressed as $\mathbf{s} = [\mathbf{I}_{N_t} \otimes \mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]\bar{\mathbf{x}}_{\text{TF}}$. Thereafter, upon substituting $\mathbf{s}$ in (63), we have

$$\begin{aligned}\bar{\mathbf{y}}_{\text{TF}} = &[\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\bar{\mathbf{H}}[\mathbf{I}_{N_t} \otimes \mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]\bar{\mathbf{x}}_{\text{TF}} \\ &+[\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{w}.\end{aligned} \tag{64}$$

The above relationship can be written in the compact form of

$$\bar{\mathbf{y}}_{\text{TF}} = \bar{\mathbf{H}}_{\text{TF}}\bar{\mathbf{x}}_{\text{TF}} + \bar{\mathbf{v}}_{\text{TF}}, \tag{65}$$

where $\bar{\mathbf{H}}_{\text{TF}} \in \mathbb{C}^{MNN_r\times MNN_t}$ is defined as $[\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\bar{\mathbf{H}}[\mathbf{I}_{N_t} \otimes \mathbf{I}_N \otimes (\boldsymbol{\zeta}_{\text{tx}}\mathbf{F}_M^H)]$ and $\bar{\mathbf{v}}_{\text{TF}} = [\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes (\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}})]\mathbf{w}$. The MMSE estimate of $\bar{\mathbf{x}}_{\text{TF}}$ in the TF-domain obeys [26]

$$\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} = \left(\bar{\mathbf{H}}_{\text{TF}}^H\mathbf{R}_{v,\text{TF}}^{-1}\bar{\mathbf{H}}_{\text{TF}} + \mathbf{I}_{MNN_t}\right)^{-1}\bar{\mathbf{H}}_{\text{TF}}^H\mathbf{R}_{v,\text{TF}}^{-1}\bar{\mathbf{y}}_{\text{TF}}, \tag{66}$$

where the covariance matrix $\mathbf{R}_{v,\text{TF}} \in \mathbb{C}^{MNN_r\times MNN_r}$ of the noise is determined as $\mathbf{R}_{v,\text{TF}} = \mathbb{E}\left[\mathbf{v}_{\text{TF}}\mathbf{v}_{\text{TF}}^H\right] = \sigma^2\big[\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes \left(\mathbf{F}_M\boldsymbol{\zeta}_{\text{rx}}\boldsymbol{\zeta}_{\text{rx}}^H\mathbf{F}_M^H\right)\big]$. For the rectangular pulse shape, the above expression can be further simplified to

$$\begin{aligned}\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} = (\mathbf{I}_{N_t} \otimes \mathbf{I}_N \otimes \mathbf{F}_M)\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}} + \sigma^2\mathbf{I}_{MNN_t}\right)^{-1}\bar{\mathbf{H}}^H \\ \left(\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes \mathbf{F}_M^H\right)\mathbf{y}_{\text{TF}}.\end{aligned} \tag{67}$$

The complexity of the MMSE estimator (67) above is on the order of $\mathcal{O}(M^3N^3N_t^3)$, which is high. To overcome this challenge, a low-complexity TF-domain detector is proposed next, wherein the matrix inversion is performed iteratively. Consider the matrix $\mathbf{G} \in \mathbb{C}^{MNN_t\times MNN_t}$ obtained as

$$\mathbf{G} = \left(\bar{\mathbf{H}}^H\bar{\mathbf{H}} + \sigma^2\mathbf{I}_{MNN_t}\right),$$

which may be expressed in terms of $N_t \times N_t$ sub-block matrices, where each sub-block is of size $(MN \times MN)$, as shown below $\mathbf{G} = \text{blkmtx}\{\mathbf{G}_{t,t'}\}_{t=1:N_t}^{t'=1:N_t}$ The (k,l)-th block $\mathbf{G}_{k,l} \in \mathbb{C}^{MN\times MN}$ of matrix $\mathbf{G}$ can be expressed as

$$\mathbf{G}_{k,l} = \begin{cases} \sum_{r=1}^{N_r} \mathbf{H}_{rk}^H\mathbf{H}_{rl}, \text{if } k \neq l \\ \sum_{r=1}^{N_r} \mathbf{H}_{rk}^H\mathbf{H}_{rl} + \sigma^2\mathbf{I}_{MN}, \text{if } k = l, \end{cases} \tag{68}$$

where $\mathbf{H}_{rk} = \text{blkdiag}(\mathbf{H}_{rk}^1, \ldots, \mathbf{H}_{rk}^N) \in \mathbb{C}^{MN\times MN}$ and $\mathbf{H}_{rl} = \text{blkdiag}(\mathbf{H}_{rl}^1, \ldots, \mathbf{H}_{rl}^N) \in \mathbb{C}^{MN\times MN}$. Thus, the above equation can be expressed as

$$\mathbf{G}_{k,l} = \begin{cases} \sum_{r=1}^{N_r} \text{blkdiag}\Big[(\mathbf{H}_{rk}^1)^H\mathbf{H}_{rl}^1, \cdots, (\mathbf{H}_{rk}^N)^H\mathbf{H}_{rl}^N\Big], \\ \qquad\qquad\qquad\qquad\qquad \text{if } k \neq l \\ \sum_{r=1}^{N_r} \text{blkdiag}\Big[(\mathbf{H}_{rk}^1)^H\mathbf{H}_{rl}^1 + \sigma^2\mathbf{I}_M, \cdots, \\ (\mathbf{H}_{rk}^N)^H\mathbf{H}_{rl}^N + \sigma^2\mathbf{I}_M\Big], \qquad \text{if } k = l. \end{cases} \tag{69}$$

The matrix $\mathbf{G}_{k,l}$ is block-diagonal, because it is formulated by the addition of block-diagonal matrices, as shown in (69). Additionally, the underlining structure of each sub-block of $\mathbf{G}$, i.e., $\mathbf{G}_{k,l}$, is quasi-banded, similar to the SISO-OTFS. Furthermore, in order to express the inverse of $\mathbf{G}_{k,l}$, N matrices of size $M \times M$ must be inverted, resulting in a complexity of order $\mathcal{O}(M^3N)$. This motivates us to develop a Schur complement based iterative block matrix inversion technique, wherein the inverse of the block matrix $\mathbf{G}$ is computed iteratively as discussed next.

Consider the block matrix $\mathbf{G} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, which arises for $N_t = 2$ TAs. Its inverse can be obtained as [39]

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{G}/\mathbf{A})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{G}/\mathbf{A})^{-1} \\ -(\mathbf{G}/\mathbf{A})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{G}/\mathbf{A})^{-1} \end{bmatrix}, \tag{70}$$

where the Schur complement of $\mathbf{G}$ with respect to $\mathbf{A}$ is defined as $(\mathbf{G}/\mathbf{A}) = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. The above approach of inversion using the Schur complement can be extended to a MIMO system having any number of TAs using Algorithm 3, that has a complexity order of $\mathcal{O}(M^3NN_t)$. The detailed steps to derive the inverse of $\mathbf{G}$ are given in Appendix A. Thus, the TF-domain estimate is given as

$$\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}} = \Big(\mathbf{I}_{N_t} \otimes \mathbf{I}_N \otimes \mathbf{F}_M\Big)\big(\mathbf{G}_{(N_t-1)}^{-1}\big)\bar{\mathbf{H}}^H \Big(\mathbf{I}_{N_r} \otimes \mathbf{I}_N \otimes \mathbf{F}_M^H\Big)\mathbf{y}_{\text{TF}}. \tag{71}$$

Upon obtaining the TF-domain estimate $\widehat{\mathbf{x}}_{\text{TF}}^{\text{MMSE}}$, the corresponding DD-domain symbols can be estimated as follows. For each TA t, the matrix $\widehat{\mathbf{X}}_{\text{DD},t}$, which corresponds to the DD-domain input data, can be generated as $\widehat{\mathbf{X}}_{\text{DD},t} = \mathbf{F}_M^H\widehat{\mathbf{X}}_{\text{TF},t}\mathbf{F}_N$. By applying the vec operation to both sides, one obtains $\widehat{\mathbf{x}}_{\text{DD},t} = (\mathbf{F}_N \otimes \mathbf{F}_M^H)\widehat{\mathbf{x}}_{\text{TF},t}$, where we have $\widehat{\mathbf{x}}_{\text{DD},t} = \text{vec}(\widehat{\mathbf{X}}_{\text{DD},t}) \in \mathbb{C}^{MN\times 1}$ and $\widehat{\mathbf{x}}_{\text{TF},t} = \text{vec}(\widehat{\mathbf{X}}_{\text{TF},t}) \in \mathbb{C}^{MN\times 1}$. Similar to (64), vertically concatenating the input symbol vectors in the DD domain $\widehat{\mathbf{x}}_{\text{DD},t}$ at each TA yields the final estimate as

$$\widehat{\mathbf{x}}_{\text{DD}} = (\mathbf{I}_{N_t} \otimes \mathbf{F}_N \otimes \mathbf{F}_M^H)\widehat{\mathbf{x}}_{\text{TF}},$$

where

$$\widehat{\mathbf{x}}_{\text{DD}} = [\widehat{\mathbf{x}}_{\text{DD},1}^T, \cdots, \widehat{\mathbf{x}}_{\text{DD},N_t}^T]^T,$$
$$\widehat{\mathbf{x}}_{\text{TF}} = [\widehat{\mathbf{x}}_{\text{TF},1}^T, \cdots, \widehat{\mathbf{x}}_{\text{TF},N_t}^T]^T.$$

VI. Time-Recursive BCRLB

This section derives the BCRLBs for the MSE of estimating both SISO and MIMO-OTFS channels.

A. SISO-OTFS system

Consider the SISO OTFS system, the Fisher information matrices corresponding to the first observation vector $\mathbf{r}_{p,(1)} = \mathbf{r}_p^1$ and the channel prior are given by $\mathbf{J}_{r_1}$ and $\mathbf{J}_h$, respectively. Using the theory in [34], these can be formulated as $\mathbf{J}_{(k=1)} = \mathbf{J}_{r_1} + \mathbf{J}_h$, where

$$\mathbf{J}_{r_1} = -\mathbb{E}_{\mathbf{r}_{p,(1)},\mathbf{h}}\left\{\frac{\partial^2 \log\Big[f(\mathbf{r}_{p,(1)} \mid \mathbf{h})\Big]}{\partial\mathbf{h}\partial\mathbf{h}^H}\right\}$$
$$\mathbf{J}_h = -\mathbb{E}_{\mathbf{h}}\left\{\frac{\partial^2 \log\Big[f(\mathbf{h};\mathbf{\Lambda})\Big]}{\partial\mathbf{h}\partial\mathbf{h}^H}\right\},$$

which on simplification yields

$$\mathbf{J}_{(k=1)} = \mathbf{\Psi}_1^H\mathbf{\Psi}_1 + \mathbf{\Lambda}^{-1}.$$

Similarly, by stacking the outputs corresponding to all the k pilot vectors, one obtains the quantity

$$\mathbf{r}_{p,(k)} = [(\mathbf{r}_p^1)^T, (\mathbf{r}_p^2)^T, \cdots, (\mathbf{r}_p^k)^T]^T,$$

which can be expressed as

$$\mathbf{r}_{p,(k)} = \mathbf{\Phi}_{(k)}\mathbf{h} + \mathbf{w}_{(k)}$$

where we have $\mathbf{\Phi}_{(k)} = [(\mathbf{\Psi}_1)^T, (\mathbf{\Psi}_2)^T, \cdots, (\mathbf{\Psi}_k]^T)^T$. Thus, following some simplification, the Fisher information matrix for the above model is given by

$$\mathbf{J}_{(k)} = \bar{\mathbf{\Phi}}_{(k)}^H\bar{\mathbf{\Phi}}_{(k)} + \mathbf{\Lambda}^{-1}, \tag{72}$$

where $\mathbf{\Lambda}$ is the hyper-parameter matrix. Subsequently, the observation vector $\mathbf{r}_{p,(k+1)}$, corresponding to the $(k+1)$-st pilot vector, is given as

$$\mathbf{r}_{p,(k+1)} = \mathbf{\Phi}_{(k+1)}\mathbf{h} + \mathbf{w}_{(k+1)}.$$

The associated Fisher information for this model can be determined as

$$\mathbf{J}_{(k+1)} = \mathbf{\Phi}_{(k+1)}^H\mathbf{\Phi}_{(k+1)} + \mathbf{\Lambda}^{-1}, \tag{73}$$

where

$$\begin{aligned}\mathbf{\Phi}_{(k+1)} &= [(\mathbf{\Psi}_1)^T, (\mathbf{\Psi}_2)^T, \cdots, (\mathbf{\Psi}_k)^T, (\mathbf{\Psi}_{k+1})^T]^T \\ &= [(\mathbf{\Phi}_k)^T, (\mathbf{\Psi}_{k+1})^T]^T\end{aligned}$$

Substituting the value of $\mathbf{\Phi}_{(k+1)}$ in (73) yields

$$\begin{aligned}\mathbf{J}_{(k+1)} &= \begin{bmatrix}\mathbf{\Phi}_k^H & \mathbf{\Psi}_{k+1}^H\end{bmatrix}\begin{bmatrix}\mathbf{\Phi}_k \\ \mathbf{\Psi}_{k+1}\end{bmatrix} + \mathbf{\Lambda}^{-1} \\ &= \mathbf{\Phi}_k^H\mathbf{\Phi}_k + \mathbf{\Psi}_{k+1}^H\mathbf{\Psi}_{k+1} + \mathbf{\Lambda}^{-1} \\ &= \mathbf{J}_{(k)} + \mathbf{\Psi}_{k+1}^H\mathbf{\Psi}_{k+1}.\end{aligned} \tag{74}$$

Therefore, the time recursive BCRLB for the online estimation problem is given by

$$\text{MSE}(k+1) \geq \text{Tr}\big[\mathbf{J}_{(k+1)}\big]^{-1}. \tag{75}$$

TABLE II: Parameters used for simulation

Parameters	System 1	System 2	*System 3
Carrier frequency (GHz)	4	24	24
Subcarrier frequency (KHz)	7.5	15	15
# of Doppler-axis symbols $M\times$ Delay-axis symbols N	32 × 16	32 × 32	512 × 128
Max. doppler spread $M_\tau\times$ Max. delay spread N_ν	8 × 8	8 ×8	16 ×16
# of Dominant reflectors ρ	5	5	9
Modulation	BPSK	BPSK	4QAM
Pulse-shape	Rectangular	Rectangular	Rectangular

Note: * Considers EVA model

TABLE III: DD-profile for (a) System 1 and (b) System 2

Path-Index(i)	1	2	3	4	5
Delay $\tau_i(\mu sec)$	4.16	8.33	12.49	16.66	20.83
Doppler $\nu_i(Hz)$	0	470	940	1410	1880
Speed (Km/Hr)	0	126.9	253.6	380.4	507.2

Path-Index(i)	1	2	3	4	5
Delay $\tau_i(\mu sec)$	2.08	4.16	6.24	8.32	10.41
Doppler $\nu_i(Hz)$	0	470	940	1880	2820
Speed (Km/Hr)	0	21.1	42.2	84.4	126.6

B. MIMO-OTFS system

The output vector for the k-th transmitted pilot vector is given by (52). Furthermore, stacking the outputs corresponding to all k pilot vectors, one obtains the quantity

$$\mathbf{R}_{p,(k)} = \Big[(\mathbf{R}_p^1)^T, (\mathbf{R}_p^2)^T, \cdots, (\mathbf{R}_p^k)^T\Big]^T,$$

which can be expressed as

$$\mathbf{R}_{p,(k)} = \widetilde{\mathbf{\Phi}}_{(k)}\mathbf{H} + \mathbf{W}_{(k)},$$

where we have $\widetilde{\mathbf{\Phi}}_{(k)} = \big[\widetilde{\mathbf{\Psi}}_1^T, \widetilde{\mathbf{\Psi}}_2^T, \cdots, \widetilde{\mathbf{\Psi}}_k^T\big]^T$. The Fisher information matrix for the above model is given as

$$\mathbf{J}_{(k)} = \widetilde{\mathbf{\Phi}}_{(k)}^H \widetilde{\mathbf{\Phi}}_{(k)} + \mathbf{\Lambda}^{-1}, \tag{76}$$

where $\mathbf{\Lambda}$ is the hyperparameter matrix. Subsequently, the observation matrix $\mathbf{R}_{p,(k+1)}$ corresponding to the $(k+1)$-st pilot vector is given by

$$\mathbf{R}_{p,(k+1)} = \widetilde{\mathbf{\Phi}}_{(k+1)}\mathbf{H} + \mathbf{W}_{(k+1)}.$$

The associated Fisher information for this model can be determined as

$$\begin{aligned}\mathbf{J}_{(k+1)} &= \widetilde{\mathbf{\Phi}}_{(k+1)}^H \widetilde{\mathbf{\Phi}}_{(k+1)} + \mathbf{\Lambda}^{-1} \\ &= \widetilde{\mathbf{\Phi}}_{(k)}^H \widetilde{\mathbf{\Phi}}_{(k)} + \widetilde{\mathbf{\Psi}}_{k+1}^H \widetilde{\mathbf{\Psi}}_{k+1} + \mathbf{\Lambda}^{-1} \\ &= \mathbf{J}_{(k)} + \widetilde{\mathbf{\Psi}}_{k+1}^H \widetilde{\mathbf{\Psi}}_{k+1}. \end{aligned} \tag{77}$$

Therefore, the time recursive BCRLB for the online estimation problem is given by

$$\text{MSE}(k+1) \geq \text{Tr}\big[\mathbf{J}_{(k+1)}\big]^{-1}. \tag{78}$$

The BCRLB is a lower bound on the normalised mean squared error (NMSE) for the proposed CSI estimator. This lower bound assumes perfect knowledge of the number of multipath components, associated delay, and Doppler locations. The simulation results demonstrate that the proposed method exhibits an NMSE performance that closely approximates the theoretical bound. This observation indicates the efficiency of the proposed method, without requiring prior knowledge of the number of non-zero multipath components or their locations in the delay-Doppler grid. An important insight from the expression of the Fisher information matrix $\mathbf{J}_{(k+1)}$ is that it relies on the Fisher information $\mathbf{J}_{(k)}$ corresponding to the k previously transmitted pilot vectors. This recursive calculation remarkably reduces the computational complexity in comparison to the batch computation of the BCRLB which simultaneously considers all the symbols received in a single frame.

VII. Complexity and comparative analysis

The online channel estimation scheme proposed has the overall complexity order of $O(M_\tau^3 N_\nu^3)$ for an integer Doppler scenario, which arises because of the matrix multiplication and inversion steps involved in (26). By contrast, the computational complexity of the conventional EP-based [17] CSI estimator is observed to be of the order $O(M_\tau N_\nu)$. However, it is important to note that while the proposed scheme exhibits higher complexity, the EP-based scheme relies on a threshold-based approach that fails to exploit the sparsity. By contrast, the OSBL scheme efficiently utilizes the sparsity inherent in such systems, which significantly improves the performance at a lower pilot overhead. The comparative analysis below comprehensively examines the efficacy of these competing strategies.

The proposed online CSI estimation framework transmits MNp pilot symbols for a block of MN data symbols, which leads to a pilot overhead of $\frac{Np}{N+Np}$. On the contrary, the conventional EP technique [17] incurs a pilot overhead of $\frac{(2M_\tau+1)(2N_\nu+1)}{MN}$ for an integer Doppler index scenario. Upon substituting the appropriate parameters from Table II, the pilot overhead for the proposed CSI estimation framework is determined to be 0.2 for System-I and System-II, when $N_p = 4, 8$, respectively. By contrast, the corresponding overheads for the scheme discussed in [17] for system I and system II are 0.56 and 0.28, respectively, which are significantly higher than that of the proposed estimation framework. Moreover, the pilot overhead of the EP scheme for a scenario with fractional Doppler is further increased.

VIII. Simulation Results

In this section, we investigate the performance of the proposed online CE technique for both SISO and MIMO-

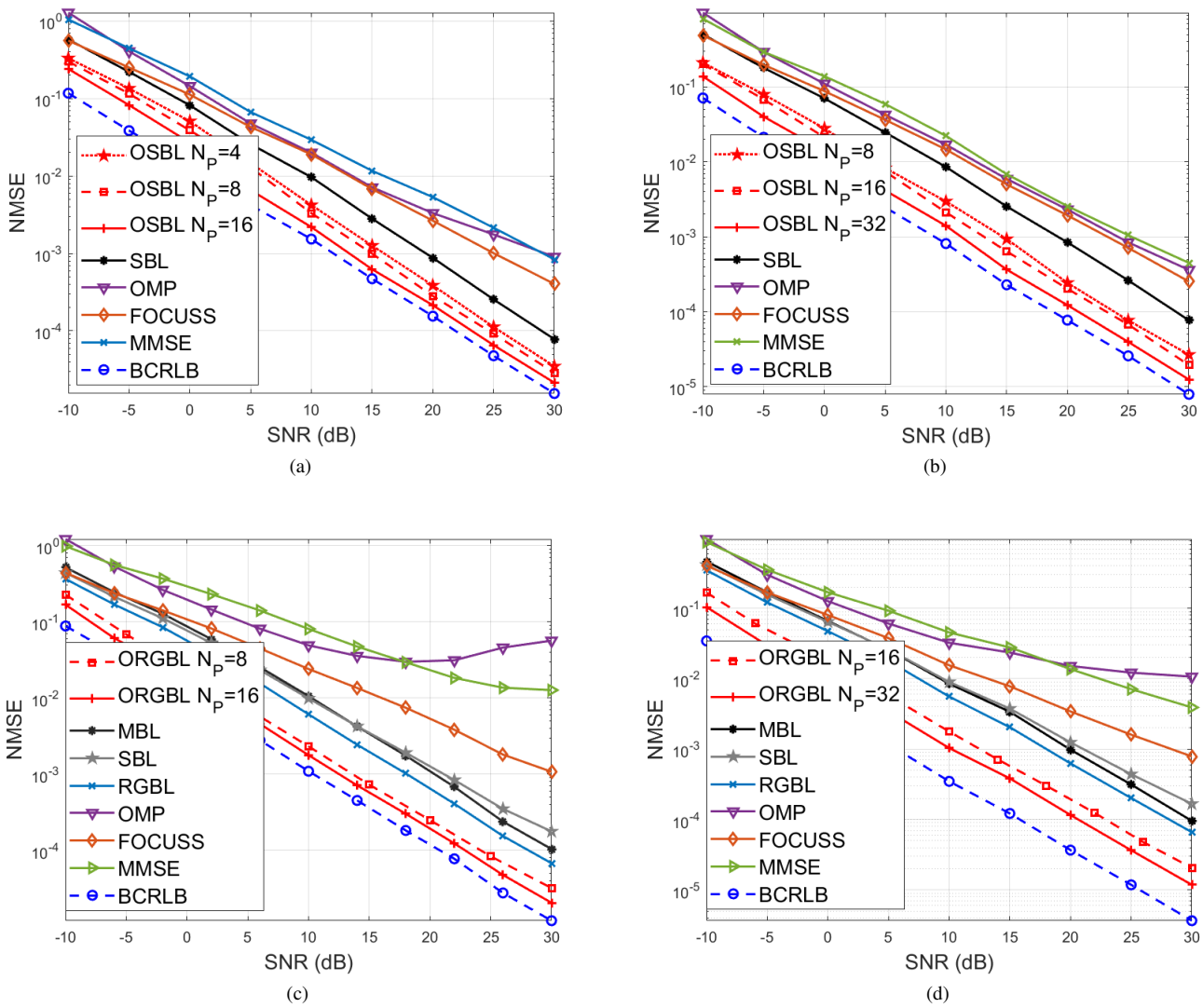


Fig. 3: NMSE vs SNR Plot for (a) SISO-OTFS System 1, (b) SISO-OTFS System 2 (c) MIMO-OTFS System 1 with $N_t = 2$, $N_r = 2$. (d) MIMO-OTFS System 2 with $N_t = 2$, $N_r = 2$.

OTFS systems. The performance metric NMSE is defined as

$$\text{NMSE} = \frac{\|\widehat{\mathbf{H}}_{\text{DD}} - \mathbf{H}_{\text{DD}}\|^2}{\|\mathbf{H}_{\text{DD}}\|^2}.$$

The various parameters used by Systems 1 and 2 are given in Table II and Table III, where the associated delay and Doppler indices are chosen randomly for each Monte Carlo iteration from Table III. Figure 3(a) and 3(b) plot the NMSE performance vs pilot SNR for SISO-OTFS systems, while Fig. 3(c) and 3(d) show the same for MIMO-OTFS systems. As expected, the NMSE progressively decreases upon increasing the pilot SNR. As shown in Fig. 3(a), 3(b), the proposed online sparse Bayesian learning (OSBL), with $N_P \in [4, 8, 16]$ number of pilots for System 1 and $N_P \in [8, 16, 32]$ for System 2, outperforms many existing sparse estimation schemes, including the SBL, OMP, FOCUSS, and the traditional MMSE algorithm for SISO-OTFS systems. This is due to fact that OSBL exploits the temporal correlation. Furthermore, the initial parameters for OSBL are computed using SBL, which guarantees an improved sparse estimate. Moreover, as the number of pilots N_P is increased from 4 to 16 for System 1, and from 8 to 32 for System 2, the performance of OSBL improves owing to the increased number of pilots available for sequential CSI estimation. The OMP technique's subpar performance can be attributed to both its sensitivity to the stopping value and the choice of dictionary matrix, whereas that of FOCUSS is due to its inability to guarantee convergence frequently and its sensitivity to the regularization parameter. Furthermore, the conventional MMSE estimator does not exploit the sparsity of the DD-domain CSI, which leads to its poor NMSE performance.

The NMSE performance of MIMO-OTFS systems with $N_t = 2$ TAs and $N_r = 2$ RAs is shown in Fig. 3(c), 3(d) for System 1 and System 2, respectively. For row-group sparse channel estimation, the proposed online row-group sparse Bayesian learning (ORGBL) scheme with $N_P \in [8, 16]$

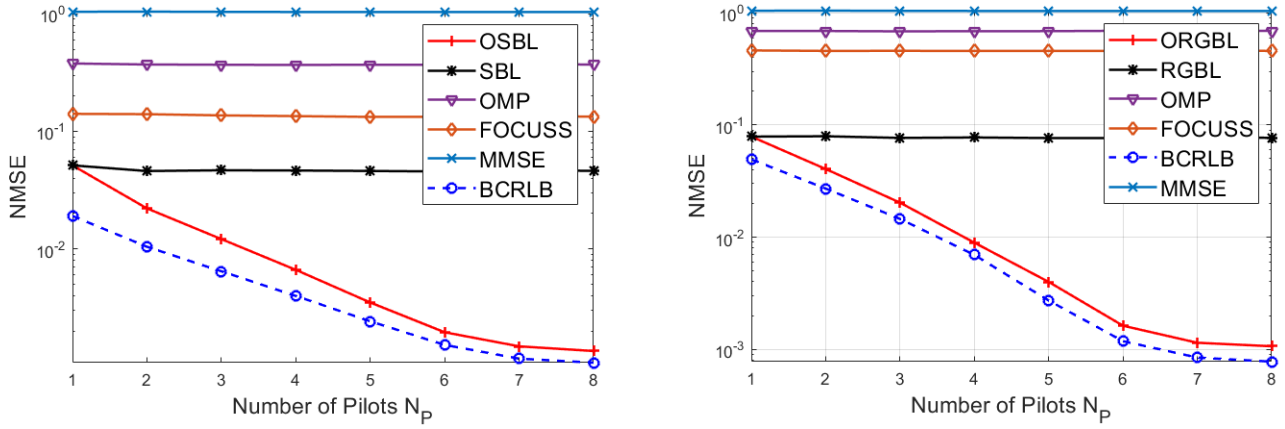


Fig. 4: NMSE vs Number of Pilots plot with $M = 64$, $N = 8$, $M\tau = 8$, $N_\nu = 8$, Pilot SNR $= 10dB$ for (a) SISO-OTFS system, (b) MIMO-OTFS system with $N_t = 2$, $N_r = 2$.

number of pilots for System 1 and $N_P \in [16, 32]$ for System 2 outperforms the competing SBL, MBL, RGBL, OMP, FOCUUSS algorithms, where MBL and RGBL denote the row sparse and row-group sparse extensions of the SBL, respectively. The performance trend is similar to that for SISO-OTFS systems, as shown in Fig. 3(a) and 3(b), with ORGBL yielding the best performance in comparison to the other competing sparse estimation schemes. Moreover, when the number of pilots N_P is increased, the performance of ORGBL improves for both the systems. The performance of SBL, which does not leverage the row or row-group sparsity, is inferior when compared to MBL and RGBL, respectively, whereas MBL performs poorly in comparison to RGBL as it only exploits row sparsity. Finally, it can be seen that the NMSEs of the OSBL and ORGBL are close to their respective BCRLBs derived in Section VI. This observation is of significant importance since it demonstrates that both these schemes can achieve their respective lower bounds even without any knowledge of the prior covariance or support of the sparse channel. This feature renders them extremely well-suited for practical implementation, where such information is frequently challenging to obtain.

Figure 4(a) and 4(b) show the NMSE performance with respect to a number of the pilot vectors for SISO and MIMO-OTFS systems, respectively, where

$$\text{NMSE} = \frac{\|\widehat{\mathbf{h}} - \mathbf{h}\|^2}{\|\mathbf{h}\|^2},$$

where $\widehat{\mathbf{h}}$ is estimated channel coefficient vector corresponding to each transmitted pilot vector and $\mathbf{h}$ is the underlying channel coefficient vector. For SISO systems, $\mathbf{h} \in \mathcal{C}^{M_\tau N_\nu} \times 1$ is given by (17) and $\widehat{\mathbf{h}} \in \mathbb{C}^{M_\tau N_\nu} \times 1$. For MIMO systems we have $\mathbf{h} = \text{vec}(\mathbf{H}) \in \mathbb{C}^{M_\tau N_\nu N_t N_r \times 1}$, where $\mathbf{H}$ is given by (53) and $\widehat{\mathbf{h}} = \text{vec}(\widehat{\mathbf{H}}) \in \mathcal{C}^{M_\tau N_\nu N_t N_r \times 1}$. The various parameter values are $M = 64$, $N = 8$, and N_p is varied from 1 to 7 for a fixed pilot SNR of 10 dB. For the MIMO-OTFS system, the number of TAs and RAs are set to $N_t = 2$, $N_r = 2$, respectively. The NMSE performance of the OSBL and ORGBL schemes is obtained by considering only the last transmitted pilot vector, whereas that for the other schemes is averaged over the entire block of pilot vectors. It can be readily noted that as the number of pilot vectors increases, the performance of OSBL and ORGBL improves, since these algorithms leverage the previous estimates of the channel, hyperparameter and covariance matrices, and update the estimate of each. Furthermore, the NMSE performance closely follows the recursive BCRLB determined in Section VI. By contrast, the NMSEs of the SBL, OMP, FOCUSS for SISO-OTFS systems and those of the RGBL, OMP, FOCUSS for MIMO-OTFS systems in Fig. 4(a) and 4(b), respectively, are seen to be constant with the increase in the amount of pilot vectors, since the performance of these estimation algorithms is independent of the number of pilot vectors transmitted.

The bit error rate (BER) of the data detected using the estimated CSI obtained via various approaches, followed by employing the proposed low complexity linear MMSE detectors described in Section III-B for SISO and V-B for MIMO-OTFS systems, is shown in Fig. 5(a), 5(b) for SISO-OTFS Systems 1, 2, respectively, and in Fig. 5(c), 5(d) for MIMO-OTFS Systems 1, 2, respectively. The resultant BER is also benchmarked against that of receivers with perfect CSI for both the cases. The BER of all the schemes improves upon increasing the data SNR for all the schemes, with the OSBL and ORGBL schemes yielding the best BER performance for SISO and MIMO-OTFS systems, respectively. This is in line with the trend for the NMSE of CE seen in the previous figures, which vindicates the hypothesis that the improved accuracy of CE indeed yields improved BER performance. Furthermore, the BERs of OSBL and ORGBL approach that of the respective hypothetical ideal receivers having perfect CSI for both systems, which demonstrates their efficiency. In summary, the proposed online Bayesian learning schemes OSBL and ORGBL, which leverage channel sparsity without requiring prior knowledge of the number of dominating re-

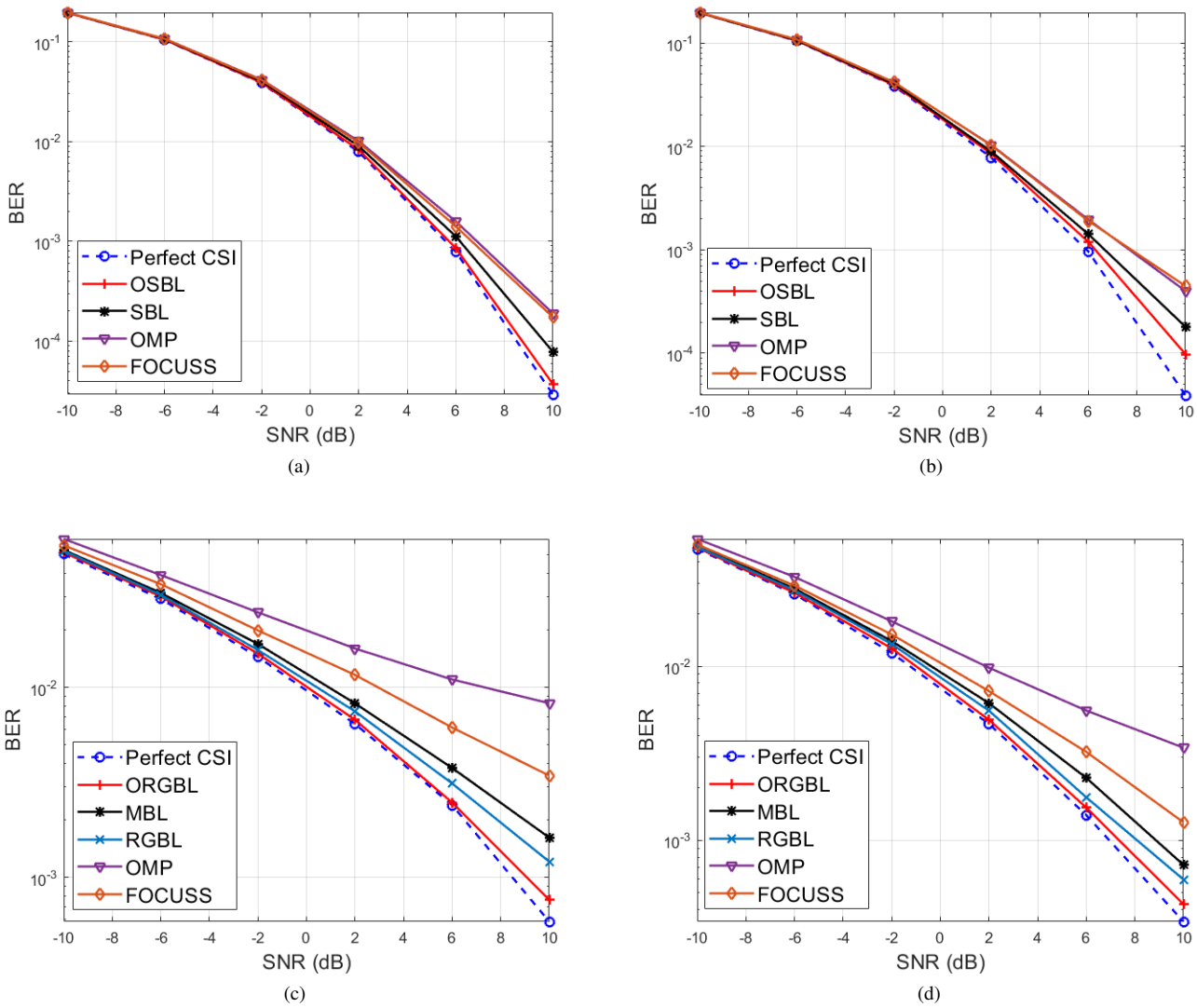
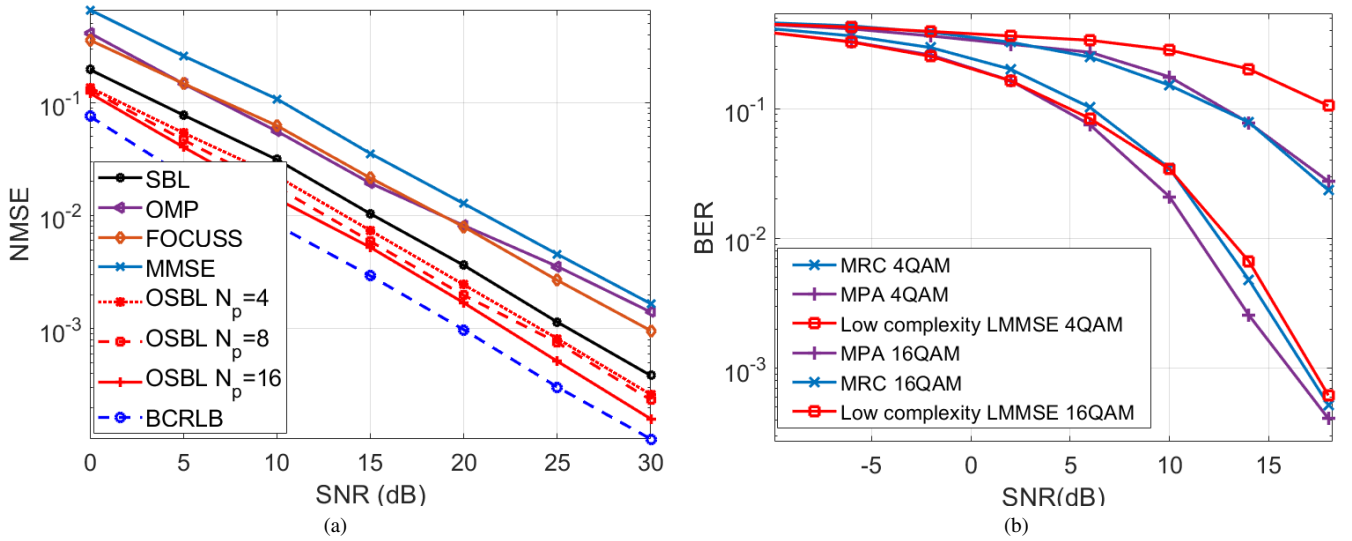


Fig. 5: BER vs SNR plot (a) SISO-OTFS System 1 with $N_p = 4$, (b) SISO-OTFS System 2 with $N_p = 8$, (c) MIMO-OTFS System 1 with $N_t = 2$, $N_r = 2$, $N_p = 4$, (d) MIMO-OTFS System 2 with $N_t = 2$, $N_r = 2$, $N_p = 8$.

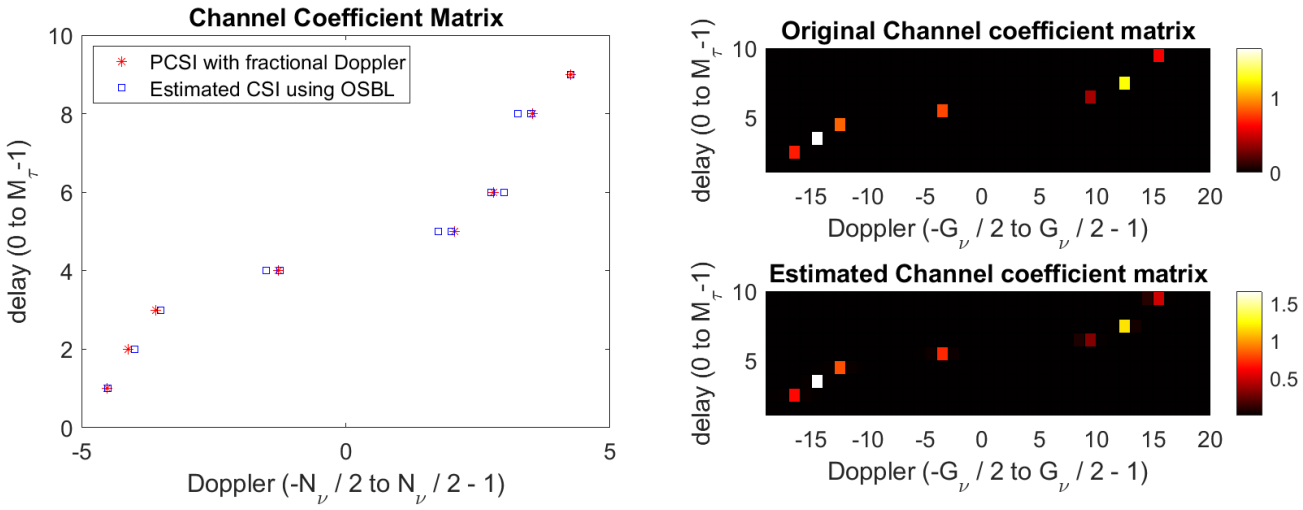
flectors or channel characteristics, outperform several other algorithms in terms of both NMSE and BER, making the suggested schemes ideal for practical application in SISO- and MIMO-aided OTFS systems.

The NMSE and BER performance is also presented for a practical EVA [40] model, which considers fractional Doppler and a maximum speed of 500Km/Hr. The simulation parameters are as per System 3 in Table II. It can be observed from Fig. 6(a) that the proposed online estimation framework yields an improved NMSE performance for the practical channel model, which corroborates the previously observed results and reinforces the efficacy of the proposed framework. Moreover, in Fig. 6(b) the performance of the low-complexity LMMSE detector is compared to that of the standard MPA detector [19] employing 50 iterations and the MRC detector [33] using 15 iterations. It is assumed that perfect CSI is available for the MRC and MPA detectors. It can be observed from Fig. 6(b) that the proposed detector for SISO systems achieves a bit error rate (BER) closer to low-complexity MRC and MP detectors for 4QAM modulation scheme. Additionally, if 16-QAM modulation is chosen performance of MRC and MP detectors are better and close to each other. However, the proposed low-complexity LMMSE detector does not require prior knowledge of the number of unique delay indices, Doppler indices or the number of paths. Fig. 7(a) displays a scatter plot for the channel coefficient matrix. This plot serves as a visual comparison between the estimated CSI using OSBL and PCSI with fractional Doppler. The parameters used are $M = 128, N = 32, M_\tau = 10, N_\nu = 10, G_\nu = 40$, $N_p = 4$, $P = 8$. The purpose of this plot is to assess the accuracy of the delay and Doppler estimates within the system. It can be inferred that the proposed OSBL scheme estimates the respective delay and Doppler coefficients efficiently. Fig. 7(b) portrays a heatmap representation of the effective delay-Doppler channel coefficient matrix. This heatmap is constructed using Doppler bins spanning from -20 to 20 after subdividing the integer



(a) (b)

Fig. 6: (a) NMSE vs. SNR plot for System 3 using EVA channel model (b) BER vs. SNR plot for System 3 using EVA channel model



(a) (b)

Fig. 7: Channel coefficient matrix for system parameters $M = 128, N = 32, M_\tau = 10, N_\nu = 10, G_\nu = 40,\ N_p = 4,\ P = 8$ (a) Scatter plot (b) Heatmap representation

bins into smaller sub-bins.

IX. Conclusions

This paper proposed online Bayesian learning assisted CE procedures for both SISO as well as MIMO-OTFS systems. Initially, an end-to-end system model was derived in the DD-domain, followed by our online sparse CE framework conceived for SISO-OTFS systems. Our sequential MMSE estimator utilizes the novel EM-based SBL to set the initial parameters of the online estimation procedure, thus leading to improved estimation and faster convergence. A low-complexity detector is also derived based on the block-diagonal TF-domain channel matrix, which was shown to have a substantially reduced complexity in comparison to conventional linear data detection. Furthermore, the online CE framework was also extended to MIMO-OTFS systems, wherein the DD-domain CSI was shown to be additionally simultaneous row and group sparse. An EM-based RGBL scheme was formulated to determine the initialization parameters of the above algorithm. A low-complexity detector was also developed for MIMO-OTFS systems based on a novel iterative block matrix inversion technique. Time-recursive BCRLBs were derived to provide valuable insights into the MSE performance of the CE schemes. Finally, our simulation results clearly demonstrated that the proposed OSBL and ORGBL schemes yield a significantly improved performance in comparison to several other schemes, together with a performance that closely follows their

$$\mathbf{G}_{(1)}^{-1} = \begin{bmatrix} \mathbf{G}_{(0)}^{-1} + \mathbf{G}_{(0)}^{-1}\mathbf{B}(\mathbf{G}_{(1)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(0)}^{-1} & -\mathbf{G}_{(0)}^{-1}\mathbf{B}(\mathbf{G}_{(1)}/\mathbf{A})^{-1} \\ -(\mathbf{G}_{(1)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(0)}^{-1} & (\mathbf{G}_{(1)}/\mathbf{A})^{-1} \end{bmatrix}. \tag{80}$$

$$\mathbf{G}_{(k)}^{-1} = \begin{bmatrix} \mathbf{G}_{(k-1)}^{-1} + \mathbf{G}_{(k-1)}^{-1}\mathbf{B}(\mathbf{G}_{(k)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(k-1)}^{-1} & -\mathbf{G}_{(k-1)}^{-1}\mathbf{B}(\mathbf{G}_{(k)}/\mathbf{A})^{-1} \\ -(\mathbf{G}_{(k)}/\mathbf{A})^{-1}\mathbf{C}\mathbf{G}_{(k-1)}^{-1} & (\mathbf{G}_{(k)}/\mathbf{A})^{-1} \end{bmatrix}. \tag{82}$$

respective BCRLB benchmarks. The improved performance coupled with low-complexity render them ideally suited for practical implementation.

APPENDIX A
ITERATIVE PROCEDURE FOR BLOCK INVERSION

The iterative procedure for determining the block inverse of $\mathbf{G}$ is described below.

Iteration 1: Initially assign

$$\mathbf{G}_{(1)} = \left[\begin{array}{c|c} \mathbf{A} = \mathbf{G}_{11} & \mathbf{B} = \mathbf{G}_{12} \\ \hline \mathbf{C} = \mathbf{G}_{21} & \mathbf{D} = \mathbf{G}_{22} \end{array}\right], \tag{79}$$

where the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are of size $MN \times MN$. Compute the inverse of $\mathbf{A}$ as $\mathbf{G}_{(0)}^{-1} = \mathbf{A}^{-1} = \mathbf{G}_{11}^{-1}$. The matrix $\mathbf{A}$ is of type $\mathbf{G}_{k,l}$, which is block diagonal, as shown in equation (69). Thus, its inverse involves evaluating the inverses of N matrices of size $M \times M$, that has a complexity of the order $\mathcal{O}(M^3N)$. Moreover, the Schur complement of $\mathbf{G}_{(1)}$ with respect to $\mathbf{A}$ is also of size $MN \times MN$, and it is given as $\mathbf{G}_{(1)}/A = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. Evaluating the inverse of $\mathbf{G}_{(1)}/A$ also has a complexity of $\mathcal{O}(M^3N)$. After the first iteration corresponding to $t = 1$, one obtains $\mathbf{G}_{(1)}^{-1}$ as given by (80), incurring the computational complexity $\mathcal{O}(M^3N)$.

Iteration $t = k$: Continuing the process, during the k-th iteration,

$$\mathbf{G}_{(k)} = \left[\begin{array}{ccc|c} & \mathbf{G}_{1,1} \cdots & \mathbf{G}_{1,k} & \mathbf{G}_{1,k+1} \\ \mathbf{A} = & \vdots \ddots & \vdots & \mathbf{B} = \vdots \\ & \mathbf{G}_{k,1} \cdots & \mathbf{G}_{k,k} & \mathbf{G}_{k,k+1} \\ \hline \mathbf{C} = & \mathbf{G}_{k+1,1} \cdots & \mathbf{G}_{k+1,k} & \mathbf{D} = \mathbf{G}_{k+1,k+1} \end{array}\right],$$

where the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are of sizes $kMN \times kMN, kMN \times MN, MN \times kMN, MN \times MN$, respectively. The Schur complement $\mathbf{G}_{(k)}/A = \mathbf{D} - \mathbf{C}\mathbf{G}_{(k-1)}^{-1}\mathbf{B}$ and $\mathbf{G}_{(k)}^{-1}$ is computed as given by (82). This incurs the complexity $\mathcal{O}(M^3N)$. Proceeding in this fashion, $\mathbf{G}^{-1}$ can be obtained as $\mathbf{G}_{(N_t-1)}^{-1}$ in iteration $t = N_t - 1$. The overall computational complexity of this procedure is $\mathcal{O}(M^3NN_t)$.

REFERENCES

[1] F. Hasegawa, A. Taira, G. Noh, B. Hui, H. Nishimoto, A. Okazaki, A. Okamura, J. Lee, and I. Kim, "High-speed train communications standardization in 3GPP 5G NR," *IEEE Communications Standards Magazine*, vol. 2, no. 1, pp. 44–52, 2018.

[2] Y. Liu, C.-X. Wang, and J. Huang, "Recent developments and future challenges in channel measurements and models for 5G and beyond high-speed train communication systems," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 50–56, 2019.

[3] J. Zhang, T. Chen, S. Zhong, J. Wang, W. Zhang, X. Zuo, R. G. Maunder, and L. Hanzo, "Aeronautical Ad-Hoc networking for the Internet-above-the-clouds," *Proceedings of the IEEE*, vol. 107, no. 5, pp. 868–911, 2019.

[4] M. Ramachandran, G. Surabhi, and A. Chockalingam, "OTFS: A new modulation scheme for high-mobility use cases," *Journal of the Indian Institute of Science*, pp. 1–22, 2020.

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

[6] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. John Wiley & Sons, 2005.

[7] L. Hanzo, Y. Akhtman, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus non-coherent and cooperative turbo transceivers*. John Wiley & Sons, 2011.

[8] A. Stamoulis, S. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2451–2464, 2002.

[9] T. Wang, J. Proakis, E. Masry, and J. Zeidler, "Performance degradation of OFDM systems due to doppler spreading," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1422–1432, 2006.

[10] N. Aboutorab, W. Hardjawana, and B. Vucetic, "A new iterative doppler-assisted channel estimation joint with parallel ICI cancellation for high-mobility MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1577–1589, 2012.

[11] W.-G. Song and J.-T. Lim, "Pilot-symbol aided channel estimation for OFDM with fast fading channels," *IEEE Transactions on Broadcasting*, vol. 49, no. 4, pp. 398–402, 2003.

[12] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A. J. Goldsmith, A. F. Molisch *et al.*, "Orthogonal time frequency space modulation," *ArXiv preprint, ArXiv:1808.00519*, 2018.

[13] X. Wang, W. Shen, C. Xing, J. An, and L. Hanzo, "Joint Bayesian channel estimation and data detection for OTFS systems in LEO satellite communications," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4386–4399, 2022.

[14] S. S. Rakib and R. Hadani, "Orthogonal time frequency space modulation system," *Google Patents*, Mar. 27 2018, US Patent 9,929,783.

[15] A. Mehrotra, R. K. Singh, S. Srivastava, and A. K. Jagannatham, "Channel estimation techniques for CP-aided OTFS systems relying on practical pulse shapes," in *IEEE International Conference on Signal Processing and Communications, 2022*. IEEE, 2022.

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

[17] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4906–4917, 2019.

[18] H. B. Mishra, P. Singh, A. K. Prasad, and R. Budhiraja, "OTFS channel estimation and data detection designs with superimposed pilots," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2258–2274, 2022.

[19] 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.

[20] K. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–10.

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

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

[23] O. K. Rasheed, G. Surabhi, and A. Chockalingam, "Sparse delay-Doppler channel estimation in rapidly time-varying channels for multiuser OTFS on the uplink," in *2020 IEEE 91st Vehicular Technology Conference (VTC-Spring)*. IEEE, 2020, pp. 1–5.

[24] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided sparse channel estimation for orthogonal time frequency space modulated systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8343–8348, 2021.

[25] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7407–7426, 2022.

[26] G. D. Surabhi and A. Chockalingam, "Low complexity linear equalization for OTFS modulation," *IEEE Communications Letters*, vol. 24, no. 2, pp. 330–334, 2020.

[27] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Bayesian learning aided simultaneous row and group sparse channel estimation in orthogonal time frequency space modulated MIMO systems," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 635–648, 2022.

[28] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse, group-sparse, and online bayesian learning aided channel estimation for doubly-selective mmwave hybrid MIMO OFDM systems," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5843–5858, 2021.

[29] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse bayesian learning approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5369–5382, 2015.

[30] 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.

[31] P. Singh, A. Gupta, H. B. Mishra, and R. Budhiraja, "Low-complexity ZF/MMSE MIMO-OTFS receivers for high-speed vehicular communication," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 209–227, 2022.

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

[33] T. Thaj and E. Viterbo, "Low complexity iterative rake decision feedback equalizer for zero-padded OTFS systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 606–15 622, 2020.

[34] A. Mehrotra, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Data-aided CSI estimation using affine-precoded superimposed pilots in orthogonal time frequency space modulated MIMO systems," *IEEE Transactions on Communications*, vol. 71, no. 8, pp. 4482–4498, 2023.

[35] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[36] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 957–961, 2018.

[37] T. De Mazancourt and D. Gerlic, "The inverse of a block-circulant matrix," *IEEE Transactions on Antennas and Propagation*, vol. 31, no. 5, pp. 808–810, 1983.

[38] H. Qu, G. Liu, M. A. Imran, S. Wen, and L. Zhang, "Efficient channel equalization and symbol detection for MIMO OTFS systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6672–6686, 2022.

[39] F. Zhang, "The Schur complement and its applications," *Springer Science and Business Media*, vol. 4, 2005.

[40] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception," 3GPP, Tech. Rep., 2011. [Online]. Available: https://www.3gpp.org