

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version 11 January, 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.011100

Gaussian Mixture Model Based Bayesian Learning for Sparse Channel Estimation in Orthogonal Time Frequency Space Modulated Systems

Surbhi Gehlot¹, (Graduate Student Member, IEEE), Suraj Srivastava¹, (Member, IEEE), Sandeep Kumar Yadav¹, (Member, IEEE), and Lajos Hanzo², (Life Fellow, IEEE)

¹Department of Electrical Engineering, Indian Institute of Technology Jodhpur, Jodhpur 342030, India

²School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

CORRESPONDING AUTHOR: Lajos Hanzo (e-mail: lh@ecs.soton.ac.uk).

The work of S. Srivastava was supported in part by IIT Jodhpur's Research Grant No. I/RIG/SUS/20240043; in part by Anusandhan National Research Foundation's PM-ECRG/2024/478/ENS and ANRF/ARG/2025/005895/ENS; and in part by Telecom Technology Development Fund (TTDF) under Grant TTDF/6G/368. Lajos Hanzo would like to gratefully acknowledge the financial support of the following Engineering and Physical Sciences Research Council (EPSRC) projects: Platform for Driving Ultimate Connectivity (TITAN) (EP/X04047X/1; EP/Y037243/1); Robust and Reliable Quantum Computing (RoQRQ, EP/W032635/1); PerCom (EP/X012301/1). S. Srivastava and L. Hanzo jointly acknowledge the funding support provided to ICON-project by DST and UKRI-EPSRC under India-UK Joint opportunity in Telecommunications Research.

ABSTRACT A novel Gaussian mixture model (GMM)-aided sparse Bayesian learning (SBL) framework is proposed for channel state information (CSI) estimation in orthogonal time-frequency space (OTFS) modulated systems. The key attribute of the proposed algorithm lies in casting CSI recovery as an SBL inference problem, where posterior distributions are iteratively refined under a hierarchical GMM prior. Using this approach, the sparsity-inducing variances beneficially promote sparsity in the delay-Doppler (DD) domain, while additionally augmenting the capability of SBL to exploit channel statistics more effectively. Moreover, to fully exploit the GMM's ability to approximate arbitrary probability density functions and model complex multipath fading scenarios, the channel statistics are represented using a complex Gaussian mixture. Simultaneously, the method leverages time-domain (TD) pilots without requiring wasteful DD domain guard intervals, thereby ensuring low pilot overhead and high spectral efficiency. The CSI recovered is subsequently applied in a linear minimum mean square error (MMSE) detector for reliable data detection. To benchmark performance, the Oracle-MMSE and the Bayesian Cramér-Rao lower bound (BCRLB) are also derived. Our simulation results demonstrate significant performance improvement over the state-of-the-art sparse estimation methods.

INDEX TERMS CSI estimation, delay-Doppler (DD) domain channel, GMM aided SBL, OTFS, sparsity.

I. INTRODUCTION

NEXT-generation (NG) wireless networks are expected to deliver ultra-high data rates in highly dynamic environments. Typical scenarios include high-speed rail, vehicular, and aerial communications, where extreme user mobility induces significant delay spread from multipath propagation and severe Doppler shifts from high relative velocity [1], [2]. Notably, under such doubly selective channel conditions, the performance of the ubiquitous orthogonal frequency-division

multiplexing (OFDM) severely degrades again primarily due to the grave inter-carrier interference induced by Doppler shifts at high mobility and carrier frequencies [3]. To address this limitation, the Doppler-resilient orthogonal time-frequency space (OTFS) modulation proposed by Hadani *et al.* has emerged as a promising alternative candidate [4]. By mapping information symbols onto the delay-Doppler (DD) domain instead of the conventional time-frequency (TF) domain, OTFS offers a channel representation that remains

nearly invariant under high mobility [5]–[7]. However, the performance gains of OTFS critically depend on the availability of highly accurate DD domain channel state information (CSI) [8]. Consequently, as discussed next, numerous studies have investigated efficient CSI estimation methods designed for OTFS systems.

A. Literature Review

Channel estimation in OTFS systems has been widely explored, with the earliest approaches belonging to impulse-based CSI estimation. Representative treatises in this category are by Hadani and Monk [9], as well as the solutions in [10], [11], which developed an end-to-end DD domain input–output relationship expressed as a two-dimensional circular convolution between the DD domain signal and the DD domain channel. While these designs were conceptually simple, their major drawback was the need for a full frame of pilots, leading to excessive pilot overhead and severely reduced spectral efficiency. To alleviate this shortcoming, Raviteja *et al.* [5] proposed a threshold-based CSI estimator and low complexity data detector within the same frame. The method embedded a known high-power pilot symbol at a chosen DD coordinate with guard bands around the pilot to isolate its interference from the data. However, it typically requires high transmit power, which raises the peak-to-average power ratio (PAPR), and large guard regions in the DD grid that reduce spectral efficiency; additionally, its threshold-based detection is signal-to-noise ratio (SNR) sensitive and requires careful tuning.

The subsequent literature of OTFS has increasingly focused on exploiting the intrinsic sparsity of wireless channels in the DD domain. Leveraging this sparse representation not only provides a structurally informed estimation framework but also helps reduce pilot overhead and improve accuracy compared to conventional pilot-based schemes. Building on this idea, in [12], [13] conventional compressed sensing techniques such as orthogonal matching pursuit (OMP) have been refined for exploiting the associated structured multi-dimensional sparsity, jointly capturing the sparsity in delay, Doppler and angular domains. Beyond greedy compressed sensing (CS) techniques, a parallel line of work has developed Bayesian learning (BL) based estimators for CSI estimation in OTFS systems. Zhao *et al.* [14] formulated a sparse Bayesian learning (SBL) aided framework for sparse channel estimation in OTFS systems. They suggested a novel DD domain pilot pattern that eliminated the need for guard symbols and maintained equal power for pilots and data, thereby reducing both the PAPR and pilot overhead. However, a fundamental limitation of the DD domain piloting strategy is that the pilot arrangement must be carefully designed to avoid interference with data symbols, which can complicate the frame structure. Addressing this limitation, Srivastava *et al.* [8] introduced a more flexible TF domain piloting strategy. The SBL framework using this pilot design significantly reduced the pilot overhead, while approaching the Bayesian

Cramér-Rao lower bound (BCRLB). Building upon these foundations, subsequently an off-grid SBL approach was developed in [15]. This method formulates channel estimation as a 1D off-grid sparse signal recovery problem within the SBL framework. It effectively separates the estimation of on-grid and off-grid delay/Doppler components, modeling the latter as hyperparameters to be estimated via expectation–maximization (EM). A group-sparse Bayesian learning framework was developed for exploiting structural sparsity in the DD domain [16]. The model effectively captures the sparsity profile across DD-grid bins, significantly enhancing estimation accuracy even in pilot-limited OTFS regimes. As a further advancement a parametric estimation-based algorithm was proposed in [17]. This proposal goes beyond sparse recovery algorithms for CSI in OTFS by advocating a modified maximum likelihood estimator and a two-step estimator, which rely on fine DD resolution to decouple the joint estimation of channel gains into independent estimation tasks using the delay and Doppler bins of each path.

Recent years have also witnessed the emergence of deep-learning (DL) techniques for channel estimation in OTFS systems. A number of these treatises [18]–[20] adopt a hybrid estimate-and-denoise paradigm, where a conventional estimator produces a coarse DD CSI estimator that is then refined by a neural network. Furthermore, several papers have conceived alternative DL based learning techniques for OTFS channel estimation. Payami *et al.* [21] introduced convolutional neural network (CNN) based support-learning architectures that directly infer the active DD support without an explicit sensing matrix. Their approach first identifies the support locations of the dominant channel taps and then estimates the corresponding amplitudes. A supervised neural network is proposed in [22] to learn the mapping from received TF-domain pilots to the underlying DD domain channel, whereas a long short-term memory (LSTM) based model in [23] leverages temporal correlations across pilot sequences to learn the underlying DD domain channel dynamics. Despite the notable estimation accuracy achieved by DL-based approaches, their employment remains constrained by the need for extensive labelled training datasets and frequent retraining. Moreover, their black-box nature reduces interpretability, motivating the development of robust, model-driven estimation frameworks over purely data-driven ones.

Existing research gaps in CSI estimation methods for OTFS systems as discussed in Table 1 underscore the need for a more efficient, interpretable, and statistically robust DD domain estimation framework. Towards this end, we propose a novel Gaussian mixture model–aided sparse Bayesian learning (GMM-SBL) framework, whose key contributions are articulated in the following section.

B. Contributions of the Paper

- 1) To enhance the capability of conventional SBL, a novel GMM-SBL framework is proposed for OTFS chan-

Table 1: Contrasting Key Features of Proposed and Existing OTFS CSI Estimation Methods (R = Required, NR = Not Required)

	[10]	[5]	[12]	[14]	[8]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	Proposed
DD domain sparsity			✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
CSI labels	NR	NR	NR	NR	NR	NR	NR	R	R	R	R	R	NR
Flexible pilot overhead			✓		✓	✓	✓		✓		✓	✓	✓
BCRLB			✓	✓	✓	✓							✓
DD-guard	R	R	NR	R	NR	NR	NR	R	R	R	R	R	NR
Rectangular pulse shaping			✓		✓		✓				✓		✓
Offline training	NR	NR	NR	NR	NR	NR	NR	R	R	R	R	R	NR
Multiple training snapshots					✓								✓
GMM channel modeling													✓
Sparsity under GMM prior													✓
Unified GMM-SBL framework													✓

nel estimation. Conventional SBL relies on a single Gaussian prior, which limits its expressiveness in modelling clustered sparsity as observed in practical wireless channels. To address this, the proposed learning strategy preserves the fundamental sparsity-promoting mechanism of SBL in the DD domain, while enhancing the statistical modelling flexibility through Gaussian mixture-based priors. The approach is supported by a theoretical analysis of sparsity guarantees under mixture priors and it is implemented via a unified EM algorithm that jointly updates the adaptive mixture weights and variance estimates directly from the received pilot observations.

- 2) By representing the channel distribution as a weighted combination of Gaussian components, the prior harnesses the full potential of the proposed GMM-SBL method. Furthermore, the learned mixture weights and component covariances strengthen the capability of SBL to exploit channel statistics beyond the restrictive zero-mean complex Gaussian assumption, offering a more practical and flexible representation of multipath channels.
- 3) By transmitting pilots in the time domain, the framework enables flexible pilot placement without requiring precise DD grid locations. This facilitates full utilisation of the DD grid for data transmission, thereby substantially reducing pilot overhead and improving spectral efficiency. Following the sparse channel estimation, the recovered CSI is incorporated into a linear minimum mean square error (MMSE) detector, which leverages the full posterior statistics for reliable symbol detection.
- 4) To establish performance references, an Oracle-based MMSE and the BCRLB of GMM priors are derived. Furthermore, the robustness of the proposed algorithm is validated through extensive simulations under varying mixture weights, cluster means, number of clusters, pilot overhead and training sample sizes, consistently

demonstrating superior performance over state-of-the-art sparse estimation methods.

The remainder of this paper is organised as follows. Section II introduces the OTFS system model. Section III develops the proposed GMM-SBL based sparse CSI estimation framework. Section IV describes the performance benchmarks, and Section V presents extensive simulation results, comparisons with existing methods, and the computational complexity assessment of the proposed algorithm. Finally, Section VI concludes the paper.

Notation - Boldface lowercase and uppercase letters denote column vectors and matrices, respectively. The vectorization of a matrix \mathbf{A} is denoted by $\text{vec}(\mathbf{A})$, while $\text{vec}^{-1}(\mathbf{a})$ denotes the inverse operation that reconstructs the original matrix. A standard identity of the vectorization operator $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$, is used in the paper, where \otimes denotes the Kronecker product. The Hermitian transpose of a matrix \mathbf{A} is denoted by \mathbf{A}^H , and the expectation operator by $\mathbb{E}[\cdot]$.

II. OTFS System Description

OTFS is a 2D modulation technique in which information symbols are mapped in the DD domain (Fig. 1). Consider a DD domain grid and the corresponding time frequency (TF) grid. Let the OTFS frame duration and the bandwidth occupied be $T_f = NT$ and $B = M\Delta f$, respectively, where N and M denote the number of time and frequency samples on the TF-grid. Furthermore, T (seconds) denotes the symbol interval, while Δf (Hz) is the subcarrier spacing, obeying $T\Delta f = 1$. Next, the signal processing operations underlying OTFS are described.

A. OTFS Modulation

Let $\mathbf{X}_{\text{DD}} \in \mathbb{C}^{M \times N}$ denote the matrix of DD domain information symbols, where $\mathbf{X}_{\text{DD}}(\ell, c)$ is placed at delay index ℓ and Doppler index c . Let $\mathbf{X}_{\text{TF}}(m, n)$ be the TF-domain symbol transmitted on the m th subcarrier during the n th symbol interval. The DD domain symbols are mapped to the TF-domain using the inverse symplectic finite Fourier

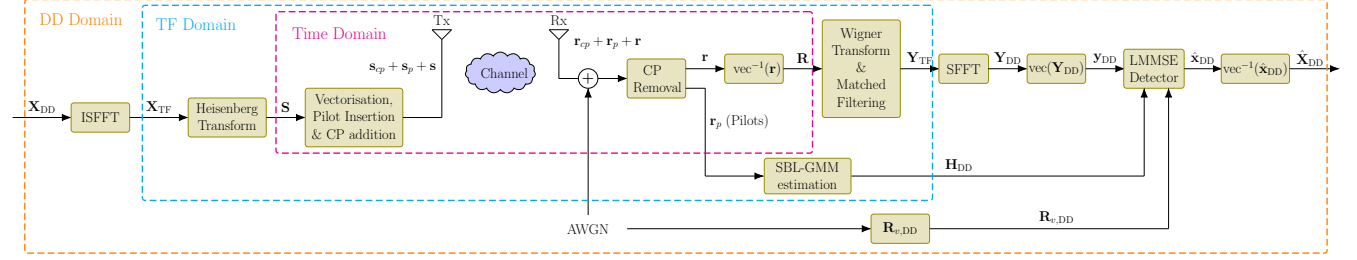


Fig. 1: Schematic diagram of GMM-SBL scheme for CP-aided OTFS transceiver.

transform (ISFFT) as described below:

$$\mathbf{X}_{\text{TF}}(m, n) = \frac{1}{\sqrt{MN}} \sum_{\ell=0}^{M-1} \sum_{c=0}^{N-1} \mathbf{X}_{\text{DD}}(\ell, c) e^{j2\pi(\frac{n\ell}{N} - \frac{m\ell}{M})}. \quad (1)$$

In matrix form this is compactly written as

$$\mathbf{X}_{\text{TF}} = \mathbf{F}_M \mathbf{X}_{\text{DD}} \mathbf{F}_N^H, \quad (2)$$

where $\mathbf{F}_M \in \mathbb{C}^{M \times M}$ and $\mathbf{F}_N \in \mathbb{C}^{N \times N}$ are the unitary discrete Fourier transform (DFT) matrices.

Let $g_{\text{tx}}(t)$ be a rectangular transmit pulse of duration T repeated N times over the OTFS frame. The Heisenberg transform produces the continuous-time transmit signal represented as

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

Sampling $s(t)$ at rate M/T (i.e., with interval T/M) yields MN discrete samples given by $s(g) = s(t)|_{t=gT/M}$, for $g = 0, \dots, MN - 1$. The transmitted sample matrix $\mathbf{S} \in \mathbb{C}^{M \times N}$ may then be expressed as

$$\mathbf{S} = \mathcal{G}_{\text{tx}} \mathbf{F}_M^H \mathbf{X}_{\text{TF}} = \mathcal{G}_{\text{tx}} \mathbf{X}_{\text{DD}} \mathbf{F}_N^H, \quad (4)$$

where $\mathcal{G}_{\text{tx}} \in \mathbb{C}^{M \times M}$ is a diagonal pulse matrix with M samples of the transmit pulse. The vectorized transmit block is obtained as $\mathbf{s} = \text{vec}(\mathbf{S}) \in \mathbb{C}^{MN \times 1}$, which can be written as $\mathbf{s} = (\mathbf{F}_N^H \otimes \mathcal{G}_{\text{tx}}) \mathbf{x}_{\text{DD}}$, where $\mathbf{x}_{\text{DD}} = \text{vec}(\mathbf{X}_{\text{DD}}) \in \mathbb{C}^{MN \times 1}$. To mitigate inter-frame interference, a cyclic prefix (CP) is appended to the transmit vector \mathbf{s} before transmission.

B. Clustered DD domain Channel Model

The DD domain representation of the frequency and time-varying propagation channel is modelled as a superposition of a small number of scatterers and may be written as [4], [11]

$$h(\tau, \nu) = \sum_{i=1}^{L_p} h_i \delta(\tau - \tau_i) \delta(\nu - \nu_i), \quad (5)$$

where L_p is the number of dominant multipath components, τ_i and ν_i are the delay and Doppler shifts of the i th path, $h_i \in \mathbb{C}$ denotes its complex path gain and $\delta(\cdot)$ is the Dirac-delta function. For typical OTFS parameter choices, the continuous DD shifts associated with the multipath

components are mapped onto a discrete DD grid. The delay sampling interval $1/(M\Delta f)$ is typically much smaller than the dominant delay spreads present in wideband systems, which allows path delays to be accurately represented using integer delay taps so that $\tau_i = l_i/(M\Delta f)$ [5], [10], [16]. For high-mobility scenarios, however, the Doppler resolution $1/(NT)$ cannot always be assumed to produce Doppler shifts that coincide with discrete grid points. Consequently, the Doppler shift of the i th multipath component is modeled as $\nu_i = \frac{c_i}{NT}$, where $c_i = k_{\nu_i} + \kappa_{\nu_i}$, $k_{\nu_i} = \text{round}(c_i)$ denotes the nearest integer Doppler tap and κ_{ν_i} represents the associated fractional Doppler component satisfying $|\kappa_{\nu_i}| < \frac{1}{2}$.

Although the exact channel distribution may vary across diverse propagation environments, prior studies have shown that clustered multipath propagation is a fundamental characteristic of wideband and high-mobility wireless channels. In such environments, scattering arises from groups of reflectors with distinct physical characteristics, such as spatial location, reflectivity, and mobility, leading to heterogeneous fading behaviours across multipath components. To account for this statistical heterogeneity, GMMs have been widely adopted as a more statistically flexible approximation to practical multipath channels [24]–[26]. To the best of our knowledge, in existing OTFS channel modelling, per-path gains are typically assumed to follow a single complex Gaussian distribution for analytical simplicity, which corresponds to the special case of a GMM with $K = 1$. Therefore, extending this framework to better capture the clustered multipath statistics, the per-path complex gains h_i are modelled as a complex Gaussian-mixture [27], [28].

$$p(h_i) = \sum_{k=1}^K \rho_k \mathcal{CN}(\mu_k, \sigma_k^2) \quad \text{s.t. } \rho_k \geq 0, \quad \sum_{k=1}^K \rho_k = 1, \quad (6)$$

where $\mathcal{CN}(\mu_k, \sigma_k^2)$ denotes a circularly-symmetric complex Gaussian with mean μ_k and variance σ_k^2 , while K is the number of mixture components. Furthermore, ρ_k denotes the mixture weight associated with the k th Gaussian component. Therefore, with a more flexible statistical description of the path gains characterised by these Gaussian-mixture parameters $\{\rho_k, \mu_k, \sigma_k^2\}$ in (6), the received continuous-time signal

is given as

$$\begin{aligned}
 r(t) &= \iint h(\tau, \nu) s(t - \tau) e^{j2\pi\nu(t-\tau)} d\tau d\nu + \eta(t) \\
 &= \sum_{i=1}^{L_p} h_i s(t - \tau_i) e^{j2\pi\nu_i(t-\tau_i)} + \eta(t),
 \end{aligned} \tag{7}$$

where $\eta(t)$ denotes additive white Gaussian noise with mean zero and variance σ^2 .

Sampling (7) at $t = gT/M$ and discarding the initial CP samples yields the discrete-time samples $r(p) = r(t)|_{t=gT/M}$, $g = 0, \dots, MN - 1$, the sampled input-output relationship becomes

$$\begin{aligned}
 r(p) &= \sum_{i=1}^{L_p} h_i s([p - l_i]_{MN}) e^{j\frac{2\pi c_i}{MN}(p-l_i)} \\
 &\quad + \eta(p), \quad p = 0, \dots, MN - 1,
 \end{aligned} \tag{8}$$

where $[\cdot]_{MN}$ denotes modulo- MN indexing (circular shift). Equation (8) is the sampled input-output law used for constructing the TD channel matrix \mathbf{H} that maps the transmit sample vector $\mathbf{s} \in \mathbb{C}^{MN \times 1}$ to the receive vector $\mathbf{r} \in \mathbb{C}^{MN \times 1}$ via

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \boldsymbol{\eta}, \tag{9}$$

where we have $\mathbf{H} \in \mathbb{C}^{MN \times MN}$ and the noise process obeys $\boldsymbol{\eta} \in \mathbb{C}^{MN \times 1}$. Furthermore, to exploit the DD structure, it is convenient to introduce the standard permutation matrix $\boldsymbol{\Pi} \in \mathbb{C}^{MN \times MN}$ and a diagonal matrix $\boldsymbol{\Delta} \in \mathbb{C}^{MN \times MN}$, where we have

$$\boldsymbol{\omega} = e^{j2\pi c_i/(MN)}, \quad \text{and} \quad \boldsymbol{\Delta} = \text{diag}(1, \boldsymbol{\omega}, \boldsymbol{\omega}^2, \dots, \boldsymbol{\omega}^{MN-1}). \tag{10}$$

Note that $\boldsymbol{\Delta}^{c_i}$ imposes a linear phase shift corresponding to Doppler index c_i , while $\boldsymbol{\Pi}^{l_i}$ represents the l_i -sample forward circular shift [8] upon using these operators, the TD channel matrix is composed as

$$\mathbf{H} = \sum_{i=1}^{L_p} h_i \boldsymbol{\Pi}^{l_i} \boldsymbol{\Delta}^{c_i}. \tag{11}$$

The equations (5)–(11) preserve the DD parameterisation, i.e. delays, Dopplers, and complex gains, forming the basis for the associated dictionary construction and sparse DD domain CSI estimation subsequently.

C. OTFS Demodulation

At the receiver, the continuous-time signal is filtered by a matched filter corresponding to $g_{\text{rx}}(t)$ and sampled in the TF domain. Let $\mathbf{R} = \text{vec}^{-1}(\mathbf{r}) \in \mathbb{C}^{M \times N}$ denote the received sample matrix. The discrete Wigner transform yields the TF-domain receive matrix

$$\mathbf{Y}_{\text{TF}} = \mathbf{F}_M \mathcal{G}_{\text{rx}} \mathbf{R}, \tag{12}$$

where $\mathcal{G}_{\text{rx}} = \text{diag} \left\{ g_{\text{rx}} \left(\frac{gT}{M} \right) \right\}_{g=0}^{M-1}$. Furthermore, the SFFT maps TF-domain samples to the DD domain

$$\mathbf{Y}_{\text{DD}}(\ell, c) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{Y}_{\text{TF}}(m, n) e^{-j2\pi \left(\frac{n\ell}{N} - \frac{m\ell}{M} \right)}, \tag{13}$$

where (13) can also be written as:

$$\mathbf{Y}_{\text{DD}} = \mathbf{F}_M^H \mathbf{Y}_{\text{TF}} \mathbf{F}_N = \mathcal{G}_{\text{rx}} \mathbf{R} \mathbf{F}_N. \tag{14}$$

Vectorizing \mathbf{Y}_{DD} as $\mathbf{y}_{\text{DD}} = \text{vec}(\mathbf{Y}_{\text{DD}}) = (\mathbf{F}_N \otimes \mathcal{G}_{\text{rx}}) \mathbf{r}$, yields the discrete DD domain linear model

$$\mathbf{y}_{\text{DD}} = \mathbf{H}_{\text{DD}} \mathbf{x}_{\text{DD}} + \mathbf{v}_{\text{DD}}, \tag{15}$$

with

$$\mathbf{H}_{\text{DD}} = (\mathbf{F}_N \otimes \mathcal{G}_{\text{rx}}) \mathbf{H} (\mathbf{F}_N^H \otimes \mathcal{G}_{\text{tx}}), \quad \mathbf{v}_{\text{DD}} = (\mathbf{F}_N \otimes \mathcal{G}_{\text{rx}}) \boldsymbol{\eta}. \tag{16}$$

The covariance of \mathbf{v}_{DD} is formulated as:

$$\mathbf{R}_{\mathbf{v}, \text{DD}} = \mathbb{E}[\mathbf{v}_{\text{DD}} \mathbf{v}_{\text{DD}}^H] = \sigma^2 [\mathbf{I}_N \otimes (\mathcal{G}_{\text{rx}} \mathcal{G}_{\text{rx}}^H)]. \tag{17}$$

Furthermore, for unit-power transmitted symbols, the linear MMSE detector formulated in the DD domain is given by

$$\hat{\mathbf{x}}_{\text{MMSE}}^{\text{DD}} = (\mathbf{H}_{\text{DD}}^H \mathbf{R}_{\mathbf{v}, \text{DD}}^{-1} \mathbf{H}_{\text{DD}} + \mathbf{I}_{MN})^{-1} \mathbf{H}_{\text{DD}}^H \mathbf{R}_{\mathbf{v}, \text{DD}}^{-1} \mathbf{y}_{\text{DD}}. \tag{18}$$

Equations (15)–(18) thus establish the discrete DD domain input-output model of the OTFS system. This compact representation directly exploits the inherent sparsity of the DD domain channel, providing the foundation for the sparse CSI estimation framework discussed in the next section.

III. GMM-SBL Based Sparse CSI Estimation for OTFS Systems

Let $\mathbf{s}_p \in \mathbb{C}^{N_p \times 1}$ denote the time-domain (TD) pilot vector inserted between OTFS frames for channel estimation.

The DD channel has a sparse representation

$$h(\tau, \nu) = \sum_{i=0}^{M_\tau-1} \sum_{j=0}^{G_\nu-1} h_i^j \delta(\tau - \tau_i) \delta(\nu - \nu_j), \tag{19}$$

where $\tau_i = \frac{i}{M\Delta f}$, and $\nu_j = \frac{jN_\nu}{G_\nu NT}$. Here, M_τ and N_ν denote the maximum delay and integer-Doppler spreads of the channel satisfying $M_\tau \ll M$ and $N_\nu \ll N$ for a typical under-spread channel. As discussed before, since the delay resolution is sufficiently high, it is adequate to represent delays using integer delay taps. Whereas, to accurately capture fractional Doppler shifts, the Doppler grid size is chosen as $G_\nu \gg N_\nu$, thereby refining the Doppler sampling resolution. In this formulation, the integer Doppler tap $k_{\nu_j} = \text{round} \left(\frac{jN_\nu}{G_\nu} \right)$, and the fractional Doppler is represented by $\kappa_{\nu_j} = \frac{jN_\nu}{G_\nu} - \text{round} \left(\frac{jN_\nu}{G_\nu} \right)$. Since only a few dominant reflectors exist, the DD channel remains sparse over the $M_\tau G_\nu$ grid, with $L_p \ll M_\tau G_\nu$ non-zero coefficients.

After CP removal, the received TD pilot vector is expressed as

$$\mathbf{r}_p = \mathbf{H}\mathbf{s}_p + \boldsymbol{\eta}_p, \tag{20}$$

where $\boldsymbol{\eta}_p \in \mathbb{C}^{N_p \times 1}$ denotes additive white Gaussian noise.

The effective channel $\bar{\mathbf{H}}$ can be written as

$$\bar{\mathbf{H}} = \sum_{i=0}^{M_\tau-1} \sum_{j=0}^{G_\nu-1} h_i^j (\bar{\mathbf{\Pi}})^i (\bar{\mathbf{\Delta}}_i)^j, \quad (21)$$

with $\bar{\mathbf{\Pi}} \in \mathbb{C}^{N_p \times N_p}$ a permutation matrix and $\bar{\mathbf{\Delta}}_i \in \mathbb{C}^{N_p \times N_p}$ a diagonal matrix given by

$$\bar{\mathbf{\Delta}}_i = \begin{cases} \text{diag}(\omega^0, \omega^1, \dots, \omega^{N_p-i-1}, \omega^{-i}, \dots, \omega^{-1}), & \text{if } i \neq 0, \\ \text{diag}(\omega^0, \omega^1, \dots, \omega^{N_p-1}), & \text{if } i = 0, \end{cases} \quad (22)$$

where $\omega = e^{j2\pi \frac{N_\nu}{G_\nu M_\tau N}}$. Each ω_i^j represents the contribution of the pilot sequence \mathbf{s}_p as it is shifted in delay by i samples (via $(\bar{\mathbf{\Pi}})^i$) and modulated in Doppler by j shifts (via $(\bar{\mathbf{\Delta}}_i)^j$). Thus, ω_i^j can be interpreted as the effective measurement vector corresponding to the (i, j) th DD bin. By stacking all such vectors column-wise, we obtain the dictionary matrix $\mathbf{\Omega}$ and the sparse vector \mathbf{h} as

$$\mathbf{\Omega} = [\boldsymbol{\omega}_0^0, \boldsymbol{\omega}_0^1, \dots, \boldsymbol{\omega}_0^{N_\nu-1}, \boldsymbol{\omega}_1^0, \dots, \boldsymbol{\omega}_{M_\tau-1}^{G_\nu-1}] \quad (23)$$

$$\mathbf{h} = [h_0^0, h_0^1, \dots, h_0^{N_\nu-1}, h_1^0, \dots, h_{M_\tau-1}^{G_\nu-1}]^T. \quad (24)$$

The pilot observation model then becomes

$$\mathbf{r}_p = \mathbf{\Omega} \mathbf{h} + \boldsymbol{\eta}_p, \quad (25)$$

which forms the basis for the subsequent GMM-SBL-based sparse DD domain channel estimation.

The DD domain channel response in OTFS systems is characterised by a sparse structure arising from a finite number of dominant multipath components. However, in practical scenarios, the complex gains associated with these components may exhibit non-stationary statistics, often organised into distinct clusters due to the presence of grouped scattering objects. This gives rise to a multimodal distribution of tap gains, a feature that is not captured by conventional estimators relying on a homogeneous single Gaussian prior.

To better model this statistical heterogeneity and inherent sparsity in the DD domain, a GMM-SBL framework is proposed for sparse CSI estimation. The proposed framework introduces a more expressive, hierarchical prior that models the channel as a mixture of K Gaussian components

$$p(\mathbf{h}) = \sum_{k=1}^K \rho_k \mathcal{CN}(\boldsymbol{\mu}_k, \mathbf{\Gamma}_k), \quad \sum_{k=1}^K \rho_k = 1, \quad \rho_k \geq 0, \quad (26)$$

where $\mathbf{\Gamma}_k = \text{diag}(\gamma_k)$ is the diagonal covariance matrix for the k th component of the channel \mathbf{h} . Also, the corresponding marginal covariance matrix is \mathbf{A}_k . The posterior mean and covariance under component k are denoted by $\boldsymbol{\mu}_k$ and $\mathbf{\Sigma}_k$, respectively. The responsibility is written as $\pi_k(\mathbf{r}_P) = \Pr(k | \mathbf{r}_P)$, and $\pi_{i,k}$ in the multi-snapshot setting.

The GMM-SBL approach represents a significant paradigm shift from imposing a sparse structure to enabling data-driven learning, as it allows the model to capture diverse channel statistics within a unified Bayesian framework. A critical advantage of this formulation is its strong theoretical

foundation [29], which exhibits sparsity-inducing properties, as shown in Theorem 1.

Theorem 1 (Sparsity of the GMM-SBL Prior): Let the latent variable $z \in \{1, \dots, K\}$ indicate the mixture component with probability $\Pr(z=k) = \rho_k$. Consider the hierarchical prior

$$p(\mathbf{h}) = \sum_{k=1}^K \rho_k \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma}_k),$$

where $\mathbf{\Gamma}_k = \text{diag}(\gamma_k)$ with $\gamma_k > 0$. Then there exists a constant $C > 0$ so that

$$p(\mathbf{h}) \leq C \cdot \prod_{r=1}^{M_\tau N_\nu} \frac{1}{|h_r|^2}. \quad (27)$$

Proof: Given in Appendix A.

This result ensures that the GMM-SBL prior is bounded by a sparsity-promoting function [30]. For our problem, let the function be $t(h)$ obey $t(h) \propto \prod_r |h(r)|^{-2}$. Consequently, evidence maximisation within the proposed framework is guaranteed to promote the robust recovery of sparse solutions, while providing the flexibility to learn the parameters γ_k, ρ_k and adapt to the underlying channel structure. The GMM-SBL framework models the DD domain channel vector with a hierarchical prior that adaptively learns sparsity patterns from data. Given the received pilots, the posterior distribution conditioned on mixture component k is Gaussian, and it is defined as

$$p(\mathbf{h} | \mathbf{r}_p, k) = \mathcal{CN}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k), \quad (28)$$

with the posterior mean $\boldsymbol{\mu}_k$ and posterior covariance $\mathbf{\Sigma}_k$ under component k defined as

$$\mathbf{A}_k = \sigma^2 \mathbf{I} + \mathbf{\Omega} \mathbf{\Gamma}_k \mathbf{\Omega}^H, \quad (29)$$

$$\boldsymbol{\mu}_k = \mathbf{\Gamma}_k \mathbf{\Omega}^H \mathbf{A}_k^{-1} \mathbf{r}_p, \quad (30)$$

$$\mathbf{\Sigma}_k = \mathbf{\Gamma}_k - \mathbf{\Gamma}_k \mathbf{\Omega}^H \mathbf{A}_k^{-1} \mathbf{\Omega} \mathbf{\Gamma}_k. \quad (31)$$

The marginal likelihood under component k is

$$p(\mathbf{r}_p | k) = \mathcal{CN}(\mathbf{0}, \mathbf{A}_k).$$

The responsibility $\pi_k(\mathbf{r}_P)$ is the posterior probability that the k th mixture component generated the received pilot vector \mathbf{r}_p ; according to the Bayes' rule

$$\pi_k(\mathbf{r}_P) = \Pr(k | \mathbf{r}_P) = \frac{\rho_k p(\mathbf{r}_P | k)}{\sum_{\ell=1}^K \rho_\ell p(\mathbf{r}_P | \ell)},$$

so $0 \leq \pi_k(\mathbf{r}_P) \leq 1$ and $\sum_{k=1}^K \pi_k(\mathbf{r}_P) = 1$. The responsibilities provide a probabilistic assignment of observations to components and quantify each component's contribution to the posterior distribution [31]. As will be seen in subsequent parameter updates, they act as adaptive weights for learning $\mathbf{\Gamma}_k$ and ρ_k from the aggregated posterior means and variances.

Also, the conditional-mean estimate (CME) is given by

$$\hat{\mathbf{h}} = \sum_{k=1}^K \pi_k(\mathbf{r}_p) \boldsymbol{\mu}_k.$$

Theorem 2 (EM updates for GMM-SBL): Given L pilot observations $\{\mathbf{r}_{p,i}\}_{i=1}^L$, the EM updates at iteration t are given by

E-step. For each snapshot i and mixture component k compute the posterior mean

$$\boldsymbol{\mu}_{i,k}^{(t)} = \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H (\mathbf{A}_k^{(t)})^{-1} \mathbf{r}_{p,i},$$

and the posterior covariance

$$\boldsymbol{\Sigma}_k^{(t)} = \boldsymbol{\Gamma}_k^{(t)} - \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H (\mathbf{A}_k^{(t)})^{-1} \boldsymbol{\Omega} \boldsymbol{\Gamma}_k^{(t)},$$

where we have

$$\mathbf{A}_k^{(t)} = \sigma^2 \mathbf{I} + \boldsymbol{\Omega} \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H.$$

Form the responsibility

$$\pi_{i,k}^{(t)} = \frac{\rho_k^{(t)} p(\mathbf{r}_{p,i} | k; \mathbf{A}_k^{(t)})}{\sum_{\ell=1}^K \rho_\ell^{(t)} p(\mathbf{r}_{p,i} | \ell; \mathbf{A}_\ell^{(t)})},$$

where we have

$$p(\mathbf{r}_{p,i} | k; \mathbf{A}_k^{(t)}) = \mathcal{CN}(\mathbf{0}, \mathbf{A}_k^{(t)}), \quad \mathbf{A}_k^{(t)} = \sigma^2 \mathbf{I} + \boldsymbol{\Omega} \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H.$$

M-step. For $r = 1, \dots, M_\tau N_\nu$ update the hyperparameters $\gamma_{k,r}$ and the mixture weights ρ_k , where we have:

$$\gamma_{k,r}^{(t+1)} = \frac{\sum_{i=1}^L \pi_{i,k}^{(t)} (|\mu_{i,k,r}^{(t)}|^2 + \Sigma_{k,r,r}^{(t)})}{\sum_{i=1}^L \pi_{i,k}^{(t)}}, \quad (32)$$

$$\rho_k^{(t+1)} = \frac{1}{L} \sum_{i=1}^L \pi_{i,k}^{(t)}. \quad (33)$$

Proof: Given in Appendix B.

After convergence (at iteration T), the CME for each snapshot is obtained as:

$$\hat{\mathbf{h}}^{(i)} = \sum_{k=1}^K \pi_{i,k}^{(T)} \boldsymbol{\mu}_{i,k}^{(T)}, \quad i = 1, \dots, L.$$

Although the proposed formulation employs a Gaussian mixture prior, sparsity is enforced through the relevance determination mechanism inherent to sparse Bayesian learning. In conventional SBL, a single Gaussian prior is combined with hyperparameter-driven variance adaptation to promote sparse solutions. By contrast, mixture-prior approaches typically incorporate mixture structures directly at the coefficient level. The proposed GMM-SBL framework preserves the explicit hyperparameter based sparsity control of SBL while introducing mixture modeling to capture statistical heterogeneity across channel realizations. The posterior responsibilities act as adaptive weights in the hyperparameter updates and do not impose coefficientwise hard clustering of DD taps. Consequently, the method extends the modeling capability of SBL, while retaining its underlying EM-based learning structure. Algorithm-1 details the steps of GMM-SBL based channel estimation for the OTFS system considered.

Algorithm 1: GMM-SBL Based CSI Estimation in OTFS

```

1 Input: Snapshots  $\mathbf{R}' = \{\mathbf{r}_{p,i}\}_{i=1}^L$ , dictionary matrix  $\boldsymbol{\Omega} \in \mathbb{C}^{N_p \times M_\tau N_\nu}$ ,
   noise variance  $\sigma^2$ , mixture components  $K$ , EM iterations  $T$ .
2 Output:  $\{\rho_k^{(T)}\}_{k=1}^K$ ,  $\{\boldsymbol{\Gamma}_k^{(T)}\}_{k=1}^K$ ,  $\{\hat{\mathbf{h}}_i\}_{i=1}^L$ .
3 Initialize:  $\rho_k^{(0)} = 1/K$ ,  $\boldsymbol{\Gamma}_k^{(0)} = \mathbf{I}_{M_\tau N_\nu}$ .
4 for  $t \leftarrow 0$  to  $T - 1$  do
5   for  $k \leftarrow 1$  to  $K$  do
6      $\mathbf{A}_k^{(t)} = \sigma^2 \mathbf{I}_{N_p} + \boldsymbol{\Omega} \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H$ .
7     for  $i \leftarrow 1$  to  $L$  do
8        $\mathbf{u}_{i,k}^{(t)} = \mathbf{A}_k^{(t)-1} \mathbf{r}_{p,i}$ .
9        $\boldsymbol{\mu}_{i,k}^{(t)} = \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H \mathbf{u}_{i,k}^{(t)}$ .
10       $\boldsymbol{\Sigma}_k^{(t)} = \boldsymbol{\Gamma}_k^{(t)} - \boldsymbol{\Gamma}_k^{(t)} \boldsymbol{\Omega}^H \mathbf{A}_k^{(t)-1} \boldsymbol{\Omega} \boldsymbol{\Gamma}_k^{(t)}$ .
11       $\ell_{i,k}^{(t)} = -(\mathbf{r}_{p,i}^H \mathbf{A}_k^{(t)-1} \mathbf{r}_{p,i} + \log \det \mathbf{A}_k^{(t)} + N_p \log(\pi))$ .
12   for  $i \leftarrow 1$  to  $L$  do
13      $\tilde{\pi}_{i,k}^{(t)} = \rho_k^{(t)} e^{\ell_{i,k}^{(t)}}$ ,  $k = 1, \dots, K$ .
14      $\pi_{i,k}^{(t)} = \tilde{\pi}_{i,k}^{(t)} / \sum_{r=1}^K \tilde{\pi}_{i,r}^{(t)}$ .
15   for  $k \leftarrow 1$  to  $K$  do
16      $N_k^{(t)} = \sum_{i=1}^L \pi_{i,k}^{(t)}$ .
17      $\mathbf{e}_{i,k}^{(t)} = |\boldsymbol{\mu}_{i,k}^{(t)}|^2 + \text{diag}(\boldsymbol{\Sigma}_k^{(t)})$ .
18      $\boldsymbol{\Gamma}_k^{(t+1)} = \text{diag}\left(\frac{1}{N_k^{(t)}} \sum_{i=1}^L \pi_{i,k}^{(t)} \mathbf{e}_{i,k}^{(t)}\right)$ .
19      $\rho_k^{(t+1)} = N_k^{(t)} / L$ .
20 for  $i \leftarrow 1$  to  $L$  do
21    $\hat{\mathbf{h}}_i = \sum_{k=1}^K \pi_{i,k}^{(T)} \boldsymbol{\mu}_{i,k}^{(T)}$ .

```

IV. Performance Benchmarks

To assess the estimator's performance, a pair of reference bounds is employed. First is an Oracle-MMSE (genie estimator) assumes perfect knowledge of the DD domain support set \mathcal{H} . Let $\boldsymbol{\Omega}_O = \boldsymbol{\Omega}(\cdot, \mathcal{H})$ denote the Oracle sensing matrix. The Oracle-MMSE estimate is formulated as:

$$\hat{\mathbf{h}}_{\text{O-MMSE}} = (\boldsymbol{\Omega}_O^H \mathbf{R}_v^{-1} \boldsymbol{\Omega}_O + \mathbf{I})^{-1} \boldsymbol{\Omega}_O^H \mathbf{R}_v^{-1} \mathbf{r}_p.$$

The simulated Oracle implements the analytic bound by applying MMSE on the true support to \mathbf{r}_p and averaging the normalised squared error over Monte Carlo trials. The second benchmark considered is the BCRLB, which captures fundamental limits when a prior $p(\mathbf{h})$ is available [8], [32]. Given L snapshots, the Bayesian information decomposes as $\mathbf{J} = \mathbf{J}_{\text{data}} + \mathbf{J}_{\text{prior}}$ with $\mathbf{J}_{\text{data}} = L \boldsymbol{\Omega}^H \mathbf{R}_v^{-1} \boldsymbol{\Omega}$ and $\mathbf{J}_{\text{prior}} = -\mathbb{E}_{\mathbf{h}}[\nabla_{\mathbf{h}}^2 \log p(\mathbf{h})]$. For a general K -component complex Gaussian mixture prior, $\mathbf{J}_{\text{prior}}$ has no closed form [33]. The BCRLB for a single-Gaussian prior $p(\mathbf{h}) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma})$ can be expressed as (see Appendix C for derivation)

$$\text{MSE}(\hat{\mathbf{h}}) \geq \text{tr}\left([\mathbf{J}_{\text{data}} + \mathbf{J}_{\text{prior}}]^{-1}\right). \quad (34)$$

The prior term reduces to $\mathbf{J}_{\text{prior}} = \boldsymbol{\Gamma}^{-1}$, and the BCRLB admits the closed form shown above when the mixture components share uniform variance. Moreover, when L is large, the data term dominates, and the BCRLB coincides with the Oracle reference; otherwise the Monte Carlo-evaluated $\mathbf{J}_{\text{prior}}$ captures the effect of the GMM prior on the fundamental MSE limit.

Table 2: Simulation Parameters

Parameter (Symbol)	Value
Carrier Frequency in GHz (f_c)	4
Subcarrier spacing in kHz (Δf)	15
No. of symbols along delay-axis (M)	32
No. of symbols along Doppler-axis (N)	32
Max. spread across delay-axis (M_τ)	16
Max. spread across Doppler-axis (N_ν)	10
Doppler grid size (G_ν)	20
No. of samples in CP (P)	16
No. of pilots (N_p)	80
No. of dominant reflectors (L_p)	5
Modulation scheme	QPSK
No. of training snapshots (L)	10
Pulse-shape	Rectangular

Table 3: DD-Profile of the Wireless Channel

Path-index (i)	1	2	3	4	5
Delay in μs (τ_i)	2.08	4.164	6.246	8.328	10.42
Doppler in Hz (ν_i)	0	470	940	1410	1880
Speed in km/h	0	126.9	253.8	380.7	507.6

V. Results and Discussion

This section presents a comprehensive performance evaluation of the proposed GMM-SBL scheme for sparse DD domain channel estimation. The normalised mean square error (NMSE) and symbol error rate (SER) are used as performance metrics, where the NMSE is defined as

$$\text{NMSE} = \frac{\|\hat{\mathbf{H}}_{\text{DD}} - \mathbf{H}_{\text{DD}}\|^2}{\|\mathbf{H}_{\text{DD}}\|^2}, \quad (35)$$

where $\hat{\mathbf{H}}_{\text{DD}}$ and \mathbf{H}_{DD} denote the estimated and the true DD domain channel matrices, respectively, calculated from (16).

A. Simulation Parameters

The OTFS system parameters (including M , N , M_τ , and N_ν) are selected to reflect the high-mobility scenarios and to satisfy the underspread channel assumption commonly adopted in OTFS literature [8], [16], [34], [35]. The grid sizes $M \in \{32, 64\}$ chosen are appropriate for the considered bandwidth and delay spread, enabling accurate representation of multipath delays on the integer delay grid. Similarly, $N = 32$, together with the fractional grid formulation, enables accurate modeling of fractional Doppler effects. The delay and Doppler supports M_τ and N_ν are selected to capture the effective channel spread while maintaining $M_\tau \ll M$ and $N_\nu \ll N$, consistent with practical OTFS system design.

To ensure fairness, the delay and Doppler indices are randomly generated for each Monte Carlo trial, and the pilot power is normalised across all schemes. The maximum

number of EM iterations is set to 100 for both GMM-SBL and SBL. For Least absolute shrinkage and selection operator (LASSO) a fixed sparsity regularisation parameter of 10^{-3} is used. Focal underdetermined system solver (FOCUSS) employs an ℓ_p norm with $p = 0.8$, a noise variance-based regularization, a stopping tolerance of 10^{-6} , and a maximum of 500 iterations. The OMP algorithm uses a residual-based stopping rule, terminating once the change in residual error falls below a noise-dependent threshold, which we set to 10^{-2} .

These parameter values were selected after preliminary evaluation across the considered SNR range to identify stable operating settings for each algorithm. In particular, the regularisation parameter for LASSO, the ℓ_p -norm parameter p and stopping tolerance for FOCUSS, as well as the residual-based threshold for OMP were identified to ensure reliable convergence behaviour and consistent estimation performance under the specified channel conditions. For SBL and GMM-SBL, the maximum number of EM iterations was set identically to provide comparable convergence conditions. For each benchmark method, the results reported correspond to the configuration that demonstrated stable performance across Monte Carlo trials, thereby enabling a meaningful comparison with the proposed approach.

B. Performance Comparison

The method is compared to state-of-the-art compressed sensing techniques, including OMP [36], FOCUSS [37], SBL [30], and LASSO [38]. In addition, performance is evaluated against the embedded pilot [EP] based OTFS channel estimation scheme [5], which serves as a classical benchmark, as well as a recent DL-based approach that combines OMP with a denoiser [39]. As observed in Fig. 2a, the EP based OTFS scheme achieves reliable performance; however, it relies on DD domain impulse pilots, guard regions, and threshold based detection. By contrast, the improved GMM-SBL framework conceived employs TD pilots and performs Bayesian inference under a learned statistical prior, thereby avoiding threshold tuning, while adaptively exploiting clustered DD sparsity.

The DL-based method achieves competitive performance in the low-SNR regime, but at the cost of substantial training overhead, which is not possible in high mobility scenarios. Nevertheless, its performance saturates at higher SNR values, as the denoiser is primarily trained for noise suppression rather than for modelling the underlying sparse channel structure, limiting further improvement when the noise becomes negligible. Classical sparse recovery methods exhibit additional limitations. OMP is sensitive to the stopping criterion, FOCUSS may suffer from slow or unstable convergence, and LASSO experiences estimation bias under DD domain sparsity. Conventional SBL improves robustness through hyperparameter learning; however, its single-Gaussian prior restricts statistical flexibility. By incorpo-

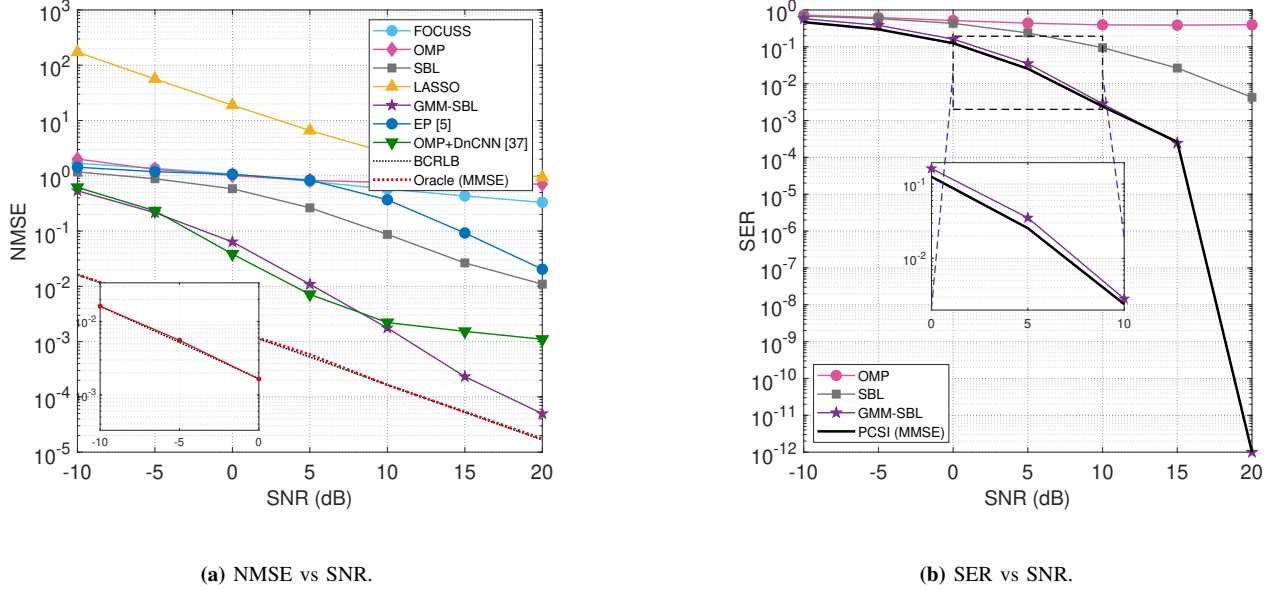


Fig. 2: Performance comparison of proposed GMM-SBL algorithm with benchmarks and existing state-of-the-art methods.

rating mixture-level statistical modelling within the SBL framework, the proposed GMM-SBL achieves consistently lower NMSE across the evaluated SNR range.

Furthermore, the proposed estimator is benchmarked against the Oracle-MMSE and the BCRLB derived in Section IV. As shown in Fig. 2a and Fig. 2b, GMM-SBL consistently outperforms all baselines, and at high SNR its NMSE approaches the Oracle-MMSE and BCRLB, demonstrating near-optimal estimation accuracy. These NMSE improvements directly translate into SER gains, where GMM-SBL achieves near-perfect CSI detection while OMP and FOCUSS exhibit significant performance degradation

C. Parameter Sensitivity Analysis

The robustness of GMM-SBL is further evaluated under varying pilot length and number of snapshots. Fig. 3a shows that increasing the number of pilot symbols N_p significantly reduces the NMSE. We observe more than 6 dB performance improvement upon increasing N_p from 80 to 140. Moreover, it can be seen in Fig. 3b that additional snapshots provide improved NMSE performance. Explicitly, at 0 dB the NMSE decreases from 6.39×10^{-1} with $L = 1$ to 7.13×10^{-2} with $L = 10$.

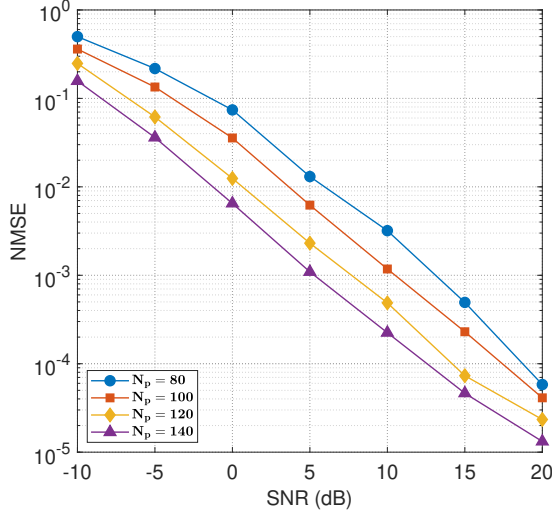
Furthermore, to evaluate the sensitivity of the proposed method under realistic mobility conditions, fractional Doppler shifts are incorporated in the channel model. Specifically, the normalized Doppler index of each multipath component includes an integer Doppler tap k_{ν_i} , while the fractional Doppler term κ_{ν_i} is uniformly generated in the range $(-\frac{1}{2}, \frac{1}{2})$. To accurately capture these off-grid components, the grid resolution was increased such that

$G_\nu \gg N_\nu$, which refines the Doppler sampling. Because the physical Doppler frequencies do not lie exactly on the grid, the channel energy spreads across adjacent Doppler bins, which makes the estimation task more challenging, particularly when the number of pilot observations is limited. Nevertheless, as shown in Fig. 4a, the proposed estimator maintains stable performance under fractional Doppler, and increasing N_p significantly improves accuracy by providing additional observations for the expanded DD representation.

Next, the impact of increasing the number of subcarriers M is examined. In OTFS, the delay resolution is given by $\Delta\tau = 1/(M\Delta f)$; thus, increasing M refines the physical delay grid. However, the size of the inverse problem depends on the modeled DD grid $M_\tau N_\nu$ and the pilot length N_p , and does not depend directly on M . As seen in Fig. 4b, it can be concluded that increasing M from 32 to 64 with fixed M_τ results in a small variation in estimation accuracy. Furthermore, when M_τ is moderately increased (while still satisfying $M_\tau \ll M$), the performance is effectively restored. Therefore, increasing M enhances delay resolution without increasing computational complexity since $M_\tau \ll M$, demonstrating that the proposed method remains scalable without additional computational overhead.

D. Impact of Mixture Order Selection

To study the effect of model-order mismatch, channels were generated with $K_{\text{true}} = 1, 2, 4$ and the GMM-SBL method was evaluated for $K_{\text{model}} \in \{1, 2, 4\}$. These values were selected to cover $K = 1$, which corresponds to a single



(a) NMSE vs SNR for different pilot overhead.

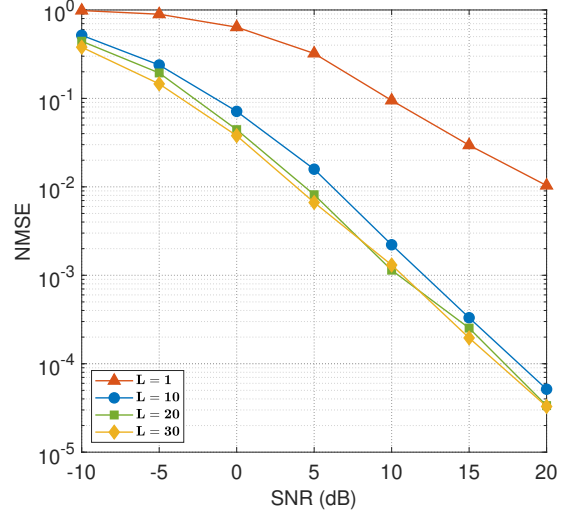
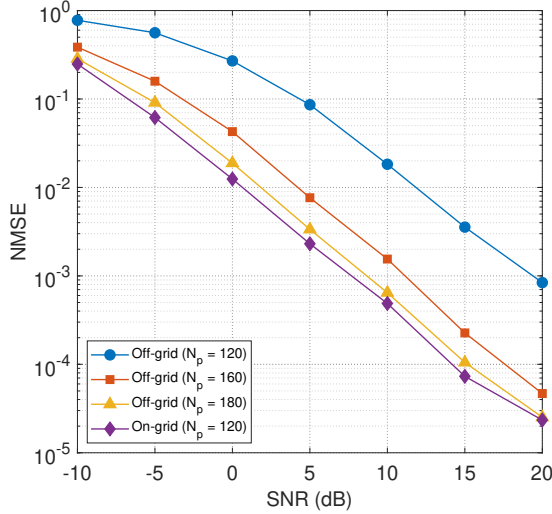
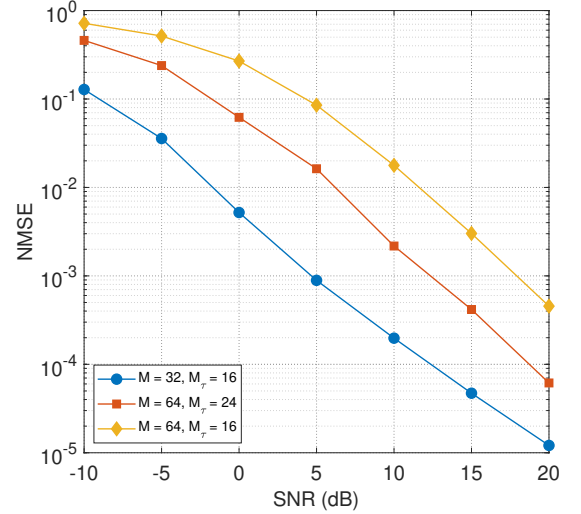

 (b) NMSE vs SNR for different No. of snapshots L .

Fig. 3: Analysis of proposed GMM-SBL algorithm with different parameters.



(a) NMSE vs SNR with fractional Doppler conditions.



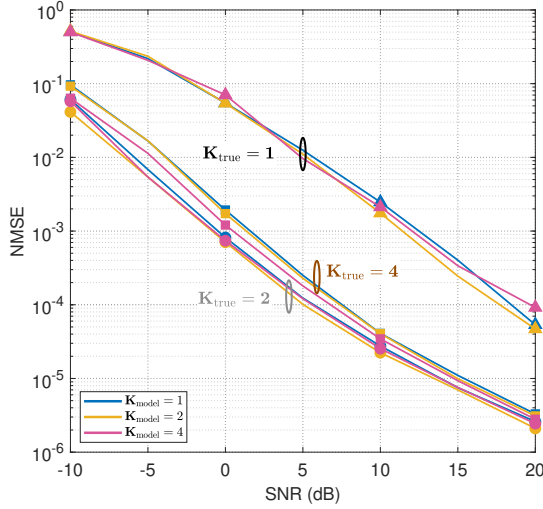
(b) NMSE vs SNR for different delay resolutions.

Fig. 4: Impact of DD grid resolution on GMM-SBL channel estimation performance.

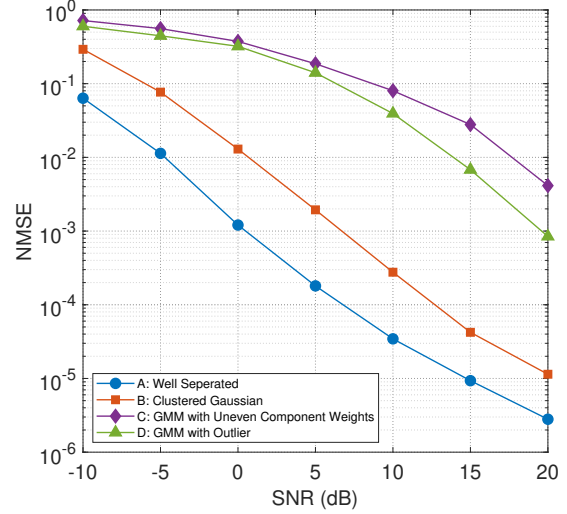
Gaussian prior (commonly used in channel modeling). Furthermore, $K = 2$ represents the simplest Gaussian mixture case, and $K = 4$ allows the examination of a higher-order mixtures. The mixture weights (ρ_k), mean (μ_k) and variances (σ_k^2) are chosen to yield balanced component contributions along with well-separated clusters, thereby identifying the number of clusters K to be the central objective. As observed in Fig. 5a, $K_{\text{true}} = 1$, $K_{\text{model}} = 1$

corresponds to the standard single Gaussian prior, which is equivalent to the baseline assumption in SBL.

In particular, when the underlying channel follows a two-component mixture ($K_{\text{true}} = 2$), the proposed GMM-SBL with $K_{\text{model}} = 2$ achieves approximately 13% NMSE reduction at 0 dB and nearly 18% at 5 dB compared to conventional single Gaussian SBL ($K_{\text{model}} = 1$). For more heterogeneous channels ($K_{\text{true}} = 4$), the improvement



(a) NMSE vs SNR across different K .



(b) NMSE vs SNR for GMM component separation cases.

Fig. 5: Analysis of the proposed GMM-SBL algorithm with clustered channel.

becomes more pronounced; at 0 dB, the NMSE reduction exceeds 35% relative to the single Gaussian SBL baseline. When the true channel distribution is unimodal ($K_{\text{true}} = 1$), the single Gaussian configuration ($K_{\text{model}} = 1$) achieves the lowest NMSE among the evaluated models, while higher order mixture configurations result in only minor performance variation. These results demonstrate that the performance gain of the proposed GMM-SBL framework arises from improved statistical alignment between the assumed prior and the underlying channel distribution.

The improved performance for $K_{\text{model}} = 2$, can be attributed to the enhanced flexibility to capture variations in the channel coefficient distribution that are not well represented by a single Gaussian. By contrast, $K_{\text{model}} = 4$ leads to a slight degradation due to over-parameterisation. For $K_{\text{true}} = 2$, the matched case with $K_{\text{model}} = 2$ yields the best NMSE, since the estimator correctly reflects the underlying two-component Gaussian mixture. Here, $K_{\text{model}} = 1$ underfits by imposing a single-Gaussian approximation, while $K_{\text{model}} = 4$ redistributes weights across unnecessary components and introduces estimation noise. For $K_{\text{true}} = 4$, the estimator with $K_{\text{model}} = 4$ best aligns with the four-component prior, while $K_{\text{model}} = 2$ provides a reasonable approximation by grouping components, and $K_{\text{model}} = 1$ shows consistently higher NMSE, especially at low SNR. Overall, fixing $K = 2$ represents a robust and computationally efficient compromise; it exactly captures the two-component case, while remaining sufficiently expressive for unimodal and higher-order mixtures, thus generalising well across practical sparse CSI estimation scenarios in OTFS systems, making $K = 2$ the most suitable choice for our work.

E. Effect of Mixture Configurations

To explore the full potential of the GMM-SBL algorithm, we further tested diverse scenarios having channel distributions following different Gaussian mixtures, as described in Table 4. Specifically, four representative cases having $K = 4$ are examined to span well-separated, partially overlapping, variance-differentiated, and outlier-dominated conditions, by suitably choosing their mean and variance.

As depicted in Fig. 5b, in Case A, the best performance is achieved, since well separated means and small variances ensure minimal overlap, enabling the estimator to identify the correct channel taps with high precision. In Case B, a moderate degradation is observed because, closely spaced means cause interference between taps within each pair. Case C is the most challenging one, having the worst performance, because identical means and differing variances produce a distribution that resembles a single dominant cluster at low SNR, forcing the estimator to purely rely on variance-based separation. Finally, Case D highlights noise sensitivity with a loss in performance since infrequent, high-amplitude channel taps are difficult to detect under noisy conditions, but become recoverable at higher SNR.

Table 4: Parameterisation of GMMs for Channel Generation

Case	GMM Description	Mixture Weights ρ_k
A	Well-Separated	[0.25, 0.25, 0.25, 0.25]
B	Clustered Gaussian	[0.25, 0.25, 0.25, 0.25]
C	GMM with Uneven Component Weights	[0.4, 0.1, 0.4, 0.1]
D	Gaussian Mixture with Outlier	[0.7, 0.15, 0.1, 0.05]

Table 5: Computational Complexity of GMM-SBL

Operation	Complex multiplications	Complex additions
Formation of $\mathbf{A}_k = \sigma^2 \mathbf{I} + \Omega \mathbf{\Gamma}_k \Omega^H$	$K(N_p^2 M_\tau N_\nu + N_p^3)$	$K(N_p^2 M_\tau N_\nu + N_p^3)$
Update of $\mathbf{u}_{i,k}$ $\mathbf{u}_{i,k} = \mathbf{A}_k^{-1} \mathbf{r}_{p,i}$	$KL N_p^2$	$KL N_p(N_p - 1)$
Posterior mean $\boldsymbol{\mu}_{i,k} = \mathbf{\Gamma}_k \Omega^H \mathbf{u}_{i,k}$	$KL M_\tau N_\nu (N_p + 1)$	$KL M_\tau N_\nu (N_p - 1)$
Diagonal covariance $\boldsymbol{\Sigma}_k = \mathbf{\Gamma}_k - \mathbf{\Gamma}_k \Omega^H \mathbf{A}_k^{-1} \Omega \mathbf{\Gamma}_k$	$K M_\tau N_\nu N_p^2$	$K M_\tau N_\nu N_p^2$
Hyperparameter updates ($\mathbf{\Gamma}_k, \rho_k$)	$K M_\tau N_\nu (L + 1)$	$K(L - 1)(1 + M_\tau N_\nu)$

F. Computational Complexity and Pilot Overhead

It can be observed from Table 5 that the per-iteration complexity scales linearly with the maximum delay and Doppler support (M_τ, N_ν) and with the mixture order K , while the cubic term arises from the inversion of covariance matrices of dimension $N_p \times N_p$. Therefore, the computational burden depends primarily on the pilot dimension N_p and the sparse DD grid parameters (M_τ, N_ν), rather than on the full OTFS frame size (M, N). Since typically $M_\tau \ll M$ and $N_\nu \ll N$, the effective inverse problem dimension remains significantly smaller than the full OTFS grid. Additionally, for the configurations considered in this work ($N_p = 80$, $K = 2$), accurate estimation performance is achieved without requiring excessive pilot dimensions or high mixture orders, resulting in moderate computational complexity.

Further, computational cost in the work is a justified trade-off, because GMM-SBL learns a data-driven, multimodal prior that captures complex channel structures. Consequently, GMM-SBL achieves significantly lower NMSE and BER, providing robust performance even in low-SNR regimes.

The proposed method inserts pilots directly in the TD, giving a pilot overhead of $\alpha_t = \frac{N_p}{MN + N_p}$. With the parameters of Table 2, this evaluates to about 0.0725. For context and fair comparison to the proposed method, both DD and TF domain pilot schemes from existing SBL-based CSI estimation are considered. A DD domain pilot scheme without guard symbols and with approximate overhead of $\alpha_{\text{DD}} \approx \frac{(4N_\nu + 1)M_\tau}{MN}$ is discussed in [14]. Although efficient, this expression depends on (N_ν, M_τ) , therefore leading to enhanced pilot overhead in highly dispersive channels. A TF domain pilot scheme proposed in [8] gives $\alpha_{\text{TF}} = \frac{N_p}{N + N_p}$. Although this scheme avoids dependency on (N_ν, M_τ) , it requires MN_p pilot symbols and involves additional linear transforms and structured matrix operations for pilots, leading to increased receiver and transmitter complexity. By contrast, the proposed TD scheme transmits only N_p raw pilot symbols, avoids any DD guard interval, and yields a direct input-output relationship for sparse DD domain CSI recovery, thereby providing a simpler transceiver structure and a lower effective pilot overhead.

VI. Conclusions

To enhance the channel estimation performance of OTFS systems, a GMM-SBL framework was proposed, which exploits the essential DD domain sparsity, while significantly improving statistical modeling using a mixture prior. A unified EM algorithm jointly refines mixture weights and diagonal variances directly from the raw TD pilot observations. The framework incorporates key innovations such as a hierarchical prior structure that readily adapts to both simple and complex channel conditions through learned mixture components. It also provides a practical implementation using rectangular pulse shaping with flexible pilot placement, which eliminates DD guard interval requirements, substantially reducing overhead while preserving spectral efficiency. The estimated CSI is subsequently utilised in an LMMSE detector for reliable data detection. Moreover, our theoretical analysis has established sparsity guarantees for the proposed mixture priors, while the Oracle-MMSE and BCRLB benchmarks derived provide rigorous performance references. Comprehensive simulations across diverse channel conditions, including single-Gaussian and multi-component mixtures with varying parameters, demonstrate consistent and substantial improvements in both estimation accuracy and detection reliability over the state-of-the-art sparse methods.

Appendix A

Proof of Theorem 1

For each component k , the Gaussian prior $p(\mathbf{h} | z=k) = \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma}_k)$ has zero mean along with diagonal covariance. For each such component, there exists a constant $C > 0$ such that:

$$\mathcal{CN}(\mathbf{0}, \mathbf{\Gamma}_k) \leq C \cdot \prod_{r=1}^{M_\tau N_\nu} \frac{1}{|h_r|^2}. \quad (36)$$

The mixture prior is a convex combination of these components:

$$\begin{aligned} p(\mathbf{h}) &= \sum_{k=1}^K \rho_k p(\mathbf{h} | z=k) \leq \left(\sum_{k=1}^K \rho_k C_k \right) \cdot \prod_{r=1}^{M_\tau N_\nu} \frac{1}{|h_r|^2} \\ &= C \cdot \prod_{r=1}^{M_\tau N_\nu} \frac{1}{|h_r|^2}, \quad \text{where } C = \sum_{k=1}^K \rho_k C_k. \end{aligned} \quad (37)$$

Appendix B

Proof of Theorem 2

Let $\mathbf{R}^t = \{\mathbf{r}_{p,i}\}_{i=1}^L$ denote the observed snapshots. For $k \in \{1, \dots, K\}$ assume component priors

$$p(\mathbf{h} | k) = \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma}_k), \quad \mathbf{\Gamma}_k = \text{diag}(\gamma_{k,1}, \dots, \gamma_{k,M_\tau N_\nu}),$$

and mixture weights ρ_k , with $\sum_{k=1}^K \rho_k = 1$. Define responsibilities at EM iteration t by

$$\pi_{i,k}^{(t)} := p^{(t)}(k | \mathbf{r}_{p,i}) = \frac{\rho_k^{(t)} p(\mathbf{r}_{p,i} | k; \mathbf{\Gamma}_k^{(t)})}{\sum_{\ell=1}^K \rho_\ell^{(t)} p(\mathbf{r}_{p,i} | \ell; \mathbf{\Gamma}_\ell^{(t)})}, \quad (38)$$

where the Gaussian marginal is

$$p(\mathbf{r}_{p,i} | k) = \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I} + \mathbf{\Omega} \mathbf{\Gamma}_k \mathbf{\Omega}^H). \quad (39)$$

Let the conditional posterior under component k for snapshot i be $p(\mathbf{h} | \mathbf{r}_{p,i}, k) = \mathcal{CN}(\boldsymbol{\mu}_{i,k}, \boldsymbol{\Sigma}_k)$ where $\boldsymbol{\mu}_{i,k}, \boldsymbol{\Sigma}_k$ represents the a-posteriori mean and covariance, after substituting $\mathbf{\Gamma} \leftarrow \mathbf{\Gamma}_k$, the M-step yields

$$\gamma_{k,r}^{(t+1)} = \frac{\sum_{i=1}^L \pi_{i,k}^{(t)} (|\mu_{i,k,r}|^2 + [\boldsymbol{\Sigma}_k]_{rr})}{\sum_{i=1}^L \pi_{i,k}^{(t)}}, \quad (40a)$$

$$\rho_k^{(t+1)} = \frac{1}{L} \sum_{i=1}^L \pi_{i,k}^{(t)}. \quad (40b)$$

To prove this, let the complete data for the snapshot i be $p(\mathbf{r}_{p,i}, \mathbf{h}, k) = p(\mathbf{r}_{p,i} | \mathbf{h}) p(\mathbf{h} | k) \rho_k$. Only the prior $p(\mathbf{h} | k)$ and the mixing weight ρ_k depend on $\mathbf{\Gamma}_k$ and ρ , respectively. Thus, the EM functional at iteration t is

$$\mathcal{Q}^{(t)}(\{\mathbf{\Gamma}_k\}, \rho) = \sum_{i=1}^L \sum_{k=1}^K \pi_{i,k}^{(t)} \left(- \sum_{r=1}^{M_\tau N_\nu} \log \gamma_{k,r} - \sum_{r=1}^{M_\tau N_\nu} \frac{\mathbb{E}_{\mathbf{h}|\mathbf{r}_{p,i},k} [|h_r|^2]}{\gamma_{k,r}} + \log \rho_k \right) + C, \quad (41)$$

where C is independent of $\mathbf{\Gamma}_k$ and ρ . By using the identity

$$\mathbb{E}_{\mathbf{h}|\mathbf{r}_{p,i},k} [|h_r|^2] = |\mu_{i,k,r}|^2 + [\boldsymbol{\Sigma}_k]_{rr}, \quad (42)$$

we isolate the scalar function of $\gamma_{k,r}$

$$Q_{k,r}(\gamma_{k,r}) = - \sum_{i=1}^L \pi_{i,k}^{(t)} \left(\log \gamma_{k,r} + \frac{|\mu_{i,k,r}|^2 + [\boldsymbol{\Sigma}_k]_{rr}}{\gamma_{k,r}} \right).$$

We differentiate $Q_{k,r}$ with respect to $\gamma_{k,r}$ and set it to zero:

$$\frac{\partial Q_{k,r}}{\partial \gamma_{k,r}} = - \sum_{i=1}^L \pi_{i,k}^{(t)} \left(\frac{1}{\gamma_{k,r}} - \frac{|\mu_{i,k,r}|^2 + [\boldsymbol{\Sigma}_k]_{rr}}{\gamma_{k,r}^2} \right) = 0.$$

Upon multiplying both sides by $\gamma_{k,r}^2$ and rearranging, we obtain

$$\gamma_{k,r}^{(t+1)} \sum_{i=1}^L \pi_{i,k}^{(t)} = \sum_{i=1}^L \pi_{i,k}^{(t)} (|\mu_{i,k,r}|^2 + [\boldsymbol{\Sigma}_k]_{rr}),$$

which yields (40a).

To update ρ retain only the ρ -dependent part of $\mathcal{Q}^{(t)}$ in (41):

$$\mathcal{Q}_\rho^{(t)} = \sum_{i=1}^L \sum_{k=1}^K \pi_{i,k}^{(t)} \log \rho_k.$$

Then introduce the Lagrange multiplier β for $\sum_{k=1}^K \rho_k = 1$ and form $\Lambda(\rho, \beta) = \mathcal{Q}_\rho^{(t)} + \beta (\sum_{k=1}^K \rho_k - 1)$. Differentiating with respect to ρ_k yields

$$\frac{\partial \Lambda}{\partial \rho_k} = \frac{\sum_{i=1}^L \pi_{i,k}^{(t)}}{\rho_k} + \beta = 0 \implies \rho_k = -\frac{1}{\beta} \sum_{i=1}^L \pi_{i,k}^{(t)}.$$

Normalization $\sum_{k=1}^K \rho_k = 1$ gives

$$-\frac{1}{\beta} \sum_{k=1}^K \sum_{i=1}^L \pi_{i,k}^{(t)} = 1.$$

Since $\sum_{k=1}^K \pi_{i,k}^{(t)} = 1$ for every i , the double sum equals L , hence we have $-1/\beta = 1/L$ and therefore $\rho_k^{(t+1)} = \frac{1}{L} \sum_{i=1}^L \pi_{i,k}^{(t)}$, which is (40b).

Appendix C BCRLB under GMM Prior

Let the dictionary be $\mathbf{\Omega} \in \mathbb{C}^{N_p \times (M_\tau N_\nu)}$ and the DD domain channel be $\mathbf{h} \in \mathbb{C}^{M_\tau N_\nu \times 1}$. With L TD pilot snapshots $\{\mathbf{r}_{p,i}\}_{i=1}^L$ under additive noise $\boldsymbol{\eta}^{(i)} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$, the Bayesian Fisher information matrix (FIM) decomposes as in [33]

$$\mathbf{J} = \mathbf{J}_{\text{data}} + \mathbf{J}_{\text{prior}}, \quad (43)$$

where, $\mathbf{J}_{\text{data}} = \frac{L}{\sigma^2} \mathbf{\Omega}^H \mathbf{\Omega}$ represents the data Fisher information matrix, while $\mathbf{J}_{\text{prior}}$ denotes the contribution from the prior, which can be derived as follows.

Assume a K -component complex Gaussian mixture prior

$$p(\mathbf{h}) = \sum_{k=1}^K \rho_k \mathcal{CN}(\boldsymbol{\mu}_k, \mathbf{\Gamma}_k), \quad \rho_k \geq 0, \quad \sum_{k=1}^K \rho_k = 1, \quad (44)$$

where ρ_k are the constant mixture weights. Let us define the local weights as

$$w_k(\mathbf{h}) = \frac{\rho_k \mathcal{CN}(\boldsymbol{\mu}_k, \mathbf{\Gamma}_k)}{p(\mathbf{h})}, \quad \sum_{k=1}^K w_k(\mathbf{h}) = 1. \quad (45)$$

Then differentiation of the mixture log-density yields the Hessian identity:

$$\nabla_{\mathbf{h}}^2 [\log p(\mathbf{h})] = \sum_{k=1}^K w_k(\mathbf{h}) \left[-\mathbf{\Gamma}_k^{-1} + \mathbf{\Gamma}_k^{-1} (\mathbf{h} - \boldsymbol{\mu}_k) (\mathbf{h} - \boldsymbol{\mu}_k)^H \mathbf{\Gamma}_k^{-1} \right] - \mathbf{b}(\mathbf{h}) \mathbf{b}(\mathbf{h})^H. \quad (46)$$

where we have

$$\mathbf{b}(\mathbf{h}) = \sum_{k=1}^K w_k(\mathbf{h}) \left[-\mathbf{\Gamma}_k^{-1} (\mathbf{h} - \boldsymbol{\mu}_k) \right]. \quad (47)$$

Taking the negative expectation with respect to $p(\mathbf{h})$ yields the prior Fisher information matrix as follows:

$$\begin{aligned} \mathbf{J}_{\text{prior}} &= -\mathbb{E}_{\mathbf{h}} [\nabla_{\mathbf{h}}^2 \log p(\mathbf{h})] = \mathbb{E}_{\mathbf{h}} \left[\sum_{k=1}^K w_k(\mathbf{h}) \mathbf{\Gamma}_k^{-1} \right] \\ &\quad - \mathbb{E}_{\mathbf{h}} \left[\sum_{k=1}^K w_k(\mathbf{h}) \mathbf{\Gamma}_k^{-1} (\mathbf{h} - \boldsymbol{\mu}_k) (\mathbf{h} - \boldsymbol{\mu}_k)^H \mathbf{\Gamma}_k^{-1} \right] \\ &\quad + \mathbb{E}_{\mathbf{h}} [\mathbf{b}(\mathbf{h}) \mathbf{b}(\mathbf{h})^H]. \end{aligned} \quad (48)$$

For $K = 1$, (48) reduces to $\mathbf{J}_{\text{prior}} = \mathbf{\Gamma}_1^{-1}$. The total Bayesian information matrix is then obtained by substituting (48) into (43), giving

$$\mathbf{J} = \frac{L}{\sigma^2} \mathbf{\Omega}^H \mathbf{\Omega} + \mathbf{J}_{\text{prior}}. \quad (49)$$

Then the BCRLB of the average MSE is expressed as

$$\text{MSE}(\hat{\mathbf{h}}) \geq \text{tr}(\mathbf{J}^{-1}) = \text{tr}\left(\left[\frac{L}{\sigma^2}\mathbf{\Omega}^H\mathbf{\Omega} + \mathbf{J}_{\text{prior}}\right]^{-1}\right). \quad (50)$$

For $K > 1$, the expectations in (48) do not admit closed-form solutions and are therefore approximated using Monte-Carlo simulation. In practice, L i.i.d. draws $\mathbf{h}^{(i)} \sim p(\mathbf{h})$ are generated, and the local weights $w_k(\mathbf{h}^{(i)})$ are evaluated for each sample, and the resultant per-sample contributions are averaged to form an empirical prior Fisher information matrix, $\hat{\mathbf{J}}_{\text{prior}}$, which is then substituted into (50) to obtain the BCRLB.

References

- [1] 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.
- [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. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, pp. 450–476, 2016.
- [4] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.
- [5] 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.
- [6] A. Monk, R. Hadani, M. Tsatsanis, and S. Rakib, "OTFS-orthogonal time frequency space," *arXiv preprint arXiv:1608.02993*, 2016.
- [7] S. K. Mohammed, "Derivation of OTFS modulation from first principles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7619–7636, 2021.
- [8] 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.
- [9] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," *arXiv preprint arXiv:1802.02623*, 2018.
- [10] 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.
- [11] M. Ramachandran, G. Surabhi, and A. Chockalingam, "OTFS: A new modulation scheme for high-mobility use cases," *Journal of the Indian Institute of Science*, vol. 100, no. 2, pp. 315–336, 2020.
- [12] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Delay-Doppler and angular domain 4D-sparse CSI estimation in OTFS aided MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 12, pp. 13 447–13 452, 2022.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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, 2021.
- [17] I. A. Khan and S. K. Mohammed, "Low complexity channel estimation for ofts modulation with fractional delay and doppler," *arXiv preprint arXiv:2111.06009*, 2021.
- [18] Q. Li, Y. Gong, F. Meng, Z. Li, L. Miao, and Z. Xu, "Residual learning based channel estimation for OTFS system," in *2022 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. IEEE, 2022, pp. 275–280.
- [19] C. Guo and L. Zhu, "OTFS fractional Doppler channel estimation method based on super-resolution networks," in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2024, pp. 1–6.
- [20] X. Zhang, C. Liu, W. Yuan, J. A. Zhang, and D. W. K. Ng, "Sparse prior-guided deep learning for OTFS channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 12, pp. 19 913–19 918, 2024.
- [21] M. Payami and S. D. Blostein, "Sparse signal recovery neural network with application to high-mobility massive MIMO-OTFS communication systems," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 8, pp. 12 175–12 188, 2025.
- [22] S. R. Mattu and A. Chockalingam, "Learning in time-frequency domain for fractional delay-Doppler channel estimation in OTFS," *IEEE Wireless Communications Letters*, vol. 13, no. 5, pp. 1245–1249, 2024.
- [23] A. F. dos Reis, B. S. Chang, Y. Medjahdi, G. Brante, and F. Bader, "LSTM-based time-frequency domain channel estimation for OTFS modulation," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 10, pp. 15 049–15 060, 2024.
- [24] N. Turan, B. Böck, K. J. Chan, B. Fesl, F. Burmeister, M. Joham, G. Fettweis, and W. Utschick, "Wireless channel prediction via Gaussian mixture models," in *2024 27th International Workshop on Smart Antennas (WSA)*. IEEE, 2024, pp. 1–5.
- [25] T. Ballal, T. Y. Al-Naffouri, and S. F. Ahmed, "Low-complexity Bayesian estimation of cluster-sparse channels," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4159–4173, 2015.
- [26] B. Böck, F. Weißer, M. Baur, and W. Utschick, "Sparse Bayesian generative modeling for joint parameter and channel estimation," *arXiv preprint arXiv:2502.18369*, 2025.
- [27] L. Kong, X. Zhang, H. Zhao, and J. Wei, "Variational sparse Bayesian learning for estimation of Gaussian mixture distributed wireless channels," *Entropy*, vol. 23, no. 10, p. 1268, 2021.
- [28] B. Selim, O. Alhussein, S. Muhaidat, G. K. Karagiannidis, and J. Liang, "Modeling and analysis of wireless channels via the mixture of Gaussian distribution," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8309–8321, 2015.
- [29] B. Böck, S. Syed, and W. Utschick, "Sparse Bayesian generative modeling for compressive sensing," *Advances in Neural Information Processing Systems*, vol. 37, pp. 4629–4659, 2024.
- [30] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [31] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning*. Cambridge University Press, 2020.
- [32] H. L. V. Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. Wiley-IEEE press, 2007.
- [33] E. S. Crafts, X. Zhang, and B. Zhao, "Bayesian Cramér-Rao bound estimation with score-based models," *IEEE Transactions on Information Theory*, 2024.
- [34] Y. Hong, T. Thaj, and E. Viterbo, *Delay-Doppler communications: principles and applications*. Academic Press, 2022.
- [35] S. K. Mohammed, R. Hadani, and A. Chockalingam, *OTFS Modulation: Theory and Applications*. John Wiley & Sons, 2024.
- [36] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.
- [37] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 2002.
- [38] J. Ranstam and J. A. Cook, "Lasso regression," *Journal of British Surgery*, vol. 105, no. 10, pp. 1348–1348, 2018.
- [39] B. He, Z. Bai, Y. Ma, H. Xu, A. Mohamed, Y. Yang, and K. Kwak, "Denosing CNN based channel estimation for vehicular OTFS communication system," in *2023 25th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2023, pp. 54–58.