Bayesian Learning Aided Sparse Channel Estimation for Orthogonal Time Frequency Space Modulated Systems

Suraj Srivastava, *Graduate Student Member, IEEE,* Rahul Kumar Singh, Aditya K. Jagannatham, *Member, IEEE,* and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—A novel sparse channel state information (CSI) estimation scheme is proposed for orthogonal time frequency space (OTFS) modulated systems, in which the pilots are directly transmitted over the time-frequency (TF)-domain grid for estimating the delay-Doppler (DD)-domain CSI. The proposed CSI estimation model leads to a reduction in the pilot overhead as well as the training duration required. Furthermore, it does not require a DD-domain guard interval between the pilot and data symbols, hence increasing the bandwidth efficiency. A novel Bayesian learning (BL) framework is proposed for CSI acquisition, which exploits the DD-domain sparsity for improving the estimation accuracy in comparison to the conventional minimum mean squared error (MMSE)-based scheme. A low-complexity linear MMSE detector is used in the subsequent data detection phase. Our simulation results demonstrate the performance improvement of the proposed BL-based scheme over the conventional MMSE-based scheme as well as over other existing sparse estimation schemes.**

***Index Terms*—OTFS, delay-Doppler domain channel, sparsity, channel estimation, BCRLB, high-mobility**

I. Introduction

Given the continuously evolving diverse range of applications, it is of paramount importance to explore novel modulation techniques that are resilient to both the delay-spread and to the Doppler-shifts introduced by the wireless propagation medium. To this end, a novel delay-Doppler (DD)-domain modulation technique, originally proposed by R. Hadani *et al.* in [1]–[3], termed as the orthogonal time frequency space (OTFS) arrangement, deserves further exploration in high-mobility scenarios. An important aspect of OTFS modulation is that it relies on the DD-domain representation of the wireless channel [1], [4], [5], which leads to a substantial reduction in the number of DD-domain channel taps. However, its significantly enhanced performance and the improved end-to-end bit-error-rate (BER) depend critically on the availability of accurate DD-domain CSI at the receiver for reliable symbol detection, which has therefore been the focus of [6]–[8].

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). A. K. Jagannatham would like to acknowledge the research supported in part by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, in part by the Space Technology Cell, IIT Kanpur, in part by the IIMA IDEA Telecom Centre of Excellence, in part by the Qualcomm Innovation Fellowship, and in part by the Arun Kumar Chair Professorship.

S. Srivastava, R. K. Singh and A. K. Jagannatham are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, UP 208016, India (e-mail: ssrivast@iitk.ac.in, rahulkz@iitk.ac.in, adityaj@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).

A. Literature Review

The early treatises [3], [9] proposed an impulse-based CSI estimation scheme for single-input single-output (SISO) OTFS systems. A key drawback of the proposed CSI estimation schemes therein is that they require an entire OTFS frame for pilot transmission, which significantly reduces the bandwidth efficiency. To circumvent this, Raviteja *et al.* [7] proposed embedded pilot (EP)-based CSI estimation schemes, where the DD-domain pilot, data and guard symbols are specifically arranged for avoiding interference, followed by a threshold-based technique for CSI acquisition. Alternatively, a few recent innovative propositions, such as those by Shen *et al.* [6] and Rasheed *et al.* [8], exploited the DD-domain sparsity of the wireless channel via a compelling formulation of the CSI estimation paradigm as a sparse signal recovery problem. These schemes have demonstrated superior CSI estimation accuracy in comparison to the previously reviewed training impulse and embedded pilot techniques as they additionally leverage the sparsity of the underlying wireless channel. Prominently, almost all of the existing treatises employ a DD-domain guard for pilot placement, which leads to a reduction of the resultant bandwidth efficiency. Most of the DD-domain CSI estimation techniques consider either an *ideal-biorthogonal* [2], [10] signalling pulse or an approximate model for the practical rectangular pulse that relies on the relationship developed in [4]. The above-mentioned impediments motivate us to develop a novel DD-domain sparse CSI estimation scheme for OTFS systems that can overcome the shortcomings of the approaches presented in the open literature. The various contributions of this paper are summarized next.

B. Contributions

- A novel sparse channel estimation model is derived for a SISO OTFS system having arbitrary transceiver pulse shaping filters. Furthermore, as per this model, the pilots are directly transmitted in the TF-domain, hence reducing

both the pilot overhead as well as the training duration and the pre-processing complexity.
- The proposed channel estimation model has the additional advantage that it does not require a DD-domain guard interval between the pilot and data symbols, thereby increasing the bandwidth efficiency.
- Subsequently, a Bayesian learning (BL) framework and its multiple measurement vector (MMV) extension M-BL are proposed, which exploit the DD-domain sparsity for improved CSI estimation.
- The performance of the proposed BL-based schemes is also benchmarked by deriving both the Oracle-MMSE and the Bayesian Cramer-Rao lower bound (BCRLB).

II. OTFS System Model

Consider a SISO OTFS system having a frame duration of $T_f = NT$ and bandwidth of $B = M\Delta f$, where T denotes the symbol duration and Δf represents the subcarrier spacing, so that $T\Delta f = 1$. The quantities N and M represent the number of symbols along the time and frequency axes in the TF-grid. OTFS places the information symbols in an equivalent DD-grid, where the delay and Doppler axes are sampled at integer multiples of $\Delta\tau = \frac{1}{B}$ and $\Delta\nu = \frac{1}{T_f}$, respectively.

A. OTFS Modulation

Let $\mathbf{X}_{\text{DD}} \in \mathbb{C}^{M\times N}$ denote a matrix of information symbols. The transmitter first maps these DD-domain symbols to the TF-domain by using the inverse symplectic finite Fourier transform (ISFFT) as $\mathbf{X}_{\text{TF}} = \mathbf{F}_M \mathbf{X}_{\text{DD}} \mathbf{F}_N^H$, where $\mathbf{F}_N$ represents the discrete Fourier transform (DFT) matrix. Subsequently, the time-domain signal is obtained by performing the Heisenberg transform of the TF-domain symbol matrix $\mathbf{X}_{\text{TF}}$. Let $p_{\text{tx}}(t)$ denote the transmit pulse of duration T, which is repeated N-times in the OTFS frame duration T_f. The transmit signal matrix $\mathbf{S} \in \mathbb{C}^{M\times N}$ is then formulated as

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

where $\mathbf{P}_{\text{tx}} = \text{diag}\left\{p_{\text{tx}}\left(\frac{pT}{M}\right)\right\}_{p=0}^{M-1} \in \mathbb{C}^{M\times M}$. Furthermore, the MN-samples of the transmit signal $\mathbf{s} \in \mathbb{C}^{MN\times 1}$ expressed in the vectorized form can be formulated as

$$\mathbf{s} = \text{vec}\left(\mathbf{P}_{\text{tx}} \mathbf{X}_{\text{DD}} \mathbf{F}_N^H\right) = \left(\mathbf{F}_N^H \otimes \mathbf{P}_{\text{tx}}\right) \mathbf{x}_{\text{DD}}, \tag{2}$$

where we have $\mathbf{x}_{\text{DD}} = \text{vec}\left(\mathbf{X}_{\text{DD}}\right) \in \mathbb{C}^{MN\times 1}$. Finally, prior to transmission, similar to OFDM, a cyclic prefix (CP) of length L is appended to remove the inter-frame interference.

B. DD-domain Wireless Channel Model

Let $h(\tau, \nu)$ denote the DD-domain representation of the wireless channel, which is a 2D function of the delay variable τ and Doppler variable ν. Since typically, only a few dominant reflectors are exhibited by a wireless channel, the DD-domain channel can be formulated as [1], [4], [10]

$$h(\tau, \nu) = \sum_{i=1}^{L_p} h_i \delta(\tau - \tau_i)\delta(\nu - \nu_i), \tag{3}$$

where τ_i and ν_i represent the delay- and Doppler-shifts, while h_i denotes the complex path gain introduced by the ith reflector. Furthermore, L_p denotes the number of dominant reflectors/ multipath components and $\delta(\cdot)$ represents the Dirac-delta function. Finally, for a suitable choice of M and N, the delay-shift τ_i and Doppler-shift ν_i associated with the ith multipath component can be approximated as $\tau_i = \frac{l_i}{M\Delta f}$, $\nu_i = \frac{k_i}{NT}$, where the integers $l_i < M$ and $k_i < N$ are the indices corresponding to τ_i and ν_i.

Let $\mathbf{r} \in \mathbb{C}^{MN\times 1}$ and $\mathbf{w} \in \mathbb{C}^{MN\times 1}$ denote the MN-samples of the received signal and the noise process. Then the received signal vector $\mathbf{r}$ can be expressed as [11]

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w}, \tag{4}$$

where the matrix $\mathbf{H} \in \mathbb{C}^{MN\times MN}$ is formulated as $\mathbf{H} = \sum_{i=1}^{L_p} h_i \left(\mathbf{\Pi}\right)^{l_i} \left(\mathbf{\Delta}\right)^{k_i}$. Here, $\mathbf{\Pi} \in \mathbb{C}^{MN\times MN}$ denotes a permutation matrix [11] and $\mathbf{\Delta} = \text{diag}\left\{\omega^i\right\}_{i=0}^{MN-1} \in \mathbb{C}^{MN\times MN}$, where $\omega = e^{j2\pi\frac{1}{MN}}$.

C. OTFS Demodulation

Let $\mathbf{Y}_{\text{TF}} \in \mathbb{C}^{M\times N}$ denote the TF-demodulated symbol matrix. This is given by the discrete Wigner transform of the received sample matrix $\mathbf{R} = \text{vec}^{-1}(\mathbf{r}) \in \mathbb{C}^{M\times N}$ as $\mathbf{Y}_{\text{TF}} = \mathbf{F}_M \mathbf{P}_{\text{rx}} \mathbf{R}$, where $\mathbf{P}_{\text{rx}} = \text{diag}\left\{p_{\text{rx}}^*\left(\frac{pT}{M}\right)\right\}_{p=0}^{M-1}$. Here, $p_{\text{rx}}(t)$ of duration T represents the receiver pulse. The OTFS-demodulated signal $\mathbf{Y}_{\text{DD}} \in \mathbb{C}^{M\times N}$ in the DD-domain is obtained by applying the SFFT to the TF-domain demodulated signal $\mathbf{Y}_{\text{TF}}$ as $\mathbf{Y}_{\text{DD}} = \mathbf{F}_M^H \mathbf{Y}_{\text{TF}} \mathbf{F}_N = \mathbf{P}_{\text{rx}} \mathbf{R} \mathbf{F}_N$, and its equivalent vector $\mathbf{y}_{\text{DD}} = \text{vec}\left(\mathbf{Y}_{\text{DD}}\right)$ is expressed as

$$\mathbf{y}_{\text{DD}} = \left(\mathbf{F}_N \otimes \mathbf{P}_{\text{rx}}\right) \mathbf{r} \in \mathbb{C}^{MN\times 1}. \tag{5}$$

Upon substituting $\mathbf{r}$ from (4), and $\mathbf{s}$ from (2) into (5), the DD-domain system model of the OTFS system is obtained as

$$\mathbf{y}_{\text{DD}} = \mathbf{H}_{\text{DD}} \mathbf{x}_{\text{DD}} + \mathbf{v}_{\text{DD}}, \tag{6}$$

where we have $\mathbf{H}_{\text{DD}} = \left(\mathbf{F}_N \otimes \mathbf{P}_{rx}\right) \mathbf{H} \left(\mathbf{F}_N^H \otimes \mathbf{P}_{\text{tx}}\right) \in \mathbb{C}^{MN\times MN}$ and $\mathbf{v}_{\text{DD}} = \left(\mathbf{F}_N \otimes \mathbf{P}_{\text{rx}}\right) \mathbf{w} \in \mathbb{C}^{MN\times 1}$. For symbols having an average power of unity, the MMSE-based linear detector is formulated by

$$\widehat{\mathbf{x}}_{\text{DD}}^{\text{MMSE}} = \left(\mathbf{H}_{\text{DD}}^H \mathbf{R}_{v,\text{DD}}^{-1} \mathbf{H}_{\text{DD}} + \mathbf{I}_{MN}\right)^{-1} \mathbf{H}_{\text{DD}}^H \mathbf{R}_{v,\text{DD}}^{-1} \mathbf{y}_{\text{DD}}, \tag{7}$$

where $\mathbf{R}_{v,\text{DD}} \in \mathbb{C}^{MN\times MN}$ denotes the covariance matrix of the noise $\mathbf{v}_{\text{DD}}$, which obeys $\mathbf{R}_{v,\text{DD}} = \sigma^2\left[\mathbf{I}_N \otimes \left(\mathbf{P}_{\text{rx}} \mathbf{P}_{\text{rx}}^H\right)\right]$. As mentioned earlier, accurate CSI estimation plays a key role in exploiting the full benefits of the OTFS system, which forms the focus of the subsequent section.

III. Sparse CSI Estimation Model for OTFS Systems

Let $\mathbf{X}_{\text{TF},\mathcal{P}} \in \mathbb{C}^{M\times N_p}$ denote the TF-domain pilot symbol matrix, where N_p denotes the number of pilot symbols placed along the time-axis. Thus, according to (1), the matrix $\mathbf{S}_{\mathcal{P}} \in \mathbb{C}^{M\times N_p}$ comprising the time-domain pilot samples may be expressed as

$$\mathbf{S}_{\mathcal{P}} = \mathbf{P}_{\text{tx}} \mathbf{F}_M^H \mathbf{X}_{\text{TF},\mathcal{P}}. \tag{8}$$

Let h_i^j denote the complex path gain associated with the ith delay-tap, $0 \leq i \leq M_\tau - 1$, and jth Doppler-tap, $0 \leq j \leq N_\nu - 1$. Here, M_τ and N_ν represent grid-sizes along the delay- and Doppler-axes, respectively, considered for sparse representation of the DD-domain CSI, which obey $\max(l_i) < M_\tau << M$ and $\max(k_i) < N_\nu << N$. Thus, the sparse representation of the DD-domain channel $h(\tau, \nu)$ of (3) can be formulated as

$$h(\tau,\nu) = \sum_{i=0}^{M_\tau-1}\sum_{j=0}^{N_\nu-1} h_i^j \delta(\tau-\tau_i)\delta(\nu-\nu_j), \tag{9}$$

where $\tau_i = \frac{i}{M\Delta f}$ and $\nu_j = \frac{j}{NT}$. Note that since there are only a few dominant reflectors in the wireless channel, only a few coefficients $L_p(<< M_\tau N_\nu)$ are non-zero among the $M_\tau N_\nu$ coefficients of $h_i^j, 0 \leq i \leq M_\tau - 1, 0 \leq j \leq N_\nu - 1$. After removing the CP, the received time-domain pilot signal $\mathbf{r}_\mathcal{P} \in \mathbb{C}^{MN_p \times 1}$ is given by

$$\mathbf{r}_\mathcal{P} = \bar{\mathbf{H}}\mathbf{s}_\mathcal{P} + \mathbf{w}_\mathcal{P}, \tag{10}$$

where $\mathbf{s}_\mathcal{P} = \mathrm{vec}\,(\mathbf{S}_\mathcal{P}) \in \mathbb{C}^{MN_p \times 1}$, and $\mathbf{w}_\mathcal{P} \in \mathbb{C}^{MN_p \times 1}$ represents the noise vector. The matrix $\bar{\mathbf{H}} \in \mathbb{C}^{MN_p \times MN_p}$ can be formulated as

$$\bar{\mathbf{H}} = \sum_{i=0}^{M_\tau-1}\sum_{j=0}^{N_\nu-1} h_i^j \left(\bar{\mathbf{\Pi}}\right)^i \left(\bar{\mathbf{\Delta}}_i\right)^j, \tag{11}$$

where $\bar{\mathbf{\Pi}}$ denotes a permutation matrix of size $MN_p \times MN_p$, whereas $\bar{\mathbf{\Delta}}_i \in \mathbb{C}^{MN_p \times MN_p}$ is defined as

$$\bar{\mathbf{\Delta}}_i = \begin{cases} \mathrm{diag}\left\{1, \omega, \cdots, \omega^{MN_p-i-1}, \omega^{-i}, \cdots, \omega^{-1}\right\}, \text{if } i \neq 0, \\ \mathrm{diag}\left\{1, \omega, \cdots, \omega^{MN_p-1}\right\}, \text{for } i = 0. \end{cases}$$

Let the output pilot matrix $\mathbf{R}_\mathcal{P} \in \mathbb{C}^{M \times N_p}$ be constructed as $\mathbf{R}_\mathcal{P} = \mathrm{vec}^{-1}(\mathbf{r}_\mathcal{P})$. Then the TF-domain pilot symbol matrix $\mathbf{Y}_{\text{TF},\mathcal{P}} \in \mathbb{C}^{M \times N_p}$ is given as: $\mathbf{Y}_{\text{TF},\mathcal{P}} = \mathbf{F}_M \mathbf{P}_{\text{rx}} \mathbf{R}_\mathcal{P}$. The equivalent vectorized representation is given by

$$\mathbf{y}_{\text{TF},\mathcal{P}} = \mathrm{vec}(\mathbf{Y}_{\text{TF},\mathcal{P}}) = \left[\mathbf{I}_{N_p} \otimes (\mathbf{F}_M \mathbf{P}_{\text{rx}})\right] \mathbf{r}_\mathcal{P}. \tag{12}$$

Upon substituting $\mathbf{r}_\mathcal{P}$ from (10) into (12), and in turn substituting $\mathbf{s}_\mathcal{P}$ using (8), $\mathbf{y}_{\text{TF},\mathcal{P}}$ above can be reformulated as

$$\mathbf{y}_{\text{TF},\mathcal{P}} = \left[\mathbf{I}_{N_p} \otimes (\mathbf{F}_M \mathbf{P}_{\text{rx}})\right] \bar{\mathbf{H}} \left[\mathbf{I}_{N_p} \otimes \left(\mathbf{P}_{\text{tx}} \mathbf{F}_M^H\right)\right] \mathbf{x}_{\text{TF},\mathcal{P}} + \mathbf{v},$$

where $\mathbf{x}_{\text{TF},\mathcal{P}} = \mathrm{vec}(\mathbf{X}_{\text{TF},\mathcal{P}}) \in \mathbb{C}^{MN_p \times 1}$ and $\mathbf{v} = \left[\mathbf{I}_{N_p} \otimes (\mathbf{F}_M \mathbf{P}_{\text{rx}})\right] \mathbf{w}_\mathcal{P} \in \mathbb{C}^{MN_p \times 1}$. Now, substituting $\bar{\mathbf{H}}$ from (11) into the above, we obtain

$$\mathbf{y}_{\text{TF},\mathcal{P}} = \sum_{i=0}^{M_\tau-1}\sum_{j=0}^{N_\nu-1} \boldsymbol{\omega}_i^j h_i^j + \mathbf{v}, \tag{13}$$

where $\boldsymbol{\omega}_i^j = \left[\mathbf{I}_{N_p} \otimes (\mathbf{F}_M \mathbf{P}_{\text{rx}})\right] \bar{\mathbf{\Pi}}^i \bar{\mathbf{\Delta}}_i^j \left[\mathbf{I}_{N_p} \otimes \left(\mathbf{P}_{\text{tx}} \mathbf{F}_M^H\right)\right] \mathbf{x}_{\text{TF},\mathcal{P}}$. Eq. (13) can now be formulated as the standard sparse signal recovery problem:

$$\mathbf{y}_{\text{TF},\mathcal{P}} = \mathbf{\Omega}\mathbf{h} + \mathbf{v}, \tag{14}$$

where the dictionary matrix $\mathbf{\Omega} \in \mathbb{C}^{MN_p \times M_\tau N_\nu}$ and the vector $\mathbf{h} \in \mathbb{C}^{M_\tau N_\nu \times 1}$ obey $\mathbf{\Omega} = \left[\boldsymbol{\omega}_0^0, \boldsymbol{\omega}_0^1, \cdots, \boldsymbol{\omega}_0^{N_\nu-1}, \cdots, \boldsymbol{\omega}_{M_\tau-1}^0, \boldsymbol{\omega}_{M_\tau-1}^1, \cdots, \boldsymbol{\omega}_{M_\tau-1}^{N_\nu-1}\right]$, $\mathbf{h} = \left[h_0^0, h_0^1, \cdots, h_0^{N_\nu-1}, \cdots, h_{M_\tau-1}^0, h_{M_\tau-1}^1, \cdots, h_{M_\tau-1}^{N_\nu-1}\right]^T$. For the linear model of (14), the conventional MMSE estimate, denoted by $\widehat{\mathbf{h}}_{\text{MMSE}}$, is given by

$$\widehat{\mathbf{h}}_{\text{MMSE}} = \left(\mathbf{\Omega}^H \mathbf{R}_v^{-1} \mathbf{\Omega} + \mathbf{R}_h^{-1}\right)^{-1} \mathbf{\Omega}^H \mathbf{R}_v^{-1} \mathbf{y}_{\text{TF},\mathcal{P}}, \tag{15}$$

where $\mathbf{R}_h \in \mathbb{C}^{M_\tau N_\nu \times M_\tau N_\nu}$ and $\mathbf{R}_v \in \mathbb{C}^{MN_p \times MN_p}$ denote the covariance matrices of the channel vector $\mathbf{h}$ and noise $\mathbf{v}$, respectively. The noise covariance $\mathbf{R}_v$ obeys $\mathbf{R}_v = \sigma^2\left[\mathbf{I}_{N_p} \otimes \left(\mathbf{F}_M \mathbf{P}_{\text{rx}} \mathbf{P}_{\text{rx}}^H \mathbf{F}_M^H\right)\right]$, whereas the true channel covariance $\mathbf{R}_h$ is typically unknown and set to $\mathbf{I}_{M_\tau N_\nu}$ in (15) for the sparsity-agnostic (SA)-MMSE. It is important to note that although the conventional SA-MMSE channel estimation scheme is popular, it fails to exploit the sparse nature of the CSI vector $\mathbf{h}$ having a limited number of the multipath components, which is a unique characteristic of the wireless channel in the DD-domain. To exploit the sparsity, this work develops a Bayesian method for sparse signal recovery. In this context, a sparse Bayesian learning (SBL) framework, which employs a parameterized Gaussian-prior together with a Gamma-hyperprior, has been derived in [12], [13] for finding robust solutions to regression and classification problems. SBL has then further been extended in [14] to the basis-selection problem using a non-informative hyperprior. Recently, priors constructed using mixtures of Bernoulli and continuous distributions have also been employed for sparse recovery. For example, the authors of [15] employed a Bernoulli-truncated exponential prior, whereas a Bernoulli-generalized Gaussian prior has been conceived in [16]. However, in all these works, the fundamental philosophy remains the same, i.e., combining these priors with the likelihood to derive the *a posteriori* distribution, followed by deriving the maximum a posteriori (MAP) or MMSE estimators. Thus, for simplicity, following the procedure described in [14], the next section develops a Bayesian learning (BL) based sparse CSI estimation framework for exploiting the sparsity of the DD-domain wireless channel in OTFS systems.

IV. BL-based Sparse CSI Estimation

The proposed BL framework is based on the Bayesian principle, which assigns the following parameterized Gaussian prior to the DD-domain channel $\mathbf{h}$ [14]:

$$f(\mathbf{h};\mathbf{\Gamma}) = \prod_{i=0}^{M_\tau N_\nu-1} \frac{1}{(\pi\gamma_i)} \exp\left(-\frac{|\mathbf{h}(i)|^2}{\gamma_i}\right). \tag{16}$$

Here, the quantity $\gamma_i, 0 \leq i \leq M_\tau N_\nu - 1$, denotes the hyperparameter and the matrix $\mathbf{\Gamma} = \mathrm{diag}\{\gamma_i\}_{i=0}^{M_\tau N_\nu-1} \in \mathbb{R}^{M_\tau N_\nu \times M_\tau N_\nu}$ represents the unknown hyperparameter matrix. The MMSE estimate $\boldsymbol{\mu} \in \mathbb{C}^{M_\tau N_\nu \times 1}$ of the DD-domain channel $\mathbf{h}$ can be expressed as [17]

$$\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{\Omega}^H \mathbf{R}_v^{-1} \mathbf{y}_{\text{TF},\mathcal{P}}, \tag{17}$$

where the error covariance matrix $\mathbf{\Sigma} \in \mathbb{C}^{M_\tau N_\nu \times M_\tau N_\nu}$ is given by $\mathbf{\Sigma} = \left(\mathbf{\Omega}^H \mathbf{R}_v^{-1} \mathbf{\Omega} + \mathbf{\Gamma}^{-1}\right)^{-1}$. Thus, it follows from (17) that the MMSE estimate $\boldsymbol{\mu}$ necessitates the estimation of the hyperparameter matrix $\mathbf{\Gamma}$. To this end, it is desirable to select the matrix $\widehat{\mathbf{\Gamma}}$ that maximizes the log-Bayesian

evidence $\log\left[f(\mathbf{y}_{\text{TF},\mathcal{P}};\mathbf{\Gamma})\right]$. Since the log-Bayesian evidence maximization with respect to the hyperparameter matrix $\mathbf{\Gamma}$ is a non-concave problem [14], and hence intractable, the expectation-maximization (EM) method is an ideal tool for maximizing the cost function iteratively. Therefore, the proposed BL framework employs the EM technique for DD-domain sparse CSI estimation in OTFS systems. Let $\widehat{\mathbf{\Gamma}}^{(j-1)} = \text{diag}\left\{\widehat{\gamma}_i^{(j-1)}\right\}_{i=0}^{M_\tau N_\nu - 1} \in \mathbb{R}^{M_\tau N_\nu \times M_\tau N_\nu}$ denote the estimate of $\mathbf{\Gamma}$ in the $(j-1)$st EM-iteration. The procedure to update the estimate $\widehat{\mathbf{\Gamma}}^{(j)}$ in the jth EM-iteration is described in Theorem-1 presented next.

Theorem 1. *Given $\widehat{\gamma}_i^{(j-1)}, 0 \leq i \leq M_\tau N_\nu - 1$, the hyperparameter update $\widehat{\gamma}_i^{(j)}$ in the jth EM iteration, which maximizes the conditional-expectation of the log-likelihood function of the complete information set $\left\{\mathbf{y}_{TF,\mathcal{P}}, \mathbf{h}\right\}$, denoted by*

$$\mathcal{L}\left(\mathbf{\Gamma}|\widehat{\mathbf{\Gamma}}^{(j-1)}\right) = \mathbb{E}_{\mathbf{h}|\mathbf{y}_{TF,\mathcal{P}};\widehat{\mathbf{\Gamma}}^{(j-1)}}\left\{\log\left[f(\mathbf{y}_{TF,\mathcal{P}}, \mathbf{h};\mathbf{\Gamma})\right]\right\}, \tag{18}$$

is given by

$$\widehat{\gamma}_i^{(j)} = \mathbf{\Sigma}^{(j)}(i,i) + |\boldsymbol{\mu}^{(j)}(i)|^2, \tag{19}$$

where we have $\boldsymbol{\mu}^{(j)} = \mathbf{\Sigma}^{(j)}\mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{y}_{TF,\mathcal{P}}$ and $\mathbf{\Sigma}^{(j)} = \left[\mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{\Omega} + \left(\widehat{\mathbf{\Gamma}}^{(j-1)}\right)^{-1}\right]^{-1}$.

Proof. The proof is given in our technical report [18]. □

The EM procedure described above is repeated for a maximum of $N_{\max}$ iterations or until $\| \widehat{\mathbf{\Gamma}}^{(j)} - \widehat{\mathbf{\Gamma}}^{(j-1)} \|_F^2 \leq \epsilon$, whichever is achieved earlier, where the stopping parameters ϵ and $N_{\max}$ are suitably chosen. The BL-based sparse estimate $\widehat{\mathbf{h}}_{\text{BL}}$ of the DD-domain channel is obtained as the converged *a-posteriori* mean, i.e., $\widehat{\mathbf{h}}_{\text{BL}} = \boldsymbol{\mu}^{(j)}$. Subsequently, $\widehat{\mathbf{h}}_{\text{BL}}$ is used for constructing the estimate $\widehat{\mathbf{H}}_{\text{DD}}$ for detection as follows:

$$\widehat{\mathbf{H}}_{\text{DD}} = \left(\mathbf{F}_N \otimes \mathbf{P}_{rx}\right)\left[\sum_{i,j}\widehat{h}_{\text{BL},i}^j \mathbf{\Pi}^i\mathbf{\Delta}^j\right]\left(\mathbf{F}_N^H \otimes \mathbf{P}_{\text{tx}}\right), \tag{20}$$

where $\widehat{h}_{\text{BL},i}^j$ represents the estimate of the component h_i^j of the vector $\mathbf{h}$ defined after Eq. (14).

A. Multiple Measurement Vector (MMV)-Extension

Consider now a scenario with multiple pilot output vectors, denoted by $\mathbf{y}_{\text{TF},\mathcal{P},k}, 1 \leq k \leq K$, which are obtained using the identical transmit pilot symbol matrix $\mathbf{X}_{\text{TF},\mathcal{P}}$. Thus, the kth pilot output $\mathbf{y}_{\text{TF},\mathcal{P},k}$ can be expressed in terms of the dictionary matrix $\mathbf{\Omega}$ and the corresponding DD-domain CSI $\mathbf{h}_k$ as $\mathbf{y}_{\text{TF},\mathcal{P},k} = \mathbf{\Omega}\mathbf{h}_k + \mathbf{v}_k$, where $\mathbf{v}_k$ denotes the noise. Defining now the concatenated matrices $\widetilde{\mathbf{Y}}_{\text{TF},\mathcal{P}} = [\mathbf{y}_{\text{TF},\mathcal{P},1} \cdots, \mathbf{y}_{\text{TF},\mathcal{P},K}] \in \mathbb{C}^{MN_p \times K}, \widetilde{\mathbf{H}} = [\mathbf{h}_1, \cdots, \mathbf{h}_K] \in \mathbb{C}^{M_\tau N_\nu \times K}$ and $\widetilde{\mathbf{V}} = [\mathbf{v}_1, \cdots, \mathbf{v}_K]$, the MMV model for the DD-domain CSI estimation can be expressed as

$$\widetilde{\mathbf{Y}}_{\text{TF},\mathcal{P}} = \mathbf{\Omega}\widetilde{\mathbf{H}} + \widetilde{\mathbf{V}}. \tag{21}$$

Considering these K snapshots of the DD-domain CSI $\mathbf{h}_k$ well within the *geometric-coherence time* [10], for which their support-set, i.e., the L_p-indices of the non-zero locations of the sparse DD-domain CSI $\mathbf{h}_k$, do not change, the resultant CSI matrix $\widetilde{\mathbf{H}}$ exhibits *simultaneous-sparsity*, since its columns $\mathbf{h}_k$ share an identical sparsity profile. As described in our technical report [18], one can readily derive an efficient MMV-BL (M-BL) framework, which exploits the multiple pilot outputs and the simultaneous-sparse nature of the CSI matrix $\widetilde{\mathbf{H}}$, yielding a superior estimation performance. The update equations of the proposed M-BL framework for the jth EM iteration is summarized below [18], [19]:

$$\mathbf{\Sigma}^{(j)} = \left[\mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{\Omega} + \left(\widehat{\mathbf{\Gamma}}^{(j-1)}\right)^{-1}\right]^{-1}, \tag{22}$$

$$\widehat{\mathbf{H}}^{(j)} = \mathbf{\Sigma}^{(j)}\mathbf{\Omega}^H\mathbf{R}_v^{-1}\widetilde{\mathbf{Y}}_{\text{TF},\mathcal{P}}, \tag{23}$$

$$\widehat{\gamma}_i^{(j)} = \mathbf{\Sigma}^{(j)}(i,i) + \frac{1}{K}\sum_{k=1}^{K}\left|\widehat{\mathbf{H}}^{(j)}(i,k)\right|^2. \tag{24}$$

The key results of our complexity analysis are summarized below. The computational complexity order of the BL, M-BL and sparsity-agnostic MMSE techniques is seen as $\mathcal{O}\left(M_\tau^3 N_\nu^3\right)$, which arises due to matrix inversion of size-$[M_\tau N_\nu \times M_\tau N_\nu]$. On the other hand, the worst-case complexity order of the orthogonal matching pursuit (OMP) scheme is seen to be $\mathcal{O}\left(M^3 N_p^3\right)$, which arises due to the intermediate LS estimate required in each iteration. Finally, the computational cost of the conventional EP-based [7] CSI estimator is seen to be of the order $\mathcal{O}\left(M_\tau N_\nu\right)$, since it does not involve any matrix multiplication/ inversion operations. Due to lack of space, the detailed derivations for the computational complexities of various schemes have been moved to our technical report [18].

V. Performance Benchmarks

In this subsection, we first derive a hypothetical Oracle-MMSE estimator, which assumes a perfect knowledge of the DD-domain channel profile. Let $\mathcal{H}$ denote the support-set and $\mathbf{\Omega}_{\text{O}} = \mathbf{\Omega}(:,\mathcal{H})$ represent the Oracle-sensing matrix comprising of the columns indexed by the support-set $\mathcal{H}$. The Oracle-MMSE estimate can be derived as

$$\widehat{\mathbf{h}}_{\text{O-MMSE}} = \left(\mathbf{\Omega}_{\text{O}}^H\mathbf{R}_v^{-1}\mathbf{\Omega}_{\text{O}} + \mathbf{I}_{L_p}^{-1}\right)^{-1}\mathbf{\Omega}_{\text{O}}^H\mathbf{R}_v^{-1}\mathbf{y}_{\text{TF},\mathcal{P}}. \tag{25}$$

Next, we derive the BCRLB for the MSE of the estimated CSI $\widehat{\mathbf{h}}$. To this end, the Bayesian Fisher information matrix (FIM), denoted by $\mathbf{J}_{\text{B}} \in \mathbb{C}^{M_\tau N_\nu \times M_\tau N_\nu}$, for the DD-domain sparse channel vector $\mathbf{h}$ can be expressed as [20]: $\mathbf{J}_{\text{B}} = \mathbf{J}_{\text{D}} + \mathbf{J}_{\text{P}}$, where the FIMs $\mathbf{J}_{\text{D}}$ and $\mathbf{J}_{\text{P}}$, both of size $M_\tau N_\nu \times M_\tau N_\nu$, are determined as $\mathbf{J}_{\text{D}} = -\mathbb{E}_{\mathbf{y}_{\text{TF},\mathcal{P}},\mathbf{h}}\left\{\frac{\partial^2 \log[f(\mathbf{y}_{\text{TF},\mathcal{P}}|\mathbf{h})]}{\partial\mathbf{h}\partial\mathbf{h}^H}\right\}, \mathbf{J}_{\text{P}} = -\mathbb{E}_{\mathbf{h}}\left\{\frac{\partial^2 \log[f(\mathbf{h};\mathbf{\Gamma})]}{\partial\mathbf{h}\partial\mathbf{h}^H}\right\}$. Employing the above relationships, we obtain $\mathbf{J}_{\text{D}} = \mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{\Omega}$ and $\mathbf{J}_{\text{P}} = \mathbf{\Gamma}^{-1}$. Thus, the Bayesian FIM $\mathbf{J}_{\text{B}}$ evaluates to $\mathbf{J}_{\text{B}} = \mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{\Omega} + \mathbf{\Gamma}^{-1}$. Finally, the BCRLB for the MSE of the CSI estimate $\mathbf{h}$ is expressed as [20]

$$\text{MSE}\left(\widehat{\mathbf{h}}\right) \geq \text{Tr}\left(\mathbf{J}_{\text{B}}^{-1}\right) = \text{Tr}\left(\left[\mathbf{\Omega}^H\mathbf{R}_v^{-1}\mathbf{\Omega} + \mathbf{\Gamma}^{-1}\right]^{-1}\right). \tag{26}$$

Note that the BCRLB derived above also assumes the perfect knowledge of the DD-domain channel profile, since it employs the true hyperparameter matrix $\mathbf{\Gamma}$ to lower bound the MSE.

TABLE I
SIMULATION PARAMETERS OF SYSTEM-I AND SYSTEM-II

Parameter	**System-I**	**System-II**
Carrier frequency (GHz)	4	28
Subcarrier spacing (KHz) (Δf)	15	78.125
# of symbols along delay-axis (M)	32	128
# of symbols along Doppler-axis (N)	32	128
# of pilots along time-axis (N_p)	8	16
# of samples in CP (L)	6	32
# of dominant reflectors (L_p)	5	5
Max. spread along delay-axis (M_τ)	16	32
Max. spread along Doppler-axis (N_ν)	8	16
Modulation scheme	8-PSK	4-PSK
Pulse-shape	Rect.	Rect.

TABLE II
DD-PROFILE OF THE WIRELESS CHANNEL OF SYSTEM-I

Path-index(i)	**1**	**2**	**3**	**4**	**5**
Delay (τ_i) (μs)	2.08	4.164	6.246	8.328	10.41
Doppler (ν_i) (Hz)	0	470	940	1410	1880
Speed (Km/h)	0	126.9	253.8	380.7	507.6

VI. SIMULATION RESULTS

This section quantifies the performance of the proposed BL-based schemes and compares them to that of the state-of-the-art OMP- and FOCUSS-based [21] techniques for sparse estimation of the DD-domain CSI in terms of the normalized MSE (NMSE) defined as $\text{NMSE} = \frac{\left\|\widehat{\mathbf{h}}-\mathbf{h}\right\|_F^2}{\|\mathbf{h}\|_F^2}$. Furthermore, the performance is also benchmarked with respect to the BCRLB and the MMSE-Oracle estimator derived in Section-V, and also to the conventional state-of-the-art EP-based estimator of [7]. Note that for a fair comparison, the power of the pilot-impulse considered in [7] is set to MN_p, which is identical to the total pilot power employed in our proposed framework. This is followed by characterizing the BER performance obtained with the aid of the CSI estimates derived. For the BL-based approaches, the stopping parameters ϵ and $N_{\max}$ are set to 10^{-6} and 50, respectively. The stopping criterion for the OMP is set for ensuring that the algorithm terminates when the residual error $R(i)$ between the consecutive iterations obeys $|R(i) - R(i-1)| < 0.1\sigma^2$. The regularization parameter for the FOCUSS is set to σ^2, the l_p-norm parameter to $p = 0.8$ and the stopping threshold is set to 10^{-5} with the maximum number of iterations set to 800. The SNR in decibels (dB) is defined as $10\log_{10}\left(\frac{1}{\sigma^2}\right)$. Table-I shows the detailed parameters of a sub-6 GHz band and of a high-frequency millimeter wave system, termed System-I and System-II, respectively, considered in our simulations. Table-II and Table-III summarize the channel parameters in a high-mobility scenario for System-I and in the low and average mobility scenarios for System-II, respectively.

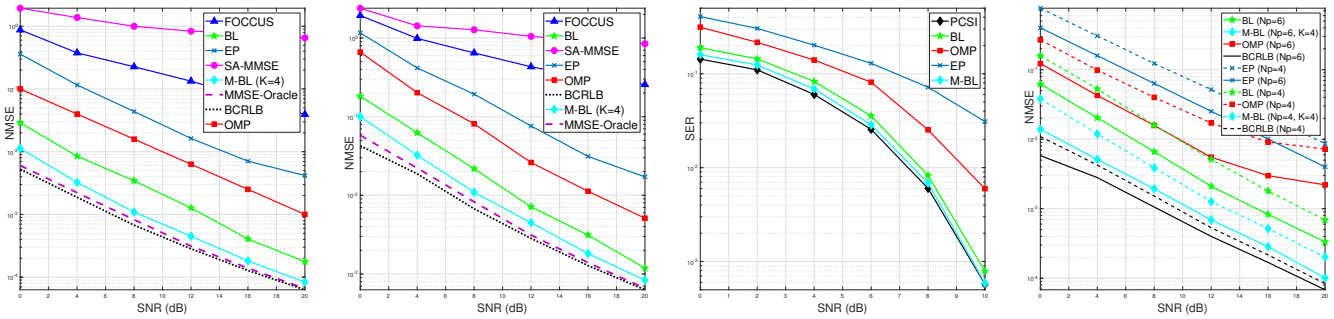
TABLE III
DD-PROFILE OF THE WIRELESS CHANNEL OF SYSTEM-II

Path-index(i)	**1**	**2**	**3**	**4**	**5**
Delay (τ_i) (μs)	0.3	1	1.7	2.4	3.1
Doppler (ν_i) (Hz)	0	610	1220	2440	3660
Speed (Km/h)	0	23.5	47	94	141

Fig. 1(a) compares the NMSE performance of the proposed BL-based schemes to that of the other competing approaches for the OTFS setup of System-I. We can readily observe the significant NMSE improvement of the proposed BL scheme over the SA-MMSE, FOCUSS, OMP and EP-based approaches. For the sparsity-agnostic SA-MMSE estimate of Eq. (15), the covariance matrix $\mathbf{R}_h$ is set to an identity matrix. Since the SA-MMSE and EP-based schemes do not exploit the sparsity of the DD-domain CSI, they yield the poor NMSE. The poor performance of the OMP can be attributed to its sensitivity to both the stopping parameter as well as to the dictionary matrix. By contrast, the limitation of the FOCUSS arises due to its convergence deficiencies and sensitivity to the regularization parameter [14]. Thus, the performance of OMP, FOCUSS and EP-based schemes is not competitive. On the other hand, the NMSEs of the BL-based schemes are seen to be close to that of the MMSE-Oracle and BCRLB. In fact, the M-BL technique associated with $K = 4$ pilot snapshots, matches the benchmarks derived. Fig. 1(b) compares the NMSE of the various CSI estimation techniques for the OTFS setup of System-II. A similar trend has been observed here also, where the BL schemes yield the best overall performance. Finally, Fig. 1(c) portrays the BER of all the contending schemes. It can be observed from the figure that owing to its improved estimation accuracy, the BER achieved using the detectors derived from the BL-based channel estimates is markedly better than that of its OMP and FOCUSS counterparts. Moreover, the former is also seen to approach the BER of a receiver with perfect CSI (PCSI), which demonstrates its enhanced capability of sparse CSI recovery.

It is worth noting that the proposed CSI estimation framework transmits MN_p pilot symbols, followed by an OTFS data frame comprising MN symbols. Thus, the pilot overhead is $\frac{N_p}{N+N_p}$. On the other hand, the pilot overhead for the conventional EP-based technique [7] is given by $\frac{(2M_\tau+1)(2N_\nu+1)}{MN}$. Upon substituting the various parameters from Table-I, the pilot overhead of the proposed CSI estimation framework becomes 0.2 and 0.11, for System-I and System-II, respectively. By contrast, the corresponding overheads for the schemes in [7] are 0.54 and 0.13, respectively, which are higher than that of the proposed estimation framework.

Let us now consider an interesting scenario, where the maximum delay and Doppler spread parameters of the underlying wireless channel of System-I are set to $M_\tau = 16$ and $N_\nu = 16$, respectively, whereas the number of pilot symbols N_p is set to $\{4, 6\}$. It can be readily observed that the channel estimation model of (14) becomes *'ill-posed'*, since one has to estimate an $M_\tau N_\nu = 256$-dimensional sparse vector $\mathbf{h}$ using only $MN_p \in \{128, 192\}$ pilot outputs. The NMSE versus SNR performance of these schemes is presented in Fig. 1(d). It is important to note that the proposed sparse estimation frameworks can yield a sufficiently low NMSE even in this challenging scenario, thanks to the sparse signal recovery guarantees available from the small number of measurements using compressive sensing techniques [12]–[16], [21], [22]. Interestingly, for this scenario, the conventional EP-based technique [7] becomes rather inefficient, since its pilot overhead



(a) (b) (c) (d)

Fig. 1. (a) NMSE versus SNR performance of the OTFS System-I; (b) NMSE versus SNR performance of the OTFS System-II; (c) BER versus SNR performance of the OTFS System-I; (d) NMSE versus SNR performance of the OTFS System-I with $M_\tau = 16$ and $N_\nu = 16$.

tends to 1. This is due to the fact that it cannot accommodate any data symbol along with the pilot and guard symbols in the same OTFS frame. By contrast, the pilot overhead of the proposed CSI estimation framework becomes 0.11 for $N_p = 4$, and 0.15 for $N_p = 6$, which is significantly lower, since our framework does not require any guard symbols.

VII. Conclusions

A novel sparse CSI estimation model is proposed for OTFS modulation based wireless systems, where the pilots are directly transmitted over the TF-domain grid. The proposed CSI estimation model reduces the pilot overhead, training duration and pre-processing burden, despite dispensing with the insertion of a DD-domain guard interval between the pilot and data symbols. The latter trait results in a significantly reduced pilot overhead and an increased bandwidth efficiency. Our novel BL framework developed using the aforementioned model exploits the DD-domain sparsity as well as simultaneous-sparsity for improved CSI estimation. The DD-domain CSI obtained was then directly employed in the subsequent data detection phase without assuming any further knowledge with regards to the number of dominant multipath components. Our simulation results demonstrated the performance improvement of the proposed BL-based schemes for OTFS channel estimation over the conventional EP-based technique as well as over the OMP, FOCUSS and the MMSE estimators, both in terms of the MSE, BER and pilot overhead.

References

[1] 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.

[2] 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:1808.00519*, 2018.

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

[4] 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.

[5] Z. Wei, W. Yuan, S. Li, J. Yuan, G. Bharatula, R. Hadani, and L. Hanzo, "Orthogonal time-frequency space modulation: A full-diversity next generation waveform," *arXiv e-prints*, pp. arXiv–2010, 2020.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] M. E. Tipping, "The relevance vector machine," in *Advances in neural information processing systems*, 2000, pp. 652–658.

[13] ——, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. Jun, pp. 211–244, 2001.

[14] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[15] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE transactions on Image Processing*, vol. 18, no. 9, pp. 2059–2070, 2009.

[16] L. Chaari, J.-Y. Toumeret, and C. Chaux, "Sparse signal recovery using a Bernoulli generalized Gaussian prior," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1711–1715.

[17] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[18] S. Srivastava, R. K. Singh, A. K. Jagannatham, and L. Hanzo, "Technical report: Bayesian learning aided sparse channel estimation for orthogonal time frequency space modulated systems." IIT Kanpur, Tech. Rep., 2021. [Online]. Available: http://www.iitk.ac.in/mwn/documents/MWNLab_TR_SISO_OTFS.pdf.

[19] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[20] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.

[21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.