Joint Bayesian Channel Estimation and Data Detection for OTFS Systems in LEO Satellite Communications

Xueyang Wang, Wenqian Shen, Chengwen Xing, *Member, IEEE*, Jianping An, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*— Lower earth orbit (LEO) satellites play an important role in the integration of space and terrestrial communication networks, which typically encounter high-mobility scenarios. It has been shown that orthogonal time frequency space (OTFS) modulation performs well in such high-mobility scenarios by transforming the time-varying channels into the delay-Doppler domain. In this paper, we develop a joint channel estimation and data detection algorithm for OTFS-based LEO satellite communications. Firstly, we adopt the powerful variational Bayesian inference (VBI) method for estimating the delay-Doppler channel vector, which contains the channel gain, the delay and the Doppler. Secondly, we exploit the unknown data symbols in an OTFS frame as 'virtual pilots' for improving the accuracy of channel estimation and detect them simultaneously. Our simulation results demonstrate that the proposed algorithm achieves improved channel estimation mean square error and bit error rate performance than its conventional counterparts.**

***Index Terms*— LEO satellites, high-mobility, orthogonal time frequency space (OTFS), VBI, joint channel estimation, data detection.**

I. Introduction

LOWER earth orbit (LEO) satellites play an important role in the integration of space and terrestrial communication networks, which are regarded as one of the key techniques of next-generation wireless systems thanks to their seamless wireless coverage [1]. LEO satellites tend to orbits at altitudes between 500 and 2000 km [2], in support of global wireless access at enhanced data rates. Although LEO satellites have the twin benefits of low delay and path-loss [3], there are still numerous challenging problems to be solved, especially their high Doppler frequency, since they move at a high speed of 5 to 10 km/s [3]. One of the challenging problems is channel estimation in high Doppler LEO scenarios. Impressive studies [4]–[6] have focusd on this problem. For example, Zhang *et al.* [4] proposed a deep learning (DL)-based channel prediction for obtaining accurate instaneous channel state information. However, the DL-based method is of high complexity. Pu *et al.* [5] developed a Kalman filtering-based channel estimation method for orthogonal frequency division multiplexing (OFDM)-aided LEO systems, while Zhang *et al.* [6] proposed a two-stage Bernoulli-Rician message passing relying on the expectation-maximization algorithm for addressing both user activity detection and channel estimation problems of LEO systems. However, they did not consider the severe impact of high Doppler frequencies in LEO scenarios. In such high-mobility scenarios, the orthogonality of the OFDM subcarriers is eroded due to the high Doppler frequency. Therefore, inter-carrier interference appears [7], which makes OFDM almost unsuitable for high-mobility LEO satellite communications [8].

Recently, R. Hadani *et al.* proposed the novel orthogonal time frequency space (OTFS) modulation technique [9], [10], which transforms a time-varying channel into an almost time-invariant two-dimensional channel in the delay-Doppler domain. Therefore, the data symbols can be multiplexed in the near-constant delay-Doppler domain. Therefore, OTFS is capable of addressing the above problems caused by excessive Doppler in LEO scenarios. This advantage makes OTFS eminently suitable for high-mobility scenarios such as LEO satellite communications [11]. However, both channel estimation and data detection become more challenging than in OFDM, since in OTFS the data symbols have a more complex two-dimensional convolutional relationship with the delay-Doppler domain channel response. Therefore, the channel estimation and data detection methods used in OFDM system cannot be directly applied in OTFS system.

As for channel estimation in OTFS systems, many authors have proposed sophisticated solutions. For example, Muriali *et al.* [12] conceived an impulse-based channel estimation method, which transmits an impulse surrounded by several guard symbols used as pilots and performs

Manuscript received 16 October 2021; revised 22 March 2022; accepted 23 May 2022. Date of publication 30 May 2022; date of current version 15 July 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61901034, and in part by the Open Research Fund of the Shaanxi Province Key Laboratory of Information Communication Network and Security under Grant ICNS201905. Lajos Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council project EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). The associate editor coordinating the review of this article and approving it for publication was M. A. A. Imran. *(Corresponding author: Wenqian Shen.)*

Xueyang Wang, Wenqian Shen, Chengwen Xing, and Jianping An are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: xywang1103@gmail.com; shenwq@bit.edu.cn; xingchengwen@gmail.com; an@bit.edu.cn).

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3179389.

Digital Object Identifier 10.1109/TCOMM.2022.3179389

TABLE I
CONTRASTING OUR CONTRIBUTIONS TO THE STATE-OF-THE-ART

	[13]	[14]	[15]	[16]	[19]	[22]	Our Proposed Method
OTFS System	✓	✓	✓	✓		✓	✓
OFDM System					✓		
Non-CS based	✓					✓	
CS Based		✓	✓	✓	✓		✓
Data Aided					✓	✓	✓
Pilot Based	✓	✓	✓	✓			

deconvolution at the receiver [9], since the received signal is obtained by the two-dimensional convolution between the transmitted signal and the channel. Raviteja *et al.* [13] poposed an embedded pilot-based channel estimation method which beneficially arranged the pilot, the guard and data symbols in an OTFS frame. To take full advantage of the delay-Doppler channel's sparsity, algorithms based on compressed sensing (CS) have also been proposed. Zhao *et al.* [14] formulated the OTFS channel estimation problem as a sparse Bayesian learning (SBL) framework and utilized the powerful expectation maximization (EM) method for estimating the channel parameters by assuming the Laplace prior of pilot signals. Shen *et al.* [15] firstly formulated a three-dimensional (3D) structured channel model in the delay-Doppler-angle domain and proposed a 3D-structured orthogonal matching pursuit (3D-SOMP) algorithm for channel estimation, which exploited sparsity of multiple-input multiple-output (MIMO) OTFS channels. Furthermore, Liu *et al.* [16] proposed an uplink-aided downlink channel estimator which exploited the parameters estimated in the uplink for assisting the downlink channel estimation. An expectation maximization variational Bayesian (EM-VB) algorithm was proposed by Liu *et al.* [16] for calculating the uplink channel parameters. However, the above algorithms only used the pilot symbols, but our proposition is that the performance of channel estimation can be further improved by additionally exploiting the data symbols.

Data-aided channel estimation algorithms have indeed been proposed for conventional OFDM systems in the literature. For example, the SBL framework is utilized for joint channel estimation and data detection of OFDM systems in [17]–[21]. Specifically, Prasad *et al.* [17] proposed a low-complexity recursive joint SBL algorithm (J-SBL) for quasi-static scenarios, complemented by a low-complexity Kalman filtering-based SBL (K-SBL) algorithm conceived for time-varying scenarios. In [18], Prasad *et al.* extended the algorithms in [17] to MIMO-OFDM systems and exploited the group-approximately sparse (ga-sparse) and the group-approximately-cluster-sparse (gac-sparse) properties of MIMO-OFDM channels for joint channel estimation and data detection. Joint channel estimation and data detection algorithms have also been applied in near-capacity coherent space-time shift keying (CSTSK) based MIMO systems by Zhang *et al.* [20] and in space-time trellis coded MIMO-OFDM systems by Mishra *et al.* [21], respectively. Furthermore, Ding *et al.* [19] proposed a data-aided quantized Bayesian channel estimation algorithm, which harnessed the transmitted data symbols as virtual pilots for improving the performance of channel estimation without any additional pilot symbols. However, these methods cannot be directly used in OTFS systems because of the complex two-dimensional convolutional relationship between the transmitted signal and the channel. Yuan *et al.* [22] for the first time proposed a data-aided channel estimation algorithm for OTFS systems based on a superimposed pilot and data transmission scheme, where they used the sum-product algorithm (SPA) for data detection. They achieved improved spectral efficiency by using data-aided channel estimation. However, they have not exploited the sparsity of the delay-Doppler channel, which can be fully exploited by the SBL framework for obtaining more accurate channel estimates. Therefore, by exploiting the sparsity of the delay-Doppler channel, we propose a joint channel estimation and data detection algorithm based on the sophisticated SBL framework for operation in high-mobility scenarios, which can simultaneously achieve accurate channel estimation and data detection.

The main contributions of this paper are summarized as follows:

- We formulate the joint channel estimation and data detection problem of OTFS systems as an SBL framework, where the unknown data information is introduced as a hidden variable for improving the accuracy of channel estimation without increasing the pilot overhead. Moreover, the data symbols can be detected simultaneously under the data-aided SBL framework advocated.
- We develop a joint channel estimation and data detection algorithm for OTFS systems based on variational Bayesian inference (VBI) [23]–[26]. Since the transmitted OTFS frame is convolved with the delay-Doppler channel for obtaining the received OTFS frame, the relationship between a single transmitted symbol and the corresponding received symbol is more complex than in conventional OFDM systems. Therefore, we calculate the *a posteriori* probability of the transmit vector, which contains the information of the entire OTFS frame, by harnessing the probability mass function (pmf) of every symbol in the frame. Then, we can update the *a posteriori* probability of the channel vector based on both the estimated data symbols and known pilot symbols. Moreover, for simplifying the calculation of the *a posteriori* probability of transmit vectors when the dimension of the OTFS frames is large, we approximate the *a priori* pmf of transmit data symbols by the Gaussian distribution.
- Our simulation results demonstrate that in high-mobility scenarios the proposed algorithm outperforms the

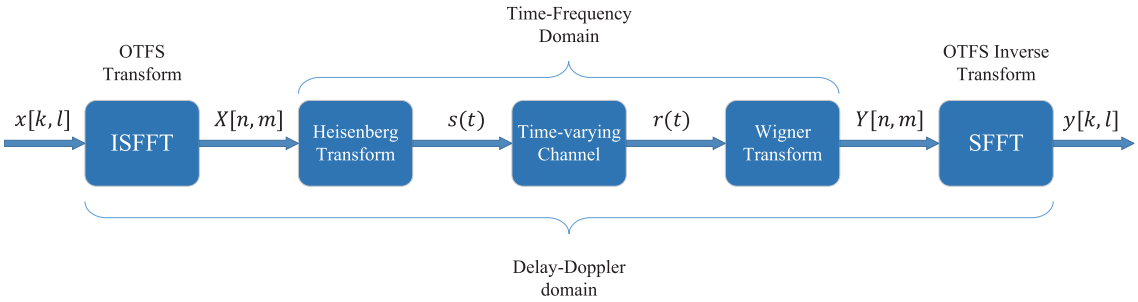


Fig. 1. OTFS communication system.

conventional OTFS channel estimation algorithms. Moreover, the bit error rate (BER) of our proposed algorithm is better than that of the conventional MP algorithms when the size of OTFS frame is large. Finally, we show that the Gaussian approximation of the data symbols' pmf is efficient despite its reduced complexity.

To boldly and explicitly clarify our contributions, we contrast our proposed solution to the state-of-the-art in Table I.

The rest of this paper is organized as follows. In Section II, we present the system model of OTFS based on both ideal bi-orthogonal waveforms and on practical rectangular waveforms. In Section III, we formulate the OTFS channel estimation problem as an SBL framework and harness the VBI method for solving it. In Section IV, we propose a joint channel estimation and data detection method. Finally, in Section V, we provide our simulation results, followed by our conclusions in Section VI.

Notations: We use the following notations throughout this paper. We let a, $\mathbf{a}$, $\mathbf{A}$ represent a scalar, vector, and matrix, respectively; $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{-1}$ denote the transpose, conjugate transpose, and inverse of a matrix, respectively; $\delta(\cdot)$ is the Dirac delta function; $[\cdot]_N$ denotes modulo N operation; $\lceil a \rceil$ represents the smallest integer no less than a; $\lfloor a \rfloor$ is the lagest integer no greater than a; a^* denotes the conjugate of complex numbers; $\propto$ indicates equality up to a multiplicative constant; $q^*(\mathbf{x})$ denotes the optimal solution; $\langle f(\mathbf{x}) \rangle_{p(\mathbf{x})}$ is the expectation with respect to $p(\mathbf{x})$. Finally, $\mathbf{0}$ denotes the zero matrix, $\mathbf{I}$ denotes identity matrix.

II. System Model

A. Input-Output Relationship of OTFS System Based on Bi-Orthogonal Waveform

As shown in Fig. 1, we assume that $x[k,l]$ denotes symbols in a $M \times N$ information block in the delay-Doppler domain, where $k = 0, \ldots, N-1$ and $l = 0, \ldots, M-1$. Based on [27], we transform $x[k,l]$ from the delay-Doppler domain (dDD) to $X[n,m]$ in the time-frequency domain (TFD) by inverse symplectic finite Fourier transform (ISFFT), which is defined as

$$X[n,m] = \frac{1}{\sqrt{NM}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k,l] e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)}, \quad (1)$$

where $n = 0, \ldots, N-1$, $m = 0, \ldots, M-1$. The OTFS frame in the TFD has a duration of NT and bandwidth of $M\Delta f$. From another perspective, we discretize the TFD by sampling the time axes at intervals T and the frequency axes at intervals Δf. Therefore, there are N grids in the time domain (TD) and M grids in the frequency domain (FD).

Next, we transform $X[n,m]$ from the TFD into the transmit signal $s(t)$ of the TD. We assume that the transmit waveform is $g_{tx}(t)$ and receive waveform is $g_{rx}(t)$, which satisfy bi-orthogonality in the ideal case. Therefore, we can obtain the transmit signal $s(t)$ by transforming $X[n,m]$ in the TFD into the TD [9]

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n,m] g_{tx}(t - nT) e^{j2\pi m\Delta f(t-nT)}. \quad (2)$$

This transform from the TFD to the TD is termed as the Heisenberg transform in the mathematics literature [9]. The received signal $r(t)$ after passing through the channel is expressed as [27]

$$r(t) = \iint h(\tau,\nu) s(t-\tau) e^{j2\pi\nu(t-\tau)} d\tau d\nu + w(t), \quad (3)$$

where

$$h(\tau,\nu) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i) \delta(\nu - \nu_i), \quad (4)$$

is the sparse representation of the time-varying channel in the dDD[1] and $w(t)$ is noise. In the representation of $h(\tau,\nu)$ of (4), P is the number of propagation paths, while h_i, τ_i and ν_i denote the channel gain, delay and Doppler of the i-th path, respectively. Moreover, the delay and Doppler taps of the i-th path can be represented as $\tau_i = l_i/M\Delta f$, $\nu_i = k_i/NT$, where l_i and k_i denote the integer indices of the delay τ_i and Doppler ν_i, respectively.

We can further reformat the received signal $Y(t,f)$ in the TFD as

$$Y(t,f) = \int g_{rx}^*(t'-t) r(t') e^{-j2\pi f(t'-t)} dt'. \quad (5)$$

After sampling at $t = nT$, $f = m\Delta f$, $Y(t,f)$ can be expressed as

$$Y[n,m] = Y(t,f)|_{t=nT, f=m\Delta f}. \quad (6)$$

[1](4) is equivalent to the channel model of LEO systems in [1].

The transformation from $r(t)$ to $Y[n,m]$ represented by (5) and (6) is known as the Wigner transform [9], which transforms $r(t)$ from the TD into $Y[n,m]$ in the TFD and is the inverse of the Heisenberg transform. We obtain the received signal $y[k,l]$ in the dDD by applying the symplectic finite Fourier transform (SFFT), which is defined as

$$y[k,l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y[n,m] e^{-j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)}. \quad (7)$$

In the ideal case, when the transmit waveform $g_{tx}(t)$ and the receive waveform $g_{rx}(t)$ satisfy the bi-orthogonality property, the input-output relationship in the dDD can be expressed as [10], [27], [28]

$$y[k,l] = \sum_{i=1}^{P} h_i e^{-\frac{j2\pi l_i k_i}{MN}} x\left([k-k_i]_N, [l-l_i]_M\right) + w[k,l], \quad (8)$$

where k is an integer in $[0, N-1]$, l is an integer in $[0, M-1]$, $l_i \in [0, l_{max}]$, $k_i \in [-k_{\max}, k_{max}]$, while l_{max} and k_{max} are the maximum delay and Doppler indices, respectively. Furthermore, $w[k,l]$ in (8) is approximated as Gaussian noise [27].

For the sake of simplicity, we use the integers t and d to denote delay index and Doppler index in (8), where $t \in [0, l_{\max}]$ and $d \in [-k_{\max}, k_{\max}]$. The size of pilot symbols is assumed to be $M_p \times N_p$, which will be described in detail in the next subsection. By only considering the pilot symbols in an OTFS frame, (8) can be expressed as

$$y[k,l] = \sum_{t=0}^{l_{\max}} \sum_{d=-k_{\max}}^{k_{\max}} h_{t,d} e^{-\frac{j2\pi td}{MN}} x\left([k-d]_N, [l-t]_M\right) + w[k,l], \quad (9)$$

where $k \in \left[N/2 - \left\lceil \frac{N_p}{2} \right\rceil - k_{max}, N/2 + \left\lfloor \frac{N_p}{2} \right\rfloor + k_{max}\right]$, $l \in \left[M/2 - \left\lceil \frac{M_p}{2} \right\rceil + 1, M/2 + \left\lfloor \frac{M_p}{2} \right\rfloor + l_{max}\right]$.

We can rewrite (9) in a matrix form as follows:

$$\mathbf{y}_p^{bi} = \mathbf{\Phi}_p^{bi} \mathbf{h} + \mathbf{w}_p, \quad (10)$$

where $\mathbf{y}_p^{bi} \in \mathbb{C}^{Z\times 1}$ is obtained by stacking $y[k,l]$, $\mathbf{w} \in \mathbb{C}^{Z\times 1}$ is obtained by stacking $w[k,l]$. Furthermore, we have $Z = (l_{\max} + M_p)(N_p + 2k_{\max})$, $\mathbf{h} = [h_{0,-k_{\max}}, \ldots, h_{l_{\max},-k_{\max}}, \ldots, h_{t,d}, \ldots, h_{0,k_{\max}}, \ldots, h_{l_{\max},k_{\max}}]^T \in \mathbb{C}^{(l_{\max}+1)(2k_{\max}+1)\times 1}$ and $\mathbf{\Phi}_p^{bi} \in \mathbb{C}^{Z\times(l_{\max}+1)(2k_{\max}+1)}$ of (10) depends on the pilots, with the $(l(N_p + 2k_{max}) + k + 1)$-st row of $\boldsymbol{\phi}_p[k,l] = [\phi_{0,-k_{\max}}[k,l], \ldots, \phi_{l_{\max},-k_{\max}}[k,l], \ldots, \phi_{t,d}[k,l], \ldots, \phi_{0,k_{\max}}[k,l], \ldots, \phi_{l_{\max},k_{\max}}[k,l]]$, where $\phi_{t,d}[k,l] = x([k-d]_N, [l-t]_M) e^{-j2\pi td/MN}$.

Due to the limited number of paths in the dDD, vector $\mathbf{h}$ is usually sparse [14]. Hence, the channel estimation problem (10) can be solved by powerful sparse signal recovery algorithms.

B. Input-Output Relationship of OTFS System Based on Rectangular Waveform

Having an ideal bi-orthogonal waveform is not practically feasible because of Heisenberg's uncertainty principle [29]. Therefore, we use a rectangular transmit waveform and receive waveform. The input-output relationship based on these rectangular waveforms in the dDD can be expressed as [27]

$$y[k,l] = \sum_{i=1}^{P} h_i e^{-\frac{j2\pi(l-li)k_i}{MN}} \beta_i(k,l) x\left([k-k_i]_N, [l-l_i]_M\right) + w[k,l], \quad (11)$$

where we have

$$\beta_i(k,l) = \begin{cases} 1 & l_i \le l < M \\ \frac{N-1}{N} e^{-j2\pi\left(\frac{[k-k_i]_N}{N}\right)} & 0 \le l < l_i, \end{cases} \quad (12)$$

with k being an integer in $[0, N-1]$ and l being an integer in $[0, M-1]$.

Similarly to the previous subsection, we use the integers t and d to represent the delay index and Doppler index in (11), where $t \in [0, l_{\max}]$ and $d \in [-k_{\max}, k_{\max}]$. By only considering the pilot symbols of an OTFS frame, we have $k \in [N/2 - \lceil \frac{N_p}{2} \rceil - k_{max}, N/2 + \lfloor \frac{N_p}{2} \rfloor + k_{max}]$, $l \in [M/2 - \lceil \frac{M_p}{2} \rceil + 1, M/2 + \lfloor \frac{M_p}{2} \rfloor + l_{max}]$. We can have $l \ge l_{max}$ by carefully arranging the pilots, as discussed in the next subsection. Hence, (11) can be simplified to

$$y[k,l] = \sum_{t=0}^{l_{\max}} \sum_{d=-k_{\max}}^{k_{\max}} h_{t,d} e^{-\frac{j2\pi(l-t)d}{MN}} x\left([k-d]_N, [l-t]_M\right) + w[k,l]. \quad (13)$$

Rewriting (13) in a matrix-form, we have

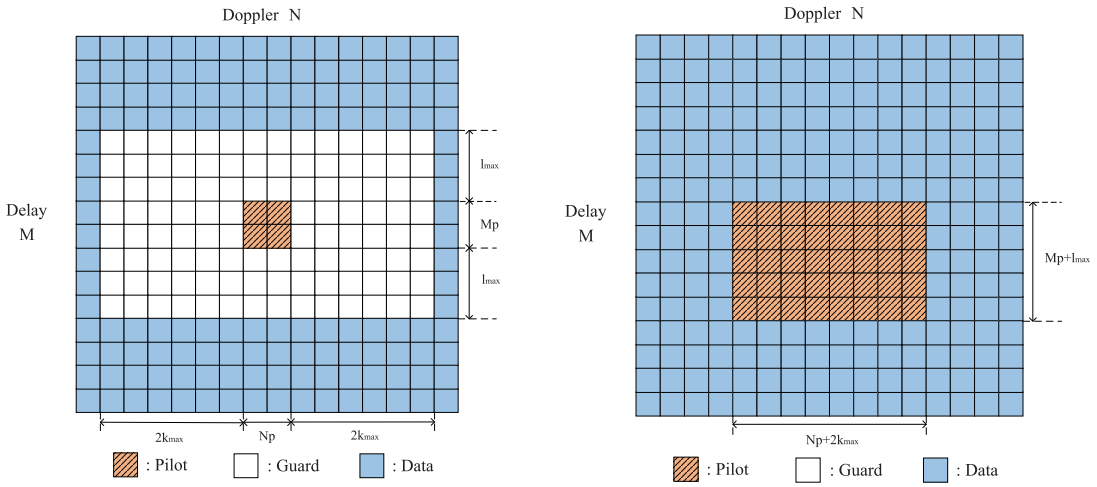
$$\mathbf{y}_p = \mathbf{\Phi}_p \mathbf{h} + \mathbf{w}_p, \quad (14)$$

where $\mathbf{y}_p \in \mathbb{C}^{Z\times 1}$ is obtained by stacking $y[k,l]$, $\mathbf{w}_p \in \mathbb{C}^{Z\times 1}$ is formulated by stacking $w[k,l]$, while $\mathbf{\Phi}_p \in \mathbb{C}^{Z\times(l_{\max}+1)(2k_{\max}+1)}$ depends on the pilots, with the $(k(N_p + 2k_{max}) + l + 1)$-st row given by $[\phi_{0,-k_{max}}[k,l], \phi_{1,-k_{max}}[k,l], \ldots, \phi_{t,d}[k,l], \ldots, \phi_{l_{max},k_{max}}[k,l]]$ and $\phi_{t,d}[k,l] = x([k-d]_N, [l-t]_M) e^{j2\pi(l-t)d/MN}$.

C. Pilot Arrangement

Fig. 2 shows the pilot arrangement. Explicitly, Fig. 2(a) presents the arrangement of a transmit OTFS frame in the dDD, where the orange upper right slash squares denote the pilot symbols, the white squares represent the guard symbols and the blue squares are data symbols. The dimension of pilots is of $M_p \times N_p$, while the frame size is of $M \times N$. To avoid the interference between data and pilot symbols, guard symbols of length l_{max} are required in the delay domain and of length $2k_{max}$ in the Doppler domain [13]. Moreover, we can set $M = 16$, $M_p = 2$ and $l_{max} = 3$, which makes the delay indices of the pilot symbols to be $l = 7, 8$. Hence, we have $l > l_{max}$ to simplify the input-output relationship.

Fig. 2 (b) shows the received OTFS frame. Observe that the pilot symbols are spread to l_{max} intervals into delay



(a) Pilot arrangement at transmitter.

(b) Pilot arrangement at receiver.

Fig. 2. The pilot arrangement at the transmitter and receiver.

domain and to $\pm k_{max}$ in the Doppler domain. Thus, the pilot symbols at the receiver have the dimension of $(M_p + l_{max}) \times (N_p + 2k_{max})$.

III. OTFS Channel Estimation Using SBL

In this section, we adopt the SBL framework and use variational Bayesian inference for recovering the sparse vector $\mathbf{h}$ introduced in the previous section. Theoretically, we have to calculate the *a posteriori* marginal distribution after observing the received symbols $\mathbf{y}_p$. However, this causes the calculation of high-dimensional integration, which is excessively complex. Therefore, we resort to an alternative method called VBI [24]–[26], [30], [31] to obtain approximate solution instead of the *a posteriori* marginal distribution.

A. SBL Framework

In the conventional SBL framework [24], [25], [32], the channel gains $h_l, l = 0, \ldots, L-1$ obey the Gaussian distribution, where we have

$$p(\mathbf{h} \mid \boldsymbol{\gamma}) = \prod_{l=0}^{L-1} p\left(h_l \mid \gamma_l\right) = \prod_{l=0}^{L-1} \mathcal{CN}\left(h_l; 0, \gamma_l^{-1}\right), \quad (15)$$

where $L = (l_{max}+1)(2k_{max}+1)$, γ_l is the precision of h_l and the hyperparameter $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_{L-1}]^T$ is the precision vector. We assume that $\boldsymbol{\gamma}$ obeys the Gamma distribution, i.e. [24]

$$p(\boldsymbol{\gamma}) = \text{Gamma}(\boldsymbol{\gamma}; a, b), \quad (16)$$

where a is the shape parameter and b is the inverse scale parameter of Gamma distribution. By assuming that γ_l obeys the Gamma distribution, $p(\gamma_l)$ is the conjugate *a priori* of the likelihood function $p(h_l|\gamma_l)$. This property can simplify the calculation of the *a posteriori* distribution. Furthermore, by completing the integration of $p(h_l) = \int p(\gamma_l)p(h_l|\gamma_l)d\gamma_l$, the marginal distribution of h_l becomes the Student-t distribution, which explains the sparsity of $\mathbf{h}$ [25].

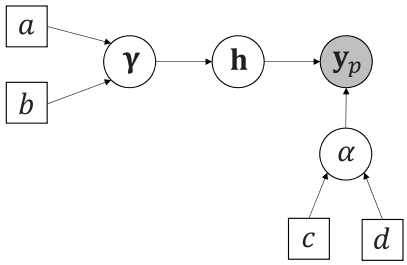


Fig. 3. Factor graph of pilot-based VBI.

We assume that the noise $\mathbf{w}_p$ obeys the zero-mean complex Gaussian distribution of

$$p(\mathbf{w}_p) = \mathcal{CN}\left(0, \alpha^{-1}\mathbf{I}\right), \quad (17)$$

where α is the inverse of the noise power, which is called the precision in SBL framework. Hence, we have

$$p(\mathbf{y}_p \mid \mathbf{h}, \alpha) = \mathcal{CN}\left(\boldsymbol{\Phi}_p\mathbf{h}, \alpha^{-1}\mathbf{I}\right). \quad (18)$$

It is assumed that α obeys the Gamma distribution of

$$p(\alpha) = \text{Gamma}(\alpha; c, d), \quad (19)$$

where c is the shape parameter and d is the inverse scale parameter of Gamma distribution.

B. Variational Bayesian Inference Principles

Fig. 3 shows the factor graph of the pilot-based VBI method, where the squares represent constant values, the circles denote hidden variables and the shaded circle denotes the observed value. We use $\boldsymbol{\Theta} = \{\alpha, \mathbf{h}, \boldsymbol{\gamma}\}$ to denote the hidden variables to be estimated. Based on the hierarchical prior model, the joint prior of hidden variables can be expressed as

$$p(\mathbf{y}_p, \boldsymbol{\Theta}) = p(\mathbf{y}_p \mid \mathbf{h}, \alpha)p(\mathbf{h} \mid \boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\alpha). \quad (20)$$

We now aim for calculating the maximum *a posteriori* (MAP) estimator of $p(\mathbf{\Theta} \mid \mathbf{y}_p)$, which relies on high-dimensional integration. But since the calculation for high-dimensional integration is excessively complex, we resort to the VBI method [24]–[26], [33] for solving this problem by alternate updates.

The Kullback-Leibler (KL) divergence [24]–[26] is introduced for quantifying the difference between two probability functions, which is expressed as follows,

$$\mathrm{KL}(q(\mathbf{\Theta})\|p(\mathbf{\Theta} \mid \mathbf{y}_p)) = -\int q(\mathbf{\Theta}) \ln \frac{p(\mathbf{\Theta} \mid \mathbf{y}_p)}{q(\mathbf{\Theta})} d\mathbf{\Theta}. \quad (21)$$

$\mathrm{KL}(q(\mathbf{\Theta})\|p(\mathbf{\Theta} \mid \mathbf{y}_p))$ quantifies the difference between the real *a posteriori* marginal distribution $p(\mathbf{\Theta} \mid \mathbf{y}_p)$ and the variational approximation $q(\mathbf{\Theta})$. Therefore, we have to find the optimal probability function $q(\mathbf{\Theta})$ for minimizing the KL divergence, which makes $q(\mathbf{\Theta})$ as similar as possible to $p(\mathbf{\Theta} \mid \mathbf{y}_p)$. Hence, we formulate the problem as follows,

$$q^*(\mathbf{\Theta}) = \arg\max_{q(\mathbf{\Theta})} \int q(\mathbf{\Theta}) \ln \frac{p(\mathbf{y}_p, \mathbf{\Theta})}{q(\mathbf{\Theta})} d\mathbf{\Theta}. \quad (22)$$

According to mean field theory [24]–[26], $q(\mathbf{\Theta})$ can be factorized as the product of the hidden variables' probabilities formulated as:

$$q(\mathbf{\Theta}) = q(\alpha) q(\mathbf{h}) q(\boldsymbol{\gamma}). \quad (23)$$

Based on [23], problem (22) can be transformed into an optimization problem as follows:

$$\ln q^*(\mathbf{\Theta}_n) = \langle \ln p(\mathbf{y}_p, \mathbf{\Theta}) \rangle_{\Pi_{i \neq n} q^*(\mathbf{\Theta}_i)} + \text{const} \quad n = 1, 2, 3, \quad (24)$$

where $\mathbf{\Theta}_n$ denotes the n-th element in $\mathbf{\Theta}$ and $\langle \ln p(\mathbf{y}_p, \mathbf{\Theta}) \rangle_{\Pi_{i \neq n} q^*(\Theta_i)}$ denotes an expectation with respect to all factors in $\mathbf{\Theta}$ except $q^*(\mathbf{\Theta}_n)$. Since the optimal probability functions $q^*(\mathbf{\Theta}_i), i = 1, 2, 3$ and $q^*(\mathbf{\Theta}_j), j \neq i$ are inter-dependent, it is an open challenge to obtain a closed-form solution. Hence, we can calculate a stable solution by alternately updating the probability functions as follows:

$$q^{(i+1)}(\alpha) \propto \exp\left(\langle \ln p(\mathbf{y}, \mathbf{\Theta}) \rangle_{q^{(i)}(\mathbf{h}) q^{(i)}(\boldsymbol{\gamma})} \right), \quad (25)$$

$$q^{(i+1)}(\mathbf{h}) \propto \exp\left(\langle \ln p(\mathbf{y}, \mathbf{\Theta}) \rangle_{q^{(i+1)}(\alpha) q^{(i)}(\boldsymbol{\gamma})} \right), \quad (26)$$

$$q^{(i+1)}(\boldsymbol{\gamma}) \propto \exp\left(\langle \ln p(\mathbf{y}, \mathbf{\Theta}) \rangle_{q^{(i+1)}(\alpha) q^{(i+1)}(\mathbf{h})} \right), \quad (27)$$

where i denotes the iteration index and $q^{(i)}(\cdot)$ represents the probability function of variables in the i-th iteration.

In the next subsection, the detailed process of solving (25)–(27) is provided.

C. Update Details for SBL

1) Update of $q(\alpha)$: Based on (18) and (19), we can derive (25) as

$$\begin{aligned} &\ln q^{(i+1)}(\alpha) \\ &\quad \propto \langle \ln p(\mathbf{y}_p, \mathbf{\Theta}) \rangle_{q^{(i)}(\mathbf{h}) q^{(i)}(\boldsymbol{\gamma})} \\ &\quad \propto \langle \ln p(\mathbf{y}_p \mid \mathbf{h}, \alpha) \rangle_{q^{(i)}(\mathbf{h})} + \ln p(\alpha) \\ &\quad \propto (a + Z - 1) \ln(\alpha) \\ &\qquad - \alpha \left[b + \left\| \mathbf{y}_p - \mathbf{\Phi}_p \mathbf{u}_{\mathbf{h}}^{(i)} \right\|_2^2 + \mathrm{tr}\left(\mathbf{\Phi}_p \mathbf{\Sigma}_{\mathbf{h}}^{(i)} (\mathbf{\Phi}_p)^H \right) \right], \end{aligned} \quad (28)$$

where Z is the dimension of $\mathbf{y}_p$, $\mathbf{u}_{\mathbf{h}}^{(i)} = \langle \mathbf{h} \rangle_{q^{(i)}(\mathbf{h})}$, $\mathbf{\Sigma}_{\mathbf{h}}^{(i)} = \left\langle (\mathbf{h} - \mathbf{u}_{\mathbf{h}}^{(i)})(\mathbf{h} - \mathbf{u}_{\mathbf{h}}^{(i)})^H \right\rangle_{q^{(i)}(\mathbf{h})}$.

Based on the expression of the Gamma distribution, we have $a_\alpha^{(i+1)} = a + Z$, $b_\alpha^{(i+1)} = b + \left\| \mathbf{y}_p - \mathbf{\Phi}_p \mathbf{u}_{\mathbf{h}}^{(i)} \right\|_2^2 + \mathrm{tr}\left(\mathbf{\Phi}_p \mathbf{\Sigma}_{\mathbf{h}}^{(i)} \mathbf{\Phi}_p^H \right)$, and α follows the Gamma distribution of

$$q^{(i+1)}(\alpha) = \mathrm{Gamma}\left(\alpha \mid a_\alpha^{(i+1)}, b_\alpha^{(i+1)} \right). \quad (29)$$

Thus, the mean of α can be expressed as

$$\hat{\alpha}^{(i+1)} = \langle \alpha \rangle_{q^{(i+1)}(\alpha)} = \frac{a_\alpha^{(i+1)}}{b_\alpha^{(i+1)}}. \quad (30)$$

2) Update of $q(\mathbf{h})$:

$$\begin{aligned} &\ln q^{(i+1)}(\mathbf{h}) \\ &\quad \propto \langle \ln p(\mathbf{y}_p, \mathbf{\Theta}) \rangle_{q^{(i+1)}(\alpha) q^{(i)}(\boldsymbol{\gamma})} \\ &\qquad \propto \langle \ln p(\mathbf{y}_p \mid \mathbf{h}, \alpha) \rangle_{q^{(i+1)}(\alpha)} + \langle \ln p(\mathbf{h} \mid \boldsymbol{\gamma}) \rangle_{q^{(i)}(\boldsymbol{\gamma})} \\ &\quad \propto -\hat{\alpha}^{(i+1)} \left\| \mathbf{y}_p - \mathbf{\Phi}_p \mathbf{h} \right\|_2^2 \\ &\qquad - \mathbf{h}^H \left\langle \mathrm{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1} \right) \right\rangle_{q^{(i)}(\boldsymbol{\gamma})} \mathbf{h}. \end{aligned} \quad (31)$$

Thus, $\mathbf{h}$ follows the Gaussian distribution of

$$q^{(i+1)}(\mathbf{h}) = \mathcal{CN}\left(\mathbf{h} \mid \mathbf{u}_{\mathbf{h}}^{(i+1)}, \mathbf{\Sigma}_{\mathbf{h}}^{(i+1)} \right), \quad (32)$$

where we have:

$$\mathbf{\Sigma}_{\mathbf{h}}^{(i+1)} = \left(\hat{\alpha}^{(i+1)} \mathbf{\Phi}_p^H \mathbf{\Phi}_p + \left\langle \mathrm{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1} \right) \right\rangle_{q^{(i)}(\boldsymbol{\gamma})} \right)^{-1}, \quad (33)$$

$$\mathbf{u}_{\mathbf{h}}^{(i+1)} = \hat{\alpha}^{(i+1)} \mathbf{\Sigma}_{\mathbf{h}}^{(i+1)} \mathbf{\Phi}_p^H \mathbf{y}_p. \quad (34)$$

3) Update of $q(\boldsymbol{\gamma})$: Based on (15) and (18), we can reformulate (27) as

$$\begin{aligned} &\ln q^{(i+1)}(\boldsymbol{\gamma}) \\ &\quad \propto \langle \ln p(\mathbf{y}_p, \mathbf{\Theta}) \rangle_{q^{(i+1)}(\alpha) q^{(i+1)}(\mathbf{h})} \\ &\quad \propto \langle \ln p(\mathbf{h} \mid \boldsymbol{\gamma}) \rangle_{q^{(i+1)}(\mathbf{h})} + \ln p(\boldsymbol{\gamma}) \\ &\quad \propto \sum_{l=0}^{L-1} (\underbrace{a + 1}_{a_l^{(i+1)}} - 1) \ln \gamma_l - \gamma_l \underbrace{(b + \langle h_l^* h_l \rangle_{q^{(i+1)}(\mathbf{h})})}_{b_l^{(i+1)}}, \end{aligned} \quad (35)$$

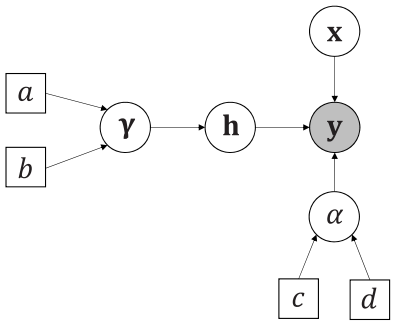


Fig. 4. Factor graph of data-aided VBI.

where $\langle h_l^* h_l \rangle_{q^{(i+1)}(\mathbf{h})} = \mathbf{\Sigma}_{\mathbf{h},(l,l)}^{(i+1)} + \left|\mathbf{u}_{\mathbf{h},(l)}^{(i+1)}\right|^2$. $\mathbf{\Sigma}_{\mathbf{h},(l,l)}^{(i+1)}$ denotes the (l,l)-th element in $\mathbf{\Sigma}_{\mathbf{h}}^{(i+1)}$ and $\mathbf{u}_{\mathbf{h},(l)}^{(i+1)}$ denotes the l-th element in $\mathbf{u}_{\mathbf{h}}^{(i+1)}$. Based on the expression of Gamma distribution function, we have $a_l^{(i+1)} = a + 1$, $b_l^{(i+1)} = b + \langle h_l^* h_l \rangle_{q^{(i+1)}(\mathbf{h})}$ and $\boldsymbol{\gamma}$ follows the Gamma distribution, leading to:

$$q^{(i+1)}(\gamma_l) = \Gamma\left(\gamma_l \mid a_l^{(i+1)}, b_l^{(i+1)}\right), \quad l = 0, \ldots, L-1, \tag{36}$$

where the mean of $\boldsymbol{\gamma}$ is

$$\hat{\gamma}_l^{(i+1)} = \langle \gamma_l \rangle_{q^{(i+1)}(\gamma_l)} = \frac{a_l^{(i+1)}}{b_l^{(i+1)}}. \tag{37}$$

Hence, we have

$$\begin{aligned} &\langle \operatorname{diag}(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}) \rangle_{q^{(i+1)}(\boldsymbol{\gamma})} \\ &\quad = \operatorname{diag}\left(\frac{a_0^{(i+1)}}{b_0^{(i+1)}}, \ldots, \frac{a_{L-1}^{(i+1)}}{b_{L-1}^{(i+1)}}\right). \end{aligned} \tag{38}$$

By iteratively updating (28), (31) and (35), we can obtain the approximate distributions to estimate the channel gains, the delay and the Doppler.

IV. Joint Channel Estimation and Data Detection Using Pilots and Data

In this section, we propose a data-aided SBL algorithm based on the VBI method in the previous section for improving the performance of channel estimation of our OTFS system by exploiting the unknown data symbols as 'virtual pilots'. Simultaneously, the unknown data symbols are detected by calculating the probability functions and estimating the mean of data symbols. This way, we jointly estimate the channel and detect the data.

A. Data-Aided SBL Framework

Since we convert the signals to the dDD and $y[k,l]$ is the superposition of the cyclic shifting of $x[k,l]$ based on (11), $y[k,l]$ and $x[k,l]$ do not correspond one-to-one. Therefore, the data-aided channel estimation of conventional OFDM systems cannot be applied directly in OTFS system. In this subsection, we present the relationship of the transmitted vector $\mathbf{x}$ and received vector $\mathbf{y}$ to calculate the *a posteriori* probability of vector $\mathbf{x}$ through the VBI method [24]–[26], [30], [31] and formulate this data-aided channel estimation problem into the proposed SBL framework. Specifically, we obtain the *a priori* pmf of vector $\mathbf{x}$ based on the pmf of element $x[k,l]$ and calculate the probability function by alternative updates.

In order to calculate the *a posteriori* probability of vector $\mathbf{x}$ to complete the data-aided channel estimation, we rewrite the relationship between the transmitted vector $\mathbf{x}$ and the received vector $\mathbf{y}$. Based on the system model (11), the received signal can be expressed in a matrix-form as [29]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{39}$$

where $\mathbf{x} \in \mathbb{C}^{MN\times 1}$ is based on stacking $x[k,l]$ and $\mathbf{H}$ can be expressed as

$$\begin{aligned} \mathbf{H} &= \sum_{i=1}^{P} h_i \underbrace{\left[(\mathbf{F}_N \otimes \mathbf{I}_M)\, \mathbf{\Pi}^{l_i} \left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_M\right)\right]}_{\mathbf{P}^{(i)}} \\ &\quad \cdot \underbrace{\left[(\mathbf{F}_N \otimes \mathbf{I}_M)\, \mathbf{\Delta}^{k_i} \left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_M\right)\right]}_{\mathbf{Q}^{(i)}} \\ &= \sum_{i=1}^{P} h_i \mathbf{P}^{(i)} \mathbf{Q}^{(i)}, \end{aligned} \tag{40}$$

where $\mathbf{F}_N$ is the n-point DFT matrix and $\mathbf{\Pi}$ is the forward permutation matrix of

$$\mathbf{\Pi} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}_{MN\times MN}, \tag{41}$$

while $\mathbf{\Delta}$ is a $MN \times MN$ diagonal matrix given by

$$\mathbf{\Delta} = \operatorname{diag}\left[z^0, z^1, \ldots, z^{MN-1}\right], \tag{42}$$

where $z = e^{\frac{j2\pi}{MN}}$.

In order to jointly estimate the channel and detect the data symbols, we utilize (39) for calculating the *a posteriori* probability of $\mathbf{x}$ and then use $\mathbf{x}$ to construct the measurement matrix $\mathbf{\Phi}$, which depends on both the pilots and on the estimated data symbols. Therefore, we extend the range of k, l in (14) to the entire frame, i.e. $k \in [0, N-1]$, $l \in [0, M-1]$ and formulate the sparse signal recovery problem as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{h} + \mathbf{w}, \tag{43}$$

where $\mathbf{y} \in \mathbb{C}^{MN\times 1}$ is created by stacking $y[k,l]$, while $\mathbf{\Phi} \in \mathbb{C}^{MN\times(2k_{max}+1)(l_{max}+1)}$ depends on all symbols of an OTFS frame and $\mathbf{w} \in \mathbb{C}^{MN\times 1}$ is formed by stacking $w[k,l]$. Compared to (14), the dimension of $\mathbf{\Phi}$ is higher than that of $\mathbf{\Phi}_p$, which corresponds to more measurement data. However, the data part of $\mathbf{\Phi}$ is unknown, hence we cannot use it for channel estimation directly.

As shown in Fig. 4, the data is introduced into the SBL framework as random variables for improving the performance of channel estimation. Therefore, the hidden variable set

becomes $\boldsymbol{\Theta}' = \{\mathbf{x}, \alpha, \mathbf{h}, \boldsymbol{\gamma}\}$. In the SBL framework, the unknown part of $\boldsymbol{\Phi}$ is estimated by calculating the probability density function (pdf) of $\mathbf{x}$. Since we use 4QAM, the pmf of $x[k,l]$ can be expressed as:

$$p(x[k,l]) = \begin{cases} p_1, & x[k,l] = \mathcal{S}_1 \\ p_2, & x[k,l] = \mathcal{S}_2 \\ p_3, & x[k,l] = \mathcal{S}_3 \\ p_4, & x[k,l] = \mathcal{S}_4, \end{cases} \tag{44}$$

where we assume that the four 4QAM symbols appear at equal probability, i.e. $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. Assuming that each $x[k,l]$ is independently and identically distributed, the pmf of vector $\mathbf{x}$ can be expressed as

$$p(\mathbf{x}) = \prod_{k=0}^{N-1} \prod_{l=0}^{M-1} p(x[k,l]), \tag{45}$$

where (k,l) denotes the range of data. The number of data symbols is $Q = MN - (M_p + 2l_{max})(N_p + 4k_{max})$. Hence, the *a priori* pmf of $\mathbf{x}$ is

$$p(\mathbf{x}) = \begin{cases} p'_1, & \mathbf{x} = \mathcal{X}_1 \\ p'_2, & \mathbf{x} = \mathcal{X}_2 \\ & \cdots \\ p'_{4^Q}, & \mathbf{x} = \mathcal{X}_{4^Q}, \end{cases} \tag{46}$$

where $p'_j = \frac{1}{4}^Q (j = 1, \ldots, 4^Q)$, $\mathcal{X}_j (j = 1, \ldots, 4^Q)$ is a legitimate vector of $\mathbf{x}$ after Q data symbols take values from $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$.

Hence, we can obtain a stable solution by alternately updating the probability functions as follows:

$$q^{(i+1)}(\mathbf{x}) \propto \exp\left(\left\langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \right\rangle_{q^{(i)}(\alpha) q^{(i)}(\mathbf{h}) q^{(i)}(\boldsymbol{\gamma})}\right), \tag{47}$$

$$q^{(i+1)}(\alpha) \propto \exp\left(\left\langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \right\rangle_{q^{(i+1)}(\mathbf{x}) q^{(i)}(\mathbf{h}) q^{(i)}(\boldsymbol{\gamma})}\right), \tag{48}$$

$$q^{(i+1)}(\mathbf{h}) \propto \exp\left(\left\langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \right\rangle_{q^{(i+1)}(\mathbf{x}) q^{(i+1)}(\alpha) q^{(i)}(\boldsymbol{\gamma})}\right), \tag{49}$$

$$q^{(i+1)}(\boldsymbol{\gamma}) \propto \exp\left(\left\langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \right\rangle_{q^{(i+1)}(\mathbf{x}) q^{(i+1)}(\alpha) q^{(i+1)}(\mathbf{h})}\right). \tag{50}$$

In the next subsection, we will provide the details of the update process (47)–(50).

B. Update Details for Data-Aided SBL

1) Update of $q(\mathbf{x})$: Based on the system model, the received signal can be expressed as (39). Hence, we have

$$\begin{aligned} q^{(i+1)}(\mathbf{x}) &\propto \exp\left\{\langle \ln p(\mathbf{y} \mid \mathbf{x}, \alpha, \mathbf{H}) \rangle_{q^{(i)}(\alpha) q^{(i)}(\mathbf{h})} + \ln p(\mathbf{x})\right\} \\ &\propto \exp\left\{\left\langle -\alpha \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \right\rangle_{q^{(i)}(\alpha) q^{(i)}(\mathbf{h})}\right\} p(\mathbf{x}) \\ &\propto \exp\left\{-(\mathbf{x} - \mathbf{u}_x)^H \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \mathbf{u}_x)\right\} p(\mathbf{x}) \\ &= \begin{cases} mn_1 p'_1, & \mathbf{x} = \mathcal{X}_1 \\ mn_2 p'_2, & \mathbf{x} = \mathcal{X}_2 \\ & \cdots \\ mn_{4^Q} p'_{4^Q}, & \mathbf{x} = \mathcal{X}_{4^Q}, \end{cases} \end{aligned} \tag{51}$$

where $\mathbf{u}_\mathbf{x} = \hat{\alpha}^{(i)} \boldsymbol{\Sigma}_\mathbf{x} \langle \mathbf{H} \rangle^H_{q^{(i)}(\mathbf{h})} \mathbf{y}$, $\boldsymbol{\Sigma}_\mathbf{x} = \left(\hat{\alpha}^{(i)} \langle \mathbf{H}^H \mathbf{H} \rangle_{q^{(i)}(\mathbf{h})}\right)^{-1}$, $m = 1/(n_1 p'_1 + n_2 p'_2 + \ldots + n_{4^Q} p'_{4^Q})$ and $n_j = \exp\left\{-(\mathcal{X}_j - \mathbf{u}_\mathbf{x})^H \boldsymbol{\Sigma}_\mathbf{x}^{-1} (\mathcal{X}_j - \mathbf{u}_\mathbf{x})\right\}$. The derivation of $\langle \mathbf{H} \rangle^H_{q^{(i)}(\mathbf{h})}$ and $\langle \mathbf{H}^H \mathbf{H} \rangle_{q^{(i)}(\mathbf{h})}$ is detailed in Appendix A.

Therefore, the mean and covariance matrix of vector $\mathbf{x}$ are

$$\mathbf{u}'_\mathbf{x} = \sum_{j=1}^{4^Q} mn_j p'_j \mathcal{X}_j, \tag{52}$$

$$\boldsymbol{\Sigma}'_\mathbf{x} = \sum_{j=1}^{4^Q} [\mathcal{X}_j - \mathbf{u}'_x]^2 p'_j. \tag{53}$$

Hence, we can obtain $\langle \boldsymbol{\Phi} \rangle_{q^{(i)}(\mathbf{x})}$, $\langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \rangle_{q^{(i)}(\mathbf{x})}$ and $\mathrm{tr}\left\langle \boldsymbol{\Phi} \boldsymbol{\Sigma}^{(i)}_\mathbf{h} \boldsymbol{\Phi}^H \right\rangle_{q^{(i)}(\mathbf{x})}$. The derivation is provided in Appendix B.

2) Update of $q(\alpha)$: Based on (18) and (19),we arrive at:

$$\begin{aligned} &\ln q^{(i+1)}(\alpha) \\ &\propto \langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \rangle_{q^{(i+1)}(\mathbf{x}) q^{(i)}(\mathbf{h}) q^{(i)}(\boldsymbol{\gamma})} \\ &\propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \alpha, \mathbf{h}) \rangle_{q^{(i+1)}(\mathbf{x}) q^{(i)}(\mathbf{h})} + \ln p(\alpha) \\ &\propto (a + Z - 1) \ln(\alpha) - \alpha \Big[b + \mathrm{tr}\left\langle \boldsymbol{\Phi} \boldsymbol{\Sigma}^{(i)}_\mathbf{h} \boldsymbol{\Phi}^H \right\rangle_{q^{(i+1)}(\mathbf{x})} + \mathbf{y}^H \mathbf{y} \\ &\quad - 2\,\mathrm{Re}(\mathbf{y}^H \langle \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_h) + (\mathbf{u}^{(i)}_\mathbf{h})^H \langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_\mathbf{h} \Big]. \end{aligned} \tag{54}$$

Compared to the update of $q(\alpha)$ in (25) in the previous section, we use $\langle \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})}$, $\langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})}$ and $\mathrm{tr}\left\langle \boldsymbol{\Phi} \boldsymbol{\Sigma}^{(i)}_\mathbf{h} \boldsymbol{\Phi}^H \right\rangle_{q^{(i+1)}(\mathbf{x})}$ to replace $\boldsymbol{\Phi}_p$, $\boldsymbol{\Phi}_p^H \boldsymbol{\Phi}_p$ and $\mathrm{tr}(\boldsymbol{\Phi}_\mathrm{p} \boldsymbol{\Sigma}^{(i)}_\mathbf{h} \boldsymbol{\Phi}^H_p)$. Hence, we have $a^{(i+1)}_\alpha = a + Z$, $b^{(i+1)}_\alpha = b + \mathrm{tr}\left\langle \boldsymbol{\Phi} \boldsymbol{\Sigma}^{(i)}_\mathbf{h} \boldsymbol{\Phi}^H \right\rangle_{q^{(i+1)}(\mathbf{x})} + \mathbf{y}^H \mathbf{y} - 2\,\mathrm{Re}(\mathbf{y}^H \langle \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_\mathbf{h}) + (\mathbf{u}^{(i)}_\mathbf{h})^H \langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_\mathbf{h}$, and α follows the Gamma distribution, yielding:

$$q^{(i+1)}(\alpha) = \mathrm{Gamma}\left(\alpha \mid a^{(i+1)}_\alpha, b^{(i+1)}_\alpha\right). \tag{55}$$

Thus, the mean of α can be expressed as

$$\hat{\alpha}^{(i+1)} = \langle \alpha \rangle_{q^{(i+1)}(\alpha)} = \frac{a^{(i+1)}_\alpha}{b^{(i+1)}_\alpha}. \tag{56}$$

3) Update of $q(\mathbf{h})$:

$$\begin{aligned} &\ln q^{(i+1)}(\mathbf{h}) \\ &\propto \langle \ln p(\mathbf{y}, \boldsymbol{\Theta}') \rangle_{q^{(i+1)}(\mathbf{x}) q^{(i+1)}(\alpha) q^{(i)}(\boldsymbol{\gamma})} \\ &\propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \alpha, \mathbf{h}) \rangle_{q^{(i+1)}(\mathbf{x}) q^{(i+1)}(\alpha)} + \langle \ln p(\mathbf{h} \mid \boldsymbol{\gamma}) \rangle_{q^{(i)}(\boldsymbol{\gamma})} \\ &\propto -\hat{\alpha}^{(i+1)} \Big(b + \mathbf{y}^H \mathbf{y} - 2\,\mathrm{Re}(\mathbf{y}^H \langle \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_\mathbf{h}) \\ &\quad + (\mathbf{u}^{(i)}_\mathbf{h})^H \langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{u}^{(i)}_\mathbf{h} \Big) \\ &\quad - \mathbf{h}^H \langle \mathrm{diag}(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}) \rangle_{q^i(\gamma)} \mathbf{h}. \end{aligned} \tag{57}$$

Therefore, $\mathbf{h}$ follows the Gaussian distribution, yielding:

$$q^{(i+1)}(\mathbf{h}) = \mathcal{CN}\left(\mathbf{h} \mid \mathbf{u}^{(i+1)}_\mathbf{h}, \boldsymbol{\Sigma}^{(i+1)}_\mathbf{h}\right), \tag{58}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{h}}^{(i+1)} = \Big(\hat{\alpha}^{(i+1)} \left\langle \boldsymbol{\Phi}^H \boldsymbol{\Phi} \right\rangle_{q^{(i+1)}(\mathbf{x})} + \left\langle \operatorname{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\right)\right\rangle_{q^{(i)}(\boldsymbol{\gamma})}\Big)^{-1}, \quad (59)$$

$$\mathbf{u}_{\mathbf{h}}^{(i+1)} = \hat{\alpha}^{(i+1)} \boldsymbol{\Sigma}_{\mathbf{h}}^{(i+1)} \left\langle \boldsymbol{\Phi}^H \right\rangle_{q^{(i+1)}(\mathbf{x})} \mathbf{y}. \quad (60)$$

4) Update of $q(\boldsymbol{\gamma})$: The update of $q(\boldsymbol{\gamma})$ is the same as that in the previous section.

Algorithm 1 The Proposed Joint Channel Estimation and Data Detection Algorithm

Input: Received signal $\mathbf{y}$, measurement matrix $\boldsymbol{\Phi}$, the precision of noise $\hat{\alpha}^{(0)}$, the mean of channel vector $\mathbf{u}_{\mathbf{h}}^{(0)}$, the covariance matrix of channel vector $\boldsymbol{\Sigma}_{\mathbf{h}}^{(0)}$ and the precision of channel vector $\boldsymbol{\gamma}^{(0)}$.

Initialization: Error $\zeta_1 = 10^{-6}$ for inner iteration, error $\zeta_2 = 10^{-6}$ for outer iteration, maximum iteration $N_{maxiter1} = 50$ and $N_{maxiter2} = 125$ inner iteration counter $i = 1$ and outer iteration counter $t = 1$.

repeat
  repeat
    Update $\mathbf{u}_{\mathbf{x}}^{\prime(i+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(i+1)}$ by (52) and (53) respectively.
    Update $\hat{\alpha}^{(i+1)}$ by (56), where the matrix $\boldsymbol{\Phi}$ is constructed by $\mathbf{u}_{\mathbf{x}}^{\prime(t)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t)}$.
    Update $\mathbf{u}_{\mathbf{h}}^{(i+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{h}}^{(i+1)}$ by (60) and (59) respectively, where the matrix $\boldsymbol{\Phi}$ is constructed by $\mathbf{u}_{\mathbf{x}}^{\prime(t)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t)}$.
    Update $\left\langle \operatorname{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\right)\right\rangle_{q^{(i+1)}(\boldsymbol{\gamma})}$ by (38).
    Set $i = i + 1$.
  until $\frac{\left\|\mathbf{u}_{\mathbf{x}}^{\prime(i+1)} - \mathbf{u}_{\mathbf{x}}^{\prime(i)}\right\|_2^2}{\left\|\mathbf{u}_{\mathbf{x}}^{\prime(i)}\right\|_2^2} \le \zeta_1$ or $i \ge N_{maxiter1}$
  Determine elements in $\mathbf{u}_{\mathbf{x}}^{\prime(i+1)}$ as $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4\}$ and assign pilot and guard values in $\mathbf{u}_{\mathbf{x}}^{\prime(i+1)}$.
  Update $\mathbf{u}_{\mathbf{x}}^{\prime(t+1)} = \mathbf{u}_{\mathbf{x}}^{\prime(i+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t+1)} = \boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(i+1)}$.
  Update $\hat{\alpha}^{(t+1)}$ by (56), where the matrix $\boldsymbol{\Phi}$ is constructed by $\mathbf{u}_{\mathbf{x}}^{\prime(t+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t+1)}$.
  Update $\mathbf{u}_{\mathbf{h}}^{(t+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{h}}^{(t+1)}$ by (60) and (59) respectively, where the matrix $\boldsymbol{\Phi}$ is constructed by $\mathbf{u}_{\mathbf{x}}^{\prime(t+1)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t+1)}$.
  Update $\left\langle \operatorname{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\right)\right\rangle_{q^{(t+1)}(\boldsymbol{\gamma})}$ by (38).
  Set $t = t + 1$.
until $\frac{\left\|\boldsymbol{\gamma}^{(t+1)} - \boldsymbol{\gamma}^{(t)}\right\|_2^2}{\left\|\boldsymbol{\gamma}^{(t)}\right\|_2^2} \le \zeta_2$ or $t \ge N_{maxiter2}$
Output: $\mathbf{u}_{\mathbf{x}}^{\prime(t+1)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t+1)}, \mathbf{u}_{\mathbf{h}}^{(t+1)}$

To summarize our joint channel estimation and data detection algorithm at a glance, we present it in **Algorithm 1**. We use (30),(34),(33) and (37) of Section III for the initial values $\hat{\alpha}^{(0)}$, $\mathbf{u}_{\mathbf{h}}^{(0)}$, $\boldsymbol{\Sigma}_{\mathbf{h}}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$. Furthermore, the known part of the initial measurement matrix $\boldsymbol{\Phi}$ is set to the pilot symbols' values and to the guard symbols' zero-values, while the unknown part is set to zero values. Since there are lots of unknown elements in $\mathbf{u}_{\mathbf{x}}^{\prime}$, $\mathbf{u}_{\mathbf{x}}^{\prime}$ cannot converge in a single iteration. We need multiple iterations, where the other hidden variables are updated with the aid of the measurement matrix $\boldsymbol{\Phi}$ constructed by $\mathbf{u}_{\mathbf{x}}^{\prime(t)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime(t)}$ obtained in the previous outer iteration. When the error $\frac{\left\|\mathbf{u}_{\mathbf{x}}^{\prime(i+1)} - \mathbf{u}_{\mathbf{x}}^{\prime(i)}\right\|_2^2}{\left\|\mathbf{u}_{\mathbf{x}}^{\prime(i)}\right\|_2^2}$ is lower than a certain threshold or the number of iterations reaches the maximum affordable value, the inner iterations end, as shown in step 2–10. Then we update $\hat{\alpha}$, $\mathbf{u}_{\mathbf{h}}$, $\boldsymbol{\Sigma}_{\mathbf{h}}$ and $\left\langle \operatorname{diag}\left(\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\right)\right\rangle_{q(\boldsymbol{\gamma})}$. When the error $\frac{\left\|\boldsymbol{\gamma}^{(t+1)} - \boldsymbol{\gamma}^{(t)}\right\|_2^2}{\left\|\boldsymbol{\gamma}^{(t)}\right\|_2^2}$ is less than a certain threshold or the number of iterations reaches the maximum, the outer iterations end.

C. Gaussian Approximation of $q(\mathbf{x})$

In the last subsection, we utilize the true pmf of $\mathbf{x}$ to update the probability functions. However, Q would become excessive when M and N take large values, which makes 4^Q too large to calculate $p(\mathbf{x})$. Therefore, we use the Gaussian distribution to approximate the pmf of $x[k,l]$ for simplifying the calculations [19].

We assume that $x[k,l]$ follows the zero-mean complex Gaussian distribution, i.e. $q(x[k,l]) \sim \mathcal{CN}(0, \sigma_x^2)$, where σ_x^2 is average symbol energy. Since every symbol is independent identically distributed, $q(\mathbf{x})$ is expressed as

$$q(\mathbf{x}) = \prod_{l=0}^{M-1} \prod_{k=0}^{N-1} p(x[k,l]). \quad (61)$$

Thus, the update of $q^{(i+1)}(\mathbf{x})$ can be formulated as

$$\begin{aligned} \ln q^{(i+1)}(\mathbf{x}) &\propto \left\langle \ln p(\mathbf{y} \mid \mathbf{H}, \alpha, \mathbf{x})\right\rangle_{q^{(i)}(\alpha) q^{(i)}(\mathbf{h})} + \ln p(\mathbf{x}) \\ &\propto \left\langle -\alpha \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \right\rangle_{q^{(i)}(\alpha) q^{(i)}(\mathbf{h})} + \ln p(\mathbf{x}) \\ &= -\left(\mathbf{x} - \mathbf{u}_{\mathbf{x}}\right)^H \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \left(\mathbf{x} - \mathbf{u}_{\mathbf{x}}\right) \\ &\quad + \sum_{l=0}^{M-1} \sum_{k=0}^{N-1} \left[\ln \pi \sigma_x^2 - \frac{x^2[k,l]}{\sigma_x^2}\right] \\ &\propto -(\mathbf{x} - \mathbf{u}_{\mathbf{x}}^{\prime})^H (\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime})^{-1} (\mathbf{x} - \mathbf{u}_{\mathbf{x}}^{\prime}), \end{aligned} \quad (62)$$

where

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{\prime} = \left(\hat{\alpha}^{(i)} \left\langle \mathbf{H}^H \mathbf{H} \right\rangle_{q^{(i)}(\mathbf{h})} + \frac{1}{\sigma_x^2} \mathbf{I}\right)^{-1}, \quad (63)$$

$$\mathbf{u}_{\mathbf{x}}^{\prime} = \hat{\alpha}^{(i)} \boldsymbol{\Sigma}_{\mathbf{x}}^{\prime} \left\langle \mathbf{H} \right\rangle_{q^{(i)}(\mathbf{h})}^H \mathbf{y}. \quad (64)$$

We use this approximation to replace the update of $q(\mathbf{x})$ in (51) for reducing the computational complexity. Our performance comparisons are discussed in Section V.

V. Simulation Results

In this section, we present our simulation results for investigating the performance of our proposed algorithms. The simulation parameters are as follows. The carrier frequency is $f_c = 137$MHz [34]. The maximum delay tap index is $l_{max} = 3$, the maximum Doppler tap index is $k_{max} = 3$ and the number of paths is $P = 4$. The dimension of data block is $M = 16$, $N = 16$. The dimension of pilots is

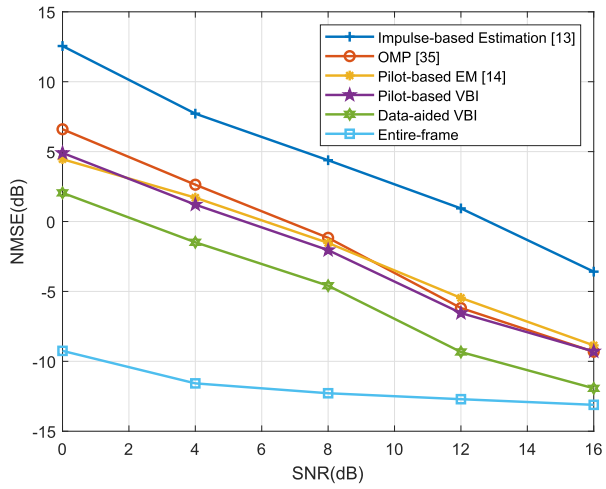


Fig. 5. The NMSE vs. SNR performance of channel estimation in the proposed OTFS system based on bi-orthogonal waveforms. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The delay dimension of pilots is $Mp = 2$. The Doppler dimension of pilots is $Np = 2$.

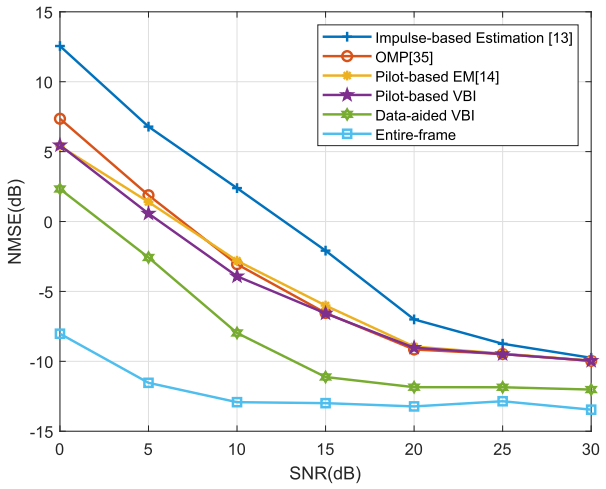


Fig. 6. The NMSE vs. SNR performance of channel estimation in our OTFS system based on rectangular waveforms. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The delay dimension of pilots is $Mp = 2$. The Doppler dimension of pilots is $Np = 2$.

$M_p = 2$, $N_p = 2$. We define the normalized mean squared error (NMSE) of the channel vector as

$$\mathrm{NMSE}_{\mathbf{h}} = \frac{\left\|\mathbf{h} - \hat{\mathbf{h}}\right\|_2^2}{\|\mathbf{h}\|_2^2}, \tag{65}$$

where $\hat{\mathbf{h}}$ denotes the estimate of $\mathbf{h}$.

As for the initial values of the VBI channel estimation algorithm, we assume that $\mathbf{u}_{\mathrm{h}}^0 = \boldsymbol{\Sigma}_h^0 \left(\boldsymbol{\Phi}_p\right)^H \mathbf{y}_p$, $\boldsymbol{\Sigma}_h^0 = \left(\left(\boldsymbol{\Phi}_p\right)^H \boldsymbol{\Phi}_{\mathrm{p}} + \mathbf{I}\right)^{-1}$, $a_l^0 = b_l^0 = 1$ [30]. The maximum error is $\frac{\left\|\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}\right\|_2^2}{\left\|\boldsymbol{\gamma}^{(i)}\right\|_2^2} < 10^{-5}$ or the maximum number of iterations is 125. The data-aided channel estimation algorithm's simulation parameters can be found in **Algorithm 1**.

A. Bi-orthogonal Waveform

Fig. 5 shows the NMSE performance of our OTFS channel estimator based on bi-orthogonal waveforms. The curves labelled as *Impulse-based Estimation* [13], *OMP* [35] and *Pilot-based EM* [14] characterize the NMSE performance of conventional methods. Since the threshold-based method of [13] requires a high pilot SNR, which is not consistent with our pilot arrangement, we simplify this method to the *Impulse-based Estimation* method relying on a single pilot symbol and take the P largest paths to estimate channel. The curve *Pilot-based VBI* shows the performance of the algorithm proposed in Section III, the curve *Data-aided VBI* shows the performance of the algorithm proposed in Section IV and the curve *Entire-frame* utilizes the entire frame (both the data and pilot are assumed to be known) to estimate channel which represents a performance upper bound for *Data-aided VBI*. In Fig. 5, we oberve that the *Pilot-based VBI* method has better NMSE performance than the *Impulse-based Estimation*, *OMP* and *Pilot-based EM* thanks to the reasonable *a priori* assumed by the SBL framework. Moreover, the *Data-aided VBI* proposed in this paper exhibits a substantial performance enhancement. This is because *Data-aided VBI* exploits the data symbols as 'virtual pilots', which amounts to harnessing more observed values for improving its NMSE performance.

As for complexity, the *Impulse-based Estimation* [13] has a complexity order of $\mathcal{O}(L)$ and *OMP* [35] has a complexity order of $\mathcal{O}\left(LZ^2P\right)$, where $L = (l_{max} + 1)(2k_{max} + 1)$, $Z = (l_{max} + M_p)(2k_{max} + N_p)$ and P is the number of propagation paths. The complexity of *Pilot-based EM* [14] is $\mathcal{O}\left(L^2 Z N_{EM}\right)$, where N_{EM} is the number of iterations for *Pilot-based EM*. The complexity of *Pilot-based VBI* is $\mathcal{O}\left(L^2 Z N_{VBI}\right)$, where N_{VBI} is the number of iterations for *Pilot-based VBI*. With regard to our *Data-aided VBI*, the complexity of channel estimation is $\mathcal{O}\left((M^2L^3)N_{outer}\right)$ and the complexity of data detection is $\mathcal{O}\left(M^3N^3N_{inner}N_{outer}\right)$, where N_{outer} is the maximum number of outer iterations and N_{inner} is the maximum number of inner iterations in **Algorithm 1**. Compared to other methods, *Data-aided VBI* has higher computational complexity, but it estimates the channel more accurately, hence yielding a better NMSE performance and detects $M \times N - (M_p + 2l_{max})(N_p + 4k_{max})$ symbols at the same time. These advantages make the computational complexity moderate.

B. Rectangular Waveform

Fig. 6 shows the NMSE performance of channel estimation in our OTFS system based on rectangular waveforms. We observe that the trend of curves is consistent with that in Fig. 5. Moreover, the *Data-aided VBI* proposed in this paper is close to its upper bound, i.e. to the performance of *Entire frame*, especially at high SNRs.

Fig. 7 shows the NMSE performance of channel estimation in our OTFS system based on rectangular waveforms, where the size of the OTFS frame is 32×32 and the size of pilots remains unchanged. This means that the pilot overhead is lower than that in Fig 6. We observe that the channel estimation algorithms only using pilots have better performance

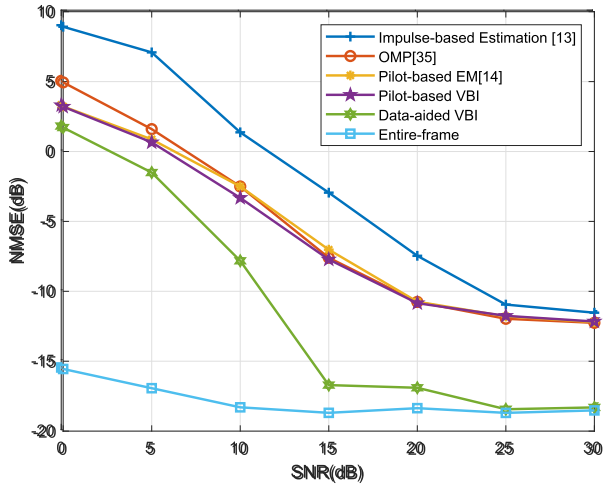


Fig. 7. The NMSE vs. SNR performance of channel estimation in our OTFS system based on rectangular waveforms. The delay dimension of data block is $M = 32$. The Doppler dimension of data block is $N = 32$. The delay dimension of pilots is $Mp = 2$. The Doppler dimension of pilots is $Np = 2$.

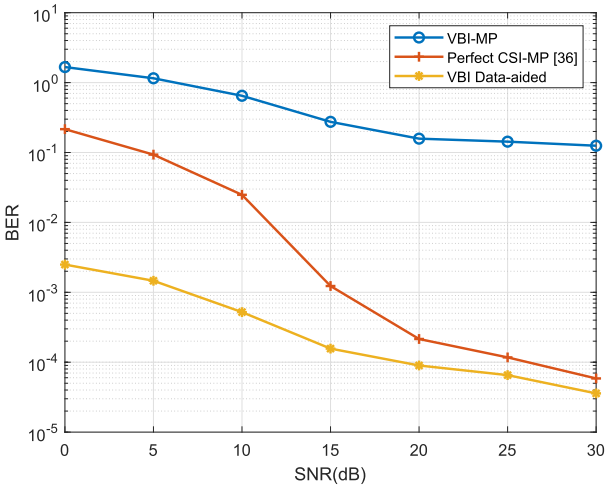


Fig. 9. The BER performance of different algorithms against SNR. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The Doppler dimension of pilots is $Np = 2$. The delay dimension of pilots is $Mp = 2$.

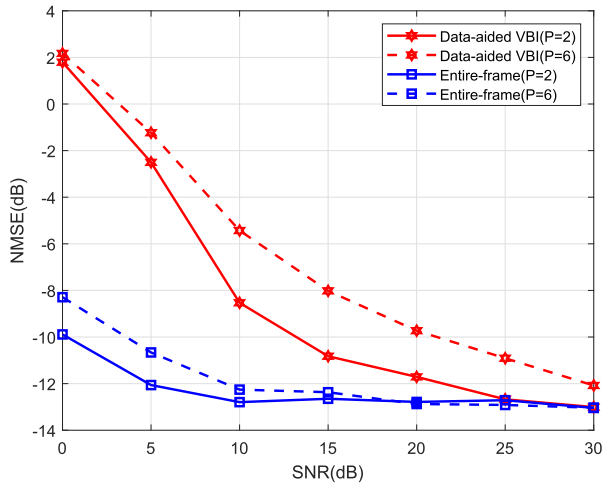


Fig. 8. The NMSE vs. SNR performance of channel estimation in our OTFS system with different path numbers. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The Doppler dimension of pilots is $Np = 2$. The delay dimension of pilots is $Mp = 2$. The path numbers are 2 and 6 respectively.

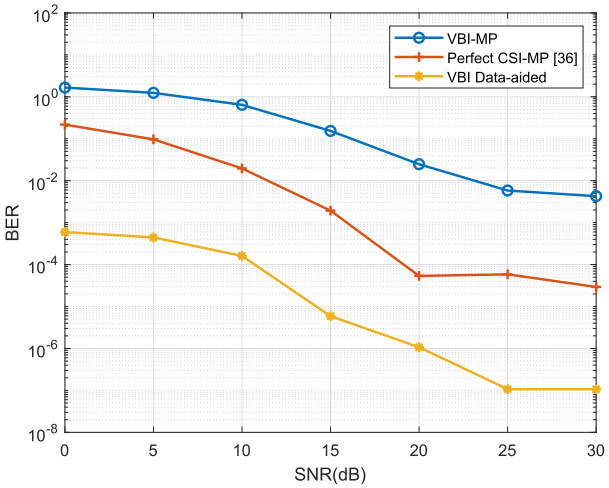


Fig. 10. The BER performance of different algorithms against SNR. The delay dimension of data block is $M = 32$. The Doppler dimension of data block is $N = 32$. The Doppler dimension of pilots is $Np = 2$. The delay dimension of pilots is $Mp = 2$.

at lower pilot overheads, especially at SNR$= 25$ dB, when the NMSE of *Pilot-based VBI* reduces to -12 dB. This is because as M and N become larger and the maximum Doppler as well as delay tap index are fixed, both the maximum Doppler shift of k_{max}/NT and the maximum delay of $l_{max}/M\Delta f$ become smaller. Furthermore, our proposed *Data-aided VBI* has an improved performance gain, because more observation values are harnessed.

Fig. 8 shows the impact of different number of paths on our channel estimator. We observe in Fig. 8 that as the number of paths increases, the NMSE is increased, because the number of observed values remains unchanged, while the number of parameters to be estimated is increased.

Fig. 9 and Fig. 10 show our BER performance comparisons of different algorithms for different OTFS frame sizes. The curve *VBI-MP* represents the performance of the algorithm utilizing the VBI method for channel estimation and the MP algorithm of [36] for data detection. The curve *Perfect CSI-MP* utilizes the MP algorithm of [36] for data detection using perfect channel state information (CSI). The curve *VBI Data-aided* represents our proposed algorithm of Section IV. We observe that the proposed algorithm has better BER performance than the MP algorithm.

Fig. 11 shows our NMSE performance comparison for a practical pmf and for the Gaussian approximation of data. We observe that the NMSE of the Gaussian approximation has only a modest performance erosion compared to that of the practical pmf scenario. This means that the Gaussian approximation is effective, despite reducing the computational complexity. Fig. 12 shows our BER performance comparison of the practical pmf and Gaussian approximation. We observe that the BER of the Gaussian approximation

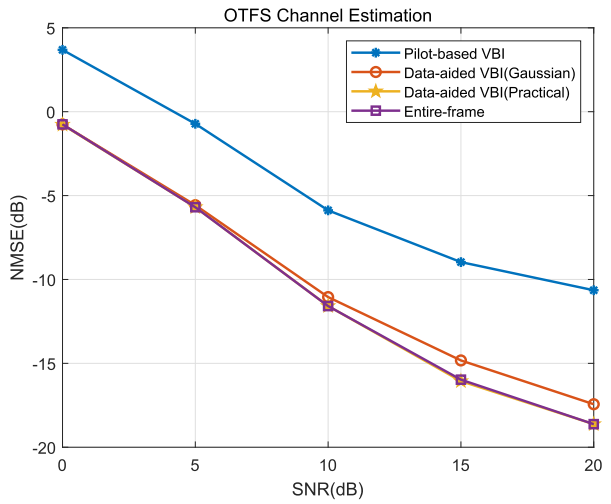


Fig. 11. The NMSE comparison of different data prior assumption. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The Doppler dimension of pilots is $Np = 3$. The delay dimension of pilots is $Mp = 2$.

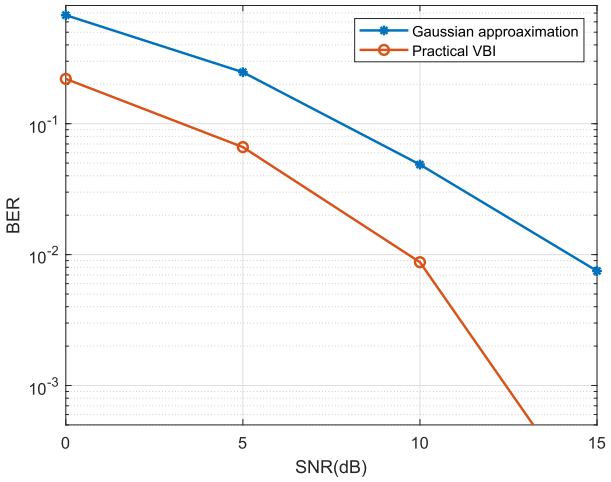


Fig. 12. The BER comparison of different data prior assumption. The delay dimension of data block is $M = 16$. The Doppler dimension of data block is $N = 16$. The Doppler dimension of pilots is $Np = 3$. The delay dimension of pilots is $Mp = 2$.

imposes a performance degradation compared to the practical pmf scenario, because the approximation also affects the data detection process. In other words, the BER performance may be degraded even if similar CSIs are used. Despite its performance reduction, the Gaussian approximation is efficient becasuse of its modest complexity.

VI. Conclusion

Joint channel estimation and data detection algorithms were developed for OTFS systems operating in high-mobility scenarios. Firstly, we adopted the VBI method for estimating the channel vector, which contains the channel gain, the delay and the Doppler. Secondly, we proposed a joint channel estimation and data detection algorithm by exploiting the data symbols as 'virtual pilots' to generate more measurement information. Our simulation results demonstrate that the pilot-based VBI method outperforms the existing OTFS channel estimators. Furthermore, our proposed data-aided VBI algorithm outperforms the pilot-based VBI method and its BER performance is better than that of the MP algorithm relying on perfect CSI.

Appendix A

In this part, we derive $\langle \mathbf{H} \rangle^H_{q^{(i)}(\mathbf{h})}$ and $\langle \mathbf{H}^H \mathbf{H} \rangle_{q^{(i)}(\mathbf{h})}$.

Based on (39), we have

$$\begin{aligned}\langle \mathbf{H} \rangle^H_{q^{(i)}(\mathbf{h})} &= \left\langle \sum_{i'=1}^{P} h_{i'} \mathbf{P}^{(i')} \mathbf{Q}^{(i')} \right\rangle^H_{q^{(i)}(\mathbf{h})} \\ &= \sum_{i'=1}^{P} \langle h_{i'} \rangle^*_{q^{(i)}(\mathbf{h})} \left(\mathbf{P}^{(i')} \mathbf{Q}^{(i')} \right)^H \\ &= \sum_{i'=1}^{P} \langle h_{i'} \rangle^*_{q^{(i)}(\mathbf{h})} \left((\mathbf{Q}^{(i')})^H (\mathbf{P}^{(i')})^H \right), \end{aligned} \tag{66}$$

where $h_{i'}$ is the channel gain of the i'-th path, which is obtained from $\mathbf{h}$.

$$\begin{aligned}&\left\langle \mathbf{H}^H \mathbf{H} \right\rangle_{q^{(i)}(\mathbf{h})} \\ &= \left\langle \left(\sum_{i'=1}^{P} h_i \mathbf{P}^{(i')} \mathbf{Q}^{(i')} \right)^H \left(\sum_{i'=1}^{P} h_i \mathbf{P}^{(i')} \mathbf{Q}^{(i')} \right) \right\rangle_{q^{(i)}(\mathbf{h})} \\ &= \left\langle \left(\sum_{i'=1}^{P} h_i \mathbf{T}^{(i')} \right)^H \left(\sum_{i'=1}^{P} h_{i'} \mathbf{T}^{(i')} \right) \right\rangle_{q^{(i)}(\mathbf{h})} \\ &= \left\langle \sum_{i'=1}^{P} \sum_{j'=1}^{P} (h^*_{i'} (\mathbf{T}^{(i')})^H)(h_{j'} (\mathbf{T}^{(j')})) \right\rangle_{q^{(i)}(\mathbf{h})} \\ &= \sum_{i'=1}^{P} \sum_{j'=1}^{P} \left\langle (h^*_{i'} (\mathbf{T}^{(i')})^H)(h_{j'} (\mathbf{T}^{(j')})) \right\rangle_{q^{(i)}(\mathbf{h})} \\ &= \sum_{i'=1}^{P} \sum_{j'=1}^{P} \langle h^*_{i'} h_{j'} \rangle_{q^{(i)}(\mathbf{h})} (\mathbf{T}^{(i')})^H \mathbf{T}^{(j')}, \end{aligned} \tag{67}$$

where we have

$$\langle h^*_{i'} h_{j'} \rangle = \begin{cases} \langle h^*_{i'} \rangle \langle h_{j'} \rangle, & i' \neq j' \\ \mathbf{\Sigma}_{\mathbf{h}(i',i')} + \left| \mathbf{u}_{\mathbf{h}(i')} \right|^2, & i' = j' \end{cases} \tag{68}$$

and $\mathbf{\Sigma}_{\mathbf{h}(i',i')}$ denotes the (i',i')-th diagonal element of $\mathbf{\Sigma}_{\mathbf{h}}$ and $\mathbf{u}_{\mathbf{h}(i')}$ represents the i'-th element of $\mathbf{u}_{\mathbf{h}}$.

Appendix B

In this part, we derive $\langle \mathbf{\Phi} \rangle_{q^{(i)}(\mathbf{x})}$, $\left\langle \mathbf{\Phi}^H \mathbf{\Phi} \right\rangle_{q^{(i)}(\mathbf{x})}$ and $\mathrm{tr} \left\langle \mathbf{\Phi} \mathbf{\Sigma}^{(i)}_h \mathbf{\Phi}^H \right\rangle_{q^{(i)}(\mathbf{x})}$.

Based on the expression of $\mathbf{\Phi}$, we have

$$\mathbf{\Phi} = \mathcal{X} \mathcal{E}, \tag{69}$$

where $\mathcal{X}$ and $\mathcal{E}$ are given in (70) and (71) respectively as shown at the top of the next page.

Therefore, we have

$$\begin{aligned}\langle \mathbf{\Phi} \rangle_{q^{(i)}(\mathbf{x})} &= \langle \mathcal{X} \mathcal{E} \rangle_{q^{(i)}(\mathbf{x})} \\ &= \langle \mathcal{X} \rangle_{q^{(i)}(\mathbf{x})} \mathcal{E}, \end{aligned} \tag{72}$$

$$\begin{aligned}
\mathcal{X} &= \left[\mathbf{X}(0)^T, \mathbf{X}(1)^T, \ldots, \mathbf{X}(N-1)^T\right]^T \in \mathbb{C}^{MN \times M(l_{max}+1)(2k_{max}+1)}, \qquad (70)\\
\mathbf{X}(k) &= \operatorname{diag}\left[\mathbf{x}(k,0), \mathbf{x}(k,1), \ldots, \mathbf{x}(k,M-1)\right], k = 0,1,\ldots,N-1,\\
\mathbf{x}(k,l) &= \left[x\left([k-(-k_{max})]_N, [l-0]_M\right), \ldots x\left([k-k_{max}]_N, [l-l_{max}]_M\right)\right],\\
\mathcal{E} &= \left[\mathbf{EXP}(0)^T, \mathbf{EXP}(1)^T, \ldots, \mathbf{EXP}(M-1)^T\right]^T \in \mathbb{C}^{M(l_{max}+1)(2k_{max}+1)\times(l_{max}+1)(2k_{max}+1)},\\
\mathbf{EXP}(l) &= \operatorname{diag}\left(e^{j2\pi(l-0)(-k_{max})}, \ldots, e^{j2\pi(l-l_{max})(-k_{max})}, \ldots, e^{j2\pi(l-0)k_{max}}, \ldots, e^{j2\pi(l-l_{max})k_{max}}\right),\\
&\quad l = 0,1,\ldots,M-1. \qquad (71)
\end{aligned}$$

$$\begin{aligned}
\left\langle \mathcal{X}^H\mathcal{X}\right\rangle_{q^{(i)}(\mathbf{x})} &= \left\langle \left[\mathbf{X}(0)^H, \mathbf{X}(1)^H, \ldots, \mathbf{X}(N-1)^H\right]\left[\mathbf{X}(0)^T, \mathbf{X}(1)^T, \ldots, \mathbf{X}(N-1)^T\right]^T\right\rangle_{q^{(i)}(\mathbf{x})}\\
&= \left\langle \sum_{k=0}^{N-1}\mathbf{X}(k)^H\mathbf{X}(k)\right\rangle_{q^{(i)}(\mathbf{x})}\\
&= \sum_{k=0}^{N-1}\left\langle \begin{bmatrix}\mathbf{x}(k,0)^H & & \\ & \ddots & \\ & & \mathbf{x}(k,M-1)^H\end{bmatrix}\begin{bmatrix}\mathbf{x}(k,0) & & \\ & \ddots & \\ & & \mathbf{x}(k,M-1)\end{bmatrix}\right\rangle_{q^{(i)}(\mathbf{x})}\\
&= \sum_{k=0}^{N-1}\begin{bmatrix}\left\langle \mathbf{x}(k,0)^H\mathbf{x}(k,0)\right\rangle_{q^{(i)}(\mathbf{x})} & & \\ & \ddots & \\ & & \left\langle \mathbf{x}(k,M-1)^H\mathbf{x}(k,M-1)\right\rangle_{q^{(i)}(\mathbf{x})}\end{bmatrix}. \qquad (74)
\end{aligned}$$

where $\langle\mathcal{X}\rangle_{q^{(i)}(\mathbf{x})}$ can be obtained from $\mathbf{u}_x$. Furthermore, we have

$$\begin{aligned}
\left\langle \mathbf{\Phi}^H\mathbf{\Phi}\right\rangle_{q^{(i)}(\mathbf{x})} &= \left\langle (\mathcal{X}\mathcal{E})^H(\mathcal{X}\mathcal{E})\right\rangle_{q^{(i)}(\mathbf{x})}\\
&= \mathcal{E}^H\left\langle \mathcal{X}^H\mathcal{X}\right\rangle_{q^{(i)}(\mathbf{x})}\mathcal{E}, \qquad (73)
\end{aligned}$$

where $\left\langle \mathcal{X}^H\mathcal{X}\right\rangle_{q^{(i)}(\mathbf{x})}$ is given in (74), as shown at the top of the page.

The diagonal elements in $\left\langle \mathbf{x}(k,l)^H\mathbf{x}(k,l)\right\rangle_{q^{(i)}(\mathbf{x})}$ can be obtained from the corresponding elements in $\|\mathbf{u}_\mathbf{x}\|_2^2 + \mathbf{\Sigma}_\mathbf{x}$ and the other elements can be obtained from the product of the means of $x[k,l]$. Our experimental results demonstate that $\mathbf{\Sigma}_\mathbf{x} \approx \mathbf{0}$, thus $\mathbf{\Sigma}_\mathbf{x}$ can be neglected.

As for $\operatorname{tr}\left\langle \mathbf{\Phi}\mathbf{\Sigma}_\mathbf{h}^{(i)}\mathbf{\Phi}^H\right\rangle_{q^{(i)}(\mathbf{x})}$, we exploit the following property of trace, i.e.

$$\begin{aligned}
\operatorname{tr}\left\langle \mathbf{\Phi}\mathbf{\Sigma}_\mathbf{h}^{(i)}\mathbf{\Phi}^H\right\rangle_{q^{(i)}(\mathbf{x})} &= \operatorname{tr}\left\langle \mathbf{\Phi}^H\mathbf{\Phi}\mathbf{\Sigma}_\mathbf{h}^{(i)}\right\rangle_{q^{(i)}(\mathbf{x})}\\
&= \operatorname{tr}\left\langle \mathbf{\Phi}^H\mathbf{\Phi}\right\rangle_{q^{(i)}(\mathbf{x})}\mathbf{\Sigma}_\mathbf{h}^{(i)}, \qquad (75)
\end{aligned}$$

where $\left\langle \mathbf{\Phi}^H\mathbf{\Phi}\right\rangle_{q^{(i)}(\mathbf{x})}$ has been obtained.

REFERENCES

[1] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1851–1865, Aug. 2020.

[2] I. Leyva-Mayorga *et al.*, "LEO small-satellite constellations for 5G and beyond-5G communications," *IEEE Access*, vol. 8, pp. 184955–184964, 2020.

[3] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, "Broadband LEO satellite communications: Architectures and key technologies," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 55–61, Apr. 2019.

[4] Y. Zhang, Y. Wu, A. Liu, X. Xia, T. Pan, and X. Liu, "Deep learning-based channel prediction for LEO satellite massive MIMO communication system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1835–1839, Aug. 2021.

[5] F. Pu, J. Gong, and L. Gan, "Improved channel estimation algorithm for OFDM over LEO channels," in *Proc. IEEE Int. Symp. Microw., Antenna, Propag. EMC Technol. Wireless Commun.*, Aug. 2005, pp. 1139–1142.

[6] Z. Zhang *et al.*, "User activity detection and channel estimation for grant-free random access in LEO satellite-enabled Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8811–8825, Sep. 2020.

[7] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications With MATLAB*. Singapore: Wiley, 2010.

[8] X. Cai and G. B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2047–2056, Dec. 2003.

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

[10] R. Hadani *et al.*, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[11] Z. Wei *et al.*, "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 136–144, Aug. 2021.

[12] K. Murali and A. Chockalingam, "On OTFS modulation for high-Doppler fading channels," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–10.

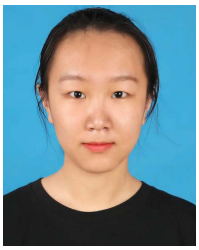
[13] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels," *IEEE Trans. Veh. Tech.*, vol. 68, no. 5, pp. 4906–4917, May 2019.

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

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

[16] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1994–2009, Sep. 2020.

[17] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.
[18] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5369–5382, Oct. 2015.
[19] Y. Ding, S.-E. Chiu, and B. D. Rao, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog–digital processing and low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 499–513, Jun. 2018.
[20] P. Zhang, S. Chen, and L. Hanzo, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1902–1913, May 2013.
[21] A. Mishra, A. K. Jagannatham, and L. Hanzo, "Sparse Bayesian learning-aided joint sparse channel estimation and ML sequence detection in space-time trellis coded MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1132–1145, Feb. 2020.
[22] W. Yuan, S. Li, Z. Wei, J. Yuan, and D. W. K. Ng, "Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1954–1958, Sep. 2021.
[23] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[24] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[25] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.
[26] C. W. Fox and S. J. Roberts, "A tutorial on variational Bayesian inference," *Artif. Intell. Rev.*, vol. 38, no. 2, pp. 85–95, Aug. 2012.
[27] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.
[28] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems*. 2nd ed. New York, NY, USA: Wiley, 1999.
[29] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Tech.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.
[30] J. Dai, A. Liu, and H. C. So, "Non-uniform burst-sparsity learning for massive MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1075–1087, Feb. 2019.
[31] J. Dai, A. Liu, and V. K. N. Lau, "Joint channel estimation and user grouping for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 622–637, Feb. 2019.
[32] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.
[33] S. S. Thoota and C. R. Murthy, "Variational bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.
[34] C. M. Rush, "How WARC '92 will affect mobile services," *IEEE Commun. Mag.*, vol. 30, no. 10, pp. 90–96, Oct. 1992.
[35] Y. C. Pati, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 1993, pp. 40–44.
[36] M. K. Ramachandran and A. Chockalingam, "MIMO-OTFS in high-Doppler fading channels: Signal detection and channel estimation," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 206–212.



Xueyang Wang received the B.S. degree in electronic engineering from the Beijing Institute of Technology, Beijing, China, in 2020, where she is currently pursuing the Ph.D. degree with the School of Information and Electronics. Her main research interests include channel estimation, orthogonal time frequency space (OTFS) modulation, and compressive sensing.

Wenqian Shen received the B.S. degree from Xi'an Jiaotong University, Shaanxi, China, in 2013, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2018. She is currently an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing. She has published several journals articles and conference papers in IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE ICC. Her research interests include massive MIMO and mmWave/THz communications. She has won the IEEE Best Paper Award from the IEEE ICC 2017.

Chengwen Xing (Member, IEEE) received the B.E. degree from Xidian University, Xi'an, China, in 2005, and the Ph.D. degree from The University of Hong Kong, Hong Kong, in 2010. From September 2012 to December 2012, he was a Visiting Scholar with the University of Macau, Macau. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor. His current research interests include machine learning, statistical signal processing, convex optimization, multivariate statistics, array signal processing, and high-frequency band communication systems.

Jianping An (Senior Member, IEEE) received the B.Eng. degree from PLA Information Engineering University, Zhengzhou, China, in 1987, and the Ph.D. degree from the Beijing Institute of Technology (BIT), Beijing, China, in 1996. From 2010 to 2011, he was a Visiting Professor with the University of California, San Diego, CA, USA. He is currently a Professor and the Dean of the School of Cyberspace Science and Technology, BIT. He has authored or coauthored more than 150 journals articles and conference papers, and holds 37 patents. His current research interests include signal processing theory and algorithms for communication systems. He was a recipient of a various awards for his academic achievements and the resultant industrial influences, including the National Award for Scientific and Technological Progress of China in 1997 and the Excellent Young Teacher Award by the China's Ministry of Education in 2000. Since 2010, he has been a Chief Reviewing Expert for the Information Technology Division, National Scientific Foundation of China.

Lajos Hanzo (Life Fellow, IEEE) received the Ph.D. degree from the Technical University (TU) of Budapest, the D.Sc. degree from the University of Southampton, U.K., in 2004, and the Honorary Doctorate degree from the University of Edinburgh in 2015. Since 1998, he has been the Chair of telecommunications at the University of Southampton, where he has directed the research of wireless communications. He has published widely, including frontier research on paving the way from classical communications to quantum communications. He has coauthored more than 2000 contributions at IEEE Xplore and 19 Wiley-IEEE PRESS monographs and nurtured 125 Ph.D. students. He is a fellow of the Royal Academy of Engineering, IET, and EURASIP. He is also a Foreign Member of the Hungarian Academy of Sciences and a Former Editor-in-Chief of the IEEE PRESS.