TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 16/20 pp1821–1828
DOI: 10.26599/TST.2023.9010093
Volume 29, Number 6, December 2024

Sparse Bayesian Learning Based Off-Grid Estimation of OTFS Channels with Doppler Squint

Xuehan Wang, Xu Shi, and Jintao Wang*

Abstract: Orthogonal Time Frequency Space (OTFS) modulation has exhibited significant potential to further promote the performance of future wireless communication networks especially in high-mobility scenarios. In practical OTFS systems, the subcarrier-dependent Doppler shift which is referred to as the Doppler Squint Effect (DSE) plays an important role due to the assistance of time-frequency modulation. Unfortunately, most existing works on OTFS channel estimation ignore DSE, which leads to severe performance degradation. In this letter, OTFS systems taking DSE into consideration are investigated. Inspired by the input-output analysis with DSE and the embedded pilot pattern, the sparse Bayesian learning based parameter estimation scheme is adopted to recover the delay-Doppler channel. Simulation results verify the excellent performance of the proposed off-grid estimation approach considering DSE.

Key words: orthogonal time frequency space modulation; Doppler squint effect; channel estimation; sparse Bayesian learning

1 Introduction

Orthogonal Time Frequency Space (OTFS) modulation has been regarded as a promising candidate to promote the reliability and capacity when it comes to the wireless communication in high-mobility scenarios[1]. By processing the data and pilot symbols in the delay-Doppler domain, full diversity over time and frequency can be utilized for each symbol, which helps mitigate the doubly-selective fading caused by the multipath channel and high mobility. So far, substantial work has been devoted to OTFS modulation to promote the performance of high-mobility communication systems[2–14].

Nevertheless, the acquisition of the wideband time-variant Channel State Information (CSI) remains a core problem due to the large dimension and fast variation. The threshold-based method was proposed in Ref. [4] to estimate the delay-Doppler channel coefficients directly, which has been proved to perform worse than the parameter estimation-based techniques due to the sparse multipath property of the channel. For example, an efficient approximated maximum likelihood algorithm was proposed in Ref. [5] and the corresponding Cramér-Rao lower bound was derived to verify the performance. Meanwhile, inspired by the development of off-grid Compressed Sensing (CS) methods, the Sparse Bayesian Learning (SBL) based approach was developed in Ref. [7] while the Message Passing (MP) based scheme was adopted in Ref. [11], where the fractional Doppler can be taken into consideration.

Most existing works on OTFS channel estimation[4–12] are developed from the input-output analysis in Ref. [2], where the Doppler shift is assumed to be frequency-independent. However, as indicated in Refs. [13, 15], non-negligible Doppler difference across the bandwidth exists in wideband systems, which is referred to as the Doppler Squint Effect

• Xuehan Wang, Xu Shi, and Jintao Wang are with Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: wang-xh21@mails.tsinghua.edu.cn; shi-x19@mails.tsinghua.edu.cn; wangjintao@tsinghua.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2023-05-06; revised: 2023-08-23; accepted: 2023-09-02

(DSE). In OTFS systems, the Doppler shift of each path varies within the bandwidth, and the frequency-dependent offset caused by DSE will be accumulated through a much longer time duration than Orthogonal Frequency Division Multiplexing (OFDM), which causes severe performance degradation if ignoring DSE[13]. Actually, DSE is caused by the time-variant delay in the baseband waveform. Reference [16] proposed to approximate the model of DSE by ignoring the impact on pulse-shaping, which does not hold due to the finite support of the pulses. On the other hand, the authors in Ref. [17] avoided the delay-Doppler modeling by multiplexing the symbols in the Mellin-Fourier domain and carrying out the scale-delay signal extraction, which is known as the Orthogonal Delay Scale Space (ODSS) modulation. However, the wideband cross-ambiguity with high resolution is not practical for typical wireless communications, it is more appropriate to adjust the OTFS characterization and system design based on the narrowband cross-ambiguity. Reference [13] considered the precise characterization of DSE in OTFS systems, however, the whole OTFS frame is employed to estimate the channel parameters and only integer Doppler can be extracted. It is impractical for realistic system design, which inspires us to reconsider the channel estimation and provide schemes with higher transmission efficiency and lower estimation loss.

In order to attain accurate CSI with less pilot overhead for practical OTFS systems with DSE and fractional Doppler, the OTFS system with DSE is investigated in this letter. Inspired by the embedded pilot[4, 11] and the input-output analysis in Ref. [13], an off-grid SBL based scheme is proposed to execute the parameter extraction and delay-Doppler CSI acquisition. Though substantial work has been devoted to OTFS channel estimation with respect to Bayesian frameworks[7–12], the pilot insertion and corresponding scheme details require more elaborate consideration due to the impact of DSE, which serves as the major contribution of this paper. Simulation results confirm the performance superiority of the proposed estimation approach considering DSE and fractional Doppler.

Notations: $\mathcal{A}$ is a set, $\boldsymbol{A}$ is a matrix, $\boldsymbol{a}$ is a column vector, a is a scalar. $\boldsymbol{A}^{\mathrm{T}}$, $\boldsymbol{A}^{\mathrm{H}}$, and $\boldsymbol{A}^{-1}$ denote its transposition, conjugate transposition, and inverse, respectively. $\boldsymbol{A}_{ij}$ and $\boldsymbol{A}_i$ are the (i, j) component and the i-th column of $\boldsymbol{A}$, respectively, while $\boldsymbol{a}_j$ represents the j-th element of vector $\boldsymbol{a}$. $\|\boldsymbol{a}\|$ denotes the l_2-norm of $\boldsymbol{a}$ and $\|\boldsymbol{A}\|_{\mathrm{F}}$ represents the Frobenius norm of $\boldsymbol{A}$. $(\cdot)^*$ denotes the conjugate operation, while $\mathrm{Re}\{\cdot\}$ returns the real part of the complex input. Random vector $\boldsymbol{x}$ obeying complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, whose Probability Density Function (PDF) is $\mathcal{CN}(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The PDF for Gamma distribution is defined as $\Gamma(x \mid a, b) = \dfrac{b^a x^{a-1} \mathrm{e}^{-bx}}{\Gamma(a)}$, where $\Gamma(\cdot)$ is the Gamma function. $\odot$ represents the point-wise Hadamard product. $\boldsymbol{u}_{-n}$ represents $\boldsymbol{u}$ without the n-th entry for vector $\boldsymbol{u}$. Finally, $\boldsymbol{I}_{\mathcal{A}}(x)$ is the indicator function for $x \in \mathcal{A}$.

2 System Model

In this section, the wideband OTFS system model is investigated. Instead of directly employing the input-output relationship offered in Refs. [2, 5], the analysis in Ref. [13] is adopted where DSE is taken into account to characterize the multipath channel more accurately. f_c and Δf denote the carrier frequency and subcarrier spacing, respectively.

2.1 OTFS transmitter and receiver

At the transmitter, a bit sequence is mapped to symbols as $\{x[k, l] \mid k = 0, 1, \ldots, N-1 \text{ and } l = 0, 1, \ldots, M-1\}$ in the discretized delay-Doppler domain, M and N represent the number of subcarriers and time slots, respectively, and k and l denote the Doppler and delay index, respectively. $x[k, l]$ is then converted into time-frequency domain symbols $X[n, m]$ by executing the Inverse Symplectic Finite Fourier Transform (ISFFT), we have

$$X[n, m] = \frac{1}{\sqrt{NM}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k, l] \cdot \mathrm{e}^{\mathrm{j}2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{1}$$

for $n = 0, 1, \ldots, N-1$ and $m = 0, 1, \ldots, M-1$, n and m are the time and subcarrier index, respectively. The Heisenberg transform employing the pulse $g_{\mathrm{tx}}(t)$ is then performed to create the continuous baseband transmitted waveform $s(t)$ as

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

where T is the duration of one time slot and we have $T\Delta f = 1$.

At the receiver, the received baseband signal $r(t)$ is processed by the Wigner transform to obtain $Y[n, m]$ as

$$Y[n,m] = \int g_{\mathrm{rx}}^{*}(t-nT)\cdot r(t)\cdot \mathrm{e}^{-\mathrm{j}2\pi m\Delta f t}\mathrm{d}t \tag{3}$$

The symplectic finite Fourier transform is then performed to attain delay-Doppler symbols as

$$y[k,l] = \frac{1}{\sqrt{NM}}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} Y[n,m]\cdot \mathrm{e}^{-\mathrm{j}2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)} \tag{4}$$

2.2 OTFS input-output analysis with DSE

At the receiver, the baseband signal $r(t)$ received via N_P paths can be derived as

$$r(t) = \sum_{i=1}^{N_P}\widetilde{\beta_i}\cdot \mathrm{e}^{-\mathrm{j}2\pi\tau_i f_c}\cdot \mathrm{e}^{\mathrm{j}2\pi\nu_i t}\cdot s\left(t-\left(\tau_i-\frac{\nu_i}{f_c}t\right)\right) \tag{5}$$

where $\widetilde{\beta_i}$, τ_i, and ν_i denote the attenuation, propagation delay and the Doppler shift at the carrier frequency for the i-th path, respectively. If the frame duration is small or the mobility is slow enough, then $\tau_i-(\nu_i/f_c)t\approx\tau_i$ holds true within a frame duration $T_s = NT$, which deduces the conventional sparse multipath channel model[2] in the delay-Doppler domain as $h(\tau,\nu)=\sum_{i=1}^{N_P}\beta_i\cdot\delta(\tau-\tau_i)\cdot\delta(\nu-\nu_i)$ by denoting $\beta_i=\widetilde{\beta_i}\cdot \mathrm{e}^{-\mathrm{j}2\pi\tau_i f_c}$. However, the offset $(\nu_i/f_c)t$ is non-negligible in OTFS channels due to the much longer frame duration than that of traditional OFDM system. As a result, if the time-variant frequency response $H(t,f)$ is adopted to characterize the equivalent channel as $r(t)=\int H(t,f)\cdot S(f)\cdot \mathrm{e}^{\mathrm{j}2\pi f t}\mathrm{d}t$, we have

$$H(t,f) = \sum_{i=1}^{N_P}\beta_i\cdot \mathrm{e}^{\mathrm{j}2\pi\frac{\nu_i}{f_c}(f_c+f)t}\cdot \mathrm{e}^{-\mathrm{j}2\pi\tau_i f} \tag{6}$$

From Eq. (6), it is clear that the Doppler frequency of each path is frequency-dependent as $\nu_i(f_c+f)/f_c$, which is referred to as DSE.

The practical rectangular pulses are employed in this letter, where we have $g_{\mathrm{rx}}(t)=g_{\mathrm{tx}}(t)=\frac{1}{\sqrt{T}}I_{[0,T]}(t)$. $\tau_i = l_i/MT$, $\nu_i = k_i\Delta f/N$, $p_i = f_c/\nu_i$, and $\beta_i' = \beta_i\cdot \mathrm{e}^{\mathrm{j}2\pi\tau_i\nu_i}$ are adopted for ease of illustration, where l_i is positive integer by assuming wideband system design, i.e., the delay resolution $1/(M\Delta f)$ is sufficient to approximate the path delays to the nearest sampling points. k_i is not necessarily integer. The input-output relationship in the delay-Doppler domain can be characterized[13] as

$$y[k,l] = \sum_{i=1}^{N_P}\sum_{k'=0}^{N-1}\sum_{l'=0}^{M-1}\beta_i'\cdot\psi_{k,l}^{i}[k',l']\cdot x[k',l'] + w[k,l] \tag{7}$$

where $w[k,l]\sim \mathcal{CN}(0,\sigma^2)$ are the white Gaussian noise, and $\psi_{k,l}^{i}[k',l']$ can be derived as Formula (10), shown at the bottom of this page. The index sets are defined as

$$\mathcal{L}_{\mathrm{ISI}}^{i} = \begin{cases}\{l'\in\mathbf{N}\mid M-l_i+1\leqslant l'\leqslant M-1\}, & p_i>0;\\ \{l'\in\mathbf{N}\mid M-l_i\leqslant l'\leqslant M-1\}, & p_i<0\end{cases} \tag{8}$$

and

$$\mathcal{L}_{\mathrm{ICI}}^{i} = \begin{cases}\{l'\in\mathbf{N}\mid 0\leqslant l'\leqslant M-l_i\}, & p_i>0;\\ \{l'\in\mathbf{N}\mid 0\leqslant l'\leqslant M-l_i-1\}, & p_i<0\end{cases} \tag{9}$$

where $\mathcal{L}_{\mathrm{ICI}}^{i}$ is derived from the relation between $Y[n,m]$ and $X[n,m]$, while $\mathcal{L}_{\mathrm{ISI}}^{i}$ embodies the property between $Y[n,m]$ and $X[n-1,m]$. The explicit definition of $\mathcal{L}_{\mathrm{ICI}}^{i}$ and $\mathcal{L}_{\mathrm{ISI}}^{i}$ varies with the mobility direction, which is different from prior analysis in Ref. [5] neglecting DSE. The approximation of Formula (10) is determined by the small value of NM/p_i. For example, when $N=128$, $M=1024$, while $v=500$ km/h, NM/p_i is less than 0.07, while the approximated error of Formula (10) is approximately bounded by 5×10^{-4}[13]. The input-output relation in Eq. (7) can also be reformulated as $\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{w}$, where we have $\boldsymbol{y}_{lN+k}=y[k,l]$, $\boldsymbol{x}_{lN+k}=x[k,l]$, $\boldsymbol{H}_{lN+k,l'N+k'}=\sum_{i=1}^{N_P}\beta_i'\cdot\psi_{k,l}^{i}[k',l']$, and $\boldsymbol{w}_{lN+k}=w[k,l]$.

$$\psi_{k,l}^{i}[k',l'] \approx \mathrm{e}^{-\mathrm{j}\pi(M-1)\left(\frac{l_i+l'-l}{M}\right)}\times \mathrm{e}^{\mathrm{j}2\pi\nu_i\frac{l'T}{M}}\times$$
$$\begin{cases}\dfrac{\sin\pi\times M\times\left(\frac{l_i+l'-l}{M}-\frac{N-2}{2p_i}\right)}{M\times\sin\pi\times\left(\frac{l_i+l'-l}{M}-\frac{N-2}{2p_i}\right)}\times\dfrac{\sin\pi\times(N-1)\times\left(\frac{k_i+k'-k}{N}+\frac{M-1}{2p_i}\right)}{N\times\sin\times\pi\left(\frac{k_i+k'-k}{N}+\frac{M-1}{2p_i}\right)}\times \mathrm{e}^{\mathrm{j}\pi(k_i+k'-k)}\times \mathrm{e}^{\mathrm{j}\pi\frac{(N-2)(M-1)}{2p_i}}\times \mathrm{e}^{-\mathrm{j}2\pi\frac{k_i+k'}{N}}, & l'\in\mathcal{L}_{\mathrm{ISI}}^{i};\\ \dfrac{\sin\pi\times M\times\left(\frac{l_i+l'-l}{M}-\frac{N-1}{2p_i}\right)}{M\times\sin\pi\times\left(\frac{l_i+l'-l}{M}-\frac{N-1}{2p_i}\right)}\times\dfrac{\sin\pi\times N\times\left(\frac{k_i+k'-k}{N}+\frac{M-1}{2p_i}\right)}{N\times\sin\pi\times\left(\frac{k_i+k'-k}{N}+\frac{M-1}{2p_i}\right)}\times \mathrm{e}^{\mathrm{j}\pi\frac{N-1}{N}(k_i+k'-k)}\times \mathrm{e}^{\mathrm{j}\pi\frac{(N-1)(M-1)}{2p_i}}, & l'\in\mathcal{L}_{\mathrm{ICI}}^{i}\end{cases} \tag{10}$$

Compared with the derivation in Refs. [2, 5] where DSE is ignored, extra delay-Doppler extension appears due to the phase modification of sinc functions in Formula (10), which can be approximately measured by NM/p_i. An extra rotation approximated as $e^{j\pi\frac{MN}{2p_i}}$ is also introduced due to DSE. As a result, the significance of DSE can be roughly evaluated by the maximum value of $NM/|p_i|$, which stands for the ratio between the size of an entire time-frequency resource block $NT \times M\Delta f = NM$ and the mobility parameter $p_i = f_c/\nu_i$. From the results of Ref. [13], significant performance degradation occurs due to the negligence of DSE, e.g., Normalized Mean Square Error (NMSE) of more than 1% and Bit Error Rate (BER) floor of about 3×10^{-4} when $N = 128$, $M = 512$ with the maximum mobility as 500 km/h even though perfect knowledge of parameters can be attained. Therefore, DSE requires elaborate consideration in OTFS receiver design more especially when NM is comparable to the mobility parameter p_i, which inspires us to develop DSE-aware schemes to exploit the potential of OTFS systems.

3 Proposed SBL-Based Channel Estimation

In this section, the delay-Doppler channel recovery considering DSE is depicted in detail, where the embedded pilot-aided scheme and SBL-based parameter estimation are employed. Since Formula (10) has indicated extra delay-Doppler spread due to DSE, additional guard space over traditional impulse-based technique in Refs. [4, 7] is required to guarantee the estimation quality, which certainly necessitates the adjustment of scheme details to employ SBL in OTFS channel estimation rather than deploy the SBL-based one in Ref. [7] directly. Let $\tau_{\max} = l_{\max}T/M$ and $\nu_{\max} = k_{\max}\Delta f/N$ denote the maximum time delay and Doppler frequency corresponding to f_c, respeclively, to simplify the notation.

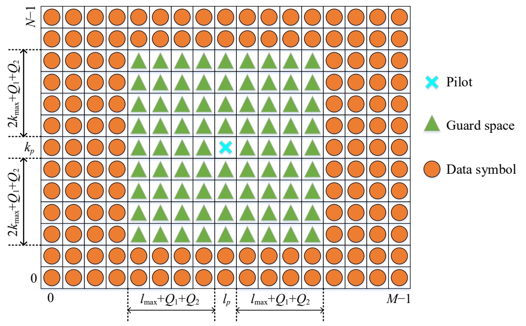
3.1 Problem formulation

Let $x_d[k, l]$ and x_p denote the data symbol and the pilot symbol, respectively. As shown in Fig. 1a, Q_1 grids are reserved to avoid the data interference between data and pilot symbols, while Q_2 grids are employed to enhance the performance of channel estimation, respectively. The transmitted symbols in delay-Doppler domain can be derived as

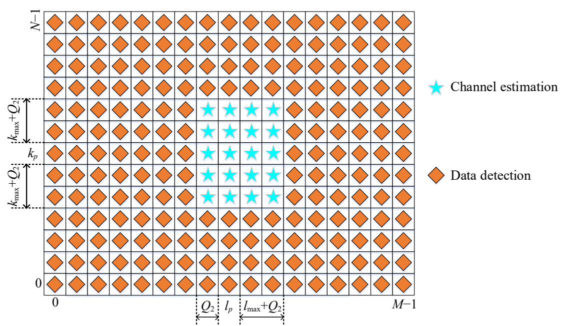
$$x[k, l] = \begin{cases} x_p, & k = k_p \text{ and } l = l_p; \\ 0, & k_p - \widetilde{k} \leqslant k \leqslant k_p + \widetilde{k} \text{ and } l_p - \widetilde{l} \leqslant l \leqslant l_p + \widetilde{l}; \\ x_d[k, l], & \text{elsewhere} \end{cases} \tag{11}$$

where we have $\widetilde{k} = 2k_{\max} + Q_1 + Q_2$ and $\widetilde{l} = l_{\max} + Q_1 + Q_2$. Q_1 grids are reserved to avoid the data interference between data and pilot symbols, while Q_2 grids are employed to enhance the performance of channel estimation. The pilot pattern is similar to that in Refs. [4, 7, 11]. However, extra interference brought by DSE forces additional guard interval compared with the prior design, which is embodied in Q_1. Taking the delay axis as an example, $Q_1 = 0$ is enough to prevent the data interference if DSE is ignored, which brings significant data interference due to the power leakage caused by DSE in Formula (10). Taking the typical value as $M = 512$, $N = 128$, $Q_1 = 10$, $Q_2 = 5$, $l_{\max} = 20$, and $k_{\max} = 16$, the pilot overhead is 10.29%, which is a little higher than prior pattern[4, 7] as 7.39%. However, the reliability is enhanced significantly considering the extra delay-Doppler extension caused by DSE, which can be shown clearly in Section 4.

As illustrated in Fig. 1b, the received symbols $y[k, l]/x_p$ for $k_p - k_{\max} - Q_2 \leqslant k \leqslant k_p + k_{\max} + Q_2$ and $l_p - Q_2 \leqslant l \leqslant l_p + l_{\max} + Q_2$ are utilized for channel



(a) Transmitter



(b) Receiver

Fig. 1 Symbol patterns of transmitter and receiver.

estimation. The interference from data symbols is absorbed into the measurement noise. According to Eq. (7) and Formula (10), the truncated region $\boldsymbol{y}_T$ can be reformulated as

$$\boldsymbol{y}_T = \boldsymbol{\Psi}_T(\boldsymbol{k}, \boldsymbol{l}) \cdot \boldsymbol{\beta} + \boldsymbol{w}_T \tag{12}$$

where $\boldsymbol{y}_T, \boldsymbol{w}_T \in \mathbf{C}^{M_T N_T \times 1}$ with $M_T = l_{\max} + 2Q_2 + 1$ and $N_T = 2k_{\max} + 2Q_2 + 1$. In this paper, the data interference is absorbed into the measured noise ω_T, which is similar to Ref. [7]. $\boldsymbol{\beta}$ is the vectorized β'_i, while the truncated measurement matrix can be derived by

$$\boldsymbol{\Psi}_T(\boldsymbol{k}, \boldsymbol{l}) = [\boldsymbol{\psi}_T(k_1, l_1), \boldsymbol{\psi}_T(k_2, l_2), \ldots, \boldsymbol{\psi}_T(k_{N_P}, l_{N_P})] \tag{13}$$

where $\boldsymbol{\psi}_T(k_i, l_i)$ denotes the vectorized $\psi^i_{k,l}[k_p, l_p]$ for the channel estimation region in Fig. 1b. Since the formulation in Eq. (12) is still non-linear due to the unknown delay and Doppler shift in $\boldsymbol{\Psi}_T$, we employ the Taylor expansion to transform the estimation problem to a linear one, which can simplify the process significantly.

Let $\widetilde{\boldsymbol{k}}_g = \{\widetilde{k}_0, \widetilde{k}_1, \ldots, \widetilde{k}_{N_v}\}$ denote the uniform sampling grid in the Doppler range $[-k_{\max}, k_{\max}]$ with the virtual Doppler resolution $r_v = 2k_{\max}/N_v$, N_v is the number of sampling grids. Suppose that $\widetilde{k}_{n_i}$ is the nearest grid point to k_i, the measurement vector $\boldsymbol{\psi}_T(k_i, l_i)$ can be approximated using first-order linear expansion as

$$\boldsymbol{\psi}_T(k_i, l_i) \approx \boldsymbol{\psi}_T\left(\widetilde{k}_{n_i}, l_i\right) + \boldsymbol{\psi}'_T\left(\widetilde{k}_{n_i}, l_i\right)\left(k_i - k_{n_i}\right) \tag{14}$$

where $\boldsymbol{\psi}_T\left(\widetilde{k}_{n_i}, l_i\right)$ and $\boldsymbol{\psi}'_T\left(\widetilde{k}_{n_i}, l_i\right)$ can be obtained from Formula (10). Meanwhile, since no fractional delay is considered, it is natural to set the virtual delay resolution as $r_\tau = 1$. Let $N_g = l_{\max}(N_v + 1)$ and $\boldsymbol{\kappa}$ denote the total number of measurement vectors and the off-grid parts of $\boldsymbol{k}$, respectively. For ease of illustration, we employ the notation as

$$\begin{aligned}
\boldsymbol{A} = [&\boldsymbol{\psi}_T(\widetilde{k}_0, 1), \ldots, \boldsymbol{\psi}_T(\widetilde{k}_{N_v}, 1), \boldsymbol{\psi}_T(\widetilde{k}_0, 2), \ldots, \\
&\boldsymbol{\psi}_T(\widetilde{k}_{N_v}, 2), \ldots, \boldsymbol{\psi}_T(\widetilde{k}_0, l_{\max}), \ldots, \boldsymbol{\psi}_T(\widetilde{k}_{N_v}, l_{\max})], \\
\boldsymbol{B} = [&\boldsymbol{\psi}'_T(\widetilde{k}_0, 1), \ldots, \boldsymbol{\psi}'_T(\widetilde{k}_{N_v}, 1), \boldsymbol{\psi}'_T(\widetilde{k}_0, 2), \ldots, \\
&\boldsymbol{\psi}'_T(\widetilde{k}_{N_v}, 2), \ldots, \boldsymbol{\psi}'_T(\widetilde{k}_0, l_{\max}), \ldots, \boldsymbol{\psi}'_T(\widetilde{k}_{N_v}, l_{\max})] \\
\boldsymbol{\Psi}_T(\boldsymbol{\kappa}) = &\boldsymbol{A} + \boldsymbol{B} \cdot \mathrm{diag}(\boldsymbol{\kappa}), \ \boldsymbol{\kappa} \in \left[-\frac{r_v}{2}, \frac{r_v}{2}\right]^{N_g}
\end{aligned} \tag{15}$$

with which the channel estimation problem in Eq. (12) can be reformulated as

$$\boldsymbol{y}_T = \boldsymbol{\Psi}_T(\boldsymbol{\kappa}) \cdot \boldsymbol{\beta} + \boldsymbol{w}_T \tag{16}$$

where $\boldsymbol{w}_T$ is the additive white Gaussian noise vector and the parameters required to be estimated are $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$.

3.2 SBL-based channel estimation scheme

The formulation in Eq. (16) can be easily solved by employing classical SBL. Since SBL has been widely employed in sparse signal recovery, we only provide a basic introduction to its key concepts and derivations here. $\boldsymbol{l}_g$ and $\boldsymbol{k}_g$ are employed to represent the delay and Doppler grids, respectively, from which $\boldsymbol{A}$ and $\boldsymbol{B}$ are expanded. The hierarchical hyper-prior distribution is employed to exploit the sparsity.

First, the noise precision α_0 is assumed to follow a Gamma hyper-prior parameterized by c and d as

$$\begin{aligned}
&p(\alpha_0; c, d) = \Gamma(\alpha_0 | c, d), \\
&p(\boldsymbol{y}_T | \boldsymbol{\kappa}, \boldsymbol{\beta}, \alpha_0) = \mathcal{CN}(\boldsymbol{y}_T | \boldsymbol{\Psi}_T(\boldsymbol{\kappa}) \cdot \boldsymbol{\beta}, \alpha_0^{-1}\boldsymbol{I})
\end{aligned} \tag{17}$$

where $\Gamma(\alpha_0 | c, d) = [\Gamma(c)]^{-1} \cdot d^c \cdot \alpha_0^{c-1} \cdot \mathrm{e}^{-d \cdot \alpha_0}$. c and d are required to be small enough to attain a broad hyper-prior.

Second, the two-stage prior for $\boldsymbol{\beta}$ is adopted as $p(\boldsymbol{\beta}; \rho) = \int p(\boldsymbol{\beta} | \boldsymbol{\alpha}) \cdot p(\boldsymbol{\alpha}; \rho)\, \mathrm{d}\boldsymbol{\alpha}$, where $\rho > 0$, $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\alpha})$ and

$$\begin{aligned}
&p(\boldsymbol{\beta} | \boldsymbol{\alpha}) = \mathcal{CN}(\boldsymbol{\beta} | \boldsymbol{0}, \boldsymbol{\Lambda}), \\
&p(\boldsymbol{\alpha}; \rho) = \prod_{n=1}^{N_g} \Gamma(\alpha_n | 1, \rho)
\end{aligned} \tag{18}$$

Since $\mathrm{Re}\{\boldsymbol{\beta}\}$ and $\mathrm{Im}\{\boldsymbol{\beta}\}$ are also Laplace distributed which is strongly peaked at the origin, the prior above tends to favor most elements of $\boldsymbol{\beta}$ being zeros.

Third, because there is no other information for $\boldsymbol{\kappa}$, the uniform prior $\boldsymbol{\kappa} \sim \mathcal{U}\left(\left[-\frac{r_v}{2}, \frac{r_v}{2}\right]^{N_g}\right)$ is assumed corresponding to its bound.

Finally, the joint probability distribution is

$$\begin{aligned}
&p(\boldsymbol{y}_T, \boldsymbol{\beta}, \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \\
&p(\boldsymbol{y}_T | \boldsymbol{\kappa}, \boldsymbol{\beta}, \alpha_0) \cdot p(\boldsymbol{\beta} | \boldsymbol{\alpha}) \cdot p(\boldsymbol{\alpha}) \cdot p(\alpha_0) \cdot p(\boldsymbol{\kappa})
\end{aligned} \tag{19}$$

where the distribution on the right can be found in the above derivation.

The Expectation-Maximization (EM) algorithm can be implemented to solve Eq. (16) with the probability analysis before, where $\boldsymbol{\beta}$ can be treated as a hidden variable. From Ref. [18], the posterior distribution of $\boldsymbol{\beta}$ can be represented as follows:

$$p(\boldsymbol{\beta} | \boldsymbol{y}_T, \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \mathcal{CN}(\boldsymbol{\beta} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{20}$$

we have

$$\boldsymbol{\Sigma} = (\alpha_0 \cdot \boldsymbol{\Psi}_T^{\mathrm{H}}(\boldsymbol{\kappa}) \cdot \boldsymbol{\Psi}_T(\boldsymbol{\kappa}) + \boldsymbol{\Lambda}^{-1})^{-1},$$
$$\boldsymbol{\mu} = \alpha_0 \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\Psi}_T^{\mathrm{H}}(\boldsymbol{\kappa}) \cdot \boldsymbol{y}_T \tag{21}$$

To obtain $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, α_0, $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$ are required. By maximizing $E(\log p(\boldsymbol{y}_T, \boldsymbol{\beta}, \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\kappa}))$, α_0 and $\boldsymbol{\alpha}$ can be updated, while the update of off-grid parts $\boldsymbol{\kappa}$ can be obtained by solving

$$\alpha_n^{\text{new}} = \frac{\sqrt{1 + 4\rho(|\boldsymbol{\mu}_n|^2 + \boldsymbol{\Sigma}_{nn})} - 1}{2\rho},$$
$$\alpha_0^{\text{new}} = \frac{c - 1 + N_g}{d + \|\boldsymbol{y}_T - \boldsymbol{\Psi}_T(\boldsymbol{\kappa}) \cdot \boldsymbol{\mu}\|^2 + \alpha_0^{-1} \sum_{n=1}^{N_g} (1 - \alpha_n^{-1} \boldsymbol{\Sigma}_{nn})} \tag{22}$$

$$\boldsymbol{\kappa}^{\text{new}} = \arg \min_{\boldsymbol{\kappa} \in \left[-\frac{r_\nu}{2}, \frac{r_\nu}{2}\right]^{N_g}} \boldsymbol{\kappa}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{\kappa} - 2\boldsymbol{v}^{\mathrm{T}} \boldsymbol{\kappa} \tag{23}$$

where we have

$$\boldsymbol{P} = \operatorname{Re}\{(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^* \odot (\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{H}} + \boldsymbol{\Sigma})\},$$
$$\boldsymbol{v} = \operatorname{Re}\{\operatorname{diag}(\boldsymbol{\mu}^*) \cdot \boldsymbol{B}^{\mathrm{H}} \cdot (\boldsymbol{y}_T - \boldsymbol{A} \cdot \boldsymbol{\mu}) - \operatorname{diag}(\boldsymbol{B}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Sigma})\} \tag{24}$$

Based on the fact that $\frac{\partial}{\partial \boldsymbol{\kappa}}\left(\boldsymbol{\kappa}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{\kappa} - 2\boldsymbol{v}^{\mathrm{T}} \boldsymbol{\kappa}\right) = 2(\boldsymbol{P}\boldsymbol{\kappa} - \boldsymbol{v})$, the computation of $\boldsymbol{\kappa}^{\text{new}}$ can be simplified. If $\boldsymbol{P}$ is invertible and $\breve{\boldsymbol{\kappa}} = \boldsymbol{P}^{-1} \cdot \boldsymbol{v} \in \left[-\frac{r_\nu}{2}, \frac{r_\nu}{2}\right]^{N_g}$, the optimal solution of Eq. (23) is $\boldsymbol{\kappa}^{\text{new}} = \breve{\boldsymbol{\kappa}}$. Otherwise, $\boldsymbol{\kappa}$ can be updated elementwise. First, we let

$$\breve{\boldsymbol{\kappa}}_n = \frac{\boldsymbol{v}_n - (\boldsymbol{P}_n)_{-n}^{\mathrm{T}} \cdot \boldsymbol{\kappa}_{-n}}{\boldsymbol{P}_{nn}} \tag{25}$$

Then we can update $\boldsymbol{\kappa}$ by carrying out

$$\boldsymbol{\kappa}_n^{\text{new}} = \begin{cases} -\frac{r_\nu}{2}, & \breve{\boldsymbol{\kappa}}_n \leqslant -\frac{r_\nu}{2}; \\ \breve{\boldsymbol{\kappa}}_n, & -\frac{r_\nu}{2} < \breve{\boldsymbol{\kappa}}_n < \frac{r_\nu}{2}; \\ \frac{r_\nu}{2}, & \breve{\boldsymbol{\kappa}}_n \geqslant \frac{r_\nu}{2} \end{cases} \tag{26}$$

Consequently, the proposed SBL-based OTFS channel estimation scheme is summarized in Algorithm 1. After initializing the hyper-parameters as $\boldsymbol{\kappa} = \boldsymbol{0}$, $\boldsymbol{\alpha} = \frac{|\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{y}_T|}{M_T \cdot N_T}$ and $\alpha_0 = \frac{100}{\operatorname{var}\{\boldsymbol{y}_T\}}$, the distribution of $\boldsymbol{\beta}$ is updated according to Eq. (21), and the hyper-parameters are updated employing Eqs. (22) and (23) iteratively. The iteration is terminated when the maximum number of iteration N_{iter} is reached or $\frac{\|\boldsymbol{\alpha}^{\text{new}} - \boldsymbol{\alpha}\|}{\|\boldsymbol{\alpha}\|}$ is smaller than a predefined tolerance ϵ. The estimation of channel parameters is provided by $\hat{\boldsymbol{k}} = \boldsymbol{k}_g + \boldsymbol{\kappa}$, $\hat{\boldsymbol{l}} = \boldsymbol{l}_g$, and $\hat{\boldsymbol{\beta}} = \boldsymbol{\mu}$. Finally, the channel matrix $\hat{\boldsymbol{H}}$ can be recovered by employing the first $\hat{N}_{\mathrm{P}}$ largest amplitude elements of $\hat{\boldsymbol{\beta}}$, where $\hat{N}_{\mathrm{P}}$ is a predefined parameter representing the maximum acceptable sparsity. The major computational load lies in the update of $\boldsymbol{\Sigma}$ for each iteration, whose complexity can be bounded as $O(M_T \cdot N_T \cdot N_g^2)$ based on efficient inversion algorithms. As a result, the total complexity order of the proposed SBL-based OTFS channel estimation scheme is $O(N_{\text{iter}} \cdot M_T \cdot N_T \cdot N_g^2)$. It is obvious that considering DSE does not increase the system complexity since the determinants keep the same as the scenarios ignoring DSE[7]. The major cost of DSE-aware schemes is the ever-decreasing transmission efficiency due to the additional guard space reserved to avoid interference caused by DSE.

Algorithm 1 SBL-based OTFS channel estimation scheme

Input: $\boldsymbol{y}_T$, r_ν, c, d, ρ, $\boldsymbol{l}_g$, $\boldsymbol{k}_g$, ϵ, N_{iter}, and $\hat{N}_P$

Output: $\hat{\boldsymbol{k}}$, $\hat{\boldsymbol{l}}$, $\hat{\boldsymbol{\beta}}$, and the recovered channel matrix $\boldsymbol{H}$

1: Generate $\boldsymbol{A}$ and $\boldsymbol{B}$ employing Eq. (15);

2: Initialize $\boldsymbol{\kappa} = \boldsymbol{0}$, $\boldsymbol{\alpha} = \frac{|\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{y}_T|}{M_T \cdot N_T}$, $\alpha_0 = \frac{100}{\operatorname{var}\{\boldsymbol{y}_T\}}$;

3: **Repeat**

4: $\boldsymbol{\Psi}_T(\boldsymbol{\kappa}) = \boldsymbol{A} + \boldsymbol{B} \cdot \operatorname{diag}(\boldsymbol{\kappa})$;

5: Update $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using Eq. (21);

6: Update $\boldsymbol{\alpha}$ and α_0 according to Eq. (22);

7: Update $\boldsymbol{\kappa}$ by solving Eq. (23);

8: **until** stopping criteria satisfied

9: $\hat{\boldsymbol{k}} = \boldsymbol{k}_g + \boldsymbol{\kappa}$, $\hat{\boldsymbol{l}} = \boldsymbol{l}_g$, $\hat{\boldsymbol{\beta}} = \boldsymbol{\mu}$;

10: Select the first $\hat{N}_P$ largest amplitude elements of $\hat{\boldsymbol{\beta}}$ to recover the channel matrix $\hat{\boldsymbol{H}}$.

11: Return $\hat{\boldsymbol{k}}$, $\hat{\boldsymbol{l}}$, $\hat{\boldsymbol{\beta}}$, and the recovered channel matrix $\hat{\boldsymbol{H}}$.

4 Simulation Result

In this section, the performance of the proposed channel estimation scheme will be evaluated by simulation results. The typical value of relevant simulation parameters is provided in Table 1. The complex gain of each path is randomly generated as $\beta_i \sim \mathcal{CN}(0, 1/N_{\mathrm{P}})$. The NMSE of the delay-Doppler channel is defined as

$$\text{NMSE} = E\left(\frac{\|\hat{\boldsymbol{H}} - \boldsymbol{H}\|_{\mathrm{F}}^2}{\|\boldsymbol{H}\|_{\mathrm{F}}^2}\right) \tag{27}$$

For the transmission frame, we have $k_p = N/2$ and $l_p = M/2$ for the location of pilot, and $Q_1 = 10$, $Q_2 = 5$ for the guard interval. We set $c = d = 10^{-4}$, $\rho = 10^{-2}$, $\epsilon = 10^{-3}$, $N_{\text{iter}} = 20$, and $\hat{N}_P = 20$ to implement Algorithm 1. The performance of the SBL-based method without

Table 1 Simulation parameters.

Parameter	Typical value
Carrier frequency (f_c)	4 GHz
Subcarrier spacing (Δf)	15 kHz
Number of subcarriers (M)	128
Number of time slots (N)	128
Modulation alphabet	QPSK
Maximum delay grid ($l_{\max}$)	20
Maximum Doppler grid ($k_{\max}$)	16
Number of paths (N_P)	4

DSE (namely SBL-NDSE) in Ref. [7], OMP-based on-grid method with DSE (OMP-DSE)[13], and OMP-based on-grid method without DSE (namely OMP-NDSE)[6] are treated as the comparison. The system Signal-to-Noise Ratio (SNR) is defined as SNR $= 1/\sigma^2$, where the power of per bit is set as 1. Similar to Refs. [6, 7, 11], the pilot power $|x_p|^2$ is assumed to be 30 dB higher than that of per bit to guarantee the quality of estimation. Though the pilot SNR denoted as $\mathrm{SNR_p}$ is usually a high value, such as 45 dB like Refs. [4, 6, 11], ISFFT will spread the power uniformly into NM time-frequency grids, e.g., $\mathrm{SNR_p} = 45$ dB leads to an extra SNR less than 3 dB under the simulation parameters in the time-frequency domain, which is quite bearable in practical system design.

In Fig. 2, we set SNR = 15 dB and illustrate the NMSE performance against the virtual Doppler resolution r_ν. It is obvious that NMSE decreases with r_ν decreasing. However, the computational complexity increases significantly with the larger dimension of $\boldsymbol{\Psi}_T$ caused by decreasing r_ν, which obtains little NMSE superiority when $r_\nu < 0.5$. So $r_\nu = 0.5$ is enough to implement the channel estimation. When $r_\nu = 0.5$, OMP-DSE even outperforms SBL-NDSE by reducing 59% of NMSE, which demonstrates the essentiality of considering DSE. SBL-DSE can reduce 37% of NMSE further by taking both DSE and fractional Doppler into account, which proves the performance superiority of this work.

Figure 3 presents the NMSE performance against SNR with $r_\nu = 0.5$. It is clear that the error floor whose level is about 10^{-2} can be diminished by considering DSE for both SBL-based and OMP-based schemes. OMP-DSE attains less NMSE than SBL-NDSE when SNR $\geqslant$ 10 dB, which motivates the development of estimation schemes taking DSE into consideration. SBL-DSE reduces about more than 30% of the channel NMSE compared with OMP-DSE when SNR = 10 dB.

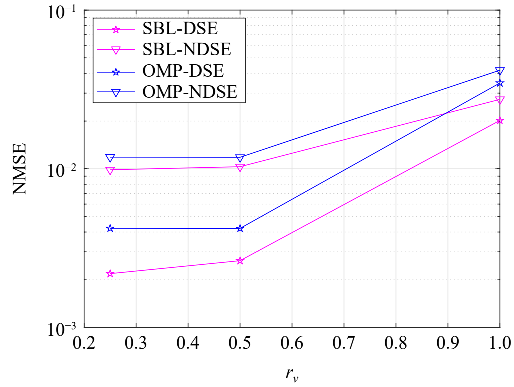


Fig. 2 NMSE performance against Doppler resolution.

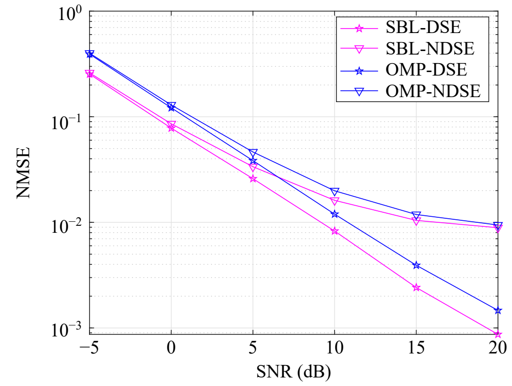


Fig. 3 NMSE performance against SNR.

5 Conclusion

In this paper, DSE in OTFS systems is taken into account, where the embedded pilot in the delay-Doppler domain is employed to estimate the channel. An off-grid SBL-based scheme is then proposed inspired by the input-output relationship with DSE and the embedded pilot. Simulation results verify the performance superiority compared with the estimation scheme ignoring DSE and grid mismatch. For future work, it is meaningful to consider the optimization of symbol patterns, pulse shapes and off-grid estimation schemes considering DSE, which is helpful for promoting the transmission efficiency and the reliability further.

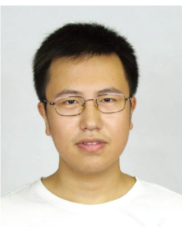
References

[1] V. S. Bhat, G. Harshavardhan, and A. Chockalingam, Input-output relation and performance of RIS-aided OTFS with fractional delay-Doppler, *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 337–341, 2023.

[2] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo,

Interference cancellation and iterative detection for orthogonal time frequency space modulation, *IEEE Trans. Wirel. Commun.*, vol. 17, no. 10, pp. 6501–6515, 2018.
[3] G. D. Surabhi and A. Chockalingam, Low-complexity linear equalization for OTFS modulation, *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, 2020.
[4] P. Raviteja, K. T. Phan, and Y. Hong, Embedded pilot-aided channel estimation for OTFS in delay-Doppler channels, *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, 2019.
[5] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, On the effectiveness of OTFS for joint radar parameter estimation and communication, *IEEE Trans. Wirel. Commun.*, vol. 19, no. 9, pp. 5951–5965, 2020.
[6] O. K. Rasheed, G. D. Surabhi, and A. Chockalingam, Sparse Delay-Doppler channel estimation in rapidly time-varying channels for multiuser OTFS on the uplink, in *Proc. 2020 IEEE 91*st *Vehicular Technology Conference* (*VTC-Spring*), Antwerp, Belgium, 2020, pp. 1–5.
[7] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, Off-grid channel estimation with sparse Bayesian learning for OTFS systems, *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 7407–7426, 2022.
[8] L. Zhao, W. J. Gao, and W. B. Guo, Sparse Bayesian learning of delay-Doppler channel for OTFS system, *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2766–2769, 2020.
[9] 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, 2020.
[10] M. Li, S. Zhang, Y. Ge, F. Gao, and P. Fan, Joint channel estimation and data detection for hybrid RIS aided millimeter wave OTFS systems, *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6832–6848, 2022.
[11] F. Liu, Z. Yuan, Q. Guo, Z. Wang, and P. Sun, Message passing-based structured sparse signal recovery for estimation of OTFS channels with fractional Doppler shifts, *IEEE Trans. Wirel. Commun.*, vol. 20, no. 12, pp. 7773–7785, 2021.
[12] L. Zhao, J. Yang, Y. Liu, and W. Guo, Block sparse Bayesian learning-based channel estimation for MIMO-OTFS systems, *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 892–896, 2022.
[13] X. Wang, X. Shi, J. Wang, and J. Song, On the Doppler squint effect in OTFS systems over doubly-dispersive channels: Modeling and evaluation, *IEEE Trans. Wirel. Commun*., vol. 22, no. 12, pp. 8781−8796, 2023.
[14] C. Jin, Z. Bie, X. Lin, W. Xu, and H. Gao, A simple two-stage equalizer for OTFS with rectangular windows, *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1158–1162, 2021.
[15] A. Liao, Z. Gao, D. Wang, H. Wang, H. Yin, D. W. K. Ng, and M. S. Alouini, Terahertz ultra-massive MIMO-based aeronautical communications in space-air-ground integrated networks, *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1741–1767, 2021.
[16] H. G. Lee, J. Kim, J. Joung, and J. Choi, Frequency-varying Doppler shift effect on wideband orthogonal time-frequency space systems, in *Proc. 2023 Int. Conf. Electronics*, *Information*, *and Communication* (*ICEIC*), Singapore, 2023, pp. 1–4.
[17] K. P. Arunkumar and C. R. Murthy, Orthogonal delay scale space modulation: A new technique for wideband time-varying channels, *IEEE Trans. Signal Process.*, vol. 70, pp. 2625–2638, 2022.
[18] Z. Yang, L. Xie, and C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, 2013.

Xuehan Wang received the BEng degree from Tsinghua University, China in 2021. He is a PhD candidate at Department of Electronic Engineering, Tsinghua University, China. His current research interests include time-variant channel estimation, orthogonal time frequency space modulation, and mobile communication.



Xu Shi received the BEng degree from Tsinghua University, China in 2019. He is a PhD candidate at Department of Electronic Engineering, Tsinghua University, China. His research interests include millimeter wave systems, RIS, extremely large-scale Multiple Input Multiple Output (MIMO), and mobile communication.

Jintao Wang received the BEng and PhD degrees in electrical engineering from Tsinghua University, China in 2001 and 2006, respectively. From 2006 to 2009, he was an assistant professor at Department of Electronic Engineering, Tsinghua University. Since 2009, he has been an associate professor and PhD supervisor. He is a standard committee member for Chinese National Digital Terrestrial Television Broadcasting Standard. He has published more than 100 journal and conference papers and holds more than 40 national invention patents. His research interests include space-time coding, MIMO, and Orthogonal Frequency Division Multiplexing (OFDM) systems.