Random Access With Massive MIMO-OTFS in LEO Satellite Communications

Boxiao Shen, *Student Member, IEEE*, Yongpeng Wu, *Senior Member, IEEE*,
Jianping An, *Senior Member, IEEE*, Chengwen Xing, *Member, IEEE*, Lian Zhao, *Senior Member, IEEE*,
and Wenjun Zhang, *Fellow, IEEE*

Abstract— This paper considers the joint channel estimation and device activity detection in the grant-free random access systems, where a large number of Internet-of-Things devices intend to communicate with a low-earth orbit satellite in a sporadic way. In addition, the massive multiple-input multiple-output (MIMO) with orthogonal time-frequency space (OTFS) modulation is adopted to combat the dynamics of the terrestrial-satellite link. We first analyze the input-output relationship of the single-input single-output OTFS when the large delay and Doppler shift both exist, and then extend it to the grant-free random access with massive MIMO-OTFS. Next, by exploring the sparsity of channel in the delay-Doppler-angle domain, a two-dimensional pattern coupled hierarchical prior with the sparse Bayesian learning and covariance-free method (TDSBL-FM) is developed for the channel estimation. Then, the active devices are detected by computing the energy of the estimated channel. Finally, the generalized approximate message passing algorithm combined with the sparse Bayesian learning and two-dimensional convolution (ConvSBL-GAMP) is proposed to decrease the computations of the TDSBL-FM algorithm. Simulation results demonstrate that the proposed algorithms outperform conventional methods.

Index Terms— Random access, OTFS, satellite communications, massive MIMO, sparse Bayesian learning.

I. Introduction

SATELLITE communication technology has been widely concerned since its inception, which will help to expand terrestrial communication networks due to its characteristics of large coverage and robustness for different geographical conditions. A lot of standardization works on satellite communications are currently underway [1], some of which hope

Boxiao Shen, Yongpeng Wu, and Wenjun Zhang are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: boxiao.shen@sjtu.edu.cn; yongpeng.wu@sjtu.edu.cn; zhangwenjun@sjtu.edu.cn).

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

Lian Zhao is with the Department of Electrical, Computer, and Biomedical Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: l5zhao@ryerson.ca).

to incorporate it into 5G and future networks [2]. Among the application requirements of 5G, the massive machine-type communications (mMTC) is an important scenario and a basic part of Internet-of-Things (IoT) applications [3]. A considerable part of the IoT devices is distributed in remote areas, such as deserts, oceans, forests, etc [4]. However, the existing cellular communication networks are mainly deployed in places where populations are concentrated, and most of them are for communications between people. As a result, supporting remote mMTC is still a huge challenge for the existing communication facilities. Fortunately, satellite communications provide a very promising solution, especially those based on low-earth orbit (LEO) satellites. The devices can access it with a low propagation delay, low path loss, and flexible elevation angle, compared with the traditional geostationary and medium earth orbit satellite [5]. Until now, there have been many companies planning to launch LEO satellites to provide seamless network coverage to the ground users, such as SpaceX [6] and OneWeb [7]. In addition, the grant-free random access (RA) is a very natural solution for satellite IoT communications [8], [9] when considering the features of the mMTC, i.e, the number of potential devices is huge while only a small part of the devices are active in any period of time and the amount of data sent by each device is small. Therefore, investigating grant-free RA technology is significant for satellite-enabled IoT systems.

Over the past few years, bunches of methods have been proposed for channel estimation and device activity detection in the grant-free RA systems, aiming at Rayleigh fading channel [10]–[15], Rician fading channel [16] and Gaussian mixture model [17]. Specifically, in [10]–[13], the authors transformed the joint channel estimation and device activity detection into a sparse signal recovery problem in compressed sensing, where the pilots of all users form a sensing matrix, and then the approximate message passing (AMP) algorithm [18] is used to solve it. The authors in [14], [15] adopted the variational Bayesian inference to solve this problem. In [16], the authors considered both the line-of-sight (LoS) and the scattering component, and then proposed the Bernoulli–Rician message passing with expectation–maximization (EM) algorithm to detect active users and estimate their channels. In [17], to perform joint channel estimation and device activity detection, the authors adopted the Bernoulli-Gaussian mixture distribution as the prior of channel, and designed a neural network architecture by unrolling the vector AMP and combining the EM algorithm. The above literature only considered sparsity

due to the sporadic device activity. When the massive multiple-input multiple-output (MIMO) is adopted in the systems, the channel will also be sparse in the angular domain. In [19], the generalized multiple measurement vector AMP (GMMV-AMP) was proposed to explore the sparsity both in the device domain and angular domain, where the sparsity ratio of each coefficient in the channel matrix not only depends on itself but also on its neighborhoods along the angular axis. As a result, the performance of the estimation and detection is improved compared with the conventional methods. The above works are based on the block fading channel which is constant during one transmission. However, the high mobility of LEO satellites inevitably leads to the rapid change of terrestrial-satellite link (TSL), and the large Doppler shift will also degrade the communication quality. Therefore, the current grant-free random access schemes can not be directly applied for the LEO satellite communications.

Orthogonal time-frequency space (OTFS) is a promising solution to tackle the time-variant channels [20]–[23], which converts the time-variant channels into the time-independent channels in the delay-Doppler domain. There has been a lot of literature focusing on the applications of OTFS for point-to-point communications [24]–[26], downlink transmissions [27]–[29], and uplink transmissions [30], [31]. Specifically, in [24] and [25], a general framework of OTFS based on ideal pulse-shaping waveforms was introduced for single-input single-output (SISO) systems. The authors in [26] analyzed the input-output relationship of OTFS with the ideal and rectangular waveforms, and adopted the message passing algorithm to detect the signal. In the applications of OTFS for the downlink transmissions, the authors in [27] derived the input-output relationship of OTFS with massive MIMO, and identified the 3D-structured sparsity of channel in the delay-Doppler-angle (DDA) domain. Then, considering the above sparsity, the 3D-structured orthogonal matching pursuit algorithm was proposed to estimate the channel. The authors in [28] considered the existence of the fractional Doppler, and derived a more accurate channel model for the MIMO-OTFS. At the same time, the deterministic pilot design method was proposed. An uplink-aided high mobility downlink channel estimation scheme for the massive MIMO-OTFS networks was proposed in [29], where the EM-based variational Bayesian framework was adopted to estimate the parameters of the uplink channel, and then the angles, the delays, and the Doppler frequencies were constructed for the downlink channels by exploring the reciprocity between the uplink and the downlink channel. In the applications of OTFS for the uplink transmission, the authors in [30] proposed the non-orthogonal multiple-access transmission protocol that incorporated OTFS modulation for the heterogeneous user mobility. A new path division multiple access for both uplink and downlink massive MIMO-OTFS was proposed in [31]. Besides, as a multicarrier modulation technology, OTFS has a lower peak-to-average power ratio (PAPR) compared with the orthogonal frequency division multiplexing (OFDM) when a small or moderate number of Doppler bins are adopted [21], [32]. This typically is the case in the LEO satellite communications that the number of Doppler bins should be small enough to guarantee

the effectiveness of the estimated channel due to the fast variations of the TSL. When a large number of Doppler bins are assumed, the PAPR of the OTFS is similar to that of the OFDM [32], and then the methods proposed in [33]–[35] can be adopted to reduce the PAPR of OTFS.

In recent years, the MIMO transmission has been widely applied in the terrestrial wireless networks [36], and massive MIMO can bring huge improvements in throughput and enhance spectral efficiency [37], [38]. However, due to the limited on-board processing capacity and instinct channel characteristics, employing the MIMO or even massive MIMO in satellite communications is still an open and challenging task [39]. Fortunately, there have been some efforts towards this goal. By exploiting the dual polarization, the two by two and four by two MIMO configurations were studied for the GEO satellite in [40]. The European Space Agency initiated the project to investigate the massive MIMO over satellite in [41]. A practical design of massive MIMO combined with the radio resource management approach was proposed in [42] to ease its implementation at the satellite payload level. The applications of massive MIMO along with full frequency reuse in the multi-user LEO satellite communications were considered in [43]–[46]. Motivated by this, we exploit the massive MIMO in the LEO satellite communications.

Overall, most of the previous grant-free random access schemes only consider the block fading channel and assume that the channel is constant during the pilot and the subsequent data transmission. However, time variation, multi-path propagation, and Doppler effects should also be involved in the TSL due to the high mobility of the LEO satellite [43], [47], which results in our considered doubly dispersive channel [48]. Furthermore, there is little literature focusing on the grant-free random access with OTFS. In this paper, we investigate the joint channel estimation and device activity detection in the LEO satellite communications with massive MIMO-OTFS, where the OTFS is adopted to suppress the doubly dispersive effect in the TSL link and the massive MIMO is integrated for improving the performance. Notice that the adoption of massive MIMO-OTFS will bring the 3D-structured sparsity in the delay-Doppler-angle domain [27], which leads to the two-dimensional (2D) burst block sparsity of the channel matrix in the grant-free random access scenarios. This new type of sparsity is not studied in the previous literature and cannot be fully addressed by conventional algorithms. Therefore, we propose to deploy the sparse Bayesian learning (SBL) framework to capture the 2D burst block sparsity. Besides, the scale of our considered problem is much larger than that in the conventional literature, due to the estimation of the doubly dispersive channel. Thus, the algorithm based on the generalized approximate message passing (GAMP) is proposed to further decrease the computations of the SBL method. Our main contributions are summarized as follows:

- We first analyze the input-output relationship of SISO-OTFS when the large delay and Doppler both exist. Then, we extend it to the grant-free random access system with massive MIMO-OTFS. We formulate the joint channel estimation and device activity detection as a sparse signal recovery problem, where the 3D-structured

sparsity of the channel in the delay-Doppler-angle domain is transformed to the 2D burst block sparsity in our channel matrix.

- We propose a 2D pattern coupled hierarchical prior with the SBL framework to capture the 2D burst block sparsity of the channel matrix, where the sparsity pattern of each coefficient is controlled not only by its own hyperparameter, but also by the hyperparameters of its neighborhoods both along the delay-Doppler axis and angular axis in the channel matrix. Then, those hyperparameters and the precision of the noise are updated by the EM algorithm, and the covariance-free method is introduced to decrease the computational complexity of the SBL, resulting in the TDSBL-FM algorithm.

- To facilitate the TDSBL-FM algorithm for solving the large-scale problems, the GAMP with the 2D convolution and SBL (ConvSBL-GMAP) is developed. Here, we consider a special case of the proposed prior in the SBL framework to capture the 2D burst block sparsity. Then, based on this prior, the GAMP is utilized to estimate the posterior mean and variance of the channel matrix. At the same time, the update of the hyperparameters in the EM algorithm can be computed through the 2D convolution.

The rest of this paper is organized as follows. Section II investigates the system model for the massive MIMO-OTFS random access and formulates the problem. Section III proposes the 2D pattern coupled SBL for joint channel estimation and device activity detection. Section IV combines the GAMP into the SBL framework to further decrease its complexity. Section V evaluates the performance of the proposed algorithms.

Notations: The superscripts $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ denote the conjugate and conjugated-transpose operations, respectively. The boldface letters denote matrices or vectors. $\bar{\jmath} = \sqrt{-1}$ denotes the imaginary unit. $\|\mathbf{X}\|_1$ and $\|\mathbf{x}\|_1$ denote the ℓ_1-norm of $\mathbf{X}$ and $\mathbf{x}$, $\|\mathbf{X}\|_{\mathrm{F}}$ denotes the Frobenius norm of $\mathbf{X}$, and $\|\mathbf{x}\|_2$ denotes the ℓ_2-norm of $\mathbf{x}$. $\lceil x \rceil$ denotes the smallest integer that is not less than x, while $\lfloor x \rfloor$ denotes the largest integer that is not greater than x. $\odot$ is the Hadamard product operator and $\oslash$ denotes element-wise division. $(\cdot)_M$ denotes mod M, and $\langle x \rangle_N$ denotes $\left(x + \left\lfloor \frac{N}{2} \right\rfloor\right)_N - \left\lfloor \frac{N}{2} \right\rfloor$. The notation $\triangleq$ is used for definitions. $\delta(\cdot)$ denotes the Dirac delta function. $\mathbf{X}[a,:]$ denotes the a-th row of $\mathbf{X}$, wihle $\mathbf{X}[:,b]$ denotes the b-th column of $\mathbf{X}$. $X[a,b]$ and $X_{a,b}$ denote the (a,b)-th element of $\mathbf{X}$. x_a denotes the a-th element of $\mathbf{x}$.

II. SYSTEM MODEL

In this section, some basic concepts of OTFS are first reviewed. Then, we analyze the input-output relationship of the SISO-OTFS system, when both the large delay and Doppler shift exist. Next, an extension of OTFS into massive MIMO systems is described. Finally, we formulate the considered problem.

A. SISO-OTFS Transmissions

1) Terrestrial Transmitter: The device in land employs the OTFS transmitter, which first maps a set of NM information

symbols $X^{\mathrm{DD}}[k,l]$ in the delay-Doppler domain to the symbols $X^{\mathrm{TF}}[n,m]$ in the time-frequency domain through the inverse symplectic finite Fourier transform (ISFFT) [23], i.e.,

$$X^{\mathrm{TF}}[n,m] = \frac{1}{\sqrt{MN}} \sum_{k=\lceil -N/2 \rceil}^{\lceil N/2 \rceil - 1} \sum_{l=0}^{M-1} X^{\mathrm{DD}}[k,l] e^{-\bar{\jmath} 2\pi \left(\frac{ml}{M} - \frac{nk}{N}\right)},$$

(1)

where $k = \lceil -N/2 \rceil, \ldots, \lceil N/2 \rceil - 1$, $l = 0, \ldots, M-1$. Next, the OFDM modulator converts the symbols $X^{\mathrm{TF}}[n,m]$ to a continuous signal $s(t)$ using a transmit waveform $g_{\mathrm{tx}}(t)$ as

$$s(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X^{\mathrm{TF}}[n,m] e^{\bar{\jmath} 2\pi m \Delta f (t - T_{\mathrm{cp}} - n T_{\mathrm{sym}})}$$

$$\times g_{\mathrm{tx}}\left(t - n T_{\mathrm{sym}}\right), \quad (2)$$

where Δf is the carrier spacing, $T_{\mathrm{cp}} = \frac{M_{\mathrm{cp}} T}{M}$ is the time duration of the cyclic prefix (CP), M_{cp} is the length of CP, $T = \frac{M T_{\mathrm{sym}}}{M + M_{\mathrm{cp}}}$ is the time duration of an OFDM symbol without CP, T_{sym} is the time duration of a complete OFDM symbol, and $g_{\mathrm{tx}}(t)$ is defined as

$$g_{\mathrm{tx}}(t) \triangleq \begin{cases} \frac{1}{\sqrt{T}}, & 0 \le t \le T_{\mathrm{sym}} \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

2) Terrestrial-Satellite Link: The TSLs will experience fast variations due to the high mobility of the LEO satellite, and the signal will be transmitted over LoS path and non-LoS (NLoS) path due to the direct signal interacting with the scatterers in the vicinity of the terrestrial device. Therefore, the time variation, multipath propagation and Doppler effects should be considered in the LEO satellite channel. Then, the channel response in the time-frequency domain can be represented as [43]–[46]

$$h(t,f) = \sqrt{\frac{\mathcal{K}}{\mathcal{K}+1}} e^{\bar{\jmath} 2\pi \left(t \nu^{\mathrm{LoS}} - f \tau^{\mathrm{LoS}}\right)}$$

$$+ \sqrt{\frac{1}{\mathcal{K}+1}} \sum_{i=1}^{P} h_i^{\mathrm{NLoS}} e^{\bar{\jmath} 2\pi \left(t \nu_i^{\mathrm{NLoS}} - f \tau_i^{\mathrm{NLoS}}\right)}, \quad (4)$$

where $\mathcal{K}$ is the Rician factor; ν^{LoS} and τ^{LoS} are the Doppler shift and propagation delay of the LoS path, respectively; h_i^{NLoS}, ν_i^{NLoS}, and τ_i^{NLoS} denote the complex gain, Doppler shift, and propagation delay of the i-th NLoS path, respectively. Note that the channel model in (4) is applicable when the relative positions of the LEO satellite and device do not change significantly, and thus the physical parameters ν^{LoS}, τ^{LoS}, h_i^{NLoS}, ν_i^{NLoS}, and τ_i^{NLoS} are assumed to be invariant [43]. When the LEO satellite or the device moves over a long distance, those parameters should be updated accordingly [49]. Next, we elaborate the key parameters as follows.

- **Doppler shift**: In the LEO satellite communications, the Doppler shift is much larger than that in the terrestrial wireless networks due to the large relative velocity between the satellite and the devices. For example, the Doppler shift can be up to 41 kHz at the 2 GHz carrier frequency for a LEO satellite operated at the 600 km altitude [50]. Generally, the Doppler shift ν^{LoS} is composed of two parts, i.e., $\nu^{\mathrm{LoS}} = \nu_S^{\mathrm{LoS}} + \nu_D^{\mathrm{LoS}}$, where ν_S^{LoS} and ν_D^{LoS}

are the Doppler shifts concerned with the movement of the satellite and the device, respectively. For the NLoS path, similarly, we have $\nu_i^{\mathrm{NLoS}} = \nu_{S,i}^{\mathrm{NLoS}} + \nu_{D,i}^{\mathrm{NLoS}}$. As a result of the high altitude of the satellite, the first part ν_S^{LoS} ($\nu_{S,i}^{\mathrm{NLoS}}$) is nearly identical for different paths of the TSL [51], and thus ν_S^{LoS} ($\nu_{S,i}^{\mathrm{NLoS}}$) can be re-written as $\nu_S^{\mathrm{LoS}} = \nu_{S,i}^{\mathrm{NLoS}} = \nu_S$. On the other hand, the Doppler shift ν_D^{LoS} ($\nu_{D,i}^{\mathrm{NLoS}}$) due to the motion of the device is typically distinct for different propagation paths [43], [44].

- **Propagation delay**: The propagation delay also exhibits much larger value than that in the terrestrial wireless networks, because of the long distance between the satellite and the device. For example, the roundtrip delay is about 17.7 ms for an LEO satellite at the 1000 km altitude and with 45° of elevation angles [52].

To convert the time-varying channel into the time-independent channel, the Fourier transform and inverse Fourier transform are applied for the channel response $h(t, f)$ along the time dimension and frequency dimension, respectively, and then we have

$$
\begin{aligned}
h(\tau, \nu) &= \iint h(t, f) e^{-j2\pi(\nu t - f\tau)} dt df \\
&= \sqrt{\frac{\mathcal{K}}{\mathcal{K}+1}} \delta\left(\tau - \tau^{\mathrm{LoS}}\right) \delta\left(\nu - \nu^{\mathrm{LoS}}\right) \\
&\quad + \sqrt{\frac{1}{\mathcal{K}+1}} \sum_{i=1}^{P} h_i^{\mathrm{NLoS}} \delta\left(\tau - \tau_i^{\mathrm{NLoS}}\right) \delta\left(\nu - \nu_i^{\mathrm{NLoS}}\right),
\end{aligned}
$$
$$(5)$$

which is the channel response in the delay-Doppler domain. We denote $\sqrt{\frac{\mathcal{K}}{\mathcal{K}+1}}$, ν^{LoS}, τ^{LoS}, $\sqrt{\frac{1}{\mathcal{K}+1}} h_i^{\mathrm{NLoS}}$, ν_i^{NLoS}, and τ_i^{NLoS} as h_0, ν_0, τ_0, h_i, ν_i, and τ_i for $i = 1 \ldots P$, respectively, and then the channel can be re-written in a more compact way, i.e.,

$$
h(\tau, \nu) = \sum_{i=0}^{P} h_i \delta\left(\tau - \tau_i\right) \delta\left(\nu - \nu_i\right). \tag{6}
$$

Finally, we define the delay and Doppler taps for the i-th path as follows

$$
\tau_i = \frac{l_i + c_i M}{M \Delta f}, \nu_i = \frac{k_i + \tilde{k}_i + b_i N}{N T_{\mathrm{sym}}}, \tag{7}
$$

where $l_i = 0, \ldots, M-1$ and $k_i = \lceil -N/2 \rceil, \ldots, \lceil N/2 \rceil - 1$ represent the indexes of the delay tap and Doppler tap corresponding to the delay τ_i and Doppler ν_i, respectively; b_i and c_i are integers and referred to as the outer Doppler and outer delay, respectively, since they represent the part exceeding one carrier spacing and one symbol duration; $\tilde{k}_i \in (-\frac{1}{2}, \frac{1}{2}]$ is the fractional Doppler shift. In addition, we define the tap index set $\mathcal{D}$ and $\mathcal{F}$ of the delay and Doppler shift of all paths as $\mathcal{D} = [l_0, \ldots, l_P]$, and $\mathcal{F} = [k_0 + \tilde{k}_0, \ldots, k_P + \tilde{k}_P]$.

3) LEO Satellite Receiver: The signal $s(t)$ is transmitted over the above TSL, and then the received signal $r(t)$ is given by [26]

$$
r(t) = \iint h(\tau, \nu) s(t - \tau) e^{j2\pi\nu(t-\tau)} d\tau d\nu + n(t), \tag{8}
$$

where $n(t)$ is the additive Gaussian noise.

At the receiver, the cross-ambiguity function $A_{g_{\mathrm{rx}}, r}(t, f)$ is computed by the matched filter as follows

$$
A_{g_{\mathrm{rx}}, r}(t, f) = \int g_{\mathrm{rx}}^*\left(t' - t\right) r\left(t'\right) e^{-j2\pi f\left(t' - T_{\mathrm{cp}} - t\right)} dt', \tag{9}
$$

where the received waveform $g_{\mathrm{rx}}(t)$ is defined as

$$
g_{\mathrm{rx}}(t) \triangleq \begin{cases} \frac{1}{\sqrt{T}}, & T_{\mathrm{cp}} \leq t \leq T_{\mathrm{sym}} \\ 0, & \text{otherwise}. \end{cases} \tag{10}
$$

Then, the received symbols $Y^{\mathrm{TF}}[n, m]$ in the time-frequency domain are sampled from $A_{g_{\mathrm{rx}}, r}(t, f)$ as

$$
Y^{\mathrm{TF}}[n, m] = A_{g_{\mathrm{rx}}, r}(t, f)\big|_{t=nT_{\mathrm{sym}}, f=m\Delta f}. \tag{11}
$$

Next, the symplectic finite Fourier transform (SFFT) is adopted to map $Y^{\mathrm{TF}}[n, m]$ to the symbols $Y^{\mathrm{DD}}[k, l]$ as

$$
Y^{\mathrm{DD}}[k, l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y^{\mathrm{TF}}[n, m] e^{j2\pi\left(\frac{ml}{M} - \frac{nk}{N}\right)}. \tag{12}
$$

Similar to [28], we define the complex gain of the time-variant channel on the delay tap l at the time ρT_s as

$$
h_{\rho, l} = \sum_{i=0}^{P} h_i e^{j2\pi(\rho-l)T_s \nu_i} \delta\left(lT_{\mathrm{s}} - (\tau_i)_T\right), \tag{13}
$$

where $T_{\mathrm{s}} = \frac{1}{M\Delta f}$ is the system sampling interval. Then the input-ouput relationship of the SISO-OTFS system is shown in the following proposition.

Proposition 1: For a SISO-OTFS system, when the time duration of CP is beyond the maximum delay of all paths, i.e., $T_{\mathrm{cp}} \geq \tau_i, i = 0, \ldots, P$, the input-output relationship in the delay-Doppler domain is given by

$$
\begin{aligned}
Y^{\mathrm{DD}}[k, l] = \sum_{l'=0}^{M-1} \sum_{k'=\lceil -N/2 \rceil}^{\lceil N/2 \rceil - 1} \phi(l, l') H^{\mathrm{DD}}\left[k', l'\right] \\
\times X^{\mathrm{DD}}\left[\langle k - k' \rangle_N, (l - l')_M\right] + Z^{\mathrm{DD}}[k, l],
\end{aligned} \tag{14}
$$

where

$$
H^{\mathrm{DD}}[k, l] = \frac{1}{N} \sum_{j=0}^{N-1} h_{M_{\mathrm{cp}} + j(M + M_{\mathrm{cp}), l}} e^{-j\frac{2\pi}{N}kj}, \tag{15}
$$

$$
\phi(l, l') = \begin{cases} e^{j2\pi \frac{(k_i + \tilde{k}_i + b_i N)(l - c_i M)}{(M + M_{\mathrm{cp}})N}}, & l' \in \mathcal{D}, l' = l_i \\ 1, & l' \notin \mathcal{D}, \end{cases} \tag{16}
$$

l_i is selected from $\mathcal{D}$, $k_i + \tilde{k}_i$ is in $\mathcal{F}$, and $Z^{\mathrm{DD}}[k, l]$ is the Gaussian noise.

Proof: See Appendix A.

We can find in (14) that the outer Doppler b_i and outer delay c_i appear in the phase compensation term $\phi(l, l')$, which means that they will cause the phase rotation of the received symbols in the delay-Doppler domain and may degrade the detection performance. Besides, if $\tau_i \ll T$ and $\nu_i \ll \Delta f$, b_i and c_i can be neglected, and then (14) is identical to that in [28]. Notice that in (14), we require $T_{\mathrm{cp}} \geq \tau_i$ to eliminate the intersymbol interference (ISI) between the OFDM symbols in one OTFS frame and to reduce the complexity of the receiver when the large differential delay exist [50]. However,
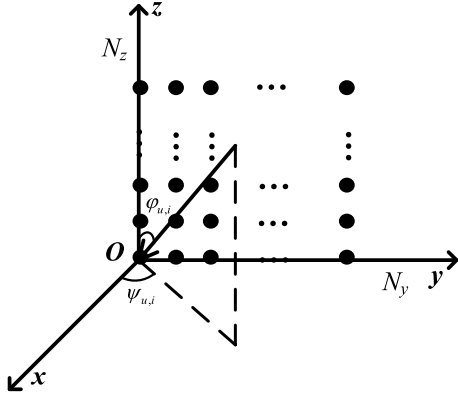
Fig. 1. Illustration of UPA in the LEO satellite. Each black dot represents one antenna.

this will lead to the low spectral efficiency since the longer CP is usually needed. For example, in the narrowband-IoT reference scenarios from 3GPP [53], the satellite operates in the S-band (2 GHz) and at the altitude 600 km. Then, the maximum differential delay and Doppler shift seen in the uplink can be up to 1397 μ s and 41 kHz, respectively. As a result, the length of CP will be several multiples of the data duration (e.g., the carrier spacing 3.75 kHz [53] is adopted) if we let $T_{\mathrm{cp}} > 1397 \mu$ s. Besides, the estimated CSI may be outdated soon due to dynamics of the TSL and the long symbol duration. Therefore, to improve the spectral efficiency further, we assumed that the pre-compensation for delay at the device level using the initial downlink time synchronization algorithms [54] is performed such that the residual delay seen in the uplink is in a small range. Then, we can adopt a relatively short CP, and select the proper symbol time duration and carrier spacing such that $\tau_i \ll T$ and $\nu_i \ll \Delta f$. In this scenario, the phase compensation term is given by

$$\phi\left(l, l'\right) = \begin{cases} e^{\bar{j}2\pi \frac{(k_i + \bar{k}_i)l}{(M + M_{\mathrm{cp}})N}}, & l' \in \mathcal{D}, l' = l_i \\ 1, & l' \notin \mathcal{D}, \end{cases} \quad (17)$$

In the following content, we focus on the case that $\tau_i \ll T$ and $\nu_i \ll \Delta f$, where τ_i represents the residual delay seen in the uplink.

B. Random Access With Massive MIMO-OTFS

We consider the uplink transmission in a massive MIMO-OTFS system, where the LEO satellite is equipped with an uniform planar array (UPA) composed of $N_a = N_y \times N_z$ antennas, as illutrated in Fig. 1, where N_y and N_z are the number of antennas along the y-axis and z-axis, respectively, and the antenna spacing is set to half wavelength. In a given time interval, there are U potential IoT devices with single antenna transmitting the data to the satellite in a sporadic traffic way, i.e., only U_a devices are active and $U_a \ll U$. In addition, the active devices share the same resources. We denote the activity indicator of the u-th device as λ_u, which is equal to 1 when it is active, and 0 otherwise. Then, the set of active IoT devices are represented as $\mathcal{U} = \{u | \lambda_u = 1, 0 \leq u \leq U - 1\}$, and the number of active IoT devices $U_a = |\mathcal{U}|$. Similar to (14) and (15), at the satellite receiver, the symbols transmitted from

the u-th device to the $(n_z + 1 + n_y N_z)$-th antenna in the delay-Doppler-space domain can be expressed as

$$Y_u^{\mathrm{DDS}}[k, l, n_y, n_z]$$
$$= \sum_{l'=0}^{M-1} \sum_{k'=\lceil -N/2 \rceil}^{\lceil N/2 \rceil - 1} \phi_u(l, l') H_u^{\mathrm{DDS}}\left[k', l', n_y, n_z\right]$$
$$\times X_u^{\mathrm{DD}}\left[\langle k - k'\rangle_N, (l - l')_M\right] + Z^{\mathrm{DD}}[k, l, n_y, n_z], \quad (18)$$

where $n_y = 0, \ldots, N_y - 1$, $n_z = 0, \ldots, N_z - 1$, $\phi_u(l, l')$ is the phase compensation factor of the u-th device, $X_u^{\mathrm{DD}}\left[\langle k - k'\rangle_N, (l - l')_M\right]$ is the transmitted symbol of the u-th device in the delay-Doppler domain, and $Z^{\mathrm{DD}}[k, l, n_y, n_z] \sim \mathcal{CN}(0, \theta^{-1})$ is the noise with the precision θ; $H_u^{\mathrm{DDS}}[k', l', n_y, n_z]$ is the delay-Doppler-space domain channel, which is defined as [49]

$$H_u^{\mathrm{DDS}}[k, l, n_y, n_z]$$
$$= \frac{1}{N} \sum_{i=0}^{P} \sum_{j=0}^{N-1} h_{u,i} e^{\bar{j}2\pi(M_{\mathrm{cp}} + j(M + M_{\mathrm{cp}}) - l)T_{\mathrm{s}}\nu_{u,i}}$$
$$\times \delta\left(lT_{\mathrm{s}} - \tau_{u,i}\right) e^{\bar{j}\pi n_y \sin \varphi_{u,i} \sin \psi_{u,i}} e^{\bar{j}\pi n_z \cos \varphi_{u,i}} e^{-\bar{j}\frac{2\pi}{N}kj}, \quad (19)$$

where $h_{u,i}$, $\nu_{u,i}$, and $\tau_{u,i}$ are the complex channel gain, Doppler shift, and propagation delay of the u-th device in the i-th path, respectively; $\psi_{u,i} \in [-\pi/2, \pi/2]$ and $\varphi_{u,i} \in [0, \pi]$ are the corresponding azimuth angle and elevation angle, respectively. Since the satellite always operates on the relatively high altitude compared with that of the scatters located around the IoT devices, the angles of all paths associated with the same device can be assumed to be identical [43]. Hence, we omit the path index of the azimuth and elevation angles in the following content.

To exploit the sparsity of the angular domain, the two-dimensional discrete Fourier transform (2D-DFT) is applied for the channel of u-th device along the space dimension in the delay-Doppler-space domain, and then the delay-Doppler-angle domain channel is given by

$$H_u^{\mathrm{DDA}}[k, l, a_y, a_z]$$
$$= \frac{1}{\sqrt{N_y N_z}} \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} H_u^{\mathrm{DDS}}[k, l, n_y, n_z] e^{-\bar{j}2\pi(\frac{a_y n_y}{N_y} + \frac{a_z n_z}{N_z})}$$
$$= \sqrt{N_y N_z} \sum_{i=0}^{P} h_{u,i} e^{\bar{j}2\pi(M_{\mathrm{cp}} - l)T_s \nu_{u,i}} \Pi_N(k - NT_{\mathrm{sym}}\nu_{u,i})$$
$$\times \Pi_{N_y}(a_y - N_y \Omega_{y_u}/2) \Pi_{N_z}(a_z - N_z \Omega_{z_u}/2) \delta\left(lT_{\mathrm{s}} - \tau_{u,i}\right), \quad (20)$$

where $a_y = 0, \ldots, N_y - 1$ and $a_z = 0, \ldots, N_z - 1$ are indexes along the angular domain; $\Omega_{y_{u,i}} = \sin \varphi_u \sin \psi_u$ and $\Omega_{z_u} = \cos \varphi_u$ are the directional cosines along the y-axis and z-axis, respectively; $\Pi_N(x) \triangleq \frac{1}{N} \sum_{i=0}^{N-1} e^{-\bar{j}2\pi \frac{x}{N}i}$. Therefore, we can find in (20) that $H_u^{\mathrm{DDA}}[k, l, a_y, a_z]$ has dominant elements only if $k \approx NT_{\mathrm{sym}}\nu_{u,i}$, $l \approx \tau_{u,i}M\Delta f$, $a_y \approx N_y \Omega_{y_u}/2$, and $a_z \approx N_z \Omega_{z_u}/2$, which means that the channel in the delay-Doppler-angle domain has the 3D-structured sparsity [27]. In practice,

the number of dominant paths $P \ll M$, and therefore, we will get very sparse channel in the delay-Doppler-angle domain. Combining (18), (19), and (20), the received symbols in the delay-Doppler-angle domain is given by

$$
\begin{aligned}
& Y_u^{\mathrm{DDA}}[k, l, a_y, a_z] \\
&= \frac{1}{\sqrt{N_y N_z}} \sum_{n_y=0}^{N_y-1} \sum_{n_z=0}^{N_z-1} Y_u^{\mathrm{DDS}}[k, l, n_y, n_z] e^{-j 2\pi\left(\frac{a_y n_y}{N_y}+\frac{a_z n_z}{N_z}\right)} \\
&= \sum_{l'=0}^{M-1} \sum_{k'=\lceil -N/2 \rceil}^{\lceil N/2 \rceil-1} \phi_u(l, l') H_u^{\mathrm{DDA}}[k', l', a_y, a_z] \\
&\quad \times X_u^{\mathrm{DD}}\left[\langle k-k'\rangle_N,(l-l')_M\right]+Z^{\mathrm{DDA}}[k, l, a_y, a_z],
\end{aligned}
\tag{21}
$$

where $Z^{\mathrm{DDA}}[k, l, a_y, a_z]$ is the Gaussian noise in the delay-Doppler-angle domain, obeying the same distribution as $Z^{\mathrm{DDS}}[k, l, n_y, n_z]$. Then, according to the superposition principle, the received symbols from all active devices can be expressed as

$$
\begin{aligned}
& Y^{\mathrm{DDA}}[k, l, a_y, a_z] \\
&= \sum_{u=0}^{U-1} \sum_{l'=0}^{M-1} \sum_{k'=\lceil -N/2 \rceil}^{\lceil N/2 \rceil-1} \lambda_u \phi_u(l, l') H_u^{\mathrm{DDA}}[k', l', a_y, a_z] \\
&\quad \times X_u^{\mathrm{DD}}\left[\langle k-k'\rangle_N,(l-l')_M\right]+Z^{\mathrm{DDA}}[k, l, a_y, a_z].
\end{aligned}
\tag{22}
$$

C. Problem Formulation

In this subsection, we formulate the joint device activity detection and channel estimation. The length of pilots along the Doppler dimension and the delay dimension are N_ν and M_τ, respectively. In addition, the guard intervals are inserted in one OTFS frame to avoid the interference between the data and pliots. Note that the channel in the delay-Doppler-angle domain has finite support $[0, M_{\max}-1]$ along the delay dimension. The length of guard intervals should be $M_g \geq M_{\max}-1$, and we assume that $M_\tau \geq M_g$. Further, to fully explore the sparsity in the Doppler domain, we propose to utilize all the available resources in the Doppler domain to transmit pilots, i.e., set $N_\nu = N$. Finally, the non-orthogonal pilot pattern is adopted, and the training pilots are the complex Gaussian random sequences, which are independent among different devices. Then, in a given time interval, the active devices transmit their pilots simultaneously. We denote the pilots of u-th device as $X_{p,u}^{\mathrm{DD}}\left[\langle k-k'\rangle_N,(l-l')_M\right]$. Then, according to (22), the received pilots in the delay-Doppler-angle domain at the satellite side is given by

$$
\begin{aligned}
& Y_p^{\mathrm{DDA}}[k, l, a_y, a_z] \\
&= \sum_{u=0}^{U-1} \sum_{l'=0}^{M_\tau-1} \sum_{k'=\lceil -N/2 \rceil}^{\lceil N/2 \rceil-1} \lambda_u \phi_u(l, l') H_u^{\mathrm{DDA}}[k', l', a_y, a_z] \\
&\quad \times X_{p,u}^{\mathrm{DD}}\left[\langle k-k'\rangle_N,(l-l')_M\right]+Z^{\mathrm{DDA}}[k, l, a_y, a_z].
\end{aligned}
\tag{23}
$$

where $k=\lceil -N/2 \rceil, \ldots, \lceil N/2 \rceil-1$, and $l=0, \ldots, M_\tau-1$.

To facilitate the following analysis, we rewrite (23) into the matrix form as

$$
\mathbf{Y}=(\Phi \odot \mathbf{X}) \Lambda \tilde{\mathbf{H}}+\mathbf{Z},
\tag{24}
$$

where the elements of $\mathbf{Z}$ are independent Gaussian noise, and the $(a_z+1+N_z a_y)$-th column of $\mathbf{Y}$ is $\operatorname{vec}\left(\mathbf{Y}_p^{DDA}[:,:, a_y, a_z]\right)$. $\Phi=\left[\Phi_0, \ldots, \Phi_{U-1}\right]$ contains the phase compensation matrices $\Phi_u \in \mathrm{C}^{M_\tau N \times U M_\tau N}$ of the u-th device composed of $\phi_u(l, l) \mathbf{1}_N$ and $l=0, \ldots, M_\tau-1$. $\mathbf{X}=\left[\mathbf{X}_0, \ldots, \mathbf{X}_{U-1}\right] \in \mathrm{C}^{M_\tau N \times U M_\tau N}$, where $\mathbf{X}_u \in \mathrm{C}^{M_\tau N \times M_\tau N}$ is the pilot matrix of the u-th device. Note that $\mathbf{X}_u$ is a doubly circulant matrix due to the 2D convolution in (23), and its sub-matrix $\mathbf{X}_{u, l}, l=0, \ldots, M_\tau-1$ is the circulant matrix, formed by $\left[X_{p,u}^{\mathrm{DD}}[0, l], X_{p,u}^{\mathrm{DD}}[1, l], \ldots, X_{p,u}^{\mathrm{DD}}[-1, l]\right]$. Λ is a diagonal matrix composed of the activity indicator of all the devices, given by

$$
\Lambda=\left[\begin{array}{cccc}
\lambda_0 \mathbf{I} & & & \\
& \lambda_1 \mathbf{I} & & \\
& & \ddots & \\
& & & \lambda_{U-1} \mathbf{I}
\end{array}\right],
\tag{25}
$$

where $\mathbf{I}$ is the identity matrix with dimension $M_\tau N \times M_\tau N$. Finally, the $(a_z+1+N_z a_y)$-th column of $\tilde{\mathbf{H}}$ is given by

$$
\tilde{\mathbf{H}}[:, a_z+1+N_z a_y]=\left[\begin{array}{c}
\operatorname{vec}\left(\mathbf{H}_0^{\mathrm{DDA}}[:,:, a_y, a_z]\right) \\
\operatorname{vec}\left(\mathbf{H}_1^{\mathrm{DDA}}[:,:, a_y, a_z]\right) \\
\vdots \\
\operatorname{vec}\left(\mathbf{H}_{U-1}^{\mathrm{DDA}}[:,:, a_y, a_z]\right)
\end{array}\right].
\tag{26}
$$

We denote the channel matrix $\Lambda \tilde{\mathbf{H}}$ as $\mathbf{H}$. Then, (24) can be written as

$$
\mathbf{Y}=(\boldsymbol{\Phi} \odot \mathbf{X}) \mathbf{H}+\mathbf{Z}.
\tag{27}
$$

Therefore, the joint device activity detection and channel estimation in the massive MIMO-OTFS systems is a sparse signal reconstruction problem. However, since the phase compensation matrix $\boldsymbol{\Phi}$ is related to the channel, and $\mathbf{H}$ cannot be directly recovered with the unknown sensing matrix. Fortunately, when choosing large enough carrier spacing Δf and the number of OFDM symbols N in one OTFS frame, as shown in our simulations, the phase compensation matrix can be approximated by all-ones matrix $\mathbf{1}_{M_\tau N \times U M_\tau N}$ while with a little performace degradation. Therefore, (27) can be rewritten as

$$
\mathbf{Y} \approx \mathbf{X} \mathbf{H}+\mathbf{Z}.
\tag{28}
$$

Note that the channel matrix $\mathbf{H}$ in the delay-Doppler-angle domian has different kinds of sparsities, compared with that in the previous literature [10]–[13], [19] which only focuses on the sparsity of the user or the sparsity in the angular domain. As shown in Fig. 2, the channel matrix $\mathbf{H}$ has the 2D burst block sparsity, i.e., not only it has the burst block sparsity [27] along the column, but also has the burst block sparsity along the row. Specifically, the row sparsity is due to the channel sparsity in the delay-Doppler domain, and the starting positions of the clusters in each column depend on the delay and Doppler shift of all the physical paths of the
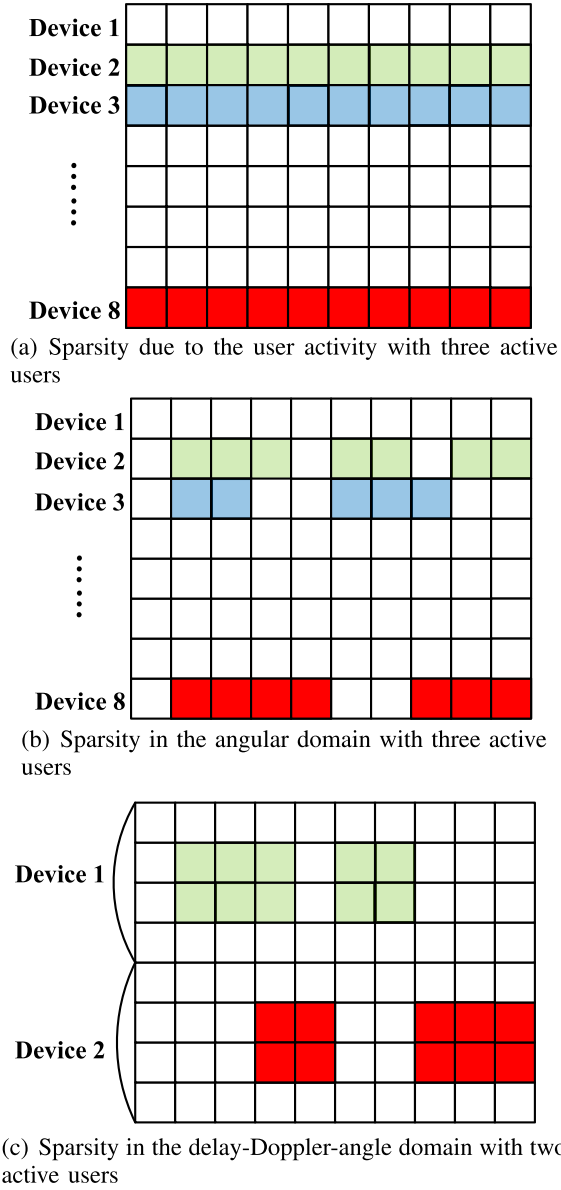
(a) Sparsity due to the user activity with three active users

(b) Sparsity in the angular domain with three active users

(c) Sparsity in the delay-Doppler-angle domain with two active users

Fig. 2. Comparison among different sparsity patterns of the channel matrix.

corrsponding users. To see this, according to (20) and (26), the $(k + 1 + lN + uM_\tau N)$-th row of $\mathbf{H}$ has the dominant elements only if $k \approx NT_{\text{sym}}\nu_{u,i}$ and $l \approx \tau_{u,i}M\Delta f$, and the magnitudes of the elements decrease when k and l move away from $NT_{\text{sym}}\nu_{u,i}$ and $\tau_{u,i}M\Delta f$; the column sparsity of $\mathbf{H}$ results from the channel sparsity in the angular domain, and the starting positions of the clusters in each row depend on the azimuth angle and elevation angle of the received signal from the corrsponding users, since according to (20) and (26), the $(a_z + 1 + N_z a_y)$-th column of $\mathbf{H}$ has dominant elements, only if $a_y \approx N_y\Omega_{y_u}/2$ and $a_z \approx N_z\Omega_{z_u}/2$, and the magnitudes of the elements decrease when a_y and a_z move away from $N_y\Omega_{y_u}/2$ and $N_z\Omega_{z_u}/2$.

III. PATTERN COUPLED SBL FRAMEWORK

In this section, a 2D pattern coupled hierarchical prior is first proposed to capture the 2D burst block sparsity in the channel matrix. Then, the SBL framework is developed to estimate
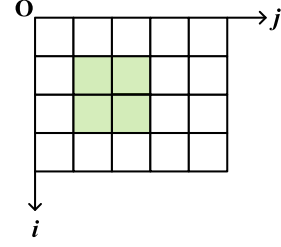
Fig. 3. An example of the sparsity pattern of $\mathbf{H}$ with one user. The non-zero elements are represented by the green boxes, and the other elements are zeros.

the channel matrix using the pilots and the received symbols. Finally, the covariance-free method is adopted to facilitate the computations of the SBL.

A. The Pattern-Coupled Hierarchical Prior

Considering that the sparsity patterns of non-zero coefficients are statistically dependent along the row and column. To explore this 2D sparsity, in the proposed prior, the precision of each coefficient involves the precisions of its neighborhoods. More precisely, the prior on the (i, j)-th element of the channel matrix $\mathbf{H}$ is given by

$$p(h_{i,j}|\mathbf{A}, \mathbf{B}_{i,j}) = \mathcal{CN}\left(h_{i,j}|0, \left(\sum_{p=i-D}^{i+D}\sum_{q=j-D}^{j+D}\beta_{i,j,p,q}\alpha_{p,q}\right)^{-1}\right),$$

(29)

where $\mathbf{A}$ is the precision matrix whose (i, j)-th element $\alpha_{i,j}$ is the local precision hyperparameter of $h_{i,j}$; $\mathbf{B}_{i,j} \in \mathrm{R}^{UM_\tau N \times N_y N_z}$ is the pattern coupled hyperparameter matrix of $h_{i,j}$, and its (p, q)-th element $\beta_{i,j,p,q}$ represents the relevence between $h_{i,j}$ and $h_{p,q}$, and $\beta_{i,j,p,q} \neq 0$ only if $p \in [i - D, i + D]$ and $q \in [j - D, j + D]$; D, called the range indicatior, is a hyperparameter controlling the number of neighborhoods of $h_{i,j}$ in the consideration. Note that when $\beta_{i,j,p,q} = 0$ for arbitrary p and q, except $\beta_{i,j,i,j} = 1$, the proposed prior reduces to the conventional SBL prior [55]; when $\beta_{i,j,p,q} = 0$ for arbitrary p and q, except $\beta_{i,j,i,j-1} = \beta_{i,j,i,j+1} = \zeta$ and $\beta_{i,j,i,j} = 1$, where ζ is a hyperparameter, the proposed prior degrades to the SBL prior capturing the one dimensional block sparsity [56].

It is clearly seen in (29) that the precision of $h_{i,j}$ not only depends on its local hyperparameter $\alpha_{i,j}$, but also on the hyperparameters of its neighborhoods along the row and column of $\mathbf{H}$, where the relevence between them is controlled by the pattern coupled hyperparameter $\mathbf{B}_{i,j}$. It is desirable that $\mathbf{B}_{i,j}$ is assigned by the true sparsity pattern of its neighborhoods. For example, Fig. 3 shows the sparsity pattern of $\mathbf{H}$ with one user, where $h_{1,1}, h_{1,2}, h_{2,1}, h_{2,2}$ are non-zero elements. Suppose $D = 1$ here, and then for $h_{1,1}$, the non-zero pattern can be captured when the coupled hyperparameters $\beta_{1,1,1,1} = \beta_{1,1,1,2} = \beta_{1,1,2,1} = \beta_{1,1,2,2} = 1$, and $\beta_{1,1,0,0} = \beta_{1,1,0,1} = \beta_{1,1,0,2} = \beta_{1,1,1,0} = \beta_{1,1,2,0} = 0$. Similarly, for $h_{0,0}$, the zero pattern can be captured when $\beta_{0,0,0,0} = \beta_{0,0,0,1} = \beta_{0,0,1,0} = 1$, and $\beta_{0,0,1,1} = 0$. Therefore, such coupled prior has the potential to capture the unknown 2D sparsity structure in $\mathbf{H}$, where the pattern coupled hyperparameter $\mathbf{B}_{i,j}$ plays an important role, and we will

discuss the update of it later. Combined with (29), the prior distribution of $\mathbf{H}$ is given by

$$
\begin{aligned}
p(\mathbf{H}|\mathbf{A},\mathbf{B}) &= \prod_{i=0}^{UM_\tau N-1}\prod_{j=0}^{N_y N_z-1} p\left(h_{i,j}|\mathbf{A},\mathbf{B}_{i,j}\right) \\
&= \prod_{j=0}^{N_y N_z-1} \mathcal{CN}(\mathbf{h}_j|0,\Upsilon_j), \quad (30)
\end{aligned}
$$

where $\mathbf{B}$ contains all the pattern coupled hyperparameters, whose (i,j)-th element is $\mathbf{B}_{i,j}$; $\mathbf{h}_j = [h_{0,j},\ldots,h_{UM_\tau N-1,j}]$ and represents the j-th column of $\mathbf{H}$; $\Upsilon_j = \mathrm{diag}(\gamma_{0,j},\ldots,\gamma_{UM_\tau N-1,j})$ is the covariance matrix of $\mathbf{h}_j$, and $\gamma_{i,j} = \left(\sum_{p=i-D}^{i+D}\sum_{q=j-D}^{j+D}\beta_{i,j,p,q}\alpha_{p,q}\right)^{-1}$. Then, following the conventional SBL, we adopt the Gamma distributions as the hierarchical prior for the local precision hyperparameter $\mathbf{A}$, i.e.,

$$
p(\mathbf{A}|a,b) = \prod_{i=0}^{UM_\tau N-1}\prod_{j=0}^{N_y N_z-1} \Gamma(a)^{-1}b^a\alpha_{i,j}^a e^{-b\alpha_{i,j}}, \quad (31)
$$

where $\Gamma(c) \triangleq \int_0^\infty t^{c-1}e^{-t}dt$. This hierarchical Beysian modeling provides the analytically tractable solutions for the estimation of $\mathbf{H}$, and encourages the sparseness in the estimation since the overall prior of $\mathbf{H}$ (i.e., integrate out $\alpha_{i,j}$) is a Student-t distribution which is sharply peaked at zero. In the conventional SBL, a and b are usually assigned by the small values (e.g., 10^{-4}), called the non-informative prior [55]. In this paper, we adopt a informative prior and set $a = 1$ since as we stated above, the channel in the delay-Doppler-angle domain is very sparse and the larger a will promote the sparseness in the estimation [56].

B. SBL Framework

1) MMSE Estimation of $\mathbf{H}$: We now develop the SBL framework for the 2D burst block sparse signal recovery. Recalling that the noise obey a Guassian distribution with the precision θ, and then the likelihood function is given by

$$
p(\mathbf{Y}|\mathbf{H},\theta) = \prod_{j=0}^{N_y N_z-1} \mathcal{CN}(\mathbf{y}_j|\mathbf{X}\mathbf{h}_j,\theta^{-1}\mathbf{I}_{UM_\tau N}), \quad (32)
$$

where $\mathbf{y}_j = [y_{0,j},\ldots,y_{UM_\tau N-1}]$ is the j-th column of $\mathbf{Y}$. According to the Bayes' rule, the posterior distribution of $\mathbf{H}$ is given as

$$
p(\mathbf{H}|\mathbf{Y},\mathbf{A},\mathbf{B},\theta) = \frac{p(\mathbf{Y}|\mathbf{H},\theta)p(\mathbf{H}|\mathbf{A},\mathbf{B})}{\int p(\mathbf{Y}|\mathbf{H},\theta)p(\mathbf{H}|\mathbf{A},\mathbf{B})d\mathbf{H}}. \quad (33)
$$

Since in (33), the distributions appearing in the numerator are Gaussians, and the normalising integral is a convolution of Gaussians. Thus, the posterior distribution of $\mathbf{H}$ is also Gaussian, and is given by

$$
p(\mathbf{H}|\mathbf{Y},\mathbf{A},\mathbf{B},\theta) = \prod_{j=0}^{N_y N_z-1} \mathcal{CN}(\mathbf{h}_j|\mu_j,\Sigma_j), \quad (34)
$$

where

$$
\mu_j = \theta\Sigma_j\mathbf{X}^{\mathrm{H}}\mathbf{y}_j, \quad \Sigma_j = (\theta\mathbf{X}^{\mathrm{H}}\mathbf{X}+\Upsilon_j^{-1})^{-1}. \quad (35)
$$

Given a set of estimated hyperparameters $\{\mathbf{A},\mathbf{B},\theta\}$, the MMSE estimation of $\hat{\mathbf{H}}$ is the mean of its posterior distribution, i.e.,

$$
\hat{\mathbf{H}} = [\mu_0,\ldots,\mu_{N_y N_z-1}]. \quad (36)
$$

Then, our problem reduces to estimating the set of hyperparameters.

2) Learning the Local Hyperparameters: Follwing the conventional SBL, we place a Gamma hyperpriors over θ, i.e.,

$$
p(\theta|r,s) = \Gamma(r)^{-1}s^r\theta^r e^{-s\theta}. \quad (37)
$$

Then, the precision matrix $\mathbf{A}$ and the noise precision θ can be estimated by maximizing their log-posterior distribution with respect to $\mathbf{A}$ and θ, i.e.,

$$
\hat{\mathbf{A}},\hat{\theta} = \arg\max_{\mathbf{A},\theta} \log p(\mathbf{A},\theta|\mathbf{Y},\mathbf{B}), \quad (38)
$$

where the pattern coupled hyperprameter $\mathbf{B}$ is assumed to be known in this step.

Next, we adopt the EM algorithm to approximate the optimal solutions in (38), where the channel matrix $\mathbf{H}$ is treated as the hidden variables. Given the estimations of the hyperparameters $\mathbf{A}^t$, θ^t in the t-th iteration, and the received symbols $\mathbf{Y}$, we denote $\mathrm{E}_{\mathbf{H}|\mathbf{Y},\mathbf{A}^t,\theta^t}[\cdot]$ as $\mathrm{E}[\cdot]$ for notation simplicity, where the expection is with respect to the channel $\mathbf{H}$ over the distribution $p(\mathbf{H}|\mathbf{Y},\mathbf{A}^t,\theta^t)$. Then, the Q-function is given by

$$
\begin{aligned}
Q(\mathbf{A},\theta|\mathbf{A}^t,\theta^t) &= \mathrm{E}[\log p(\mathbf{Y},\mathbf{H},\mathbf{A},\theta|\mathbf{B})] \\
&= \mathrm{E}[\log p(\mathbf{Y}|\mathbf{H},\theta)p(\theta)] + \mathrm{E}[\log p(\mathbf{H}|\mathbf{A},\mathbf{B})p(\mathbf{A})]. \quad (39)
\end{aligned}
$$

It is observed in (39) that the parameters $\mathbf{A}$ and θ are seperated from each other. This allows the estimations of $\mathbf{A}$ and θ to be decoupled into the two independent problems, i.e.,

$$
\mathbf{A}^{t+1} = \arg\max_{\mathbf{A}} \mathrm{E}[\log p(\mathbf{H}|\mathbf{A},\mathbf{B})p(\mathbf{A})], \quad (40)
$$

$$
\theta^{t+1} = \arg\max_{\theta} \mathrm{E}[\log p(\mathbf{Y}|\mathbf{H},\theta)p(\theta)]. \quad (41)
$$

The optimization of (40) in the conventional SBL can be decoupled into a number of separate problems in which each hyperparameter is updated independently [55], and usually, the analytic solution is easy to obtain. However, it is not the case for the problem being considered here. To see this, (40) can be written as

$$
\begin{aligned}
\mathbf{A}^{t+1} = \arg\max_{\mathbf{A}} \sum_i\sum_j & \log\left(\sum_{p=i-D}^{i+D}\sum_{q=j-D}^{j+D}\beta_{i,j,p,q}\alpha_{p,q}\right) \\
& -\left(\sum_{p=i-D}^{i+D}\sum_{q=j-D}^{j+D}\beta_{i,j,p,q}\alpha_{p,q}\right)\mathrm{E}[|h_{ij}|^2] + a\log\alpha_{i,j} - b\alpha_{i,j},
\end{aligned}
$$
$$
(42)
$$

where the hyperparameters in (42) are entangled with each other due to the logarithm term. Therefore, the analytical

solution of the optimization is difficult to obtain. Besides, although the gradient descent methods can be adopted to get the optimal solution, it involves higher computational complexity as compared with an analytical update rule. Similar to [55], we derive a closed form sub-optimal solution by examining the optimality in (42), as shown in the following proposition.

Proposition 2: The range of the optimal (i, j)-th element $\alpha_{i,j}^*$ in the $\mathbf{A}^{\mathrm{t}+1}$ is given by

$$\left[\frac{a}{b + \omega_{i,j}^t}, \frac{a + (2D + 1)^2}{b + \omega_{i,j}^t} \right], \tag{43}$$

where $\omega_{i,j}^t = \sum_{p=i-D}^{i+D} \sum_{q=j-D}^{j+D} \beta_{p,q,i,j} \mathrm{E}[|h_{pq}|^2]$.

Proof: See Appendix B.

Therefore, a sub-optimal solution to (40) can be simply chosen as

$$\alpha_{i,j}^{t+1} = \hat{\alpha}_{i,j} = \frac{a}{b + \omega_{i,j}^t}. \tag{44}$$

Finally, the second moment $\mathrm{E}[|h_{pq}|^2] = \left|\mu_{p,q}^t\right|^2 + \left|\Sigma_{p,q}^t\right|^2$, where $\mu_{p,q}^t$ is the p-th element of the posterior mean μ_q^t in the t-th iteration, and $\Sigma_{p,q}^t$ is the (p, p)-th entry of the covariance matrix $\mathbf{\Sigma}_q^t$.

For the estimation of θ, (41) can be written as

$$\theta^{t+1} = \arg\max_{\theta} \sum_{j=0}^{N_y N_z - 1} \left(MN \log \theta - \theta \mathrm{E}\left[\|\mathbf{y}_j - \mathbf{X}\mathbf{h}_j\|^2 \right] \right) + r \log \theta - s\theta. \tag{45}$$

Computing the first derivative of the objective function in (45) with respect to θ and setting it equal to zero, we get

$$\theta^{t+1} = \hat{\theta} = \frac{M_\tau N N_y N_z + r}{\sum_{j=0}^{N_y N_z - 1} \mathrm{E}\left[\|\mathbf{y}_j - \mathbf{X}\mathbf{h}_j\|^2 \right] + s}, \tag{46}$$

where

$$\mathrm{E}\left[\|\mathbf{y}_j - \mathbf{X}\mathbf{h}_j\|^2 \right] = \left\| \mathbf{y}_j - \mathbf{X}\mu_j^t \right\|^2 + (\theta^t)^{-1}$$
$$\times \sum_{i=0}^{UM_\tau N - 1} (1 - \Sigma_{i,j}^t (\gamma_{i,j}^t)^{-1}), \tag{47}$$

and $\gamma_{i,j}^t$ is the prior variance of $h_{i,j}$ in the t-th iteration.

3) Updating the Coupled Hyperparameters: As discussed in the last subsection, the coupled hyperparameter matrices $\mathbf{B}_{i,j}$ of $h_{i,j}$ can be assigned by the sparsity pattern of its neighborhoods. Here, we will first estimate the support of $\mathbf{H}$ based on the current estimations of hyperparameters and then, use it to update $\mathbf{B}$. Consider the following binary hypothesis testing

$$\mathcal{H}_0 : \gamma_{i,j} = 0$$
$$\mathcal{H}_1 : \gamma_{i,j} > 0, \tag{48}$$

where $\mathcal{H}_0(\mathcal{H}_1)$ represents that $h_{i,j}$ is zero (non-zero). Similar to [57], [58], we consider the log-likelihood ratio (LLR) test, given by

$$\log \frac{p(\mathbf{y}_j|\mathcal{H}_1)}{p(\mathbf{y}_j|\mathcal{H}_0)} \geqslant \xi, \tag{49}$$

where ξ is the detection threshold. Similar to the denominator in (33), $p(\mathbf{y}_j|\mathcal{H}_1)$ as well as $p(\mathbf{y}_j|\mathcal{H}_0)$ can be written as the convolution of two Gaussians, and thus is given by

$$p(\mathbf{y}_j|\mathcal{H}_1) = \mathcal{CN}(0, \theta^{-1}\mathbf{I}_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j\mathbf{X}^{\mathrm{H}}), \tag{50}$$
$$p(\mathbf{y}_j|\mathcal{H}_0) = \mathcal{CN}(0, \theta^{-1}\mathbf{I}_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}}), \tag{51}$$

where $\mathbf{\Upsilon}_j' = [\gamma_{0,j}, \ldots, \gamma_{i-1,j}, 0, \gamma_{i+1,j}, \ldots, \gamma_{UM_\tau N - 1, j}]$. Then, the LLR test in (49) is equivalent to

$$\frac{\left| \mathbf{x}_i^{\mathrm{H}} \left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right)^{-1} \mathbf{y}_j \right|^2}{\mathbf{x}_i^{\mathrm{H}} \left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right)^{-1} \mathbf{x}_i} \geqslant \bar{\xi}. \tag{52}$$

According to the Neyman-Pearson lemma [59], the threshold $\bar{\xi} = \log \frac{1}{P_{fa}}$, and P_{fa} is false alarm probability, since the random variable on the left side of the inequality obeys the chi-square distribution with two degrees of freedom under the hypothesis $\mathcal{H}_0$. We denote the support of $\mathbf{H}$ as $\mathbf{S}_H$, and then according to (52), the (i, j)-th element of the $\mathbf{S}_H$ is given by

$$S_{H_{i,j}} = \mathbb{I}\left\{ \frac{\left| \mathbf{x}_i^{\mathrm{H}} \left(\theta^{-1}\mathbf{I}_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right)^{-1} \mathbf{y}_j \right|^2}{\mathbf{x}_i^{\mathrm{H}} \left(\theta^{-1}\mathbf{I}_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right)^{-1} \mathbf{x}_i} \geqslant \log \frac{1}{P_{fa}} \right\} \tag{53}$$

where $i = 0, \ldots, UM_\tau N - 1$, $j = 0, \ldots, N_y N_z - 1$, and $\mathbb{I}\{\cdot\}$ is the indicatior function which is 1 if the condition inside the braces is fulfilled and 0 otherwise. Based on $\mathbf{S}_H$, the pattern coupled hyperparameter $\mathbf{B}_{i,j}$ can be updated by the support of the neighborhoods of $h_{i,j}$. For $p \in [i - D, i + D]$ and $q \in [j - D, j + D]$, $\beta_{i,j,p,q}$ is given by

$$\beta_{i,j,p,q} = \begin{cases} S_{H_{p,q}}, & \text{if } S_{H_{i,j}=1}, \\ 1 - S_{H_{p,q}}, & \text{otherwise.} \end{cases} \tag{54}$$

Finally, based on the estimation of the support of $\mathbf{H}$, the device activity detection rule is given by

$$\hat{\lambda}_u = \mathbb{I}\left\{ \sum_{i=uM_\tau N}^{(u+1)M_\tau N - 1} \sum_{j=0}^{N_y N_z - 1} S_{H_{i,j}} > 0 \right\}. \tag{55}$$

4) The Covariance-Free Method: Indeed, the above method can effectively detect the device activity and estimate the channel. However, to compute the covariance matrix in (35) and (52), the costly matrix inversions are involved, which limits its application for the large-scale systems. Here, we adopt the covariance-free method [60] to facilitate the computations of the above SBL algorithm. The main idea is that transforming the matrix inversions into solving the linear equations which can be computed faster by the effective iterative algorithms. To see this, we re-express (35) and the terms in (52) involved the matrix inversions as

$$(\theta\mathbf{X}^{\mathrm{H}}\mathbf{X} + \mathbf{\Upsilon}_j^{-1})\mu_j = \theta\mathbf{X}^{\mathrm{H}}\mathbf{y}_j \tag{56}$$
$$\left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right) \mathbf{t}_{nu} = \mathbf{y}_j \tag{57}$$
$$\left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right) \mathbf{t}_{de} = \mathbf{x}_i. \tag{58}$$

Therefore, the posterior mean μ_j and the terms in (52) involved the matrix inversions are the solutions of the above linear equations, where $\mathbf{t}_{nu} = \left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\mathbf{\Upsilon}_j'\mathbf{X}^{\mathrm{H}} \right)^{-1} \mathbf{y}_j$

and $\mathbf{t}_{de} = \left(\theta^{-1}I_{M_\tau N} + \mathbf{X}\Upsilon'_j\mathbf{X}^{\mathrm{H}}\right)^{-1}\mathbf{x}_i$. Note that except for the posterior mean, the update of the hyperparameter still need the posterior variance of the channel $h_{i,j}$ (i.e., the diagonal elements of the covariance matrix Σ_j), which can be estimated by the following proposition [61].

Proposition 3: Let $\mathbf{S}$ be any square matrix of size $W \times W$. Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_K \in \mathbb{R}^W$ be K random probe vectors, where each $\mathbf{p}_k$ is comprised of independent and identically distributed components such that $\mathrm{E}[\mathbf{p}_k] = \mathbf{0}$. Consider the estimator

$$\mathbf{s}_d = \left(\sum_{k=1}^K \mathbf{p}_k \odot \mathbf{S}\mathbf{p}_k\right) \oslash \left(\sum_{k=1}^K \mathbf{p}_k \odot \mathbf{p}_k\right),$$

where $\oslash$ denotes element-wise division between two vectors. Then, the vector $\mathbf{s}_d$ is an unbiased estimator of the diagonal elements of $\mathbf{S}$.

We apply this diagonal estimation rule to Σ_j to estimate its diagonal elements vector $\mathbf{s}_{d_j}$. The simplest distribution to use in drawing probe vectors $\mathbf{p}_k$ is the Randemacher distribution, which lets each independent component of $\mathbf{p}_k$ be either -1 or $+1$ with equal probability. In this case, the diagonal estimator $\mathbf{s}_{d_j}$ simplifies to

$$\mathbf{s}_{d_j} = \frac{1}{K}\sum_{k=1}^K \mathbf{p}_k \odot \Sigma_j\mathbf{p}_k. \tag{59}$$

Similarly, the term $\Sigma_j\mathbf{p}_k$ can be re-expressed as

$$(\theta\mathbf{X}^{\mathrm{H}}\mathbf{X} + \Upsilon_j^{-1})\mathbf{t}_{d,j} = \mathbf{p}_k, \tag{60}$$

where $\mathbf{t}_{d,j} = \Sigma_j\mathbf{p}_k$. Until now, all the needed computations involved the matrix inversions have been transformed into solving the linear equations, and then the conjugate gradient (CG) algorithm [62] can be adopted to tackle these linear equations.

The TDSBL-FM[1] is summarized in Algorithm 1. Note that in the Algorithm 1, we collect all the activity indicators λ_u into a single vector $\lambda = [\lambda_0, \ldots, \lambda_{U-1}]$, and the algorithm will output its estimation $\hat{\lambda}$. The complexity of TDSBL-FM mainly depends on the complexity of iterative algorithms solving the linear equations. For example, when the CG algorithm is adopted, the complexity is $\mathcal{O}((K+1)N_yN_z(UM_\tau N)^2 + 2(M_\tau N)^3UN_yN_z)$ in each iteration, where the quadratic term is due to the CG algorithm for estimating the posterior mean and variance, and the cubic term is due to the LLR test for each element in the channel matrix. Compared with the complexity $\mathcal{O}(N_yN_z(UM_\tau N)^3 + 2(M_\tau N)^4UN_yN_z)$ of using the matrix inversions directly, the TDSBL-FM has greatly reduced the computations. However, it is still prohibitive for the very large-scale systems, since typically the number of carriers adopted in the OTFS and the potential devices are huge. To facilitate the TDSBL-FM for solving the large-scale problems, the GAMP-based algorithm is developed in the next section.

[1] For the LEO satellite communications, the complexity of the TDSBL-FM is acceptable when the number of potential devices is small or moderate, and the grant-free mechanism offers the possibility of low access latency.

Algorithm 1 TDSBL-FM Algorithm

Require: Recieved symbols $\mathbf{Y}$, pilot matrix $\mathbf{X}$; the maximum number of iterations T_{SBL}, the termination threshold η, the false alarm probability P_{fa}, the range indicatior D, and the number of probes K.

Ensure: The estimated channel matrix $\hat{\mathbf{H}}$ and the activity indicatior vector $\hat{\lambda}$.

1: Initialization: $\forall i, j$: $\mu_{i,j}^0 = 0$, $\mu_{i,j}^1 = 10^{-8}$ $\gamma_{i,j}^0 = 10^{-2}$, $\beta_{i,j,p,q}^0 = 1$ for $p \in [i-D, i+D]$, $q \in [j-D, j+D]$, and 0 otherwise; $\theta^0 = 10^3$, and $t = 0$.
2: **while** $t < T_{SBL}$ & $\sum_{j=0}^{N_yN_z-1} \frac{\|\mu_j^t - \mu_j^{t+1}\|^2}{\|\mu_j^t\|^2} > \eta$ **do**
3: $\quad t \leftarrow t + 1$
4: $\quad \forall i, j$: Update posterior mean $\mu_{i,j}^t$ and variance $\Sigma_{i,j}^t$ by solving the linear equations in (56), (59) and (60).
5: $\quad$ Update the noise precision θ^t according to (46).
6: $\quad \forall i, j$: Update the local precision parameter $\alpha_{i,j}$ according to (44).
7: $\quad \forall i, j$: Update the support matrix $S_{H_{i,j}}^t$ according to (52) and by solving the linear equations (57) and (58).
8: $\quad \forall i, j$: Update $\beta_{i,j,p,q}^t$ according to (54)
9: $\quad \forall u$: Update λ_u^t according to (55).
10: **end while**
11: **return** $\hat{\mathbf{H}} = [\mu_0^t, \ldots, \mu_{N_yN_z-1}^t]$ and $\hat{\lambda} = [\lambda_0^t, \ldots, \lambda_{U-1}^t]$

IV. LOW COMPLEXITY SBL WITH GAMP

In this section, we further decrease the complexity of the above SBL framework by adopting the GAMP algorithm to estimate the posterior mean and variance of the channel matrix, and by considering a specical case of the proposed prior without updating the coupled hyperparameters. Instead of approximating the solution of (35), the GAMP algorithm directly estimates $\mu_{i,j}$ and $\Sigma_{i,j}$ by utilizing the factorization of the prior and likelihood function of the channel matrix.

Recalling that the prior of $h_{i,j}$ in (30) and the likelihood function in (32) can be written as the product of multiple probability density functions, and then the posterior distribution of $h_{i,j}$ is given by

$$p(h_{i,j}|\mathbf{Y}, \mathbf{A}, \mathbf{B}, \theta)$$
$$\propto \int p(\mathbf{y}_j|\mathbf{g}_j, \theta)p(\mathbf{h}_j|\mathbf{A}, \mathbf{B})dh_{\sim i,j}$$
$$= \int \prod_{o=0}^{M_\tau N-1} p(y_{o,j}|g_{o,j}, \theta) \prod_{i=0}^{UM_\tau N-1} p(h_{i,j}|\mathbf{A}, \mathbf{B})dh_{\sim i,j}, \tag{61}$$

where $\int(\cdot)dh_{\sim i,j}$ represents the integral with respect to $\mathbf{h}_j$ except $h_{i,j}$, $\mathbf{g}_j = \mathbf{X}\mathbf{h}_j$, and $g_{o,j}$ is the o-th element of $\mathbf{g}_j$. This implies that the message passing algorithms can be adopted to estimate the posterior mean and variance of $h_{i,j}$, instead of directly computing the matrix inversion in (35), which will deacrease the computational complexity sharply. Here, similar to [63], we resort to the GAMP algorithm. According to [64], based on the estimations of all the hyperparameters in the t-th iteration, the approximations of the posterior mean and

variance of $h_{i,j}$ in $(t+1)$-th iteration is given by

$$\mu_{i,j}^{t+1} = \mathrm{E}[h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t}, \mathbf{A}, \mathbf{B}], \quad (62)$$

$$\Sigma_{i,j}^{t+1} = \mathrm{Var}[h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t}, \mathbf{A}, \mathbf{B}], \quad (63)$$

where the expectation is with respect to $p(h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t}, \mathbf{A}, \mathbf{B})$, which is related to the prior of $h_{i,j}$ and given by

$$p(h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t}, \mathbf{A}, \mathbf{B}) = \frac{p(h_{i,j}|\mathbf{A},\mathbf{B})\mathcal{CN}(h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t})}{\int p(h_{i,j}|\mathbf{A},\mathbf{B})\mathcal{CN}(h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t})dh_{i,j}}, \quad (64)$$

Since the prior of $h_{i,j}$ is Guassian, and thus $p(h_{i,j}|\hat{r}_{i,j}^t, \tau_{i,j}^{r,t}, \mathbf{A}, \mathbf{B})$ is also Gaussian. Then, $\mu_{i,j}^{t+1}$ and $\Sigma_{i,j}^{t+1}$ can be easily computed as

$$\mu_{i,j}^{t+1} = \frac{\gamma_{i,j}^t \hat{r}_{i,j}^t}{\gamma_{i,j}^t + \tau_{i,j}^{r,t}}, \quad (65)$$

$$\Sigma_{i,j}^{t+1} = \frac{\gamma_{i,j}^t \tau_{i,j}^{r,t}}{\gamma_{i,j}^t + \tau_{i,j}^{r,t}}, \quad (66)$$

where

$$\tau_{i,j}^{r,t} = \left(\sum_o |X_{o,i}|^2 \tau_{o,j}^{s,t} \right)^{-1}, \quad (67)$$

$$\hat{r}_{i,j}^t = \mu_{i,j}^t + \tau_{i,j}^{r,t} \sum_o X_{o,i}^* \hat{s}_{o,j}^t, \quad (68)$$

and

$$\hat{s}_{o,j}^t = \frac{1}{\tau_{o,j}^{p,t}} \left(\mathrm{E}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right] - \hat{p}_{o,j}^t \right), \quad (69)$$

$$\tau_{o,j}^{s,t} = \frac{1}{\tau_{o,j}^{p,t}} \left(1 - \frac{1}{\tau_{o,j}^{p,t}} \mathrm{Var}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right] \right). \quad (70)$$

The expectation in (69) is with respect to $p(g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t})$, which is related to the likelihood function and is given by

$$p(g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}) = \frac{p(y_{o,j}|g_{o,j})\mathcal{CN}(g_{o,j}|\hat{p}_{o,j}^t, \tau_{o,j}^{p,t})}{\int p(y_{o,j}|g_{o,j})\mathcal{CN}(g_{o,j}|\hat{p}_{o,j}^t, \tau_{o,j}^{p,t})dg_{o,j}}, \quad (71)$$

Note that the likelihood is also Gaussian, and so is $p(g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t})$. Then, we can get

$$\mathrm{E}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right] = \hat{p}_{o,j}^t + \frac{\tau_{o,j}^{p,t}}{\tau_{o,j}^{p,t} + (\theta^t)^{-1}}(y_{o,j} - \hat{p}_{o,j}^t), \quad (72)$$

$$\mathrm{Var}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right] = \frac{\theta^{-1}\tau_{o,j}^{p,t}}{(\theta^t)^{-1} + \tau_{o,j}^{p,t}}, \quad (73)$$

where

$$\tau_{o,j}^{p,t} = \sum_i |X_{o,i}|^2 \Sigma_{i,j}^t, \quad (74)$$

$$\hat{p}_{o,j}^t = \sum_i (X_{o,i}\hat{\mu}_{i,j}^t) - \hat{s}_{o,j}^{t-1}\tau_{o,j}^{p,t}, \quad (75)$$

Then, to further decrease the complexity of the proposed SBL method, we consider a special case of the proposed prior in (29), where we set the coupled parameter $\beta_{i,j,i,j} = 1$, and

$\beta_{i,j,p,q} = \beta$ for $p \in [i-D, i+D]$ and $q \in [j-D, j+D]$. Then, (29) is simplified as

$$p(h_{i,j}|\mathbf{A}, \mathbf{B}_{i,j})$$
$$= \mathcal{CN}\left(h_{i,j}|0, \left(\alpha_{i,j} + \beta\left(\sum_{p=i-D, p\neq i}^{i+D} \sum_{q=j-D, q\neq j}^{j+D} \alpha_{p,q}\right)\right)^{-1}\right), \quad (76)$$

Here, the single parameter β is adopted to represent the relevence between the $h_{i,j}$ and its neighborhoods, and will not be updated using the sparsity pattern of $\mathbf{H}$ for decreasing the computations. Then, the variance of $h_{i,j}$ in the t-th iteration can be computed by the 2D convolution between the local precison matrix $\mathbf{A}^t$ and a kernel $\mathcal{B}$, i.e.,

$$(\gamma_{i,j}^t)^{-1} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \mathbf{A}_{m,n}^t \mathcal{B}_{i-m, j-n}, \quad (77)$$

where $\mathcal{B} \in \mathrm{R}^{(2D+1)\times(2D+1)}$ and only the (D, D)-th entry of $\mathcal{B}$ is equal to one, and other entries are β. Besides, according to (44), $\alpha_{i,j}^{t+1}$ is given by

$$\alpha_{i,j}^{t+1} = \frac{a}{b + \omega^{t+1}}, \quad (78)$$

where ω^{t+1} is the 2D convolution between the kernel $\mathcal{B}$ and the second moment matrix Ψ^t whose (i, j)-th element is $\left|\mu_{i,j}^t\right|^2 + \left|\Sigma_{i,j}^t\right|^2$, and then is given by

$$\omega^{t+1} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \Psi_{m,n}^t \mathcal{B}_{i-m, j-n}, \quad (79)$$

Note that the special prior leads to the 2D-convolution-based update in the EM algorithm, considering the elements both along the delay-Doppler domain and angular domain, and thus matches the 2D burst block sparsity. Finally, the precision of noise can be updated according to (46). When the channel has been estimated, we propose to adopt the energy-based detector to get the active device, i.e.,

$$\hat{\lambda}_u = \mathbb{I}\left\{ \sum_{i=uM_\tau N}^{(u+1)M_\tau N-1} \sum_{j=0}^{N_y N_z - 1} \left|\hat{H}_{i,j}\right|^2 > \xi_{th} \right\}, \quad (80)$$

where ξ_{th} is a pre-defined threshold.

The ConvSBL-GAMP is summarized in Algorithm 2. Note that although each device transmits the independent Gaussian pilots, the elements of the pilot matrix $\mathbf{X}$ are not independent due to the 2D convolution in (23), which may lead to divergence of the conventional AMP [12]. Therefore, the damping parameter ρ is added in Algorithm 2 (see *lines* $5, 7-9$), to guarantee the convergence of ConvSBL-GAMP. The basic computation in the ConvSBL-GAMP is the matrix multiplication leading to the complexity $\mathcal{O}(U(M_\tau N)^2 N_y N_z)$ in each iteration. Compared with the complexity $\mathcal{O}((K + 1)N_y N_z(UM_\tau N)^2 + 2(M_\tau N)^3 UN_y N_z)$ of TDSBL-FM, the computations in ConvSBL-GAMP has been greatly reduced. Therefore, the ConvSBL-GAMP is more suitable for the large-scale systems.

Algorithm 2 ConvSBL-GAMP Algorithm

Require: Recieved symbols $\mathbf{Y}$, pilot matrix $\mathbf{X}$; the maximum number of iterations T_{gamp}, the termination threshold η_g, the damping parameter ρ, the detection threshold ξ_{th}, and the coupled parameter β.

Ensure: The estimated channel matrix $\hat{\mathbf{H}}$ and the activity indicatior vector $\hat{\lambda}$.

1: Initialization: $\forall i,j,o$: $\mu_{i,j}^0 = 0$, $\mu_{i,j}^1 = 10^{-8}$, $\gamma_{i,j}^0 = 10^{-2}$, $\Sigma_{i,j}^1 = \gamma_{i,j}^0$, $\hat{s}_{o,j}^0 = \tau_{o,j}^{p,0} = \bar{\mu}_{i,j}^{t-1} = 0$; $t = 0$.

2: \\ **Estimate the channel matrix**.

3: **while** $t < T_{gamp}$ & $\sum_{j=0}^{N_y N_z - 1} \frac{\|\mu_j^t - \mu_j^{t+1}\|^2}{\|\mu_j^t\|^2} > \eta_g$ **do**

4:     $t \leftarrow t + 1$

5:     $\forall o, j$: $\tau_{o,j}^{p,t} = \rho \sum_i |X_{o,i}|^2 \Sigma_{i,j}^t + (1-\rho)\tau_{o,j}^{p,t-1}$

6:     $\forall o, j$: Update $\hat{p}_{o,j}^t$ according to (75).

7:     $\forall o, j$: Update
       $\hat{s}_{o,j}^t = \frac{\rho}{\tau_{o,j}^{p,t}} \left(\mathrm{E}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right] - \hat{p}_{o,j}^t \right) + (1-\rho)\hat{s}_{o,j}^{t-1}$

8:     $\forall o, j$: Update
       $\tau_{o,j}^{s,t} = \frac{\rho}{\tau_{o,j}^{p,t}}(1 - \frac{1}{\tau_{o,j}^{p,t}}\mathrm{Var}\left[g_{o,j}|y_{o,j}, \hat{p}_{o,j}^t, \tau_{o,j}^{p,t}\right]) + (1-\rho)\tau_{o,j}^{s,t-1}$

9:     $\forall i, j$: Update $\bar{\mu}_{i,j}^t = \rho\mu_{i,j}^t + (1-\rho)\bar{\mu}_{i,j}^{t-1}$

10:    $\forall i, j$: Update $\tau_{i,j}^{r,t}$ and $\hat{r}_{i,j}^t$ according to (67) and (68), where $\mu_{i,j}^t$ is replaced for $\bar{\mu}_{i,j}^t$.

11:    $\forall i, j$: Update $\mu_{i,j}^t$ and $\Sigma_{i,j}^t$ according to (65) and (66).

12:    \\ **Update the hyperparameters**.

13:    Update the noise precision θ^{t+1} according to (46).

14:    $\forall i, j$: Update the local precision parameter $\alpha_{i,j}^{t+1}$ according to (78) and (79).

15:    $\forall i, j$: Update the $\gamma_{i,j}^t$ according to (77).

16:    $\forall u$: Update λ_u^t according to (55).

17: **end while**

18: $\hat{\mathbf{H}} \leftarrow [\mu_0^t, \ldots, \mu_{N_y N_z - 1}^t]$

19: \\ **Detect the active device**.

20: Update $\hat{\lambda} = [\lambda_0^t, \ldots, \lambda_{U-1}^t]$ according to (80).

21: **return** $\hat{\mathbf{H}}$ and $\hat{\lambda}$.

V. NUMERICAL RESULTS

In this section, simulations are utilized to demonstrate the performance of the proposed algorithms. We consider the senarios of the non-terrestrial networks recommended by the 3GPP [53] and assume that the delay of each device is precompensated in the initial downlink synchronazation so that the residual delay is in a small range. The detailed system parameters are summarized in Table I. Here, we adopt the all-ones matrix to approximate the phase compensation matrix (see (27) and (28)) and set $D = a = 1$, $b = r = s = 10^{-4}$ and $P_{fa} = 10^{-3}$ in our algorithms. Each user transmits Gaussian pliot matrix where entries in the pliot matrix are independent and obey $\mathcal{CN}(0, \frac{1}{M_\tau N})$. Moreover, the power of multi-path channel gain is normalized as 1, i.e., $\sum_{i=0}^p |h_{u,i}|^2 = 1$. Finally, we define the signal-to-noise ratio (SNR) as SNR $= 10\log_{10}\frac{\theta}{MN}$, where θ is the precision of the Gaussian noise.

To show the effectiveness of the proposed algorithms, we adopt SBL-GAMP [63] and GMMV-AMP [19] as benchmarks, where the SBL-GAMP was based on the conventional

TABLE I

SIMULATION PARAMETERS

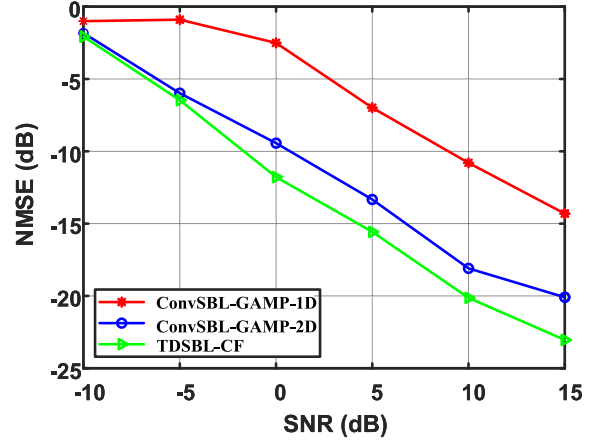| Parameter	Values
Carrier frequency (GHz)	2
Subcarrier spacing (kHz)	330
Size of OTFS symbol (M, N)	$(256, 15)$
Length of CP	85
Number of Satellite antennas	$4/16/36/64/144$
Number of user terminal antennas	1
Number of LoS paths	1
Number of Non-LoS path	3
The residual delay (μs)	$[0, 0.8]$
Doppler shift (kHz)	$[-41, 41]$
Directional cosine along the y- and z-axis	$[-1, 1]$
Rician factor (dB)	5

Fig. 4.   Performance comparison between the ConvSBL-GAMP and TDSBL-CF under different SNR values. ConvSBL-GAMP-1D (ConvSBL-GAMP-2D) represents the ConvSBL-GAMP with the one (two) dimensional kernel. $U = U_a = 3$, $N_y = N_z = 8$, and the pliot overhead is 0.3.

SBL prior only considering the single element in the channel matrix, and the GMMV-AMP explored the sparsity in the angular domain. Besides, the normalized mean square error (NMSE) and the average error probability is adopted as the metrics for the channel estimations and device activity detection, respectively, given by

$$\text{NMSE} = \frac{\left\|\mathbf{H} - \hat{\mathbf{H}}\right\|_F^2}{\|\mathbf{H}\|_F^2}, P_e = \frac{1}{U}\sum_{u=0}^{U-1}\left|\lambda_u - \hat{\lambda}_u\right|. \quad (81)$$

Firstly, as shown in Fig. 4, we compare the performance for the channel estimation between the proposed TDSBL-CF and ConvSBL-GAMP, where the ConvSBL-GAMP-1D (ConvSBL-GAMP-2D) represents that the ConvSBL-GAMP adopts the one (two) dimensional kernel $\mathcal{B}$. For the ConvSBL-GAMP-1D, we set $\beta = 0.5$, and for the ConvSBL-GAMP-2D, we set $\beta = 0.125$, which means that the neighborhoods of the channel $h_{i,j}$ have the same influence on it. Here, we set $U = U_a = 3$, $N_y = N_z = 8$, and the pilot ratio is 0.3 for facilitating the computations of TDSBL-CF. In addition, we assume that the phase compensation is known here and the numerical results are obtained by averaging over 100 channel realizations. It is observed that ConvSBL-GAMP-2D outperforms the ConvSBL-GAMP-1D for different SNR values. For example, in order to achieve the same NMSE, the SNR required by ConvSBL-GAMP-2D is about 8 dB lower than that of the ConvSBL-GAMP-1D,
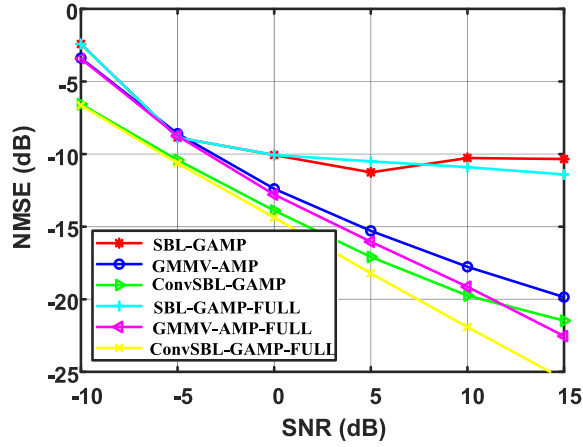
Fig. 5. Performance comparison between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP under different SNR values. $U = 50$, $U_a = 10$, $N_y = N_z = 8$, and the pliot overhead is 0.3.
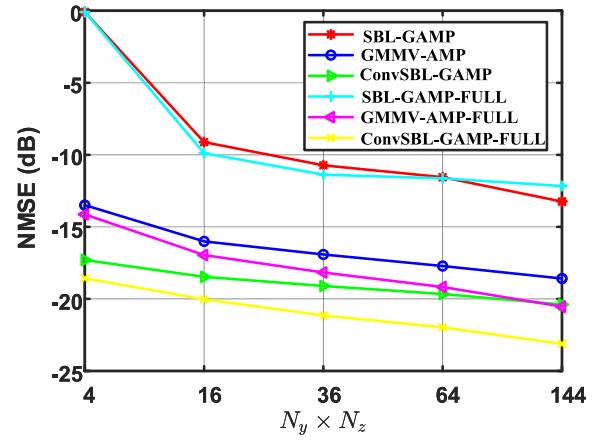

Fig. 6. Performance comparison between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP under different number of antennas. $U = 50$, $U_a = 10$, SNR = 10 dB, and the pliot overhead is 0.3.

since the 1D kernel only accounts for the sparsity in the angular domain while the 2D kernel can capture the 2D burst block sparsity. It is indicated the effectiveness of the proposed 2D kernel. Besides, the NMSE of the TDSBL-CF is always lower than the ConvSBL-GAMP-2D, since the TDSBL-CF computes the posterior mean of the channel more precisely and its coupled hyperparameter can be adjusted automatically in each iteration while ConvSBL-GAMP only adopts the single fixed coupled parameter which is less flexible. However, as we discussed earlier, with a little performance degradation, the computational complexity of ConvSBL-GAMP is far less than the TDSBL-CF, which makes it more suitable for the large-scale systems. Overall, both the TDSBL-CF and ConvSBL-GAMP are effective for estimating the channel matrix. Further, the TDSBL-CF can achieve superior performance while with more computations, and the ConvSBL-GAMP has lower computation complexity but with a little performance loss, which allows us to trade off between the complexity and performance. In the following experiments, we show the performance comparison between the ConvSBL-GAMP and benchmarks for dealing with large dimensional problems, and omit the TDSBL-CF curve since its computation will be prohibitive in those scenarios.

Fig. 5 shows the NMSE comparison between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP. The number of the potential devices and active devices are 50 and 10, respectively, i.e., $U = 50$ and $U_a = 10$, and the numerical results are obtained by averaging over 100 channel realizations. Here, we adopt the identity matrix to approximate the phase compensation matrix, and we also provide the performance comparison between the three algorithms when the phase compensation matrix is fully known (marked by "FULL" in Fig. 5). It is observed that ConvSBL-GAMP outperforms the other algorithms under different SNR values, even when other algorithms have the knowledge of the phase compensation matrix. For example, to achieve the NMSE of −15 dB, the SNR required by the ConvSBL-GAMP is about 3 dB lower than the GMMV-AMP and 1.5 dB lower than the GMMV-AMP-FULL, since the GMMV-AMP only captures

the sparsity in the angular domain, while the ConvSBL-GAMP is able to deal with the sparsity in the angular domain and delay-Doppler domain. Notice that the NMSE of SBL-GAMP is much more than other algorithms and fluctuates around −10 dB when the SNR is larger than 0 dB, since it only accounts for the sparsity of the single element in the channel matrix, and thus fails to capture the sparsities in the angular domain and the delay-Doppler domain. Therefore, utilizing the 2D burst block sparsity is favourable for the channel estimation and our proposed algorithm can capture that feature better compared with the conventional algorithms. In addition, it is observed that the performance of the three algorithms can be improved further when the phase compensation matrix is known. However, the estimation of the phase compensation will bring extra complexity to the algorithms which may limit their applications for the large-scale systems.

Next, as shown in Fig. 6, we provide the performance comparison for the channel estimation under the different number of antennas. Here, we set SNR = 10 dB, the pilot overhead is 0.3, and the numerical results here are obtained by averaging over 100 channel realizations. It is observed that the NMSE of the three algorithms can be decreased when more antennas are adopted, and the ConvSBL-GAMP always outperforms the other algorithms, which indicates its effectiveness. For example, the NMSE of ConvSBL-GAMP is about 3.5 dB less than that of the GMMV-AMP when the 36 antennas are adopted, and about 2.5 dB less than that of the GMMV-AMP-FULL (i.e., the phase compensation matrix is known). In addition, we can see that the NMSE of the ConvSBL-GAMP is about −17.5 dB when the 4 antennas are adopted, and is about 4.5 dB and 17 dB lower than those of the GMMV-AMP and SBL-GAMP, respectively. It is indicated that the ConvSBL-GAMP can estimate the channel matrix precisely even if fewer antennas are adopted.

Then, we show the performance comparison for the channel estimation under different pilot ratios in Fig. 7, where the pilot overhead is defined as the ratio between the number of pilots and total delay-Doppler domain resources, i.e., $\frac{M_\tau N_\nu}{MN}$. Here, we set SNR = 10 dB, and 64 antennas are adopted. It is observed that the NMSE of ConvSBL-GAMP decreases
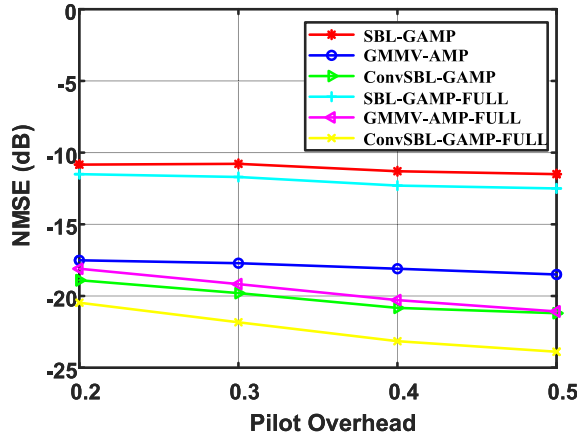
Fig. 7.　Performance comparison between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP under different pilot ratios. $U = 50$, $U_a = 10$, SNR = 10 dB, and $N_y = N_z = 8$.
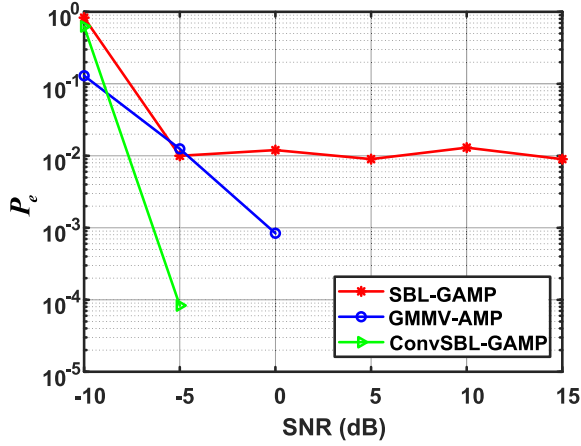

Fig. 8.　Performance comparison of device activity detection between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP under different SNR values. $U = 10$, $U_a = 2$, the pilot overhead is 0.25, and $N_y = N_z = 2$.

with the increase of the pilot overhead and outperforms other algorithms, which indicates its superiority.

Finally, we show the performance comparison for device activity detection between the ConvSBL-GAMP, GMMV-AMP, and SBL-GAMP under different SNR values. The detection threshold $\xi_{th} = 0.5$. There are 10 potential devices, two of which are active, i.e., $U = 10$ and $U_a = 2$. The number of antennas is set as $N_y = N_z = 2$, and the pilot overhead is 0.25. The numerical results here are obtained by averaging over 1200 channel realizations. It is observed that the error probability of the ConvSBL-GAMP is very close to zero when the SNR is larger than -5 dB and is lower than the other algorithms, which indicates that the ConvSBL-GAMP can successfully detect the active devices in the low SNR regime. This may be due to the time-varying channel is considered, and the channel of each device is represented by $M_\tau N$ rows of H, which are much larger than that in the conventional channel matrix [12] (i.e., the channel of each device is represented by only one row of the channel matrix). Then, the difference between the energy of the channel of the active devices and inactive devices are larger than that in the conventional channel matrix, and thus the active device detec-

tion will be accurate when the channel estimation is precise enough. In addition, the error probability of the SBL-GAMP fluctuates around 10^{-2} when the SNR is larger than -5 dB. This may be due to the inaccurate channel estimation of the SBL-GAMP, which is consistent with that in Fig. 5.

VI. CONCLUSION

This work investigated the application of OTFS for the grant-free random access with massive MIMO in LEO satellite communications. The input-output relationship in the SISO-OTFS system was firstly analyzed, and then extended to the random access with massive MIMO-OTFS. Next, by exploring the two-dimensional burst block sparsity in the delay-Doppler-angle domain, the TDSBL-FM was proposed to estimate the channel and detect the active device. Then, ConvSBL-GAMP was proposed to further decrease the complexity of the TDSBL-FM. Simulation results demonstrated that the proposed scheme could acquire accurate channel state information and device activity.

APPENDIX A
PROOF OF PROPOSITION 1

According to (11), we can get

$$
\begin{aligned}
& Y^{\text{TF}}[n, m] \\
&= \sum_{m'} \sum_{n'} X^{\text{TF}}[n', m'] \iint h(\tau, \nu) e^{\jmath 2\pi n' \nu T_{sym}} e^{\jmath 2\pi \nu T_{cp}} \\
&\quad A_{g_{rx}, g_{tx}}((n - n') T_{sym} - \tau, (m - m')\Delta f - \nu) \\
&\quad e^{\jmath 2\pi (m'\Delta f + \nu)((n-n')T_{sym} - \tau)} d\tau d\nu \\
&\stackrel{(a)}{=} \frac{1}{T} \sum_{i=0}^{P} h_i \sum_{m'} X^{\text{TF}}[n, m'] e^{-\jmath 2\pi (m'\Delta f + \nu_i)\tau_i} e^{\jmath 2\pi \nu_i T_{\text{sym}}} \\
&\quad \int_{T_{\text{cp}} - \tau_i}^{T_{\text{sym}} - \tau_i} e^{-\jmath 2\pi ((m-m')\Delta f - \nu_i)(t' + \tau_i)} dt' e^{\jmath 2\pi (m-m')\Delta f T_{\text{cp}}}
\end{aligned}
\tag{82}
$$

where (a) is due to (6) and $T_{\text{cp}} > \tau$. Thus $A_{g_{rx}, g_{tx}}((n - n')T_{sym} - \tau, (m - m')\Delta f - \nu) \neq 0$ only if $n' = n$. Then, according to (1), (7), and (12),

$$
\begin{aligned}
& Y^{\text{DD}}[k, l] \\
&= \frac{1}{NM} \sum_{i=0}^{P} h_i \sum_{k'} \sum_{l'} X^{\text{DD}}[k', l'] \frac{1}{M} \sum_{p = M_{\text{cp}} - (l_i + c_i M)}^{M + M_{\text{cp}} - 1 - (l_i + c_i M)} \\
&\quad e^{\jmath 2\pi \frac{p(k_i + \tilde{k}_i + b_i N)}{(M + M_{\text{cp}})N}} \sum_{m=0}^{M-1} e^{-\jmath 2\pi \frac{m(p - l + (l_i + c_i M) - M_{\text{cp}})}{M}} \\
&\quad \sum_{m'=0}^{M-1} e^{\jmath 2\pi \frac{m(p - l' - M_{\text{cp}})}{M}} \sum_{n=0}^{N-1} e^{-\jmath 2\pi \frac{n(k - k' - k_i - \tilde{k}_i - b_i N)}{N}} \\
&= \sum_{i=0}^{P} h_i \sum_{k'} \sum_{l'} X^{\text{DD}}[k', l'] \Pi_N(k - k' - k_i - \tilde{k}_i - b_i N) \\
&\quad \sum_{p = M_{\text{cp}} - (l_i + c_i M)}^{M + M_{\text{cp}} - 1 - (l_i + c_i M)} \delta((p - l' - M_{\text{cp}})_M) \\
&\quad \delta((p - l + l_i + c_i M - M_{\text{cp}})_M)
\end{aligned}
$$

$$= \frac{1}{N} \sum_{i=0}^{P} \sum_{k'} X^{\mathrm{DD}} \left[\langle k - k' \rangle_N , (l - l_i)_M \right] \sum_{j=0}^{N-1} h_i e^{-\bar{\jmath} 2\pi \frac{k'j}{N}}$$

$$e^{\bar{\jmath} 2\pi \frac{(M_{\mathrm{cp}} + j(M + M_{\mathrm{cp}}) - l_i)(k_i + \bar{k}_i + b_i N)}{(M + M_{\mathrm{cp}}) N}} e^{\bar{\jmath} 2\pi \frac{(k_i + \bar{k}_i + b_i N)(l - c_i M)}{(M + M_{\mathrm{cp}}) N}}$$

$$= \frac{1}{N} \sum_{i=0}^{P} \sum_{k'} X^{\mathrm{DD}} \left[\langle k - k' \rangle_N , (l - l_i)_M \right]$$

$$\sum_{j=0}^{N-1} h_{M_{\mathrm{cp}} + j(M + M_{\mathrm{cp}}), l_i} e^{-\bar{\jmath} \frac{2\pi}{N} k' j} e^{\bar{\jmath} 2\pi \frac{(k_i + \bar{k}_i + b_i N)(l - c_i M)}{(M + M_{\mathrm{cp}}) N}}$$

$$(83)$$

where $\Pi_N(x) \triangleq \frac{1}{N} \sum_{i=0}^{N-1} e^{-\bar{\jmath} 2\pi \frac{x}{N} i}$. The proof is compeleted.

APPENDIX B
PROOF OF PROPOSITION 2

Computing the first derivative of the objective function in (42) with respect to $\alpha_{i,j}$ and setting it equal to zero, we get

$$\frac{a}{\alpha_{i,j}^*} - b + \sum_{k=i-D}^{i+D} \sum_{l=j-D}^{j+D} \frac{\beta_{k,l,i,j}}{\sum_{p=k-D}^{k+D} \sum_{q=l-D}^{l+D} \beta_{k,l,p,q} \alpha_{p,q}^*}$$

$$- \sum_{k=i-D}^{i+D} \sum_{l=j-D}^{j+D} \beta_{k,l,i,j} \mathrm{E}[|h_{k,l}|^2] = 0. \tag{84}$$

Since $\beta_{k,l,i,j}, \beta_{k,l,p,q}, \alpha_{p,q}^* \geq 0$, we have

$$\sum_{k=i-D}^{i+D} \sum_{l=j-D}^{j+D} \frac{\beta_{k,l,i,j}}{\sum_{p=k-D}^{k+D} \sum_{q=l-D}^{l+D} \beta_{k,l,p,q} \alpha_{p,q}^*} \geq 0, \tag{85}$$

$$\sum_{k=i-D}^{i+D} \sum_{l=j-D}^{j+D} \frac{\beta_{k,l,i,j}}{\sum_{p=k-D}^{k+D} \sum_{q=l-D}^{l+D} \beta_{k,l,p,q} \alpha_{p,q}^*} \leq \frac{(2D+1)^2}{\alpha_{p,q}^*}. \tag{86}$$

Then, substituting (84) into (85) and (86), we get the result in Proposition 2.

REFERENCES

[1] A. Guidotti *et al.*, "Architectures and key technical challenges for 5G systems incorporating satellites," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2624–2639, Jan. 2019.

[2] Z. Xiao *et al.*, "A survey on millimeter-wave beamforming enabled UAV communications and networking," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 557–610, 2022.

[3] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE—A random access for massive machine-type communications in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 76–83, Jan. 2017.

[4] M. de Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting internet of remote things," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 113–123, Feb. 2016.

[5] B. Di, L. Song, Y. Li, and H. V. Poor, "Ultra-dense LEO: Integration of satellite access networks into 5G and beyond," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 62–69, Apr. 2019.

[6] *SpaceX Non-Geostationary Satellite System*, Fed. Commun. Commissions, Washington, DC, USA, 2016.

[7] *OneWeb Non-Geostationary Satellite System*, Fed. Commun. Commissions, Washington, DC, USA, 2016.

[8] O. Kodheli *et al.*, "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, Oct. 2021.

[9] Z. Xiao, L. Zhu, and X.-G. Xia, "UAV communications with millimeter-wave beamforming: Potentials, scenarios, and challenges," *China Commun.*, vol. 17, no. 9, pp. 147–166, Sep. 2020.

[10] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, U.K., Jun. 2015, pp. 2086–2091.

[11] Z. Chen and W. Yu, "Massive device activity detection by approximate message passing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 3514–3518.

[12] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[13] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part II: Achievable rate characterization," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2947–2959, Jun. 2018.

[14] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9631–9640, Oct. 2018.

[15] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12000–12012, Dec. 2019.

[16] 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.

[17] X. Shao, X. Chen, Y. Qiang, C. Zhong, and Z. Zhang, "Feature-aided adaptive-tuning deep learning for massive device detection," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1899–1914, Jul. 2021.

[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Oct. 2009.

[19] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.

[20] 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.

[21] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "Peak-to-average power ratio of OTFS modulation," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 999–1002, Jun. 2019.

[22] A. K. Sinha, S. K. Mohammed, P. Raviteja, Y. Hong, and E. Viterbo, "OTFS based random access preamble transmission for high mobility scenarios," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15078–15094, Dec. 2020.

[23] S. K. Mohammed, "Derivation of OTFS modulation from first principles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7619–7636, Aug. 2021.

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

[25] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "OTFS-NOMA: An efficient approach for exploiting heterogenous user mobility profiles," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7950–7965, Nov. 2019.

[26] 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.

[27] 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.

[28] D. Shi *et al.*, "Deterministic pilot design and channel estimation for downlink massive MIMO-OTFS systems in presence of the fractional Doppler," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7151–7165, Nov. 2021.

[29] 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.

[30] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "OTFS signaling for uplink NOMA of heterogeneous mobility users," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3147–3161, May 2021.

[31] M. Li, S. Zhang, F. Gao, P. Fan, and O. A. Dobre, "A new path division multiple access for the massive MIMO-OTFS networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 903–918, Apr. 2021.

[32] P. Wei, Y. Xiao, W. Feng, N. Ge, and M. Xiao, "Characterizing the peak-to-average power ratio of OTFS signals: A large system analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3705–3720, Jun. 2022.

[33] S. Gao and J. Zheng, "Peak-to-average power ratio reduction in pilot-embedded OTFS modulation through iterative clipping and filtering," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 2055–2059, Sep. 2020.

[34] C. Naveen and V. Sudha, "Peak-to-average power ratio reduction in OTFS modulation using companding technique," in *Proc. 5th Int. Conf. Devices, Circuits Syst. (ICDCS)*, Coimbatore, India, Mar. 2020, pp. 140–143.

[35] M. Liu, M.-M. Zhao, M. Lei, and M.-J. Zhao, "Autoencoder based PAPR reduction for OTFS modulation," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Norman, OK, USA, Sep. 2021, pp. 1–5.

[36] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[37] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[38] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. Xia, "Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5065–5079, Nov. 2019.

[39] L. Gao, S. Zhang, Z. Liu, J. Lin, S. Wang, and C. Xing, "An overview of multi-antenna technologies for space-ground integrated networks," *Sci. China Inf. Sci.*, vol. 59, no. 12, pp. 1–18, Nov. 2016.

[40] J. Kyröläinen *et al.*, "Applicability of MIMO to satellite communications," *Int. J. Satell. Commun. Netw.*, vol. 32, no. 4, pp. 343–357, Aug. 2014.

[41] *Large Scale Multi-Antennas Over Satellite*, Eur. Space Agency, Paris, France, 2020.

[42] P. Angeletti and R. De Gaudenzi, "A pragmatic approach to massive MIMO for broadband communication satellites," *IEEE Access*, vol. 8, pp. 132212–132236, 2020.

[43] 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.

[44] K.-X. Li *et al.*, "Downlink transmit design for massive MIMO LEO satellite communications," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1014–1028, Feb. 2022.

[45] 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.

[46] Z. Gao, A. Liu, C. Han, and X. Liang, "Sum rate maximization of massive MIMO NOMA in LEO satellite communication system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1667–1671, Aug. 2021.

[47] 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.

[48] G. D. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3049–3063, Jun. 2019.

[49] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*. New York, NY, USA: Academic, 2011.

[50] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "NB-IoT random access for non-terrestrial networks," 2021, *arXiv:2101.08079*.

[51] F. P. Fontan, M. Vazquez-Castro, C. E. Cabado, J. P. Garcia, and E. Kubista, "Statistical modeling of the LMS channel," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1549–1567, Nov. 2001.

[52] A. Guidotti *et al.*, "LTE-based satellite communications in LEO mega-constellations," *Int. J. Satell. Commun. Netw.*, vol. 37, no. 4, pp. 316–330, Jul. 2019.

[53] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Solutions for NR to Support Nonterrestrial Networks (NTN) (Release 16)*, document 3GPP TR 38.811V15.1.0, Jun. 2019.

[54] W. Wang, Y. Tong, L. Li, A.-A. Lu, L. You, and X. Gao, "Near optimal timing and frequency offset estimation for 5G integrated LEO satellite communication system," *IEEE Access*, vol. 7, pp. 113298–113310, 2019.

[55] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.

[56] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.

[57] M. Hurtado, C. H. Muravchik, and A. Nehorai, "Enhanced sparse Bayesian learning via statistical thresholding for signals in structured noise," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5430–5443, Nov. 2013.

[58] S. Khanna and C. R. Murthy, "Communication-efficient decentralized sparse Bayesian learning of joint sparse signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 617–630, Sep. 2017.

[59] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer, 1994.

[60] A. Lin, A. H. Song, B. Bilgic, and D. Ba, "Covariance-free sparse Bayesian learning," 2021, *arXiv:2105.10439*.

[61] C. Bekas, E. Kokiopoulou, and Y. Saad, "An estimator for the diagonal of a matrix," *Appl. Numer. Math.*, vol. 57, nos. 11–12, pp. 1214–1229, Nov. 2007.

[62] M. R. Hestenes *et al.*, *Methods Conjugate Gradients for Solving Linear Systems*, vol. 49, no. 1. Washington, DC, USA: NBS, 1952.

[63] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294–308, Jan. 2018.

[64] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2168–2172.

Boxiao Shen (Student Member, IEEE) received the B.E. degree in communication engineering from the Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China, in 2018, and the M.E. degree in communication and information systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2021. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University (SJTU). He has been a Visiting Scholar at the Future Network of Intelligence Institute (FNii), The Chinese University of Hong Kong, Shenzhen, China, since 2019. His research interests include satellite communications, massive machine-type communications, information theory, and the application of artificial intelligence on wireless communications.
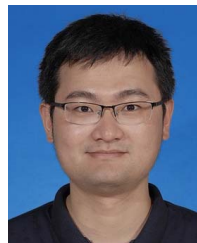
Yongpeng Wu (Senior Member, IEEE) received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in July 2007, and the Ph.D. degree in communication and signal processing from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in November 2013.

He is currently a Tenure-Track Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was a Senior Research Fellow with the Institute for Communications Engineering, Technical University of Munich, Germany, and the Humboldt Research Fellow and a Senior Research Fellow with the Institute for Digital Communications, University Erlangen-Nürnberg, Germany. During his Ph.D. studies, he conducted cooperative research at the Department of Electrical Engineering, Missouri University of Science and Technology, Rolla, MO, USA. His research interests include massive MIMO/MIMO systems, massive machine type communication, physical-layer security, and signal processing for wireless communications.

Dr. Wu was awarded the IEEE Student Travel Grants from the IEEE International Conference on Communications (ICC) in 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants for IEEE Communication Theory Workshop in 2016, the Excellent Doctoral Thesis Awards of China Communications Society in 2016, the Exemplary Editor Award of IEEE Communication Letters in 2017, the Young Elite Scientist Sponsorship Program by CAST 2017, and the Excellent Youth Science Fund Project of National Natural Science Foundation of China in 2021. He was an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2015, 2016, and 2018. He was the Lead Guest Editor for the Special Issue "Physical Layer Security for 5G Wireless Networks" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and "Safeguarding 5G-and-Beyond Networks with Physical Layer Security" of IEEE WIRELESS COMMUNICATIONS. He is currently an Editor of IEEE WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, and IEEE COMMUNICATIONS LETTERS. He has been the Symposium Chair of various conferences, including Globecom, ICC, VTC, and PIMRC.

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 at San Diego, 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 journal and conference papers, and holds or co-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 of the Information Technology Division, National Scientific Foundation of China.

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, China, in 2010. 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. From September 2012 to December 2012, he was a Visiting Scholar with the University of Macau, Macau. His current research interests include machine learning, statistical signal processing, convex optimization, multivariate statistics, array signal processing, and high-frequency band communication systems.

Lian Zhao (Senior Member, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.

She is currently a Professor at the Department of Electrical, Computer, and Biomedical Engineering, Ryerson University, Toronto, ON, Canada. Her research interests include wireless communications, resource management, mobile edge computing, caching and communications, and vehicular ad-hoc networks.

Dr. Zhao is a Senior Member of the IEEE Communication Society and Vehicular Society. She has served as a Committee Member for Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grants Evaluation Group for Electrical and Computer Engineering from 2015 to 2018. She received the Best Land Transportation Paper Award from the IEEE Vehicular Technology Society in 2016, the Top 15 Editor Award in 2016 from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the Best Paper Award from the 2013 International Conference on Wireless Communications and Signal Processing (WCSP), and the Canada Foundation for Innovation (CFI) New Opportunity Research Award in 2005. She has been serving as an Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE INTERNET OF THINGS JOURNAL. She has served as a Co-Chair of Wireless Communication Symposium for IEEE Globecom 2020 and IEEE ICC 2018, the Local Arrangement Co-Chair for IEEE VTC Fall 2017 and IEEE Infocom 2014, and the Co-Chair of Communication Theory Symposium for IEEE Globecom 2013. She is a Licensed Professional Engineer in the Province of Ontario. She has been an IEEE Communication Society (ComSoc) Distinguished Lecturer (DL).
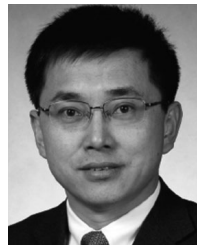
Wenjun Zhang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987, and 1989, respectively. After three years' working as an Engineer with Philips, Nuremberg, Germany, he went back to his Alma Mater in 1993, and became a Full Professor of electronic engineering in 1995. He was one of the main contributors of the Chinese DTTB Standard (DTMB) issued in 2006. He holds 142 patents and published 110 papers in international journals and conferences. He is the Chief Scientist of the Chinese Digital TV Engineering Research Centre (NERC-DTV), an industry/government consortium in DTV technology research and standardization, and the Director of the Cooperative MediaNet Innovation Center (CMIC), an excellence research cluster affirmed by the Chinese Government. His main research interests include video coding and wireless transmission, multimedia semantic analysis, and broadcast/broadband network convergence.