OFDM-Based Massive Connectivity for LEO Satellite Internet of Things

Yong Zuo\textsuperscript{a,}\textsuperscript{,} Mingyang Yue\textsuperscript{b,}\textsuperscript{,} Mingchen Zhang\textsuperscript{c,}\textsuperscript{,} Sixian Li\textsuperscript{d,}\textsuperscript{,} Shaojie Ni, and Xiaojun Yuan\textsuperscript{e}, Senior Member, IEEE

Abstract—Low earth orbit (LEO) satellite has been considered as a potential supplement for the terrestrial Internet of Things (IoT). In this paper, we consider grant-free non-orthogonal random access (GF-NORA) in the orthogonal frequency division multiplexing (OFDM) system to increase access capacity and reduce access latency for LEO satellite-IoT. We focus on the joint device activity detection (DAD) and channel estimation (CE) problem at the satellite access point. The delay and the Doppler effect of the LEO satellite channel are assumed to be partially compensated. We propose an OFDM-symbol repetition technique to better distinguish the residual Doppler frequency shifts, and present a grid-based parametric probability model to characterize channel sparsity in the delay-Doppler-user domain, as well as to characterize the relationship between the channel states and the device activity. Based on that, we develop a robust Bayesian message-passing algorithm named modified variance state propagation (MVSP) for joint DAD and CE. Moreover, to tackle the mismatch between the real channel and its on-grid representation, an expectation–maximization (EM) framework is proposed to learn the grid parameters. Simulation results demonstrate that our proposed algorithms significantly outperform the existing approaches in both activity detection probability and channel estimation accuracy.

Index Terms—GF-NORA, LEO satellite-IoT, joint DAD and CE, Bayesian message passing, expectation–maximization.

I. INTRODUCTION

In recent years, the Internet of Things (IoT) has attracted much attention, and is expected to support different applications such as smart homes, smart gateways, environmental monitoring, and smart cities [1], [2]. Different from human-type communications, the IoT spawns a new communication scenario named massive connectivity, where a huge number of devices can access a single base station (BS) in a sporadic manner. In this scenario, the conventional grant-based random access solutions designed for human-type communications become very inefficient due to severe latency caused by device collisions [3].

Grant-free non-orthogonal random access (GF-NORA) has been considered as a promising technique for massive connectivity [3], [4], where active devices are allowed to directly transmit pilots and data to the BS without waiting for a grant. As such, in each transmission frame, the BS needs to conduct device activity detection (DAD), channel estimation (CE), and data detection (DD). A possible solution to this problem is to divide each transmission frame into a pilot phase and a data phase. The pilot phase is for pilot transmission at the devices and joint DAD and CE at the BS, and the data phase is for data transmission at the devices and DD at the BS. Since the device activity patterns are sporadic, at any given time, only a small and random fraction of all devices are active. Joint DAD and CE can be cast into a compressed sensing (CS) problem, where advanced compressed sensing algorithms, such as approximate message passing (AMP) [5], sparse Bayesian learning (SBL) [6], turbo compressed sensing (Turbo-CS) [7], and variance state propagation (VSP) [8], can be used to solve the problem. For example, the authors in [9] formulated the joint DAD and CE problem as a compressed sensing multiple measurement vector (MMV) problem by assuming multiple receive antennas and used the AMP algorithm to solve the formulated problem. In [10], the turbo generalized MMV (GMMV) algorithm was proposed to solve the joint DAD and CE problem in a MIMO system with mixed analog-to-digital converters. In addition to user sparsity, the channel sparsity in the angle domain of the receiver antenna array can also help with joint DAD and CE. For example, [11] exploited user-angle-domain sparsity in a massive MIMO grant-free system and proposed a Turbo-GMMV-AMP algorithm for the problem. Moreover, machine learning technologies have been applied to further improve performance. In [12], the authors used a deep neural network to learn the weights involved in message passing to improve the convergence performance. In contrast to the two-phase approach, another line of research proposed...
a one-phase approach [13], [14], where the BS is required to conduct joint DAD, CE, and DD. As compared to the two-phase approach, the more challenging one-phase approach generally increases the computational complexity, but can achieve significant performance improvement by efficiently exploiting the structure (such as sparsity and low rank) inherent in the channels and the signals.

Due to the limited coverage of terrestrial BSs, the development of terrestrial IoT is highly restricted in extreme environments such as deserts, forests, and oceans. Recently, the satellite is considered as a potential solution for global IoT services [15], [16], [17]. In particular, low earth orbit (LEO) satellites supplement and extend terrestrial IoT systems, which can effectively solve the environmental constraints faced by terrestrial networks. In addition, the LEO satellite-IoT system has the ability to be immune to natural disasters and to guarantee all-weather communication. Compared with the geostationary orbit (GEO) satellites, LEO satellites operate in low-earth orbits with a height typically lower than 2000 km. A shorter communication distance provides a more real-time IoT service. Yet, the LEO satellite raises new features due to the high speed and the wide coverage:

- Large delay spread: Due to the wide coverage, the delay spread of the LEO satellite channel is quite large compared with that in terrestrial networks. For example, for a LEO satellite system that operates in an orbit of 500 km and with a coverage radius of 50 km, the delay spread between the devices at the center and the border of the coverage area is over 8 µs. Ref. [18] and Ref. [19] proposed to use global navigation satellite system (GNSS) based techniques to synchronize the devices and the satellite, and the delay spread of each device can be largely precompensated based on the positions of the device and the LEO satellite. In this paper, we use orthogonal frequency division multiplexing (OFDM) and cyclic prefix (CP) to handle the residual delay spread.

- Severe Doppler effect: It has been reported in [20] that in a LEO satellite communication system with a carrier frequency of Ku band, the maximum Doppler shift can be over 200 kHz. In [21], the authors considered GF-NORA in LEO satellite-IoT, and assumed that with the help of terrestrial BSs, the Doppler shifts are completely compensated. But this method is not applicable in remote areas without terrestrial BSs. In addition, with GNSS, the devices can acquire their position information and calculate the Doppler shifts. However, the compensation of the Doppler shifts at the terrestrial device is incomplete, since there is typically more than one path, and a terrestrial device can only compensate the Doppler shift of one path. The residual Doppler shifts of Ku-band signals can be over several thousand Hertz. As such, it is of pressing interest to design a grant-free random access scheme that can reliably handle the severe Doppler effect of the LEO satellite-IoT channel.

In this paper, we assume that with the aid of GNSS, the delay spread of the satellite channel can be largely compensated. Then we adopt the OFDM technique to deal with the residual delay spread, providing that the residual delays do not exceed the length of CP. Besides, since the Doppler shift compensation at the IoT devices is expensive and inaccurate, we propose to deal with the Doppler effect at the satellite by assuming that the average Doppler shift of the devices in a beam is estimated and then compensated. We focus on the joint DAD and CE problem in a GF-NORA for LEO satellite-IoT, where active devices suffer from the residual Doppler effect.

To distinguish the Doppler components with high precision, we adopt the OFDM-symbol repetition technique for the pilot design, where a super OFDM symbol is constructed by concatenating repeated regular OFDM symbols. In [22] and [23], the authors designed the long preamble sequence by concatenating the circularly shifted replicas of a short Zadoff–Chu (ZC) sequence, for random access in satellite communication. In addition, similar repetition techniques have been used for carrier frequency offset estimation for a single user, while the receiver carries out the maximum Likelihood (ML) estimation of the carrier frequency offset [24], [25], [26], [27]. The problem considered in this paper is more challenging. On one hand, due to the multi-path effect, each IoT device generally has more than one Doppler component. On the other hand, there are a large number of devices in the satellite-IoT system. As such, the ML-based estimation methods, if applied directly, may incur a prohibitively high computational complexity.

To estimate the time-varying channel of satellite-IoT, we represent the channel with a grid-based parametric model, and point out that the time-varying channel exhibits block sparsity in the delay-Doppler domain. Then, together with the sparsity in the user domain (due to sporadic transmission of the terrestrial devices), we formulate the joint DAD and CE problem for OFDM-based GF-NORA in LEO satellite-IoT as a sparse signal recovery problem. Many existing compressed sensing algorithms [5], [6], [7], [8], [28], [29], [30], [31] can be applied to provide approximate solutions to the problem. It is known that Bayesian CS algorithms, such as AMP and Turbo-CS, can achieve significant performance improvement over non-Bayesian methods in sparse signal reconstruction. But AMP and Turbo-CS generally rely on a certain randomness property of the measurement matrix to ensure convergence, and the recovery performance of these algorithms may degrade seriously when such randomness is not met.

As inspired by the robustness of the VSP algorithm to a broad family of measurement matrices, we extend the VSP algorithm to the massive connectivity scenario by appropriately handling the user sparsity prior, with the resulting algorithm termed modified VSP (MVSP). Specifically, we characterize the channel sparsity structure in the delay-Doppler-user domain with a probability model, which consists of a CE module and a DAD module. The CE module handles the linear constraint between the received signal and the unknown vector, and the DAD module handles the block-sparse prior in the delay-Doppler domain as well as the sparse prior in the user domain. Different from
the original Ising model [32], we introduce an auxiliary variable in the DAD module to characterize the relationship between the channel states and the device activity. The two modules are iterated until convergence. The proposed approach generally suffers from the energy leakage problem since the employed parametric channel model is based on a two-dimensional grid in the delay-Doppler domain. To reduce the mismatch between the actual channel and its on-grid representation, an expectation-maximization (EM) based learning method, named EM-MVSP, is proposed to update the delay-Doppler grid parameters. The contributions of this paper are summarized as follows.

- We develop a grid-based parametric system model for the OFDM-based satellite-IoT, and formulate the joint DAD and CE problem as a sparse signal recovery problem. Interestingly, we show that the measurement matrix in our considered problem can only be partially manipulated by the design of pilots, and exhibits a special correlation structure caused by the Doppler effect. Our experiments show that most existing compressed sensing algorithms including AMP and Turbo-CS behave poorly in the considered problem.

- To distinguish the Doppler components, we adopt the OFDM-symbol repetition technique to increase the frequency resolution of the OFDM system. We show that this OFDM-symbol repetition technique can efficiently improve the DAD and CE performance of the OFDM-based satellite-IoT system.

- We propose the MVSP algorithm for the joint DAD and CE problem, which is robust to the measurement matrix in our problem. Different from the original VSP algorithm, we introduce an auxiliary variable in the DAD module to characterize the relationship between the channel state and the device activity. To alleviate the mismatch of the grid-base model, we further propose the EM-MVSP algorithm to update the grid parameters using the EM method. We show that significant performance improvement can be achieved by the EM-MVSP algorithm, as compared to the counterpart algorithms including AMP, SBL, Turbo-CS, and MVSP.

The rest of the paper is organized as follows. Section II introduces the time-varying satellite-IoT channel and the GF-NORA satellite-IoT system model, and then transforms them into a parametric form. Section III formulates the DAD and CE problem, constructs a probability model, and presents the MVSP algorithm. In Section IV, the MVSP algorithm is extended to the mismatch scenario with the EM framework. Numerical results are given in Section V, and Section VI concludes this paper. The frequently used abbreviations in this paper are summarized in Table I.

**Table I: Abbreviation Table**

| Abbreviation | Explanation |
|--------------|-------------|
| AMP          | Approximate message passing |
| CE           | Channel estimation |
| CP           | Cyclic prefix |
| CS           | Compressed sensing |
| DAD          | Device activity detection |
| EM           | Expectation-maximization |
| GAMP         | Generalized approximate message passing |
| GF-NORA      | Grant-free non-orthogonal random access |
| GNSS         | Global navigation satellite system |
| IoT          | Internet of Things |
| LEO          | Low earth orbit |
| ML           | Maximum Likelihood |
| MRF          | Markov random field |
| MVSP         | Modified variance state propagation |
| NMMSE        | Normalized mean-squared error |
| OFDM         | Orthogonal frequency division multiplexing |
| OMP          | Orthogonal matching pursuit |
| FCSBL        | Pattern-coupled sparse Bayesian learning |
| SBL          | Sparse Bayesian learning |
| SNR          | Signal-noise-ratio |
| STCS         | Structured turbo compressed sensing |
| VSP          | Variance state propagation |

![Fig. 1. The LEO satellite-IoT model.](image)

**A. Time-Varying Satellite-IoT Channel**

As illustrated in Fig. 1, we consider a multi-beam LEO-satellite system where there exists $K$ potential devices within a beam coverage. We assume that the signals in different beams are orthogonal, i.e., the inter-beam interference is ignored. Then, within a beam, the noiseless baseband received scalar signal from device $k$ at the satellite can be expressed as

$$r_k(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{h}_k(\tau, \nu) s_k(t - \tau) e^{j2\pi \nu t} d\tau d\nu,$$

where $\bar{h}_k(\tau, \nu)$ is the channel impulse response at delay $\tau$ and Doppler frequency $\nu$, and $s_k(t)$ is the signal from the $k$-th device. As different devices are usually geographically separated, we assume that the channel realizations between the satellite and different devices are uncorrelated. We also assume that with the help of GNSS, the delay spread can be partially pre-compensated at the device, and the satellite uses
the Doppler frequency shift at the beam center to compensate all the devices in the beam. As such, the impact of the delay spread and Doppler frequency shift can be largely eliminated, or in other words, only the residual delay spread and the residual Doppler frequency shift need to be taken into account. Without loss of generality, we assume that there exist \( P_k \) paths for each device \( k \), where \( P_k \) is a small positive integer since the satellite communication is typically in a weak multipath transmission environment. The residual delay and the residual Doppler shift of the \( p \)-th path of device \( k \) are denoted as \( \tilde{\tau}_{k,p} \) and \( \tilde{\nu}_{k,p} \), respectively. We assume that \( \tilde{\tau}_{k,p} \) and \( \tilde{\nu}_{k,p} \) are constant during a transmission frame. The corresponding attenuation including path loss, reflection, and processing gains of each device in the \( p \)-th path is characterized by a complex coefficient \( \hat{h}_{k,p} \). Therefore, \( \hat{h}_{k}(\tau, \nu) \) can be approximated by

\[
\hat{h}_{k}(\tau, \nu) = \sum_{p=0}^{P_k-1} \hat{h}_{k,p} \delta(\tau - \tilde{\tau}_{k,p}) \delta(\nu - \tilde{\nu}_{k,p}).
\]

Substituting (2) into (1) yields

\[
r_k(t) = \sum_{p=0}^{P_k-1} \hat{h}_{k,p} s_k(t - \tilde{\tau}_{k,p}) e^{j2\pi \tilde{\nu}_{k,p} t}. \tag{3}
\]

B. Grant-Free Satellite-IoT System Model

In this paper, we adopt the grant-free non-orthogonal random access (GF-NORA) scheme, in which the devices share the physical channel resource and directly transmit their signals without requiring the permission of the satellite. Following [3], we assume that each transmission frame in GF-NORA consists of two phases, namely, the pilot phase and the data phase. Each device is preassigned with a unique non-orthogonal pilot sequence. In the pilot phase, the active devices transmit their pilots, based on which the satellite detects the active devices and estimates their channels. In the data phase, the devices transmit data without the grant from the satellite, and the satellite decodes the data based on the estimated channel of the active devices. Orthogonal frequency division multiplexing (OFDM) is employed for both the pilot and data transmission phases. In this work, we focus on device activity detection (DAD) and channel estimation (CE) in the pilot phase. The system model of the pilot phase is described as follows.

In the pilot phase, we construct a super-symbol by concatenating \( N \) repetitions of an OFDM symbol, and a cyclic prefix (CP) is applied to eliminate the inter-symbol interference, as shown in Fig. 2. The duration of such a super-symbol with a CP is \( \bar{T} = NT + T_{cp} \), where \( T \) is the length of a regular OFDM symbol, \( T_{cp} > \tau_{max} \) is the length of the CP, and \( \tau_{max} \) is the maximal residual delay for all devices. We note that the above repetitions of an OFDM symbol improve the frequency resolution of the receiver, so that the Doppler frequency shifts can be identified more clearly, and thus its adverse effect can be more efficiently mitigated. Without loss of generality, we assume that the pilot sequence of a device contains \( U \) consecutive super-symbols. The frequency spacing between any two adjacent subcarriers is set to \( \Delta f = 1/T \).

In the \( u \)-th super-symbol, the baseband modulated signal at the \( k \)-th device is given by

\[
d_{k,u}(t) = \sum_{m=0}^{M-1} x_{k,m,u} e^{j2\pi m \Delta f t} \xi(t - u\bar{T}) \quad \forall k \forall u, \tag{4a}
\]

where \( x_{k,m,u} \) is the pilot on the \( m \)-th subcarrier in the \( u \)-th super-symbol of the \( k \)-th device, and

\[
\xi(t) = \begin{cases} 1, & t \in [-T_{cp}, NT] \\ 0, & \text{otherwise} \end{cases}, \tag{4b}
\]

is the transmitted rectangular pulse. The baseband pilot signal of device \( k \) is given by

\[
s_k(t) = \sum_{u=0}^{U-1} d_{k,u}(t) \quad \forall k, \tag{4c}
\]

where \( U \) is the number of super-symbols in a pilot sequence.

In each transmission frame, only a small subset of devices are active. To characterize such sporadic transmission, the device activity is represented by an indicator function \( \alpha_k \) as

\[
\alpha_k = \begin{cases} 1, & \text{if device } k \text{ is active,} \\ 0, & \text{if device } k \text{ is inactive} \end{cases}, \quad k = 1, \ldots, K, \tag{5}
\]

with a probability \( p(\alpha_k = 1) = \rho \) where \( \rho \ll 1 \). Combining (3), (4a), (4c) and (5), the received baseband signal of all devices at the satellite can be expressed as

\[
r(t) = \sum_{k=1}^{K} \sum_{u=0}^{U-1} \sum_{m=0}^{M-1} \sum_{p=0}^{P_k-1} \alpha_k \sum_{j} x_{k,m,u} e^{j2\pi \tilde{\nu}_{k,p} t} \hat{h}_{k,p} s_k(t - \tilde{\tau}_{k,p}) e^{j2\pi m \Delta f (t - \tilde{\tau}_{k,p})} \xi(t - \tilde{\tau}_{k,p} - u\bar{T}) + w(t), \tag{6}
\]

where \( w(t) \) is an additive white Gaussian noise (AWGN).

In a conventional OFDM system with symbol duration \( T \), any two subcarriers with minimum frequency shift \( \Delta f = 1/T \) are orthogonal to each other, i.e., the frequency resolution is the subcarrier spacing 1//T. In our super-symbol system, since each super-symbol consists of \( N \) repeated regular OFDM symbols, the frequency resolution is 1//\( (NT) \). In other words, \( N \)-time oversampling in the frequency domain can be applied in our system. In the \( u \)-th super-symbol interval, the demodulated signals with \( N \)-time oversampling is given by

\[
y_{n,u} = \frac{1}{NT} \int_{u\bar{T}}^{(u+1)\bar{T}+NT} r(t) e^{-j2\pi \bar{f} n \Delta f t} dt, \quad n \in \{0, 1, \ldots, NM-1\}, \tag{7}
\]

where the boundary of the observation window of the \( u \)-th super-symbol specifies the upper and lower limits of the
Fig. 2. The structure of a pilot sequence in the pilot phase.

integral. By plugging (6) into (7), $y_{n,u}$ can be expressed as

$$y_{n,u} = \sum_{k=1}^{K} \frac{\alpha_k}{\sqrt{T}} \sum_{m=0}^{M-1} x_{k,m,u} \sum_{p=0}^{P_k-1} e^{-j2\pi m\Delta f x_{k,p}} h_{k,p}$$

$$\times \int_{uT}^{uT+NT} e^{j2\pi x_{k,p} t} e^{j2\pi (m-\frac{\pi}{2}) \Delta f t} dt + w_{n,u}$$

$$= \sum_{k=1}^{K} \sum_{m=0}^{M-1} x_{k,m,u} g_{m,n,k,u} + w_{n,u}, \quad (8a)$$

where $w_{n,u} = \int_{uT}^{uT+NT} e^{j2\pi \frac{\pi}{2} \Delta f t} u(t) dt$, and

$$g_{m,n,k,u} = \frac{\alpha_k}{\sqrt{T}} \sum_{p=0}^{P_k-1} e^{-j2\pi m\Delta f x_{k,p}} h_{k,p}$$

$$\times \int_{uT}^{uT+NT} e^{j2\pi x_{k,p} t} e^{j2\pi (m-\frac{\pi}{2}) \Delta f t} dt. \quad (8b)$$

Let $x_{m,u} = [x_{1,m,u}, \cdots, x_{K,M,u}]^T$ and $g_{m,n,u} = [g_{m,n,1,u}, \cdots, g_{m,n,K,u}]^T$. We can rewrite $y_{n,u}$ as

$$y_{n,u} = \sum_{m=0}^{M-1} x_{m,u} g_{m,n,u} + w_{n,u}. \quad (9)$$

Let $y_u = [y_{0,u}, \cdots, y_{NM-1,u}]^T$. We have

$$y_u = G_u x_u + w_u, \quad (10a)$$

where $w_u = [w_{0,u}, \cdots, w_{NM-1,u}]^T$ is the AWGN with variance $\sigma^2$, $x_u \triangleq [x_{0,u}, \cdots, x_{NM-1,u}]^T \in \mathbb{C}^{M \times 1}$, and

$$G_u \triangleq \begin{bmatrix} g_{0,0,u}^T & \cdots & g_{0,M-1,u}^T \\ \vdots & \ddots & \vdots \\ g_{NM-1,0,u}^T & \cdots & g_{NM-1,M-1,u}^T \end{bmatrix} \in \mathbb{C}^{NM \times MK}. \quad (10b)$$

$G_u$ characterizes the channels of all the $M$ subcarriers in the $u$-th super-symbol. Our goal is to estimate $\{G_u\}_{u=0}^{U-1}$ based on $\{y_u\}_{u=0}^{U-1}$ and $\{x_u\}_{u=0}^{U-1}$. A brute-force approach to this problem is infeasible since the number of the unknown variables in $\{G_u\}_{u=0}^{U-1}$ is only $UMK$, which is far less than that of the unknown variables in $\{G_u\}_{u=0}^{U-1}$, i.e., $UM^2K$. As such, we need a more elegant representation of $\{G_u\}_{u=0}^{U-1}$ with fewer unknowns, as detailed in the next subsection.

C. Grid-Based Parametric System Model

We now present a grid-based parametric model for the channel $G_u$. From (8b) and (10b), the unknown parameters in $G_u$ include each device’s path delays $\{\bar{r}_{k,p}\}_{p=0}^{P_k-1}$, path Doppler shifts $\{\bar{f}_{k,p}\}_{p=0}^{P_k-1}$ and channel gains $\{\bar{h}_{k,p}\}_{p=0}^{P_k-1}$. In practical wireless communication scenarios, each channel path may consist of many sub-paths, and the parameters of all these sub-paths are usually difficult to distinguish. The exact identification of these parameters is therefore very challenging.

To address this issue, we discretize the delay domain and the Doppler domain into a two-dimensional grid. Instead of estimating the equivalent channel matrix $G_u$, we estimate the representation of the channel on the grid. Then $G_u$ can be recovered based on the grid parameters and the corresponding channel representation. This approach avoids the estimation of each physical path or sub-path separately, but considers the overall representation of the physical channel on the delay-Doppler grid. In specific, for each device $k$, the grid parameters of the delay domain and the Doppler domain are defined as

$$\begin{aligned}
\tau_k &= \left\{ \tau_l \right\}_{l=0}^{L-1}, \quad \tau_l \in [0, \tau_{max}] , \\
\nu_k &= \left\{ \nu_j \right\}_{j=0}^{J-1}, \quad \nu_j \in [-\nu_{max}/2, \nu_{max}/2] .
\end{aligned} \quad (11)$$

As such, the channel $\bar{h}_{k}(\tau, \nu)$ can be approximated as

$$\bar{h}_{k}(\tau, \nu) \approx \sum_{l=0}^{L-1} \sum_{j=0}^{J-1} \bar{h}_{k,l,j} \delta(\tau - \tau_{l,k}) \delta(\nu - \nu_{k,j}) , \quad (11)$$

where $\bar{h}_{k,l,j}$ is device $k$’s channel representation at grid $(\tau_{l,k}, \nu_{k,j})$. We rewrite $y_{n,u}$ in (8a) as

$$y_{n,u} = \frac{1}{\sqrt{T}} \int_{uT}^{uT+NT} r(t) e^{-j2\pi \frac{\pi}{2} \Delta f t} dt$$

$$= \sum_{k=1}^{K} \frac{\alpha_k}{\sqrt{T}} \sum_{m=0}^{M-1} x_{k,m,u} \sum_{j=0}^{J-1} \int_{uT}^{uT+NT} e^{j2\pi (m-\frac{\pi}{2}) \Delta f t}$$

$$\times e^{j2\pi x_{k,p} t} dt \sum_{l=0}^{L-1} e^{-j2\pi m\Delta f \tau_{l,k}} \bar{h}_{k,l,j} + w_{n,u} . \quad (12)$$

Let $b_{k,m} = [e^{-j2\pi m\Delta f \tau_{0,k}}, \cdots, e^{-j2\pi m\Delta f \tau_{L-1,k}}]^T \in \mathbb{C}^{L \times 1}$, $h_{k,L} = [\bar{h}_{k,0,j}, \cdots, \bar{h}_{k,L-1,j}]^T \in \mathbb{C}^{L \times 1}$, and $c_{k,m,n,j} = \frac{1}{\sqrt{T}} \int_{uT}^{uT+NT} e^{j2\pi x_{k,p} t} e^{j2\pi (m-\frac{\pi}{2}) \Delta f t} dt$. We rewrite (12) as

$$y_{n,u} = \sum_{k=1}^{K} \sum_{m=0}^{M-1} x_{k,m,u} \sum_{j=0}^{J-1} c_{k,m,n,j} b_{k,m}^T h_{k,j} + w_{n,u}$$

$$= \sum_{k=1}^{K} \sum_{m=0}^{M-1} x_{k,m,u} \left( b_{k,m}^T \otimes c_{k,m,n,u} \right) \text{vec} \left( H_k^T \right) + w_{n,u} , \quad (13a)$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where \( c_{k,m,n,u} = [c_{k,m,n,0,u}, \cdots, c_{k,m,n,J-1,u}]^T \in \mathbb{C}^{J \times 1} \), \( H_k = [h_{k,0}, \cdots, h_{k,J-1}] \in \mathbb{C}^{L \times J} \), and
\[
a_{k,n,u} = \sum_{m=0}^{M-1} x_{k,m,u} (b_{k,m}^T \otimes c_{k,m,n,u})^T \in \mathbb{C}^{1 \times LJ}.
\]
(13b)

Then we have
\[
y_u = A_u h + w_u,
\]
(14a)

where \( h_u \triangleq [\alpha_1 \text{vec}(H_1^T), \cdots, \alpha_K \text{vec}(H_K^T)]^T \in \mathbb{C}^{Q \times 1} \),
\[
A_u \triangleq \begin{bmatrix}
a_{1,0,u} & \cdots & a_{K,0,u} \\
\vdots & \ddots & \vdots \\
a_{1,NM-1,u} & \cdots & a_{K,NM-1,u}
\end{bmatrix} \in \mathbb{C}^{NM \times Q},
\]
(14b)

\( R = NMU \) and \( Q = K LJ \). \( h \) combines the device activity and delay-Doppler grid parameters. Note that \( h \) is a sparse vector since the channel is sparse in the user-delay-Doppler domain, as elaborated in Section II-D.

Denote by \( y \triangleq [y_0^T, \cdots, y_{U-1}^T]^T \in \mathbb{C}^{R \times 1} \) the collection of all the \( U \) demodulated super-symbols in the pilot phase. From (14a), we obtain
\[
y = Ah + w,
\]
(15)

where \( A = [A_0 \cdots, A_{U-1}]^T \in \mathbb{C}^{R \times Q} \), and
\[
w = [w_0^T, \cdots, w_{U-1}^T]^T \in \mathbb{C}^{R \times 1}.
\]

We emphasize that \( h \) contains all the unknown channel parameters for the reconstruction of the channels \( \{G_u\}_{u=0}^{U-1} \). To be specific, we use model (10a) to generate channels and signals in simulation, and use the discretized model (15) to design DAD and CE algorithms. We then reconstruct the channel \( G_u \) by the estimated \( h \). Suppose that \( \hat{h} \) is an estimate of \( h \), \( \hat{h}_{k,i,j} \) is the estimate corresponding to the \((k,i,j)\)-th grid point, and \( \hat{\alpha}_k \) is an estimate of \( \alpha_k \). Then the recovered channel is given by
\[
\hat{G}_u \triangleq \begin{bmatrix}
\hat{g}_{0,0,u}^T & \cdots & \hat{g}_{0,M-1,0,u}^T \\
\vdots & \ddots & \vdots \\
\hat{g}_{NM-1,0,u}^T & \cdots & \hat{g}_{NM-1,M-1,u}^T
\end{bmatrix},
\]
(16a)

where \( \hat{g}_{m,n,u} = [\hat{g}_{m,n,1,u}, \cdots, \hat{g}_{m,n,K,u}]^T \), and
\[
\hat{g}_{m,n,k,u} = \hat{\alpha}_k \frac{L-1}{NT} \sum_{i=0}^{L-1} \sum_{j=0}^{J-1} e^{-j2\pi m \Delta f r_{k,i} \Delta t} e^{j2\pi (m-\frac{u}{2})} \Delta f \Delta t.
\]
(16b)

The normalized mean-squared error (NMSE) of the channel estimation is defined by
\[
\text{NMSE} = \frac{1}{U} \sum_{u=0}^{U-1} \mathbb{E} \left[ \frac{\| \hat{G}_u - G_u \|_F^2}{\| G_u \|_F^2} \right].
\]
(17)

**D. Channel Sparsity**

It is well known that channel sparsity can be exploited to significantly reduce the number of pilots required in channel estimation. In this regard, the channel in our considered satellite-IoT scenario exhibits the following sparsity structure:

1) Sparsity in the user domain: Most of the devices are inactive at any given time, i.e., most of \( \{\alpha_k\} \) are zero.

2) Sparsity in the delay-Doppler domain: Since satellite communication is in a weak multi-path transmission environment, the number of dominant paths between the satellite and the devices is limited.

3) Block-sparsity in the delay-Doppler domain: The scattering effect of the electromagnetic waves causes delay and Doppler spread in wireless channels. Besides, the grid mismatch causes additional spreading in the delay-Doppler domain. These effects make the channel coefficients appear in clusters in the delay-Doppler domain.

Fig. 3 illustrates the sparsity structure of \( h \) in the delay-Doppler-user domain. This sparsity structure is exploited in the design of the receiver.

**E. Problem Description**

Recall that the receiver of the considered grant-free satellite-IoT system carries out joint DAD and CE. With (15) and the discussions in Subsection D, the joint DAD and CE problem is to estimate a sparse vector \( h \) given the observation \( y \). Various compressed sensing algorithms have been proposed to solve the linear-inverse problem in the form of (15). However, popular Bayesian algorithms, such as AMP and Turbo-CS, suffer from significant performance degradation when applied to our problem. This is because AMP and Turbo-CS generally rely on the randomness of the measurement matrix \( A \) to ensure convergence. For example, the recovery performance of AMP is guaranteed only when the elements of \( A \) are independently and identically distributed (i.i.d.) Gaussian; as for Turbo-CS, \( A \) is required to be right-rotationally invariant. When the randomness requirement of \( A \) is not met, the recovery performance of the corresponding algorithms will be seriously degraded. In this work, the structure of \( A \) cannot be arbitrarily designed. From (13b) and (14b), each row of \( A \) is the sum of a series of Kronecker products. This introduces a correlation between the elements in the same row of \( A \).
Fig. 4(a) shows the local thermogram of \( A \) in a random experiment, which reflects the amplitudes of the elements in \( A \). It is clear that the amplitudes of the elements in the same row are correlated, rather than independent and identically distributed. As a comparison, Fig. 4(b) shows the thermogram of a matrix of the same size with each element randomly drawn from the standard complex Gaussian distribution. Fig. 4(c) shows the local thermogram of \( AF \) where \( F \) is a discrete Fourier transform (DFT) matrix (which is unitary). Clearly, \( AF \) and \( A \) show different patterns. Therefore \( A \) is not a right-rotationally invariant matrix, and thus does not meet the randomness requirements of both AMP and Turbo-CS. We will see that AMP and Turbo-CS perform poorly for this task in simulations.

Considering the special structure of \( A \) in (15), we follow the variance state propagation (VSP) framework proposed in [8] and propose the MVSP algorithm to solve this sparse signal recovery problem. Compared with AMP and Turbo-CS, the MVSP algorithm is more robust to the structure of the measurement matrix, which is derived in the next section.

### III. Receiver Design for Joint DAD and CE

In this section, we first introduce the probability model and the problem formulation, and then derive the MVSP algorithm based on a factor-graph representation of the probability model.

#### A. Probability Model

For notational convenience, we rewrite \( h \) as

\[
\mathbf{h} = [h_{1,1}, h_{1,2}, \ldots, h_{1,LJ}, \ldots, h_{K,1}, h_{K,2}, \ldots, h_{K,LJ}]^T.
\]

Similarly to [33], we assign the sparse channel \( h \) with a conditional Gaussian prior as

\[
p(h|v) = \prod_{k=1}^{K} \prod_{i=1}^{LJ} p(h_{k,i}|v_{k,i}),
\]

where \( v = [v_{1,1}, v_{1,2}, \ldots, v_{K,LJ}] \), and \( p(h_{k,i}|v_{k,i}) = \mathcal{CN}(h_{k,i}; 0, v_{k,i}) \) is a circularly symmetric complex Gaussian (CSCG) distribution with zero mean and variance \( v_{k,i} \), \( i \in \{1, \ldots, L \} \). Each \( v_{k,i} \) is assigned with a conditionally independent distribution given by

\[
p(v_{k,i}|s_{k,i}) = \text{Gamma}(v_{k,i}; \gamma_{k,1}, \gamma_{k,2}) \delta(s_{k,i} - 1) + \delta(v_{k,i}) \delta(s_{k,i} + 1),
\]

where \( s_{k,i} \in \{-1, 1\} \) is a hidden binary state; \( \text{Gamma}(v_{k,i}; \gamma_{k,1}, \gamma_{k,2}) \) is the Gamma distribution

\[
\text{Gamma}(v_{k,i}; \gamma_{k,1}, \gamma_{k,2}) = \begin{cases} 
\frac{v_{k,i}^{\gamma_{k,1}-1} e^{-v_{k,i}/\gamma_{k,2}}}{\Gamma(\gamma_{k,1})}, & v_{k,i} > 0, \\
0, & \text{otherwise},
\end{cases}
\]

with \( \Gamma(\gamma_{k,1}) = \int_0^\infty t^{\gamma_{k,1}-1} e^{-t}dt \) being the Gamma function.

Then, a Markov random field (MRF) prior is used to characterize the sparse structure of \( v \). The joint probability of the hidden state and device activity variables is modeled as

\[
p(s_k, \alpha_k) \propto \prod_{i=1}^{L_J} \left( \varphi(s_{k,z}, s_{k,i}) \right)^{1/2} \psi(\ell_k, \alpha_k) p(\alpha_k),
\]

where \( s_k = [s_{k,1}, s_{k,2}, \ldots, s_{k,LJ}] \), \( \varphi(s_{k,z}, s_{k,i}) = \exp(\beta s_{k,z} s_{k,i}) \), and \( \ell_k = \sum_{i=1}^{L_J} s_{k,i} \); \( D_i \) denotes the set includes the indexes of the left, right, top and bottom neighbors of \( s_{k,i} \), i.e., \( \{-1, i+1, i-1, i-J, i+J\} \); \( \beta \) is the parameter of the MRF corresponding to the average size of non-zero blocks. Different from the original Ising model [32], we add the constraint \( \psi(\ell_k, \alpha_k) \) that represents the conditional probability density of \( \ell_k \) given \( \alpha_k \). Note that \( \ell_k \) is discrete since each \( s_{k,i} \) is binary. However, since \( L_J \) is large, \( \ell_k \) for an active device \( k \) (i.e., \( \alpha_k = 1 \)) can be well approximated by a Gaussian random variable according to the central limit theorem [34]. Thus, we have

\[
\psi(\ell_k, \alpha_k) = \mathcal{N}(\ell_k; m_\psi, \sigma^2_\psi) \delta(\alpha_k - 1) + \delta(\ell_k + L_J) \delta(\alpha_k),
\]

where \( m_\psi \) and \( \sigma^2_\psi \) are the mean and variance of \( \ell_k \) conditioned on \( \alpha_k = 1 \), respectively given by \( m_\psi = L_J (2\rho_s - 1) \) and \( \sigma^2_\psi = 4L_J \rho_s (1 - \rho_s) \), and \( \rho_s \) is the sparsity rate of \( \{v_{k,i}\}_{i=1}^{L_J} \), i.e., \( p(s_{k,i} = 1) = \rho_s \). The joint probability of \( p(y, h, v, s, \alpha) \) can be decomposed as

\[
p(y, h, v, s, \alpha) = p(y|h) p(h|v) p(v|s) p(s, \alpha) = p(y|h) \prod_{k=1}^{K} \prod_{i=1}^{L_J} p(h_{k,i}|v_{k,i}) p(v_{k,i}|s_{k,i})
\]

\[
\times \prod_{k=1}^{K} p(s_k, \alpha_k),
\]

where \( s = [s_1, \ldots, s_K] \) and \( \alpha = [\alpha_1, \ldots, \alpha_K] \). The dependencies of the random variables in the factorization (23) can be shown by a factor graph as depicted in Fig. 5, where circles represent variable nodes and squares represent factor nodes. The factor nodes in Fig. 5 are defined as

\[
\zeta_{k,i} : p(v_{k,i}|s_{k,i}),
\]

\[
\eta_{k,i} : p(h_{k,i}|v_{k,i}) = \mathcal{CN}(h_{k,i}; 0, v_{k,i}),
\]

\[
i : p(y|h) = \mathcal{CN}(y - Ax; 0, \sigma^2 I),
\]

\[
\chi_k : \delta(\ell_k - \sum_i s_{k,i}),
\]

\[
\psi_k : \psi(\ell_k, \alpha_k).
\]

The factor graph in Fig. 5 includes two modules, namely, the CE module that handles the linear constraint in (15) and the DAD module that handles the MRF prior in (21).

#### B. MVSP Algorithm

The MVSP algorithm is a sum-product message passing algorithm defined in Fig. 5. A major difference between MVSP and the original VSP in [8] is that variable nodes \( \{\ell_k\} \) and \( \{\alpha_k\} \) are added to the factor graph for device activity
detailed factor graph characterizing the 4-connected DAD module is given in Fig. 6. For clarity, the left, right, top, and bottom neighbors to \( s_{k,i} \) are reindexed by \( \mathcal{I}_k = \{i_L, i_R, i_T, i_B\} \) (i.e., \( s_{k,i_L} = s_{k,i-1}, s_{k,i_T} = s_{k,i+1}, s_{k,i_R} = s_{k,i-J}, s_{k,i_B} = s_{k,i+J} \)). The left, right, top, and bottom incoming messages of \( s_{k,i} \), denoted as \( \omega^L_{k,i}, \omega^R_{k,i}, \omega^T_{k,i}, \) and \( \omega^B_{k,i} \), are Bernoulli distributions. By defining \( \mathcal{J} = \{L, R, T, B\} \), the incoming message of \( s_{k,i} \) from the left is given by

\[
\omega^L_{k,i} \propto \prod_{j \in \mathcal{J} \setminus R} \varphi (s_{k,i}, s_{k,i_L}) \omega^T_{k,i-L} \delta (s_{k,i} - 1) + (1 - \lambda^L_{k,i}) \delta (s_{k,i} + 1),
\]

where \( \mathcal{J} \setminus R \) denotes the set \( \mathcal{J} \) by excluding the element “R”, \( \omega^T_{k,i-L} \) is the message from \( \chi_k \) to \( s_{k,i_L} \), denoted as

\[
\omega^T_{k,i-L} \propto \prod_{j \in \mathcal{J} \setminus R} \varphi (s_{k,i}, s_{k,i_L}) \omega^R_{k,i-L} \delta (s_{k,i} - 1) + (1 - \lambda^T_{k,i}) \delta (s_{k,i} + 1),
\]

and \( \lambda^L_{k,i} \) is given by (27), shown at the bottom of the next page.

The incoming messages of \( s_{k,i} \) from the right, top, and bottom, i.e., \( \omega^R_{k,i}, \omega^T_{k,i}, \) and \( \omega^B_{k,i} \), have a form similar to \( \omega^L_{k,i} \).
The message from $\chi_k$ to $\ell_k$ is expressed as

$$\varpi_{\chi_k \rightarrow \ell_k} \propto \delta \left( \ell_k - \sum_i s_{k,i} \right) \prod_i \varpi_{\xi_i \rightarrow s_{k,i}} \prod_{j \in J} \varpi_{\Pi_{k,i}}. \quad (28)$$

Similarly to (22a), we approximate (28) with a Gaussian distribution:

$$\varpi_{\chi_k \rightarrow \ell_k} = \mathcal{N}(\ell_k; m_{\chi_k \rightarrow \ell_k}, \sigma_{\chi_k \rightarrow \ell_k}^2), \quad (29a)$$

where $m_{\chi_k \rightarrow \ell_k}$ and $\sigma_{\chi_k \rightarrow \ell_k}^2$ are the mean and variance, respectively given by

$$m_{\chi_k \rightarrow \ell_k} = \sum_i (2\pi_{s_{k,i} \rightarrow \chi_k} - 1) \quad \text{and} \quad \sigma_{\chi_k \rightarrow \ell_k}^2 = 4 \sum_i \pi_{s_{k,i}} (1 - \pi_{s_{k,i}}) - 1,$$

and

$$\pi_{s_{k,i} \rightarrow \chi_k} = \frac{\pi_{\xi_i \rightarrow s_{k,i}} \prod_{j \in J} \lambda_{k,i}^j}{\psi_{\ell_k}} + (1 - \pi_{\xi_i \rightarrow s_{k,i}}) \prod_{j \in J} (1 - \lambda_{k,i}^j), \quad (29b)$$

The message from $\psi_k$ to $\alpha_k$ is given by

$$\varpi_{\psi_k \rightarrow \alpha_k} \propto \int_{\ell_k} \psi(\ell_k; \alpha_k) \varpi_{\chi_k \rightarrow \ell_k} = \pi_{\psi_k \rightarrow \alpha_k} \delta(\alpha_k - 1) + (1 - \pi_{\psi_k \rightarrow \alpha_k}) \delta(\alpha_k), \quad (30a)$$

with $\varpi_{\ell_k \rightarrow \psi_k} = \varpi_{\chi_k \rightarrow \ell_k}$, where

$$\pi_{\psi_k \rightarrow \alpha_k} = \frac{\pi_{\psi_k \rightarrow \alpha_k}}{\pi_{\psi_k \rightarrow \alpha_k, 1} + \pi_{\psi_k \rightarrow \alpha_k, 0}}, \quad (30b)$$

$$\pi_{\psi_k \rightarrow \alpha_k, 1} = e^{-(m_{\chi_k \rightarrow \ell_k} - m_{\psi} + \sigma_{\psi}^2/2)}/(2\pi_{\chi_k \rightarrow \ell_k} + 2\pi_{\psi}^2), \quad (30c)$$

$$\pi_{\psi_k \rightarrow \alpha_k, 0} = \frac{e^{-(L_J - m_{\chi_k \rightarrow \ell_k} + \sigma_{\psi}^2/2)}/2\pi_{\chi_k \rightarrow \ell_k}}{\sqrt{2\pi}\sigma_{\psi}^2}, \quad (30d)$$

The message from $\psi_k$ to $\ell_k$ is constant and given by

$$\varpi_{\psi_k \rightarrow \ell_k} = \int_{\ell_k} p(\alpha_k) \psi(\ell_k; \alpha_k) = \rho \mathcal{N}(\ell_k; m_{\psi}, \sigma_{\psi}^2) + (1 - \rho) \delta(\ell_k + L_J), \quad (31)$$

Then we calculate the message from $\chi_k$ to $s_{k,i}$ with a similar form of (26b). To obtain $\pi_{\chi_k \rightarrow s_{k,i}}$, we first calculate

$$\lambda_{k,i}^l = \frac{e^{\beta} \pi_{\xi_i \rightarrow s_{k,i}} \pi_{\chi_k \rightarrow s_{k,i,l}} \prod_{j \in J \setminus R} \lambda_{k,i,j} - e^{-\beta} (1 - \pi_{\xi_i \rightarrow s_{k,i}}) \prod_{j \in J \setminus R} (1 - \lambda_{k,i,j})}{(e^{\beta} + e^{-\beta}) \pi_{\xi_i \rightarrow s_{k,i}} \pi_{\chi_k \rightarrow s_{k,i,l}} \prod_{j \in J \setminus R} \lambda_{k,i,j} + (1 - \pi_{\xi_i \rightarrow s_{k,i,l}}) \prod_{j \in J \setminus R} (1 - \lambda_{k,i,j})} \quad (37)$$
With \( \varpi_{v_{k,i} \rightarrow \eta_{k,i}} = \varpi_{\varsigma_{k,i} \rightarrow v_{k,i}} \), the message from \( \eta_{k,i} \) to \( h_{k,i} \) is given by
\[
\varpi_{\eta_{k,i} \rightarrow h_{k,i}} \propto \int_{v_{k,i}} p(h_{k,i}|v_{k,i}) \varpi_{v_{k,i} \rightarrow \eta_{k,i}}.
\] (38)

The message from \( h_{k,i} \) to \( i \) is \( \varpi_{h_{k,i} \rightarrow i} = \varpi_{\pi_{h_{k,i}} \rightarrow h_{k,i}} \). The message from \( i \) to \( h_{k,i} \) is
\[
\varpi_{i \rightarrow h_{k,i}} \propto \int_{h_{k,i}} p(y|h_{k,i}) \prod_{i' \neq i} \prod_{k' \neq k} \varpi_{\eta_{k,i'} \rightarrow h_{k,i'}}
\]
\[
= \int_{h_{k,i}} p(y|h_{k,i}) \prod_{i' \neq i} \int_{v_{k,i'}} p(x_{k,i'}|v_{k,i'}) \varpi_{v_{k,i'} \rightarrow \eta_{k,i'}}
\]
\[
\times \prod_{k' \neq k} \prod_{i' \neq i} p(x_{k,i'}, h_{k,i'}) \varpi_{h_{k,i'} \rightarrow \eta_{k,i'}}.
\] (39)

where \( h_{k,i} \) denotes all the entries of \( h \) except \( h_{k,i} \). Clearly, \( \varpi_{\pi_{h_{k,i}} \rightarrow h_{k,i}} \). Then, the messages \( \varpi_{\eta_{k,i} \rightarrow v_{k,i}} \) and \( \varpi_{v_{k,i} \rightarrow \varsigma_{k,i}} \) can be computed as
\[
\varpi_{v_{k,i} \rightarrow \varsigma_{k,i}} = \varpi_{\eta_{k,i} \rightarrow v_{k,i}} \propto \int_{h_{k,i}} p(h_{k,i}|v_{k,i}) \varpi_{h_{k,i} \rightarrow \eta_{k,i}}.
\] (40)

Note that the integrals involved in (39) and (40) are difficult to evaluate. From [8], we can replace the output of the CE module for node \( v_{k,i} \) by the mean \( \mu_{\eta_{k,i} \rightarrow v_{k,i}} = \mathbb{E}_{\eta_{k,i} \rightarrow v_{k,i}} [v_{k,i}] \). The above messages are updated iteratively until the algorithm converges. Then, we use the estimate \( \hat{h} = m \) to recover the channel \( G_{u} \) based on (16a), and estimate the device activity as
\[
\hat{\alpha}_k = \begin{cases} 
1, & \text{if } \rho \varpi_{\psi_{k} \rightarrow \alpha_k} \geq (1 - \rho)(1 - \varpi_{\psi_{k} \rightarrow \alpha_k}), \\
0, & \text{if } \rho \varpi_{\psi_{k} \rightarrow \alpha_k} \leq (1 - \rho)(1 - \varpi_{\psi_{k} \rightarrow \alpha_k}),
\end{cases} \quad \forall k.
\] (47)

C. Overall Algorithm and Complexity

The overall MVSP algorithm is summarized in Algorithm 1.

We now analyze the computational complexity of the MVSP algorithm. The complexity in step 5 is \( O(R^3 + R^2Q) \). The complexities in steps 8, 10, and 11 are all \( O(Q) \). The complexity in step 13 is \( O(Q + K) \). Thus, the computational complexity of the MVSP algorithm is dominated by step 6, and is given by \( O \left( T_{out}(T_{in1}(R^3 + R^2Q + Q) + K) \right) \).

Algorithm 1: MVSP Algorithm

Input: \( y, A, T_{out}, T_{in1}, T_{in2} \)

for \( t_{out} = 1 \) to \( T_{out} \) do

\( \hat{\mu} = [\mu_{v_{1,1} \rightarrow \eta_{1,1}}, \ldots, \mu_{v_{K,LJ} \rightarrow \eta_{K,LJ}}]^T \); 

for \( t_{in1} = 1 \) to \( T_{in1} \) do

\( D = \text{diag}(\hat{\mu}) \); 

Compute \( m \) and \( \Phi \) by using (42) and (43); 

Update \( \hat{\mu} = [\hat{\mu}_{v_{1,1} \rightarrow \eta_{1,1}}, \ldots, \hat{\mu}_{v_{K,LJ} \rightarrow \eta_{K,LJ}}]^T \) with (44); 

end

Compute \( \{\pi_{\varsigma_{k,i} \rightarrow \eta_{k,i}}\} \) based on (45) and 

\[ [\hat{\mu}_{v_{1,1} \rightarrow \eta_{1,1}}, \ldots, \hat{\mu}_{v_{K,LJ} \rightarrow \eta_{K,LJ}}]^T = \hat{\mu} ; \]

for \( t_{in2} = 1 \) to \( T_{in2} \) do

Compute \( \{\varpi_{v_{k,i} \rightarrow \eta_{k,i}}\} \) based on (26a);  

Compute \( \{\pi_{\chi_{k} \rightarrow \eta_{k}}\} \) based on (33)-(35); 

end

Compute \( \{\pi_{\chi_{k} \rightarrow \alpha_k}\}, \{\hat{\alpha}_k\}, \) and \( \{\mu_{v_{k,i} \rightarrow \eta_{k,i}}\} \) by 

(30b), (47) and (46), respectively;

end

Output: \( \hat{h} = m, \hat{\alpha}_k, \forall k \).

IV. EM-MVSP ALGORITHM

A. Model Mismatch Problem

It is worth noting that the true delay and Doppler frequency shift of each path may not fall onto the given grid \( \tau_k \) and \( \nu_k \).
In each outer iteration of the MVSP algorithm, the posterior distribution of the channel representation \( h \) given by the CE module is approximated by
\[
p(h | y, \omega(i)) = \mathcal{C}N(h - m(\omega(i)); 0, \Phi(\omega(i)))
\] (51)
where \( m(\omega(i)) \) and \( \Phi(\omega(i)) \) are the posterior mean and the covariance matrix of \( h \) calculated based on \( y \) and \( A(\omega(i)) \).

Plugging (50) and (51) into (49), we have
\[
\mathcal{F}(\omega, \omega(i)) = \frac{1}{\sigma^2} \int_h \left( y^H y - y^H A(\omega) h - y^H A^H(\omega) y + h^H A^H(\omega) A(\omega) h \right) p(h | y, \omega(i)) + \ln(\pi \sigma^2)
\] (52)
Remove the irrelevant terms in (52) and define a new objective function
\[
\mathcal{G}(\omega, \omega(i)) = \int_h \left( -y^H A(\omega) h - h^H A^H(\omega) y + h^H A^H(\omega) A(\omega) h \right) p(h | y, \omega(i)) + \frac{1}{\sigma} \left( y^H A(\omega) m(\omega(i)) \right)^2
\] (53)

where \( \Omega(i) = \Phi(\omega(i)) + m(\omega(i)) m^H(\omega(i)) \).

In summary, in each update of \( \omega \) through EM, our method is to calculate the posterior distribution of \( h \) and construct the objective function \( \mathcal{G}(\omega, \omega(i)) \) based on \( y \) and \( \Omega(i) \). It is not easy to obtain an analytical solution to this problem with respect to both \( \tau_k \) and \( \nu_k \). Therefore, we use the gradient descent method to minimize \( \mathcal{G}(\omega, \omega(i)) \) and then update \( \omega \) in each iteration.

Algorithm 2 EM-MVSP Algorithm

**Input:** \( y, \{x_u\}, T_{out}, T_{in}, t_{max} \).

1. Initialization:
   \[ i = 0, \omega(0) = \{\tau^{(0)}_1, \ldots, \tau^{(0)}_K, \nu^{(0)}_1, \ldots, \nu^{(0)}_K\} \]
   2. While the stopping criterion is not met do
   3. Generate \( A(\omega(i)) \) based on \( \omega(i) \) and \( \{x_u\} \);
   4. With \( y \) and \( A(\omega(i)) \), calculate \( m(\omega(i)), \Phi(\omega(i)) \) and \( \hat{\alpha}_k(\omega(i)) \) by Algorithm 1;
   5. Minimize (53) with respect to \( \tau_k \), and obtain the updated \( \tau^{(i+1)}_k \), \( \forall k \);
   6. Minimize (53) with respect to \( \nu_k \), and obtain the updated \( \nu^{(i+1)}_k \), \( \forall k \);
   7. \( \omega(i+1) = \{\tau^{(i+1)}_1, \ldots, \tau^{(i+1)}_K, \nu^{(i+1)}_1, \ldots, \nu^{(i+1)}_K\} \), \( i = i + 1 \);
8. End

**Output:** \( \hat{h} = m(\omega(i)), \hat{\alpha}_k = \hat{\alpha}_k(\omega(i)) \).
Fig. 8. Illustration of the simulated satellite-IoT system.

Fig. 9. CE and DAD performances under various SNR and on-grid settings. Related parameters are $K = 200$, $\rho = 0.1$, $N = 4$, $L = 4$, $J = 24$, $M = 32$ and $\alpha_e = 50^\circ$.

C. Overall Algorithm and Complexity

Algorithm 2 summarizes the EM-MVSP algorithm. The stopping criterion of the iteration is generally set to $\|\omega^{(i+1)} - \omega^{(i)}\|_2^2$ is less than a small positive threshold or $i \geq i_{\text{max}}$, where $i_{\text{max}}$ is the maximal number of iterations of the EM. The block diagram of the MVSP and the EM-MVSP is illustrated in Fig. 7.

We now analyze the computational complexity of the EM-MVSP algorithm. The complexity in step 3 is $O(RQ(M + LJ))$. The complexities in step 5 and step 6 are both $O(Q^3 + Q^2R)$. As mentioned in Section IV-C, the complexity in step 4 is $O(T_{\text{init}}T_{\text{in}}(R^3 + R^2Q))$. Thus, the computational complexity of the EM-MVSP algorithm is given by $O(i_{\text{max}}(Q^3 + Q^2R + T_{\text{init}}(T_{\text{in}}(R^3 + R^2Q + Q) + K)))$.

V. Simulation Results

In this section, we carry out simulations to demonstrate the effectiveness of the proposed algorithms. The CE and DAD performance are respectively measured in terms of NMSE and detection error probability $P_e = E[1/K \sum_k (p(\hat{\alpha}_k = 0|\alpha_k = 1) + p(\hat{\alpha}_k = 1|\alpha_k = 0))]$ over 100 independent trials. The signal-noise-ratio (SNR) is defined as $\|G_u x_u\|_2^2 / (R\sigma_u^2)$.

The baseline methods used for comparison include OMP [28], GAMP [29], structured TCS (STCS) [30], SBL [6], and pattern-coupled SBL (PCSBL) [31]. The iteration number of EM in EM-MVSP is 4. $\{x_{k,m,u}\}$ are generated from the standard complex Gaussian distribution. The parameters in the simulation are listed as follows. The satellite operates at a speed of 7 km/s in an orbit of 600 km, and the coverage of the beam is a circle with a diameter of 50 km. As shown in Fig. 8, device A and device B are at the border of the beam coverage, and the elevation angle between the satellite and device A is $\alpha_e$. The velocity of devices follows the uniform distribution with $[0, 120]$ km/h, the delay spreading is 0.5 µs, the length of CP is $T_{\text{cp}} = 2$ µs. We adopt the TDL-A channel model in [35]. The large-scale fading is compensated since it changes slightly over the devices in a beam. The total number of OFDM symbols in a transmission frame is $UN = 12$, and the subcarrier spacing is $\Delta f = 15$ kHz.

We first consider the scenario when there is no mismatch between the real channel and the grid-based model, that is, the
delay and the Doppler frequency shift of each channel path fall onto the delay-Doppler grid. In Fig. 9(a) we show the CE NMSE versus the SNR. The proposed MVSP outperforms other baseline methods as SNR increases. In particular, due to the special structure of the measurement matrix discussed in Section II-E, we notice that GAMP and STCS behave poorly in this task, with an NMSE of 0 dB. PCSBL outperforms SBL but still has a large performance gap with the proposed MVSP algorithm. In the SNR range of [16, 24] dB, the MVSP algorithm is at least 5 dB better than other baseline methods in NMSE. Fig. 9(b) shows the detection error probability versus the SNR. We see that as the SNR increases, the \( P_e \) of GAMP, STCS, OMP, and SBL can hardly be improved. MVSP obviously outperforms the baseline methods, especially when the SNR is over 16 dB.

We further consider the mismatch between the real channel and the grid-based model, namely, the off-grid scenario. Fig. 10(a) shows the NMSE against the number of repeated OFDM symbols \( N \) in a super-symbol. As \( N \) increases, it is interesting that different methods have different trends since more repeated OFDM symbols can improve the frequency resolution, but also results in the correlation of measurement matrix. The NMSEs of GAMP and OMP increase because they are sensitive to measurement matrix, and repeated OFMD symbols result in performance degradation. As for PCSBL and SBL, their NMSEs first slightly decrease and then increase. Then, MVSP is more robust to the measurement matrix, with significant performance gain as \( N \) increases. We see that for MVSP more than 4 dB gain can be obtained from repeated OFDM symbols. When \( N > 4 \), almost all the algorithms have a performance degradation due to the measurement matrix correlation. In Fig. 11(a), we show detection error probability against \( N \). The trade-off between frequency resolution improvement and measurement matrix correlation is also obvious, which demonstrates the effectiveness of cascaded OFDM symbols.

Then we show the NMSE performance against the SNR with \( \alpha_e = 50^\circ \) in Fig. 11(a). As the SNR increases, MVSP and EM-MVSP are at least 4 and 6 dB better than other baseline methods in NMSE, respectively. The proposed algorithms also behave well in DAD as shown in Fig. 11(b). We notice that PCSBL, MVSP, and EM-MVSP have a significant performance gap compared with other methods in the considered SNR range, and their performance is similar at a lower SNR. This is because these three methods all consider the channel block-sparsity structure, but one-dimension block-sparsity in PCSBL and two-dimension block-sparsity in MVSP and EM-MVSP. In Fig. 12, we further consider the scenario with elevation angle \( \alpha_e = 90^\circ \), where the Doppler effect is more severe. We see that MVSP and EM-MVSP still outperform the baselines. Compared with PCSBL, EM-MVSP have a performance gain of more than 3 dB in NMSE and one order of magnitude in \( P_e \). Thus, the proposed MVSP and
EM-MVSP algorithms show performance advantages under different Doppler effects.

Fig. 13 shows the CE and DAD performances of all the algorithms under varying sparsity \( \rho \). The proposed MVSP and EM-MVSP have a considerable performance gap within the considered sparsity range. Even at \( \rho = 0.2 \), i.e., about 40 active devices access to the satellite, \( P_c \) of EM-MVSP can reach \( 10^{-2} \), which demonstrates the advantage of the proposed algorithms for massive connectivity in satellite-IoT systems.

**VI. CONCLUSION**

In this paper, we studied the joint DAD and CE for GF-NORA in LEO satellite-IoT. We developed an OFDM-symbol repetition technique to better distinguish the Doppler shifts of the LEO satellite channel. We established a grid-based parametric system model, and showed that joint DAD and CE can be formulated as a CS problem. However, we pointed out that the measurement matrix of the problem exhibits a special correlation structure, so that existing Bayesian CS algorithms such as AMP and Turbo-CS do not behave well. To address this issue, we proposed the MVSP algorithm which is robust to the sensing matrix and can efficiently exploit the channel sparsity in the delay-Doppler-user domain. We then used the EM method to learn the grid parameters and further improve the performance of MVSP. Simulation results demonstrated that the proposed algorithms significantly outperform the counterparts’ methods.

**REFERENCES**

[1] L. Atzori, I. A. Iera, and M. Giacomo, “The Internet of Things: A survey,” *Comput. Netw.*, vol. 54, pp. 2787–2805, May 2010.
[2] J. Chen, K. Hu, Q. Wang, Y. Sun, Z. Shi, and S. He, “Narrowband Internet of Things: Implementations and applications,” *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2299–2314, Dec. 2017.
[3] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, “Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things,” *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.
[4] J. Ding, M. Nemati, C. Ranaweera, and J. Choi, “IoT connectivity technologies and applications: A survey,” *IEEE Access*, vol. 8, pp. 67646–67673, 2020.
[5] 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, 2009.
[6] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
[7] J. Ma, X. Yuan, and L. Ping, “Turbo compressed sensing with partial DFT sensing matrix,” *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.
[8] M. Zhang, X. Yuan, and Z.-Q. He, “Variance state propagation for structured sparse Bayesian learning,” *IEEE Trans. Signal Process.*, vol. 68, pp. 2386–2400, 2020.
[9] Z. Chen, F. Sohrabi, and W. Yu, “Sparse activity detection for massive connectivity,” *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 1890–1904, Apr. 2018.
[10] T. Liu, S. Jin, C.-K. Wen, M. Matthaiou, and X. You, “Generalized channel estimation and user detection for massive connectivity with mixed-ADC massive MIMO,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3236–3250, Jun. 2019.
[11] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Scholer, “Compressive sensing-based adaptive user detection and channel estimation: Massive access meets massive MIMO,” *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.
[12] 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.
[13] Q. Zou, H. Zhang, D. Cai, and H. Yang, “A low-complexity joint user activity, channel and data estimation for grant-free massive MIMO systems,” *IEEE Signal Process. Lett.*, vol. 27, pp. 1290–1294, 2020.
[14] T. Ding, X. Yuan, and S. C. Liew, “Sparsity learning-based multiser user detection in grant-free massive-device multiple access,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3569–3582, Jul. 2019.
[15] M. de Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, “Satellite communication supporting internet of remote things,” *IEEE Internet Things J.*, vol. 36, no. 2, pp. 280–291, Feb. 2018.
studying new radio (NR) to support non-terrestrial networks.

Introduction to Probability

D. Bertsekas and J. N. Tsitsiklis, 2008.

Variance state propagation with clustered channel sparsity, in Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops), Xiamen, China, Jul. 2021, pp. 233–238.

Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior, IEEE Trans. Veh. Technol., vol. 67, no. 5, pp. 4635–4639, Jun. 2018.

Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals, IEEE Trans. Signal Process., vol. 63, no. 2, pp. 360–372, Jan. 2015.

Approximate message passing for recovery of sparse signals with Markov-random-field support structure, in Proc. Int. Conf. Mach. Learn., Washington, DC, USA, Jul. 2011, pp. 1–15.

Variance state propagation for channel estimation in underwater acoustic massive MIMO-OFDM with clustered channel sparsity, in Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops), Xiamen, China, Jul. 2021, pp. 233–238.

Introduction to Probability, Nashua, NH, USA: Athena Scientific, 2008.

Study on New Radio (NR) to Support Non-Terrestrial Networks, document TR 38.811 V15.4.0 Release 15, 3Gpp, Sep. 2020.

Mingchen Zhang received the B.S. degree in communication engineering from Harbin Engineering University, Harbin, China, in 2017. He is currently pursuing the Ph.D. degree in information and communication engineering with the University of Electronic Science and Technology of China, Chengdu, China.

Yong Zuo received the B.E. degree in communication engineering from Zhejiang University, China, in 2007, and the Ph.D. degree in communication engineering from the Chinese Academy of Sciences (CAS), China, in 2012. From 2012 to 2015, he was a Research Assistant Professor with the Chinese Academy of Sciences, Beijing, China. From 2015 to 2017, he was an Assistant Professor with the University of Electronic Science and Technology of China, Chengdu, China. He is currently a Professor with the National Key Laboratory on Wireless Communications, University of Electronic Science and Technology of China.

Xiaojun Yuan (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of electronic science and technology of China, in 2001. He is currently a Professor with the Department of Electrical Engineering, University of electronic science and technology of China. He has been a Visiting Scholar with the National Key Laboratory on Wireless Communications, University of Electronic Science and Technology of China. His research interests include machine learning, satellite communications, and wireless communications, including but not limited to distributed learning, intelligent communications, and Bayesian approximate inference.

Mingyang Yue received the B.S. degree in communication engineering from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, in 2020. He is currently pursuing the Ph.D. degree in electrical and information engineering with the National Key Laboratory on Wireless Communications, University of Electronic Science and Technology of China. His research interests include wireless communications, compressed sensing, and machine learning.

Sixin Li received the B.S. degree from the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China, in 2018, and the M.S. degree in communication and information system from the University of Electronic Science and Technology of China, Chengdu, China, in 2021.

Shaojie Ni received the Ph.D. degree from the National University of Defense Technology. He is currently a Professor with the Department of Electronic Science, National University of Defense Technology, and also the Satellite Navigation Expert. He has long been involved in BeiDou navigation satellite system construction and application promotion work. He is also responsible for more than ten projects as the main director.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.