This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

Massive MIMO-OTFS-Based Random Access for Cooperative LEO Satellite Constellations

Boxiao Shen, *Student Member, IEEE*, Yongpeng Wu, *Senior Member, IEEE*, Shiqi Gong, *Member, IEEE*, Heng Liu, *Member, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

***Abstract*—This paper investigates joint device identification, channel estimation, and symbol detection for cooperative multi-satellite-enhanced random access, where orthogonal time-frequency space modulation with the large antenna array is utilized to combat the dynamics of the terrestrial-satellite links (TSLs). We introduce the generalized complex exponential basis expansion model to parameterize TSLs, thereby reducing the pilot overhead. By exploiting the block sparsity of the TSLs in the angular domain, a message passing algorithm is designed for initial channel estimation. Subsequently, we examine two cooperative modes to leverage the spatial diversity within satellite constellations: the centralized mode, where computations are performed at a high-power central server, and the distributed mode, where computations are offloaded to edge satellites with minimal signaling overhead. Specifically, in the centralized mode, device identification is achieved by aggregating backhaul information from edge satellites, and channel estimation and symbol detection are jointly enhanced through a structured approximate expectation propagation (AEP) algorithm. In the distributed mode, edge satellites share channel information and exchange soft information about data symbols, leading to a distributed version of AEP. The introduced basis expansion model for TSLs enables the efficient implementation of both centralized and distributed algorithms via fast Fourier transform. Simulation results demonstrate that proposed schemes significantly outperform conventional algorithms in terms of the activity error rate, the normalized mean squared error, and the symbol error rate. Notably, the distributed mode achieves performance comparable to the centralized mode with only two exchanges of soft information about data symbols within the constellation.**

***Index Terms*—Satellite communications, random access, OTFS, message passing, Doppler effect**

I. Introduction

Internet-of-Things (IoT) is one of the critical scenarios in the next-generation communications, revolutionizing the way we live and work [1]. However, terrestrial networks cover only about 20 percent of the Earth's surface, facing substantial challenges in providing ubiquitous connectivity for diverse IoT applications, such as smart agriculture, climate monitoring and intelligent transportation systems [2]. Low earth orbit (LEO) satellites have attracted significant research interest in the fifth generation (5G) and are expected to be pivotal for the advancement of ubiquitous connectivity in the forthcoming 6G era [3]–[5]. Recent efforts have focused on the deployment of satellite constellations, such as OneWeb and Starlink [6], driven by breakthroughs in aerospace, production, and communication technologies. These developments pave the way for LEO satellite communications to offer seamless global coverage. Unlike traditional human-centric communications, IoT primarily relies on machine-type communications (MTC), characterized by bursty traffic patterns due to the sporadic activity of IoT devices [7]. In this context, the random access protocol plays a key role in supporting efficient connectivity.

Grant-free random access (GFRA) is favored in MTC for its potential to reduce signaling overhead and power consumption while improving access capability [8], [9]. Over the past few years, many methods have been proposed for joint device identification and channel estimation (JDICE) in terrestrial GFRA systems. For instance, [10] formulated the JDICE as a sparse signal recovery problem in compressed sensing and solved it by the approximate message passing (AMP) algorithm. Alternatively, [11] utilized a Bernoulli-Gaussian-mixture as the channel prior and adopted the vector AMP unrolled with the expectation-maximization (EM) algorithm to design a neural network that improves the performance of original vector AMP. The integration of orthogonal frequency division multiplexing (OFDM) into GFRA systems was investigated in [12], where the AMP-based algorithm was proposed to explore the channel sparsity in the angular domain. To mitigate timing and frequency offset impacts in OFDM-based GFRA, [13] proposed to decouple and incorporate the phase shift into the measurement matrix, leading to the structured generalized AMP for JDICE. It is worth noting that these schemes in [10]–[12] are tailored for block fading channels which are assumed to remain constant during one transmission, and [13] simplifies the system model by assuming that the frequency offset tends to zero. However, the inherent high mobility of LEO satellites introduces significant Doppler shifts, leading to rapid changes in the terrestrial-satellite links (TSLs). Device movement may further contribute to considerable Doppler spread [14], resulting in outdated channel state information and severe inter-carrier interference. These issues will degrade the performance of existing algorithms. Consequently, current terrestrial GFRA frameworks, without adjustments or new designs, struggle to meet the demands of LEO satellite communications.

In light of the Doppler effect, extensive efforts have been devoted to GFRA adaptations for LEO satellite communications.

The work of Y. Wu is supported in part by the Fundamental Research Funds for the Central Universities, National Science Foundation (NSFC) under Grant 62122052 and 62071289, 111 project BP0719010, and STCSM 22DZ2229005; The work by B. Ottersten is funded in part by the FNR research grant 18014377. *(Corresponding author: Yongpeng Wu)*.

B. Shen, Y. Wu, and W. Zhang are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mails: {boxiao.shen, yongpeng.wu, zhangwenjun}@sjtu.edu.cn).

S. Gong and H. Liu are with the School of Cyberspace Science and Technology, and the School of Information and Electronics, respectively, Beijing Institute of Technology, Beijing 100081, China (e-mails: {gsqyx, heng_liu_bit_ee}@163.com).

B. Ottersten is with the Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg City, Luxembourg (e-mail: bjorn.ottersten@uni.lu).

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

In [15], Bernoulli–Rician message passing combined with the EM algorithm was proposed for the LEO satellite-based narrowband GFRA, accounting for random phase rotations in the channel. However, this method still presumes a slow time-varying channel, which may be violated in dynamic satellite environments. The work in [16] focused on the OFDM-based GFRA for LEO satellite IoT, where the grid-based parametric model was proposed to discretize the delay and Doppler shift, and a message-passing type algorithm was designed for JDICE. However, the proposed scheme relies heavily on pre-compensation for both delay and Doppler shift, which limits its application in the LEO contexts. Meanwhile, [17] leveraged structured sparsity in the delay-Doppler-time domain to propose an AMP-based algorithm for joint active device, delay, and Doppler detection, while neglecting the potential non-line-of-sight paths. Orthogonal time frequency space (OTFS) modulation has emerged as a promising solution to ensure reliable communications in high-mobility scenarios [18]–[20], which converts the time-variant channels into quasi-static channels in the delay-Doppler domain [21], [22]. In [18], OTFS is combined with tandem spreading multiple access to accommodate the differential Doppler shift characteristics in LEO satellite-enabled GFRA. Note that [15]–[18] focuses on single-antenna systems. To exploit the spatial diversity, the integration of massive multiple-input multiple-output (MIMO) with OTFS-based GFRA has been explored in [19] and [20]. Specifically, [19] introduced a two-dimensional (2D) pattern coupled hierarchical prior within sparse Bayesian learning for JDICE, exploiting channel sparsity in the delay-Doppler-angle domain. [20] investigated OTFS aided by training sequences and proposed a two-stage JDICE scheme alongside a streamlined multi-user symbol detection method.

All the above GFRA schemes focus on one satellite node. Nevertheless, the advent of densely deployed LEO satellite constellations [23], coupled with their inter-connectivity through highly reliable and low-latency inter-satellite links (ISLs) [24], opens new avenues for enhanced RA schemes through cooperative processing. In [25], the study derived closed-form expressions for spectral efficiency in ultra-dense LEO satellite networks utilizing distributed massive MIMO techniques. The theoretical analysis and simulation results, based on Rician fading with random phase modeling for TSLs, demonstrated superior performance compared to both collocated massive MIMO and conventional single-satellite connectivity. In addition, [26] proposed a DFT-s-OFDM framework tailored for multi-satellite cooperative GFRA with massive MIMO, which involves independent channel estimation by single satellite, with subsequent activity and data detection centralized at a high-power node. However, [25] and [26] neglect the Doppler effect. Moreover, they are confined to the centralized cooperative mode, which requires a central server to process the collective signal, imposing huge computational burden on this server.

Overall, most previous works focus on the single satellite-based GFRA or the centralized cooperative schemes neglecting the Doppler effect. In this paper, we investigate the joint device identification, channel estimation, and symbol detection for GFRA enhanced by cooperative LEO satellite constellations with massive MIMO, where the generalized complex exponential basis expansion model (GCE-BEM) [27] is introduced as an efficient channel parameterization, and OTFS is leveraged to mitigate the doubly dispersive effect in TSLs. Beyond the centralized mode, our study examines the distributed mode, which enables the offloading of computational tasks to edge satellites. Our main contributions are summarized as follows:

- The application of GCE-BEM for modeling angular domain TSLs is firstly studied. Through this parameterized channel model, we analyze the input-output relationship of the system. Then, by exploiting the 2D channel block sparsity in the angular domain, the Markov random field aided message passing algorithm with EM is designed for initial channel estimation.
- Centralized cooperative multi-satellite-enhanced joint device identification, channel estimation, and symbol detection are investigated. Specifically, the device identification is performed by aggregating the estimated channel energy from edge satellites. For channel estimation and symbol detection refinement, we design the approximate expectation propagation (AEP) algorithm, structured into three modules: interference cancellation module mitigates both inter-user and inter-component interference in the BEM model, providing likelihoods for channels and transmitted data symbols; channel estimation and symbol detection refinement modules utilize the soft information from the interference cancellation module to further enhance channel estimation and symbol detection.
- Distributed methods for cooperative multi-satellite-enhanced joint device identification, channel estimation, and symbol detection are investigated. Each satellite conducts the device identification by sharing the estimated channel energy. To exploit the spatial diversity of multi-satellite configurations, we propose to exchange the soft information regarding transmitted data symbols within the constellation, leading to the distributed AEP method. Simulation results indicate that only two exchanges of soft information about data symbols within the constellation are required to attain the performance comparable to that of the centralized framework.

The rest of this paper is organized as follows. Section II introduces the system model and formulates the problem. Section III proposes initial channel estimation algorithm. The refinement schemes of two cooperative modes are discussed in Section IV. Finally, Section V evaluates the performance of proposed algorithms, followed by conclusions in Section VI.

Notations: The superscripts $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ denote the conjugate and conjugated-transpose operations, respectively. The boldface capital letter $\mathbf{X}$ and lowercase letter $\mathbf{x}$ denote matrix and vector, respectively. $\mathrm{diag}(\mathbf{x})$ denotes the diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal. $\bar{\jmath} = \sqrt{-1}$ denotes the imaginary unit. $\|\mathbf{X}\|_{\mathrm{F}}$ 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. $\otimes$ is the Kronecker product, $\odot$ is the Hadamard product, and $\oslash$ denotes element-wise division. $(\cdot)_M$ denotes mod M. $\delta(\cdot)$ denotes the Dirac delta function. $X[a,b]$ and $x_{a,b}$ denote the (a,b)-th element of $\mathbf{X}$. $x[a]$ and x_a denote the a-th element of $\mathbf{x}$.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

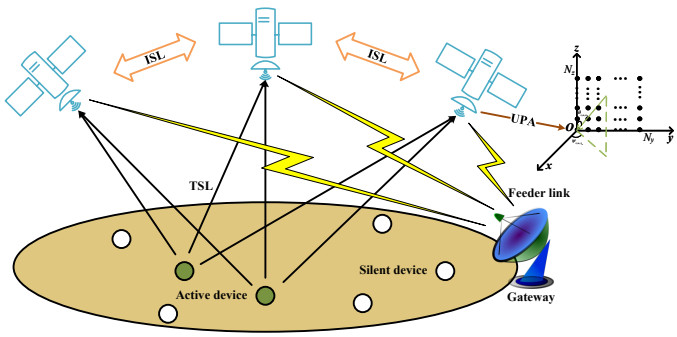


Fig. 1. Diagram for cooperative multi-satellite-enhanced random access.

II. System Model

We investigate the LEO satellite constellations-enabled grant-free random access system as illustrated in Fig. 1, where S_a LEO satellites are interconnected by the high-reliable ISLs and cooperatively offer access services to U single-antenna terrestrial devices. In a given time interval, active devices share the same time-frequency resources to transmit to the S_a satellites. Additionally, each satellite in the constellation is equipped with a regenerative payload capable of on-board processing of baseband signals and a uniform planar array (UPA) with $N_a = N_z \times N_y$ antennas, where N_z and N_y are the number of antennas along the z-axis and y-axis, respectively. To mitigate the Doppler effect inherent in the TSLs, OTFS modulation is employed. Furthermore, we assume that the propagation delay can be partially pre-compensated at the devices by coarse timing advance [28] such that the residual delay is within one symbol duration.

A. Terrestrial-satellite Link Model

Due to the high mobility of LEO satellites, the TSLs undergo rapid variations. This variability, compounded by multipath propagation, requires a doubly dispersive channel model for accurate TSLs characterization. Consequently, the channel response for the u-th device to the $(n_z + n_y N_z)$-th antenna of the s_a-th satellite at time n and tap l' is characterized as [28]–[30]

$$\tilde{h}_{u,s_a}^{n_z,n_y}[n,l'] = \sum_{i=1}^{P} h_{i,u,s_a} e^{\bar{\jmath}2\pi\nu_{i,u,s_a} nT_s} e^{\bar{\jmath}\pi n_z \Theta_{i,u,s_a}^{z}} e^{\bar{\jmath}\pi n_y \Theta_{i,u,s_a}^{y}} \times A_{g_{rx,tx}}\left(l'T_s - \tau_{i,u,s_a}\right), \quad (1)$$

where $u = 0,\ldots,U-1$, $n_z = 0,\ldots,N_z-1$, $n_y = 0,\ldots,N_y-1$, $s_a = 0,\ldots,S_a-1$, P is the number of physical paths, and T_s is the system sampling interval. The notations h_{i,u,s_a}, ν_{i,u,s_a}, and τ_{i,u,s_a} represent the path gain, Doppler shift, and residual delay for the i-th path from the u-th device to the s_a-th satellite, respectively. The directional cosines $\Theta_{i,u,s_a}^{z} = \cos\theta_{i,u,s_a}$ and $\Theta_{i,u,s_a}^{y} = \sin\theta_{i,u,s_a}\sin\psi_{i,u,s_a}$, along the z-axis and y-axis, depend on the zenith angle θ_{i,u,s_a} and azimuth angle ψ_{i,u,s_a}, respectively. Here, we assume that the matched filter is adopted, and then the band-limited pulse shaping filter response $A_{g_{rx,tx}}(t)$ is given by

$$A_{g_{rx,tx}}(t) = \int g_{rx}^{*}(t'-t)\, g_{tx}(t')dt', \quad (2)$$

where $g_{tx}(t)$ and $g_{rx}(t)$ are the transmit and receive waveform, respectively. Contrary to prior OTFS studies [19]–[21], which often require Doppler shifts or delays to be integer multiples of the frequency and delay resolutions, our framework does not rely on this assumption. The channel model in (1) accommodates fractional Doppler shifts and delays, making it more suited for realistic satellite communication scenarios. Moreover, it's reasonable to assume that the propagation paths from a single device to a satellite have identical angles due to the satellite's significantly higher altitude compared to that of scatterers near the devices [14], i.e., $\theta_{i,u,s_a} = \theta_{u,s_a}$ and $\psi_{i,u,s_a} = \psi_{u,s_a}$. Then, the channel model in (1) can be simplified as

$$\tilde{h}_{u,s_a}^{n_z,n_y}[n,l'] = e^{\bar{\jmath}\pi n_z \Theta_{u,s_a}^{z}} e^{\bar{\jmath}\pi n_y \Theta_{u,s_a}^{y}} \sum_{i=1}^{P} h_{i,u,s_a} e^{\bar{\jmath}2\pi\nu_{i,u,s_a} nT_s} \times A_{g_{rx,tx}}\left(l'T_s - \tau_{i,u,s_a}\right). \quad (3)$$

It is observed that the term $e^{j2\pi\nu_{i,u,s_a} nT_s}$ exhibits significant temporal variation due to the Doppler effect, in contrast to the other terms which remain relatively constant over time. This fact motivates us to adopt the GCE-BEM [27] to parameterize this time varying channel as

$$\tilde{h}_{u,s_a}^{n_z,n_y}[n,l'] = e^{\bar{\jmath}\pi n_z \Theta_{u,s_a}^{z}} e^{\bar{\jmath}\pi n_y \Theta_{u,s_a}^{y}} \sum_{q=0}^{Q} \tilde{c}_{q,l',u,s_a} e^{\bar{\jmath}\omega_q n}, \quad (4)$$

where Q is the order of the BEM basis functions, $\tilde{c}_{q,l',u,s_a}$ is the q-th BEM coefficient of l'-th channel tap from the u-th device to the s_a-th satellite, and ω_q is the q-th BEM modeling frequency quantizing the Doppler shift. Note that the modeling error term is omitted in (4) primarily since we focus on BEMs that provide an excellent fit. As highlighted in [31], including the modeling error term does not significantly improve the reliability of channel estimates if the BEM itself cannot accurately characterize the true channel. Furthermore, [32] shows that, under certain mild conditions, the modeling error can be approximated as a Gaussian random variable with zero mean, allowing it to be absorbed into the noise term. To exploit the sparsity of channel resulted from adoption of the massive antenna array, the two dimensional discrete Fourier transform (DFT) is applied for transforming the channel from space domain into angular domain, i.e.,

$$h_{u,s_a}^{a_z,a_y}[n,l'] = \sum_{n_z=0}^{N_z-1}\sum_{n_y=0}^{N_y-1} \tilde{h}_{u,s_a}^{n_z,n_y}[n,l'] e^{-\bar{\jmath}2\pi\left(\frac{a_z n_z}{N_z}+\frac{a_y n_y}{N_y}\right)} = \sum_{q=0}^{Q} c_{q,l',u,s_a}^{a_z,a_y} e^{\bar{\jmath}\omega_q n}, \quad (5)$$

where $a_z = 0,\ldots,N_z-1$ and $a_y = 0,\ldots,N_y-1$. The parameter $c_{q,l',u,s_a}^{a_z,a_y} = \tilde{c}_{q,l',u,s_a}\Pi_{N_z}(a_z - N_z\Theta_{u,s_a}^{z}/2)\Pi_{N_y}(a_y - N_y\Theta_{u,s_a}^{y}/2)$ is the BEM coefficient in the angular domain, where $\Pi_N(x) \triangleq \frac{1}{N}\sum_{i=0}^{N-1} e^{-\bar{\jmath}2\pi\frac{x}{N}i}$. It is observed that $c_{q,l',u,s_a}^{a_z,a_y}$ has the dominant element only if $a_z \approx N_z\Theta_{u,s_a}^{z}/2$ and $a_y \approx N_y\Theta_{u,s_a}^{y}/2$, which means that the coefficient has 2D block sparsity in the angular domain. For notation simplicity, we denote the index of n_a-th antenna as $n_a = a_z + a_y N_z$,

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

$n_a = 0, \ldots, N_z N_y - 1$, and then (5) can be re-expressed as

$$h_{u,s_a}^{n_a}[n, l'] = \sum_{q=0}^{Q} c_{q,l',u,s_a}^{n_a} e^{\bar{\jmath}\omega_q n}. \tag{6}$$

B. Transmission Signal Model

In this subsection, the transmission process is explained. Each terrestrial device employs OFDM-based OTFS modulation. Firstly, the u-th device rearranges a sequence of MN information symbols into one OTFS frame $\mathbf{X}_u^{\mathrm{DD}} \in \mathbb{C}^{M\times N}$, where M and N are the number of subcarriers and OFDM symbols within one frame, respectively. Then, by the inverse symplectic finite Fourier transform (ISFFT) [22], the 2D data frame $\mathbf{X}_u^{\mathrm{DD}}$ in the delay-Doppler domain is mapped to $\mathbf{X}_u^{\mathrm{TF}}$ in the time-frequnecy domain, i.e.,

$$\mathbf{X}_u^{\mathrm{TF}} = \mathbf{F}_M \mathbf{X}_u^{\mathrm{DD}} \mathbf{F}_N^{\mathrm{H}}, \tag{7}$$

where $\mathbf{F}_{\mathrm{M}} \in \mathbb{C}^{M\times M}$ and $\mathbf{F}_{\mathrm{N}} \in \mathbb{C}^{N\times N}$ are normalized DFT matrices. The OFDM modulator processes the 2D data frame $\mathbf{X}_u^{\mathrm{TF}}$ to get the transmitted data frame $\mathbf{B}_u^{\mathrm{TD}}$, i.e.,

$$\mathbf{B}_u^{\mathrm{TD}} = \mathbf{F}_M^{\mathrm{H}} \mathbf{X}_u^{\mathrm{TF}} = \mathbf{X}_u^{\mathrm{DD}} \mathbf{F}_N^{H}. \tag{8}$$

Then, the data frame $\mathbf{B}_u^{\mathrm{TD}}$ is transformed to the 1D transmit signal $\mathbf{b}_u^{\mathrm{TD}} = \mathrm{vec}\left(\mathbf{B}_u^{\mathrm{TD}}\right)$, where $\mathrm{vec}(\cdot)$ denotes the vectorization operator stacking the columns of one matrix into a long column vector. Note that in the traditional OTFS literature [19], [20], the guard interval will be added for each OFDM symbols to avoid inter-symbol interference. Here, we only insert one cyclic prefix (CP) with length of M_{cp} before the head of the OTFS frame for increasing spectral efficiency. Typically, M_{cp} should be more than the largest normalized residual delay $L = \lceil \tau_{\max}/T_s \rceil$, where $\tau_{\max}$ represents the largest residual delay among all the resolvable physical paths in TSLs. After CP removal, the received signal is given by

$$r_{u,s_a}^{n_a}[n] = \sum_{l'=0}^{L} h_{u,s_a}^{n_a}\left[n + M_{\mathrm{cp}}, l'\right] b_u^{\mathrm{TD}}\left[(n-l')_{MN}\right], \tag{9}$$

where $n = 0, \ldots, MN-1$ and noise is neglected here for simplicity. Then, this sequence of data $\mathbf{r}_{u,s_a}^{n_a} = [r_{u,s_a}^{n_a}[0], \ldots, r_{u,s_a}^{n_a}[MN-1]]$ is rearranged to the 2D data frame $\mathbf{R}_{u,s_a}^{n_a} = \mathrm{vec}^{-1}(\mathbf{r}_{u,s_a}^{n_a})$, where $\mathrm{vec}^{-1}(\cdot)$ is the inverse operation of $\mathrm{vec}(\cdot)$. Applying the M-point DFT on each OFDM symbol, we can obtain the received signal in the time-frequency domain as

$$\mathbf{Y}_{n_a,u,s_a}^{\mathrm{TF}} = \mathbf{F}_{\mathrm{M}} \mathbf{R}_{u,s_a}^{n_a}. \tag{10}$$

Finally, the data frame $\mathbf{Y}_{n_a,u,s_a}^{\mathrm{TF}}$ is transformed into $\mathbf{Y}_{n_a,u,s_a}^{\mathrm{DD}}$ in the delay-Doppler domain, i.e.,

$$\mathbf{Y}_{n_a,u,s_a}^{\mathrm{DD}} = \mathbf{F}_M^{\mathrm{H}} \mathbf{Y}_{n_a,u,s_a}^{\mathrm{TF}} \mathbf{F}_N = \mathbf{R}_{u,s_a}^{n_a} \mathbf{F}_N. \tag{11}$$

We denote the (l,k)-th element of $\mathbf{Y}_{n_a,u,s_a}^{\mathrm{DD}}$ and $\mathbf{X}_u^{\mathrm{DD}}$ as $Y_{n_a,u,s_a}^{\mathrm{DD}}[l,k]$ and $X_u^{\mathrm{DD}}[l,k]$ respectively, where $l = 0, \ldots, M-1$ and $k = 0, \ldots, N-1$. According to (6)-(11), the input-output relationship (without noise term) of this system is given in the following proposition.

Proposition 1: Given the q-th BEM modeling frequency $\omega_q = \frac{2\pi q'}{MNR}$ with $q' = \left(q - \left\lceil \frac{Q}{2} \right\rceil\right)$ and the resolution factor R, the received signal in the delay-Doppler-angle domain from the u-th device to the n_a-th antenna of the s_a-th satellite is represented as

$$Y_{n_a,u,s_a}^{\mathrm{DD}}[l,k] = \sum_{q=0}^{Q} \sum_{l'=0}^{L} \sum_{k'=0}^{N-1} e^{\bar{\jmath}\omega_q l} \alpha_{l,l'}\left[k,k',q'\right] c_{q,l',u,s_a}^{n_a} \times X_u^{\mathrm{DD}}\left[(l-l')_M, k'\right], \tag{12}$$

where we have

$$\alpha_{l,l'}\left[k,k',q'\right] = \begin{cases} \frac{1}{N} \frac{1-e^{\bar{\jmath}2\pi\left(k'-\left(k-\frac{q'}{R}\right)\right)}}{1-e^{\bar{\jmath}\frac{2\pi}{N}\left(k'-\left(k-\frac{q'}{R}\right)\right)}}, & l' \leqslant l \\ \frac{1}{N} e^{-\bar{\jmath}\frac{2\pi}{N}k'} \frac{1-e^{\bar{\jmath}2\pi\left(k'-\left(k-\frac{q'}{R}\right)\right)}}{1-e^{\bar{\jmath}\frac{2\pi}{N}\left(k'-\left(k-\frac{q'}{R}\right)\right)}}, & l' > l \end{cases}.$$

Proof: Please see Appendix A. ■

It is observed that the received signal in the delay-Doppler-angle domain is given by the sum of Q components. Each component is the result of phase compensated periodic convolution between the OTFS data frame and the BEM coefficient. Here, the order Q and the resolution factor R govern the modeling accuracy of the BEM. At $R = 1$, the GCE-BEM is simplified to the CE-BEM which has higher modeling error. Increasing R to 2 enhances the Doppler shift quantization precision, thereby significantly reducing the modeling error. As indicated in [27], $R = 2$ is large enough for modeling time-varying channel, with only marginal gain from further increases. Additionally, the order Q usually depends on R, given as $Q \geq 2\lceil RN\bar{\nu}_{\max}\rceil$, where $\bar{\nu}_{\max} = \nu_{\max}/\Delta f$; $\nu_{\max}$ is the largest Doppler shift across all the resolvable physical paths in TSLs and Δf is the subcarrier spacing. Under the BEM modeling, the unknown channel parameters from the u-th device to the s_a-th satellite have been reduced significantly from $MN(L+1)$ to $(Q+1)(L+1)$ at each antenna. However, a higher resolution leads to an increased BEM order, which may in turn raise the interference among the Q components in (12). This interference makes it difficult to accurately estimate channel and detect transmitted symbols. Therefore, we should make Q as small as possible while ensuring the BEM modeling accuracy.

C. Problem Formulation

According to Proposition 1 and superposition principle, the received signal at the n_a-th antenna of the s_a-th satellite is expressed as

$$Y_{n_a,s_a}^{\mathrm{DD}}[l,k] = \sum_{u=0}^{U-1} \sum_{q=0}^{Q} \sum_{l'=0}^{L} \sum_{k'=0}^{N-1} \lambda_u c_{q,l',u,s_a}^{n_a} e^{\bar{\jmath}\omega_q l} \alpha_{l,l'}\left[k,k',q'\right] \times X_u^{\mathrm{DD}}\left[(l-l')_M, k'\right] + E_{s_a}^{n_a}[l,k], \tag{13}$$

where $\lambda_u \in \{0,1\}$ is the activity indicator of the u-th device with active probability p_λ and we denote the set of active devices as $\mathcal{U}_a = \{u \mid \lambda_u = 1, u = 0, \ldots, U-1\}$; $E_{s_a}^{n_a}[l,k] \sim \mathcal{CN}(0, \sigma_{s_a}^2)$ is the noise. To facilitate algorithm design, we rewrite (13) as the matrix form, i.e.,

$$\mathbf{y}_{s_a}^{n_a} = \sum_{u=0}^{U-1} \sum_{q=0}^{Q} \left(\mathbf{F}_N \otimes \mathbf{I}_M\right) \mathrm{diag}\left(\mathbf{b}_q\right) \mathbf{F}_{MN}^{\mathrm{H}} \times$$

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

$$\operatorname{diag}\left(\mathbf{F}_{1:L+1}\mathbf{c}_{q,u,s_a}^{n_a}\right)\mathbf{F}_{MN}\left(\mathbf{F}_N^{\mathrm{H}}\otimes\mathbf{I}_M\right)\mathbf{x}_u+\mathbf{e}_{s_a}^{n_a}, \quad (14)$$

where $\mathbf{y}_{s_a}^{n_a} = \mathrm{vec}(\mathbf{Y}_{n_a,s_a}^{\mathrm{DD}})$, $\mathbf{x}_u = \mathrm{vec}(\mathbf{X}_u^{\mathrm{DD}})$, $\mathbf{e}_{s_a}^{n_a} = \mathrm{vec}(\mathbf{E}_{s_a}^{n_a})$, $\mathbf{I}_M \in \mathbb{R}^{M\times M}$ is the identity matrix, $\mathbf{b}_q = [1, e^{\bar{\jmath}\omega_q}, \ldots, e^{\bar{\jmath}\omega_q(MN-1)}]$ is the q-th BEM basis function, $\mathbf{F}_{1:L+1}$ represents the first $L+1$ columns of $\mathbf{F}_{MN}$, and $\mathbf{c}_{q,u,s_a}^{n_a} = \sqrt{MN}\lambda_u[c_{q,0,u,s_a}^{n_a}, \ldots, c_{q,L,u,s_a}^{n_a}]$ collecting the q-th BEM coefficients for L channel taps and combining the device activity indicator.

Given all the received signals of LEO satellite constellations, we aim at jointly estimating $\left(\mathbf{c}_{q,u,s_a}^{n_a}, \mathbf{x}_u\right)$ based on the maximum a posteriori probability (MAP) principle, i.e.,

$$(\hat{\mathbf{c}}_{q,u,s_a}^{n_a}, \hat{\mathbf{x}}_u) = \underset{(\mathbf{c}_{q,u,s_a}^{n_a}, \mathbf{x}_u)}{\arg\max}\ p\left(\mathbf{c}_{q,u,s_a}^{n_a}, \mathbf{x}_u \mid \mathbf{y}\right), \quad (15)$$

where $\mathbf{y} = [\mathbf{y}_0^0, \ldots, \mathbf{y}_{s_a}^{n_a}, \ldots, \mathbf{y}_{S_a}^{N_zN_y-1}]$. The problem (15) is generally non-convex and difficult to solve. Message passing algorithms offer potential solutions, but the coupling of variables into a bilinear function complicates the implementation of exact belief propagation (BP) based on the sum-product rule. Furthermore, the accuracy of bilinear inference problems is highly dependent on the initial values of the solution variables [33], necessitating a reasonable initialization method for precise channel and data estimation. To overcome these issues, we propose a two-phase algorithm. The first phase based on the GAMP [34] algorithm to generate initial estimates of the channel and transmitted symbols. The second phase resorts to the expectation propagation (EP) [35] algorithm to enhance device identification, channel estimation, and symbol detection.

III. Initial Channel Estimation for Single Satellite

In this section, we firstly illustrate the OTFS data frame structure. Then, based on the Bayesian method, a message-passing type algorithm combined with EM is designed for the initial channel estimation. For sake of low computations and pilot overhead, similar to [32], we adopt a lower resolution $R=1$ with the smaller order of BEM $Q_{\mathrm{I}} \geq 2\lceil N\bar{\nu}_{\max}\rceil$ in this initial phase, and higher resolution is utilized in the refine phase.

A. Frame Structure

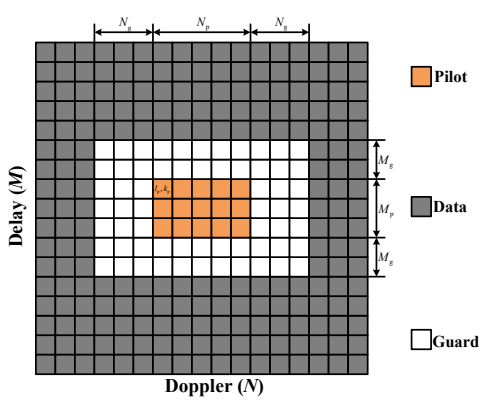


Fig. 2. OTFS frame structure.

When $R=1$, (13) can be simplified as

$$Y_{n_a,s_a}^{\mathrm{DD}}[l,k] = \sum_{u=0}^{U-1}\sum_{q=0}^{Q_{\mathrm{I}}}\sum_{l'=0}^{L}\lambda_u e^{\bar{\jmath}\omega_q l}\alpha_{l,l'}[k,q']\, c_{q,l',u,s_a}^{n_a} \times X_u^{\mathrm{DD}}\left[(l-l')_M, (k-q')_N\right] + E_{s_a}^{n_a}[l,k], \quad (16)$$

where

$$\alpha_{l,l'}[k,q'] = \begin{cases} 1, & l' \leqslant l \\ e^{-\bar{\jmath}\frac{2\pi}{N}(k-q')}, & l' > l \end{cases}. \quad (17)$$

By considering this 2D periodic convolution relationship, the OTFS frame structure in the delay-Doppler domain is depicted in Fig. 2. The position of pilots is given by $l = l_{\mathrm{p}}, \cdots, l_{\mathrm{p}} + M_{\mathrm{p}} - 1$ and $k = k_{\mathrm{p}}, \cdots, k_{\mathrm{p}} + N_{\mathrm{p}} - 1$, where $(l_{\mathrm{p}}, k_{\mathrm{p}})$ is the initial position of pilots. M_{p} and N_{p} are the length of pilots along the delay and Doppler dimension, respectively. Without loss of generality, we assume that $l_{\mathrm{p}} \geq L$. To avoid interference between pilots and data, guard intervals along delay and Doppler dimensions with length of $M_{\mathrm{g}} \geq L$ and $N_{\mathrm{g}} \geq \left\lceil \frac{Q_{\mathrm{I}}}{2} \right\rceil$ are required. Additionally, the pilots and guard symbols are the i.i.d. complex Gaussian random sequences. Note that guard symbols can be considered as one kind of pilots. Therefore, the pilot overhead of one OTFS frame is obtained as $\rho = \frac{(N_{\mathrm{p}}+2N_{\mathrm{g}})(M_{\mathrm{p}}+2M_{\mathrm{g}})}{MN}$.

Given the OTFS frame structure in Fig. 2, the received symbols with the same positions as pilots are utilized for channel estimation for each satellite and can be written as

$$\mathbf{Y}_{s_a} = \left((\mathbf{1}_U^{\mathrm{T}} \otimes \mathbf{P}^{\omega}) \odot \mathbf{X}^{\mathrm{p}}\right)\mathbf{C}_{s_a} + \mathbf{E}_{s_a}^{\mathrm{p}}, \quad (18)$$

where $\mathbf{Y}_{s_a} \in \mathbb{C}^{M_{\mathrm{p}}N_{\mathrm{p}}\times N_zN_y}$ with the $(l - l_{\mathrm{p}} + k - k_{\mathrm{p}}, n_a)$-th element $Y_{n_a,s_a}^{\mathrm{DD}}[l,k]$, $\mathbf{1}_U \in \mathbb{C}^U$ is the all-one vector, and $\mathbf{P}^{\omega} \in \mathbb{C}^{M_{\mathrm{p}}N_{\mathrm{p}}\times(Q_{\mathrm{I}}+1)(L+1)}$ with the $(l - l_{\mathrm{p}} + \tilde{l}M_{\mathrm{p}}, q + \bar{l})$-th element $e^{\bar{\jmath}\omega_q l}$, $\tilde{l} = 0, \ldots, N_{\mathrm{p}} - 1$, $\bar{l} = 0, \ldots, L$. $\mathbf{X}^{\mathrm{p}} = [\mathbf{X}_0^{\mathrm{p}}, \ldots, \mathbf{X}_{U-1}^{\mathrm{p}}]$ contains the pilots of U devices, where $\mathbf{X}_u^{\mathrm{p}} \in \mathbb{C}^{M_{\mathrm{p}}N_{\mathrm{p}}\times(Q_{\mathrm{I}}+1)(L+1)}$ with the $(l - l_{\mathrm{p}} + (k - k_{\mathrm{p}})M_{\mathrm{p}}, l' + q(L+1))$-th element $X_u^{\mathrm{DD}}\left[(l-l')_M, (k-q')_N\right]$, $\mathbf{C}_{s_a} \in \mathbb{C}^{U(Q_{\mathrm{I}}+1)(L+1)\times N_zN_y}$ with the $(u(Q_{\mathrm{I}}+1)(L+1) + q(L+1) + l', n_a)$-th element $\lambda_u c_{q,l',u,s_a}^{n_a}$, and $\mathbf{E}_{s_a}^{\mathrm{p}}$ is the Gaussian noise with the $(l - l_{\mathrm{p}} + k - k_{\mathrm{p}}, n_a)$-th element $E_{s_a}^{n_a}[l,k]$. With the definition $\mathbf{X}^{\omega} \triangleq ((\mathbf{1}_U^{\mathrm{T}} \otimes \mathbf{P}^{\omega}) \odot \mathbf{X}^{\mathrm{p}})$, we rewrite (18) as

$$\mathbf{Y}_{s_a} = \mathbf{X}^{\omega}\mathbf{C}_{s_a} + \mathbf{E}_{s_a} = \mathbf{Z}_{s_a} + \mathbf{E}_{s_a}^{\mathrm{p}}, \quad (19)$$

where $\mathbf{Z}_{s_a} \triangleq \mathbf{X}^{\omega}\mathbf{C}_{s_a}$. Then, the channel estimation turns into a sparse signal reconstruction problem. Next, the compressed sensing-based algorithm is designed for channel recovery, and simultaneously, exploiting the 2D block sparsity of the channel in the angular domain.

B. Initial Channel Estimation Algorithm

1) Prior distribution: To estimate the channel in (19), we resort to Bayesian method which needs prior distribution of the estimated variables. Firstly, we model the channel as K-components Bernoulli-Gaussian-mixture (BGM) distribution, which is a general and accurate distribution for the practical scenarios with a large-scale antenna array [11]. Additionally,

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

we assign different BGM distributions for the channel of different devices, i.e.,

$$p\left(c_{i,j}^{u,s_a} \mid s_{i,j}^{u,s_a}\right) = \delta\left(s_{i,j}^{u,s_a}+1\right)\delta\left(c_{i,j}^{u,s_a}\right) + \delta\left(s_{i,j}^{u,s_a}-1\right) \times \sum_{k=1}^{K} \omega_k^{u,s_a} \mathcal{CN}\left(c_{i,j}^{u,s_a} \mid \mu_k^{u,s_a}, \phi_k^{u,s_a}\right), \quad (20)$$

where $c_{i,j}^{u,s_a}$ is the $(u(Q_{\mathrm{I}}+1)(L+1)+i,j)$-th element of $\mathbf{C}_{s_a}$, $i=0,\ldots,(Q_{\mathrm{I}}+1)(L+1)-1$, $j=0,\ldots,N_zN_y-1$, corresponding to the channel of the u-th device, $s_{i,j}^{u,s_a} \in \{+1,-1\}$ is the corresponding support, and $\{\omega_k^{u,s_a}, \mu_k^{u,s_a}, \phi_k^{u,s_a}\}$ are the parameters of the BGM distribution. Furthermore, we adopt the Markov random field (MRF) prior to exploit 2D block sparsity of the channel, and then its support can be modeled by the classic Ising model as

$$p(\mathbf{s}_i^{u,s_a}) \propto \left(\prod_{j=0}^{N_zN_y-1} \prod_{j'\in\mathcal{D}_j} \Gamma\left(s_{i,j}^{u,s_a}, s_{i,j'}^{u,s_a}\right)\right)^{\frac{1}{2}} \times \prod_{j=0}^{N_zN_y-1} \Psi\left(s_{i,j}^{u,s_a}\right), \quad (21)$$

where $\mathbf{s}_i^{u,s_a} = [s_{i,0}^{u,s_a},\ldots,s_{i,N_zN_y-1}^{u,s_a}]$ involves the supports of $(u(Q_{\mathrm{I}}+1)(L+1)+i)$-th row in $\mathbf{C}_{s_a}$, $\mathcal{D}_j \subset \{0,\ldots,N_zN_y-1\}\backslash j$ contains the indexes of the neighbors of index j, $\Gamma(s_{i,j}^{u,s_a}, s_{i,j'}^{u,s_a}) = \exp(\beta s_{i,j}^{u,s_a} s_{i,j'}^{u,s_a})$, and $\Psi(s_{i,j}^{u,s_a}) = \exp(-\gamma s_{i,j}^{u,s_a})$. β and γ are the parameters of MRF prior, where a larger β implies a larger size of each block of nonzeros, and a larger γ encourages a sparser $\mathbf{C}_{s_a}$. Based on (19)-(21), the posterior distribution related to channel is given by

$$p(\mathbf{Z}_{s_a},\mathbf{C}_{s_a},\mathbf{S}_{s_a} \mid \mathbf{Y}_{s_a}) \propto \prod_u \prod_i p(\mathbf{s}_i^{u,s_a}) \prod_j p\left(c_{i,j}^{u,s_a} \mid s_{i,j}^{u,s_a}\right) \prod_m \prod_j p\left(y_{m,j}^{s_a} \mid z_{m,j}^{s_a}\right) \delta\left(z_{m,j} - \mathbf{c}_{s_a j}^{\mathrm{T}} \mathbf{x}_m^{\omega}\right), \quad (22)$$

where $y_{m,j}^{s_a}$ and $z_{m,j}^{s_a}$ are the (m,j)-th element of $\mathbf{Y}_{s_a}$ and $\mathbf{Z}_{s_a}$, respectively. $\mathbf{c}_{s_a j}$ is the j-th column of $\mathbf{C}_{s_a}$ and $\mathbf{x}_m^{\omega}$ is a column vector containing the elements of m-th row of $\mathbf{X}^{\omega}$. Unfortunately, the MMSE or MAP estimation of $\mathbf{C}_{s_a}$ with respect to (22) is still hard to carry out straightforwardly, since it involves marginalizing a joint distribution with high dimensions. In order to obtain a tractable alternative, the GAMP algorithm is firstly leveraged to eliminate inter-device interference in (19) and obtain the scalar likelihood for $c_{i,j}^{u,s_a}$. Then the message passing algorithm is combined with GAMP to get the approximated MMSE estimation of channel.

2) MMSE estimation: We denote $f_{i,j}^{c_{i,j}^{u,s_a}}$ as the factor corresponding to $p(c_{i,j}^{u,s_a} \mid s_{i,j}^{u,s_a})$; Δ_n^m and Δ_m^n represent the message passed from node m to n and the message in the corresponding opposite direction, respectively. Now we focus on the message passing related to the MRF prior. Note that the message $\Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{c_{i,j}^{u,s_a}}$ is approximated as the Gaussian distribution by GAMP, given as

$$\Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{c_{i,j}^{u,s_a}} = \mathcal{CN}\left(c_{i,j}^{u,s_a} \mid \hat{r}_{i,j}^{c_{i,j}^{u,s_a}}, \tau_{i,j}^{r_{i,j}^{u,s_a}}\right), \quad (23)$$

where the mean $\hat{r}_{i,j}^{c_{i,j}^{u,s_a}}$ and variance $\tau_{i,j}^{c_{i,j}^{u,s_a}}$ are updated by GAMP (see lines 9-10 of Algorithm 1). Then, the message from $f_{i,j}^{c_{i,j}^{u,s_a}}$ to $s_{i,j}^{u,s_a}$ can be derived as

$$\Delta_{s_{i,j}^{u,s_a}}^{f_{i,j}^{c_{i,j}^{u,s_a}}} \propto \int_{c_{i,j}^{u,s_a}} p\left(c_{i,j}^{u,s_a} \mid s_{i,j}^{u,s_a}\right) \Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{c_{i,j}^{u,s_a}} = \eta_{i,j}^{u,s_a}\delta\left(s_{i,j}^{u,s_a}-1\right) + (1-\eta_{i,j}^{u,s_a})\delta\left(s_{i,j}^{u,s_a}+1\right), \quad (24)$$

where $\eta_{i,j}^{u,s_a} = \frac{\eta_{i,j}^{\mathrm{A}^{u,s_a}}}{\mathcal{CN}(0|\hat{r}_{i,j}^{c_{i,j}^{u,s_a}},\tau_{i,j}^{r_{i,j}^{u,s_a}})+\eta_{i,j}^{\mathrm{A}^{u,s_a}}}$, and $\eta_{i,j}^{\mathrm{A}^{u,s_a}} = \sum_{k=1}^{K} \omega_k^{u,s_a}\mathcal{CN}(0 \mid \mu_k^{u,s_a} - \hat{r}_{i,j}^{c_{i,j}^{u,s_a}}, \phi_k^{u,s_a} + \tau_{i,j}^{r_{i,j}^{u,s_a}})$. Subsequently, MRF utilizes $\Delta_{s_{i,j}^{u,s_a}}^{f_{i,j}^{c_{i,j}^{u,s_a}}}$ to exploit the 2D block sparsity of the channel matrix, where the concrete message derivations are given in Appendix B. With the output message $\Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{s_{i,j}^{u,s_a}}$ from MRF, the refined prior of $c_{i,j}^{u,s_a}$ is given by

$$\Delta_{c_{i,j}^{u,s_a}}^{f_{i,j}^{c_{i,j}^{u,s_a}}} \propto \int_{s_{i,j}^{u,s_a}} p\left(c_{i,j}^{u,s_a} \mid s_{i,j}^{u,s_a}\right) \Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{s_{i,j}^{u,s_a}} = (1-\zeta_{i,j}^{u,s_a}) \times \delta(c_{i,j}^{u,s_a}) + \zeta_{i,j}^{u,s_a} \sum_{k=1}^{K} \omega_k^{u,s_a}\mathcal{CN}\left(c_{i,j}^{u,s_a} \mid \mu_k^{u,s_a}, \phi_k^{u,s_a}\right), \quad (25)$$

where $\zeta_{i,j}^{u,s_a}$ is given in (71). Then, combining the refined prior and the likelihood in (23), the posterior distribution of $c_{i,j}^{u,s_a}$ in this initial phase can be recognized as a BGM distribution, given as

$$\Delta_{c_{i,j}^{u,s_a}}^{\mathrm{I}} \propto \Delta_{c_{i,j}^{u,s_a}}^{f_{i,j}^{c_{i,j}^{u,s_a}}} \Delta_{f_{i,j}^{c_{i,j}^{u,s_a}}}^{c_{i,j}^{u,s_a}} = (1-\chi_{i,j}^{u,s_a})\delta(c_{i,j}^{u,s_a}) + \chi_{i,j}^{u,s_a} \times \sum_{k=1}^{K} \bar{\omega}_{i,j,k}^{u,s_a}\mathcal{CN}(c_{i,j}^{u,s_a} \mid \theta_{i,j,k}^{u,s_a}, \varphi_{i,j,k}^{u,s_a}), \quad (26)$$

where

$$\varphi_{i,j,k}^{u,s_a} = \left((\phi_k^{u,s_a})^{-1} + (\tau_{i,j}^{r_{i,j}^{u,s_a}})^{-1}\right)^{-1}, \quad (27)$$

$$\theta_{i,j,k}^{u,s_a} = \varphi_{i,j,k}^{u,s_a}\left(\frac{\mu_k^{u,s_a}}{\phi_k^{u,s_a}} + \frac{\hat{r}_{i,j}^{c_{i,j}^{u,s_a}}}{\tau_{i,j}^{r_{i,j}^{u,s_a}}}\right), \quad (28)$$

$$\bar{\omega}_{i,j,k}^{u,s_a} = \frac{\omega_k^{u,s_a}\mathcal{CN}\left(0 \mid \mu_k^{u,s_a} - \hat{r}_{i,j}^{c_{i,j}^{u,s_a}}, \phi_k^{u,s_a} + \tau_{i,j}^{r_{i,j}^{u,s_a}}\right)}{\eta_{i,j}^{\mathrm{A}^{u,s_a}}}, \quad (29)$$

$$\chi_{i,j}^{u,s_a} = \frac{\zeta_{i,j}^{u,s_a}\eta_{i,j}^{\mathrm{A}^{u,s_a}}}{(1-\zeta_{i,j}^{u,s_a})\mathcal{CN}(0 \mid \hat{r}_{i,j}^{c_{i,j}^{u,s_a}}, \tau_{i,j}^{r_{i,j}^{u,s_a}}) + \zeta_{i,j}^{u,s_a}\eta_{i,j}^{\mathrm{A}^{u,s_a}}}. \quad (30)$$

Finally, the variance and mean of $c_{i,j}^{u,s_a}$ with respect to (26) can be expressed as

$$\tau_{i,j}^{\mathrm{I}c_{i,j}^{u,s_a}} = \chi_{i,j}^{u,s_a} \sum_{k=1}^{K} \bar{\omega}_{i,j,k}^{u,s_a}\left(\left|\theta_{i,j,k}^{u,s_a}\right|^2 + \varphi_{i,j,k}^{u,s_a}\right) - \left|\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}\right|^2,$$

$$\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}} = \chi_{i,j}^{u,s_a} \sum_{k=1}^{K} \bar{\omega}_{i,j,k}^{u,s_a}\theta_{i,j,k}^{u,s_a}. \quad (31)$$

Note that the previous derived algorithm needs the full knowledge of the noise variance and the parameters in prior BGM i.e., $\bar{\boldsymbol{\theta}}_{s_a} \triangleq \{\sigma_{s_a}^2, \omega_k^{u,s_a}, \mu_k^{u,s_a}, \phi_k^{u,s_a}, \forall k, u\}$. We now adopt the EM algorithm to learn these parameters, which is an iterative technique that increases a lower bound on the likelihood at each iteration. Then, the update of $\bar{\boldsymbol{\theta}}_{s_a}$ in the $(t_{\mathrm{I}}+1)$-th

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

iteration is given as

$$\bar{\boldsymbol{\theta}}_{s_a}^{t_{\mathrm{I}}+1} = \arg\max_{\bar{\boldsymbol{\theta}}} \mathrm{E}[\log p\left(\mathbf{Z}_{s_a}, \mathbf{C}_{s_a}, \mathbf{Y}_{s_a} \mid \bar{\boldsymbol{\theta}}\right) \mid \mathbf{Y}_{s_a}, \bar{\boldsymbol{\theta}}_{s_a}^{t_{\mathrm{I}}}], \tag{32}$$

where $\mathbf{Z}_{s_a}$ and $\mathbf{C}_{s_a}$ are treated as the hidden variables and the expectation is with respect to $p\left(\mathbf{Z}_{s_a}, \mathbf{C}_{s_a} \mid \mathbf{Y}_{s_a}, \bar{\boldsymbol{\theta}}_{s_a}^{t_{\mathrm{I}}}\right)$. With the previous approximated posterior distribution, we can derive the update of the hyperparameters as

$$(\sigma_{s_a}^2)^{t_{\mathrm{I}}+1} = \frac{1}{M_{\mathrm{p}} N_{\mathrm{p}} N_z N_y} \sum_{m,j} \left| y_{m,j}^{s_a} - \hat{z}_{m,j}^{s_a} \right|^2 + \tau_{m,j}^{z^{s_a}}, \tag{33}$$

$$(\mu_k^{u,s_a})^{t_{\mathrm{I}}+1} = \frac{\sum_{i,j} \chi_{i,j}^{u,s_a} \bar{\omega}_{i,j,k}^{u,s_a} \theta_{i,j,k}^{u,s_a}}{\sum_{i,j} \chi_{i,j}^{u,s_a} \bar{\omega}_{i,j,k}^{u,s_a}}, \tag{34}$$

$$(\phi_k^{u,s_a})^{t_{\mathrm{I}}+1} = \frac{\sum_{i,j} \chi_{i,j}^{u,s_a} \bar{\omega}_{i,j,k}^{u,s_a} \left(\left| \theta_{i,j,k}^{u,s_a} - \mu_k^{u,s_a} \right|^2 + \varphi_{i,j,k}^{u,s_a} \right)}{\sum_{i,j} \chi_{i,j}^{u,s_a} \bar{\omega}_{i,j,k}^{u,s_a}}, \tag{35}$$

$$(\omega_k^{u,s_a})^{t_{\mathrm{I}}+1} = \frac{\sum_{i,j} \chi_{i,j}^{u,s_a} \bar{\omega}_{i,j,k}^{u,s_a}}{\sum_{i,j} \chi_{i,j}^{u,s_a}}. \tag{36}$$

Building upon the message expressions and EM update rules, we propose the MRF-BGM-AMP algorithm for initial channel estimation, and we denote $x_{m,i}^{\omega^u}$ as the $(m, u(Q_{\mathrm{I}}+1)(L+1)+i)$-th element of $\mathbf{X}^{\omega}$ as shown in Algorithm 1. The lines 3-10 represent the GAMP, lines 11-18 are expressions derived using message passing rules, and the line 21 is the EM update. Proper initialization of the hyperparameters is crucial for EM update which refers to [36], and the damping factor is leveraged in the GAMP part to help the convergence of the algorithm [34].

Algorithm 1 MRF-BGM-AMP

Input: Recieved signals $\mathbf{Y}_{s_a}$, sensing matrix $\mathbf{X}^{\omega}$; the maximum number of iterations T_{I} and T_{mrf}, and the termination threshold η_{I}.
Output: The estimated BEM coefficient $\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}$, variance $\tau_{i,j}^{\mathrm{I}c^{u,s_a}}$ and hyper-parameters $\bar{\boldsymbol{\theta}}_{s_a}$.

1: **Initialization**: choose $\bar{\boldsymbol{\theta}}_{s_a}$; $\forall i,j,u,s_a$: choose $\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}, \tau_{i,j}^{\mathrm{I}c^{u,s_a}}$, $\xi_{i,j}^{\mathrm{L}^{u,s_a}} = \xi_{i,j}^{\mathrm{R}^{u,s_a}} = \xi_{i,j}^{\mathrm{T}^{u,s_a}} = \xi_{i,j}^{\mathrm{B}^{u,s_a}} = 0.5$; $\forall m,j,s_a$: $\hat{s}_{m,j}^{z^{s_a}} = 0$;
2: **for** $t_{\mathrm{I}} = 1$ to T_{I} **do**
3: $\quad \forall m,j,s_a$: $\tau_{m,j}^{p^{s_a}} = \sum_{u,i} \left| x_{m,i}^{\omega^u} \right|^2 \tau_{i,j}^{\mathrm{I}c^{u,s_a}}$;
4: $\quad \forall m,j,s_a$: $\hat{p}_{m,j}^{s_a} = \sum_{u,i} x_{m,i}^{\omega^u} \hat{c}_{i,j}^{\mathrm{I}^{u,s_a}} - \tau_{m,j}^{p^{s_a}} \hat{s}_{m,j}^{z^{s_a}}$;
5: $\quad \forall m,j,s_a$: $\tau_{m,j}^{z^{s_a}} = \tau_{m,j}^{p^{s_a}} \sigma_{s_a}^2 / \left(\tau_{m,j}^{p^{s_a}} + \sigma_{s_a}^2 \right)$;
6: $\quad \forall m,j,s_a$: $\hat{z}_{m,j}^{s_a} = \left(\tau_{m,j}^{p^{s_a}} y_{m,j}^{s_a} + \sigma_{s_a}^2 \hat{p}_{m,j}^{s_a} \right) / \left(\tau_{m,j}^{p^{s_a}} + \sigma_{s_a}^2 \right)$;
7: $\quad \forall m,j,s_a$: $\tau_{m,j}^{s^{s_a}} = (\tau_{m,j}^{p^{s_a}} - \tau_{m,j}^{z^{s_a}}) / (\tau_{m,j}^{p^{s_a}})^2$;
8: $\quad \forall m,j,s_a$: $\hat{s}_{m,j}^{z^{s_a}} = (\hat{z}_{m,j}^{s_a} - \hat{p}_{m,j}^{s_a}) / \tau_{m,j}^{p^{s_a}}$;
9: $\quad \forall i,j,u,s_a$: $\tau_{i,j}^{r^{u,s_a}} = (\sum_m \left| x_{m,i}^{\omega^u} \right|^2 \tau_{m,j}^{s^{s_a}})^{-1}$;
10: $\quad \forall i,j,u,s_a$: $r_{i,j}^{c^{u,s_a}} = \hat{c}_{i,j}^{\mathrm{I}^{u,s_a}} + \tau_{i,j}^{r^{u,s_a}} \sum_m x_{m,i}^{\omega^{u^*}} \hat{s}_{m,j}^{z^{s_a}}$;
11: $\quad \forall i,j,u,s_a$: Update $\eta_{i,j}^{u,s_a}$ via (24);
12: $\quad$ \\ **MRF Module**
13: $\quad$ **for** $t_{\mathrm{mrf}} = 1$ to T_{mrf} **do**
14: $\quad\quad \forall i,j,u,s_a$: Update $\xi_{i,j}^{\mathrm{L}^{u,s_a}}, \xi_{i,j}^{\mathrm{R}^{u,s_a}}, \xi_{i,j}^{\mathrm{T}^{u,s_a}}, \xi_{i,j}^{\mathrm{B}^{u,s_a}}$ via (69);
15: $\quad$ **end for**
16: $\quad \forall i,j,u,s_a$: Update $\zeta_{i,j}^{u,s_a}$ and $\chi_{i,j}^{u,s_a}$ via (70) and (30);
17: $\quad \forall i,j,k,u,s_a$: Update $\varphi_{i,j,k}^{u,s_a}$, $\theta_{i,j,k}^{u,s_a}$, and $\bar{\omega}_{i,j,k}^{u,s_a}$ via (27)-(29);
18: $\quad \forall i,j,u,s_a$: Update $\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}(t_{\mathrm{I}}+1)$ and $\tau_{i,j}^{\mathrm{I}c^{u,s_a}}(t_{\mathrm{I}}+1)$ via (31);
19: $\quad$ **if** $\left\| \hat{\mathbf{C}}_{s_a}(t_{\mathrm{I}}+1) - \hat{\mathbf{C}}_{s_a} \right\|_{\mathrm{F}}^2 \le \eta_{\mathrm{I}} \left\| \hat{\mathbf{C}}_{s_a} \right\|_{\mathrm{F}}^2$: break; **end if**
20: $\quad$ \\ **EM Update**
21: $\quad \forall s_a$: Update $\bar{\boldsymbol{\theta}}_{s_a}$ via (33)-(36);
22: **end for**

IV. Enhancement via Cooperative Satellite Constellations

In this section, we detail our algorithm for joint device identification, channel estimation, and symbol detection, enhanced by cooperative LEO satellite constellations. We explore two cooperative modes: centralized and distributed modes. Firstly, we describe the method for identifying active devices, and then the algorithm for initial symbol detection. Finally, we introduce the joint refinement process for channel estimation and symbol detection.

A. Centralized Cooperative Mode

In the centralized mode, all the computation is performed centrally at the gateway or one specific satellite with high computation power, which requires each edge satellite to feed back their own received signal to the central server for further processing.

1) Device identification: To identify active devices, the Algorithm 1 is firstly adopted to get the initial channel estimation $\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}$. Then, by aggregating the estimated channel energy at the central node, the activity indicator of the u-th device is given by

$$\hat{\lambda}_u = \mathbb{I}\left\{ \frac{1}{S_a} \sum_{s_a,i,j} \left| \hat{c}_{i,j}^{\mathrm{I}^{u,s_a}} \right|^2 > \eta_\lambda \right\}, \tag{37}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, η_λ is an empirical predefined threshold, and we set $\eta_\lambda = 10^{-2}$ in the simulations. Then, the set of estimated active devices is given by $\hat{\mathcal{U}}_a = \{u \mid \hat{\lambda}_u = 1, u = 0, \ldots, U-1\}$.

2) Initial symbol detection: According to (7)-(11), the received signal at the n_a-th antenna of the s_a-th satellite can be written as

$$\mathbf{y}_{n_a,s_a}^{\mathrm{DD}} = \sum_{u \in \mathcal{U}_a} \bar{\mathbf{H}}_{u,s_a}^{n_a} \mathbf{x}_u + \mathbf{e}_{s_a}^{n_a}, \tag{38}$$

where $\bar{\mathbf{H}}_{u,s_a}^{n_a} = (\mathbf{F}_N \otimes \mathbf{I}_M) \mathbf{H}_{u,s_a}^{n_a} \left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_M \right)$, $\mathbf{H}_{u,s_a}^{n_a} \in \mathbb{C}^{MN \times MN}$ with the (n,m)-th element $h_{u,s_a}^{n_a}[n + M_{\mathrm{cp}}, (n-m)_{MN}]$, and $\mathbf{x}_u = \mathrm{vec}(\mathbf{X}_u^{\mathrm{DD}})$. At the central node, the received signals of all the satellites are collected in $\mathbf{y}^{\mathrm{DD}} = [\mathbf{y}_{0,0}^{\mathrm{DD}^{\mathrm{T}}}, \ldots, \mathbf{y}_{n_a,s_a}^{\mathrm{DD}^{\mathrm{T}}}, \ldots, \mathbf{y}_{N_z N_y - 1, S_a - 1}^{\mathrm{DD}^{\mathrm{T}}}]^{\mathrm{T}}$, and then we have the following relationship

$$\mathbf{y}^{\mathrm{DD}} = \bar{\mathbf{H}} \mathbf{x} + \mathbf{e}, \tag{39}$$

where $\bar{\mathbf{H}} = [\bar{\mathbf{H}}_{0,0}^{\mathrm{T}}, \ldots, \bar{\mathbf{H}}_{n_a,s_a}^{\mathrm{T}}, \ldots, \bar{\mathbf{H}}_{N_z N_y - 1, S_a - 1}^{\mathrm{T}}]^{\mathrm{T}}$, $\bar{\mathbf{H}}_{n_a,s_a} = [\bar{\mathbf{H}}_{0,s_a}^{n_a}, \ldots, \bar{\mathbf{H}}_{u,s_a}^{n_a}, \ldots, \bar{\mathbf{H}}_{U-1,s_a}^{n_a}]$ for $u \in \mathcal{U}_a$, $\mathbf{x} = [\mathbf{x}_0^{\mathrm{T}}, \ldots, \mathbf{x}_u^{\mathrm{T}} \ldots, \mathbf{x}_{U-1}^{\mathrm{T}}]^{\mathrm{T}}$ for $u \in \mathcal{U}_a$, and $\mathbf{e} = [\mathbf{e}_0^{0^{\mathrm{T}}}, \ldots, \mathbf{e}_{S_a-1}^{N_z N_y - 1^{\mathrm{T}}}]^{\mathrm{T}}$. Note that $\mathcal{U}_a$ can be estimated by $\hat{\mathcal{U}}_a$, and the channel coefficient $\hat{h}_{u,s_a}^{n_a}[n,l']$ can be reconstructed by $\hat{c}_{i,j}^{\mathrm{I}^{u,s_a}}$ according to (6). Then, the GAMP algorithm with uniform prior [34] is utilized for symbol detection. $\hat{\mathbf{x}}_u^{\mathrm{I}}$ including both pilots and estimated data symbols and the corresponding variance $\boldsymbol{\tau}_u^{x_{\mathrm{I}}}$ of active devices can be obtained.

3) Jointly refine channel estimation and symbol detection: We adopt a higher resolution $R = 2$ with the larger order of

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

BEM $Q_{\mathrm{R}} \geq 2\lceil 2N\bar{\nu}_{\max}\rceil$ to improve the modeling accuracy for TSLs. Here, the received signals in the time-angle domain are utilized to jointly refine channel estimation and symbol detection for active devices, where the only difference from (14) is that there is no demodulation matrix $\mathbf{F}_N \otimes \mathbf{I}_M$, i.e.,

$$\mathbf{y}'_{n_a,s_a} = \sum_{u\in\mathcal{U}_a}\sum_{q=0}^{Q_{\mathrm{R}}} \mathrm{diag}\left(\mathbf{b}_q\right)\mathbf{F}_{MN}^{\mathrm{H}}\,\mathrm{diag}\left(\mathbf{F}_{1:L+1}\mathbf{c}_{q,u,s_a}^{n_a}\right)\mathbf{F}_{MN} \times \left(\mathbf{F}_N^{\mathrm{H}}\otimes\mathbf{I}_M\right)\mathbf{x}_u + \mathbf{e}'_{n_a,s_a}, \tag{40}$$

where $\mathbf{e}'_{n_a,s_a}$ represents the noise in the time-angle domain. To facilitate the algorithm design, we introduce the auxiliary variables $\mathbf{x}_u^{\mathrm{F}} = \mathbf{F}_{\mathrm{b}}\mathbf{x}_u$, $\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} = \mathbf{F}_{1:L+1}\mathbf{c}_{q,u,s_a}^{n_a}$, $\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}} = \mathrm{diag}\left(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}\right)\mathbf{x}_u^{\mathrm{F}}$, $\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}} = \sum_u \mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}$, $\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}} = \mathbf{F}_{MN}^{\mathrm{H}}\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}$, $\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}} = \mathrm{diag}\left(\mathbf{b}_q\right)\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}$, and $\mathbf{g}_{n_a,s_a} = \sum_q \mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}$, where $\mathbf{F}_{\mathrm{b}} \triangleq \mathbf{F}_{MN}\left(\mathbf{F}_N^{\mathrm{H}}\otimes\mathbf{I}_M\right)$. Then, the posterior distribution of the variables is given by

$$\begin{aligned}
&p\left(\mathbf{x},\mathbf{c},\mathbf{x}^{\mathrm{F}},\mathbf{c}^{\mathrm{F}},\mathbf{d}^{\mathrm{P}},\mathbf{d}^{\mathrm{W}},\mathbf{d}^{\mathrm{F}},\mathbf{d}^{\mathrm{B}},\mathbf{g}\mid\mathbf{y}\right)\propto\\
&\prod_{s_a,n_a}\{p(\mathbf{y}'_{n_a,s_a}\mid\mathbf{g}_{n_a,s_a})p(\mathbf{g}_{n_a,s_a}\mid\mathbf{d}_{0,n_a,s_a}^{\mathrm{B}},\cdots,\mathbf{d}_{Q,n_a,s_a}^{\mathrm{B}})\\
&\prod_q[p(\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}\mid\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}})p(\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}\mid\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}})\\
&p(\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}\mid\mathbf{d}_{q,0,n_a,s_a}^{\mathrm{P}},\cdots,\mathbf{d}_{q,U-1,n_a,s_a}^{\mathrm{P}})\\
&\prod_u p(\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}\mid\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}},\mathbf{x}_u^{\mathrm{F}})p(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}\mid\mathbf{c}_{q,u,s_a}^{n_a})\\
&p(\mathbf{c}_{q,u,s_a}^{n_a})]\}\prod_u p(\mathbf{x}_u^{\mathrm{F}}\mid\mathbf{x}_u)p(\mathbf{x}_u),
\end{aligned}\tag{41}$$

where $\mathbf{x} = \{\mathbf{x}_u \mid \forall u\}$, $\mathbf{c} = \{\mathbf{c}_{q,u,s_a}^{n_a} \mid \forall q,u,n_a,s_a\}$, $\mathbf{x}^{\mathrm{F}} = \{\mathbf{x}_u^{\mathrm{F}} \mid \forall u\}$, $\mathbf{c}^{\mathrm{F}} = \{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \forall q,u,n_a,s_a\}$, $\mathbf{d}^{\mathrm{P}} = \{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}} \mid \forall q,u,n_a,s_a\}$, $\mathbf{d}^{\mathrm{W}} = \{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}} \mid \forall q,n_a,s_a\}$, $\mathbf{d}^{\mathrm{F}} = \{\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}} \mid \forall q,n_a,s_a\}$, $\mathbf{d}^{\mathrm{B}} = \{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}} \mid \forall q,n_a,s_a\}$, $\mathbf{g} = \{\mathbf{g}_{n_a,s_a} \mid \forall n_a,s_a\}$, and $\mathbf{y} = \{\mathbf{y}'_{n_a,s_a} \mid \forall n_a,s_a\}$.

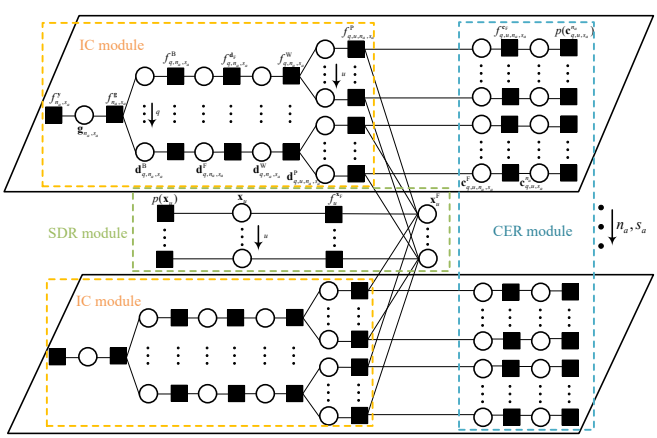


Fig. 3. Factor graph representation for refinement scheme.

The factor graph representation of (41) is shown in Fig. 3, and the corresponding factors are summarized in Table I. For sake of low computations, we assign the non-informative prior for $\mathbf{c}_{q,u,s_a}^{n_a}$ and $\mathbf{x}_u$, i.e., $p(\mathbf{c}_{q,u,s_a}^{n_a}) = \mathcal{CN}(\mathbf{c}_{q,u,s_a}^{n_a} \mid \mathbf{0}_{L+1},\infty\mathbf{I}_{L+1})$ and $p(\mathbf{x}_u) = \mathcal{CN}(\mathbf{x}_u \mid \mathbf{0}_{MN},\infty\mathbf{I}_{MN})$, which is enough for achieving superior performance as shown in the simulations. Since exact BP rule is difficult to implement in Fig. 3, we resort to EP which extends BP by enforcing an exponential family constraint on the messages. Based on the EP and central limit theorem (CLT), we aims to approximate the posterior distribution $p(\mathbf{x}_u \mid \mathbf{y})$ and $p(\mathbf{c}_{q,u,s_a}^{n_a} \mid \mathbf{y})$, leading to the MAP estimations $\hat{\mathbf{x}}_u$ and $\hat{\mathbf{c}}_{q,u,s_a}^{n_a}$. As shown in Fig. 3, we divide the whole receiver structure into three modules: interference cancellation (IC) module addresses the linear signal models to mitigate both inter-user and inter-component interference within the BEM model, subsequently providing likelihoods for $\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}$ and $\mathbf{x}_u^{\mathrm{F}}$; The channel estimation refinement (CER) and symbol detection refinement (SDR) modules further enhance channel estimation and symbol detection, respectively, utilizing soft information from the IC module. For EP, the projection of one density onto the Gaussian family is a key concept, denoted as

$$\begin{aligned}\mathrm{Proj}[p'(x)] &\triangleq \arg\min_{p\in\mathcal{G}}\mathrm{KL}[p'(x)\|p(x)]\\ &= \mathcal{CN}(x\mid\mu,\sigma^2),\end{aligned}\tag{42}$$

where $\mathrm{KL}[\cdot]$ represents the KL-divergence, $\mathcal{G}$ is the family of Gaussian densities, $\mu = \int xp'(x)\mathrm{d}x$, and $\sigma^2 = \int|x-\mu|^2p'(x)\mathrm{d}x$.

In Appendix C, we derive the messages in the IC module. It provides the likelihoods for CER module and SDR module, respectively, which are given as

$$\Delta_{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}} = \mathcal{CN}(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \overrightarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}},\mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}})) \tag{43}$$

$$\Delta_{\mathbf{x}_u^{\mathrm{F}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}} = \mathcal{CN}(\mathbf{x}_u^{\mathrm{F}} \mid \overleftarrow{\mathbf{x}}_{q,u,n_a,s_a}^{\mathrm{F}},\mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{x_{\mathrm{F}}})). \tag{44}$$

With the likelihood in (43) and $\Delta_{f_{q,u,n_a,s_a}^{\mathrm{c_F}}}^{\mathbf{c}_{q,u,s_a}^{n_a}} = \mathcal{CN}(\mathbf{c}_{q,u,s_a}^{n_a} \mid \overleftarrow{\mathbf{c}}_{q,u,s_a}^{n_a},\overleftarrow{\tau}_{q,u,s_a}^{c_{n_a}}\mathbf{I}_{L+1})$ passing to CER module, the posterior distribution of $\mathbf{c}_{q,u,s_a}^{n_a}$ can be obtained as

$$\begin{aligned}\Delta_{\mathbf{c}_{q,u,s_a}^{n_a}} &\propto \mathrm{Proj}\left[\Delta_{f_{q,u,n_a,s_a}^{\mathrm{c_F}}}^{\mathbf{c}_{q,u,s_a}^{n_a}}\int_{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}} f_{q,u,n_a,s_a}^{\mathrm{c_F}}\Delta_{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}}\right]\\ &\propto \mathcal{CN}(\mathbf{c}_{q,u,s_a}^{n_a} \mid \hat{\mathbf{c}}_{q,u,s_a}^{n_a},\tau_{q,u,s_a}^{c_{n_a}}\mathbf{I}_{L+1}),\end{aligned}\tag{45}$$

where

$$\overleftarrow{\tau}_{q,u,s_a}^{c_{n_a}} = \frac{1}{L+1}\mathrm{Tr}[\boldsymbol{\Sigma}_{q,u,s_a}^{c_{n_a}}],\tag{46}$$

$$\boldsymbol{\Sigma}_{q,u,s_a}^{c_{n_a}} = (\mathbf{F}_{1:L+1}^{\mathrm{H}}\,\mathrm{diag}^{-1}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}})\mathbf{F}_{1:L+1} + (\overleftarrow{\tau}_{q,u,s_a}^{c_{n_a}})^{-1}\mathbf{I}_{L+1})^{-1},\tag{47}$$

$$\hat{\mathbf{c}}_{q,u,s_a}^{n_a} = \boldsymbol{\Sigma}_{q,u,s_a}^{c_{n_a}}(\mathbf{F}_{1:L+1}^{\mathrm{H}}\,\mathrm{diag}^{-1}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}})\overrightarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} + \overleftarrow{\mathbf{c}}_{q,u,s_a}^{n_a}/\overleftarrow{\tau}_{q,u,s_a}^{c_{n_a}}).\tag{48}$$

Note that (48) can be recognized as the LMMSE estimate of a random vector $\mathbf{c}_{q,u,s_a}^{n_a}$ under likelihood $\mathcal{CN}(\overrightarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \mathbf{F}_{1:L+1}\mathbf{c}_{q,u,s_a}^{n_a},\mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}}))$ and prior $\mathbf{c}_{q,u,s_a}^{n_a} \sim \mathcal{CN}(\mathbf{c}_{q,u,s_a}^{n_a} \mid \overleftarrow{\mathbf{c}}_{q,u,s_a}^{n_a},\overleftarrow{\tau}_{q,u,s_a}^{c_{n_a}}\mathbf{I}_{L+1})$. Then, the message $\Delta_{\mathbf{c}_{q,u,s_a}^{n_a}}^{f_{q,u,n_a,s_a}^{\mathrm{c_F}}}$ is computed as

$$\begin{aligned}\Delta_{\mathbf{c}_{q,u,s_a}^{n_a}}^{f_{q,u,n_a,s_a}^{\mathrm{c_F}}} &\propto \frac{\Delta_{\mathbf{c}_{q,u,s_a}^{n_a}}}{\Delta_{f_{q,u,n_a,s_a}^{\mathrm{c_F}}}^{\mathbf{c}_{q,u,s_a}^{n_a}}}\\ &\propto \mathcal{CN}(\mathbf{c}_{q,u,s_a}^{n_a} \mid \overrightarrow{\mathbf{c}}_{q,u,s_a}^{n_a},\overrightarrow{\tau}_{q,u,s_a}^{c_{n_a}}\mathbf{I}_{L+1}),\end{aligned}\tag{49}$$

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

Table I: Notations of Factor Nodes

Factor	Distribution	Function
$f^{\mathbf{y}}_{n_a,s_a}$	$p(\mathbf{y}'_{n_a,s_a} \mid \mathbf{g}_{n_a,s_a})$	$\mathcal{CN}(\mathbf{y}'_{n_a,s_a} \mid \mathbf{g}_{n_a,s_a}, \sigma^2_{s_a}\mathbf{I}_{MN})$
$f^{\mathbf{g}}_{n_a,s_a}$	$p(\mathbf{g}_{n_a,s_a} \mid \mathbf{d}^{\mathrm{B}}_{0,n_a,s_a}, \cdots, \mathbf{d}^{\mathrm{B}}_{Q,n_a,s_a})$	$\delta(\mathbf{g}_{n_a,s_a} - \sum_q \mathbf{d}^{\mathrm{B}}_{q,n_a,s_a})$
$f^{\mathrm{B}}_{q,n_a,s_a}$	$p(\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a} \mid \mathbf{d}^{\mathrm{F}}_{q,n_a,s_a})$	$\delta(\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a} - \mathrm{diag}\,(\mathbf{b}_q)\,\mathbf{d}^{\mathrm{F}}_{q,n_a,s_a})$
$f^{\mathbf{d}_{\mathrm{F}}}_{q,n_a,s_a}$	$p(\mathbf{d}^{\mathrm{F}}_{q,n_a,s_a} \mid \mathbf{d}^{\mathrm{W}}_{q,n_a,s_a})$	$\delta(\mathbf{d}^{\mathrm{F}}_{q,n_a,s_a} - \mathbf{F}^{\mathrm{H}}_{MN}\mathbf{d}^{\mathrm{W}}_{q,n_a,s_a})$
$f^{\mathrm{W}}_{q,n_a,s_a}$	$p(\mathbf{d}^{\mathrm{W}}_{q,n_a,s_a} \mid \mathbf{d}^{\mathrm{P}}_{q,0,n_a,s_a}, \cdots, \mathbf{d}^{\mathrm{P}}_{q,U-1,n_a,s_a})$	$\delta(\mathbf{d}^{\mathrm{W}}_{q,n_a,s_a} - \sum_u \mathbf{d}^{\mathrm{P}}_{q,u,n_a,s_a})$
$f^{\mathrm{P}}_{q,u,n_a,s_a}$	$p(\mathbf{d}^{\mathrm{P}}_{q,u,n_a,s_a} \mid \mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}, \mathbf{x}^{\mathrm{F}}_u)$	$\delta(\mathbf{d}^{\mathrm{P}}_{q,u,n_a,s_a} - \mathrm{diag}\left(\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}\right)\mathbf{x}^{\mathrm{F}}_u)$
$f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}$	$p(\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a} \mid \mathbf{c}^{n_a}_{q,u,s_a})$	$\delta(\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a} - \mathbf{F}_{1:L+1}\mathbf{c}^{n_a}_{q,u,s_a})$
$f^{\mathbf{x}_{\mathrm{F}}}_u$	$p(\mathbf{x}^{\mathrm{F}}_u \mid \mathbf{x}_u)$	$\delta(\mathbf{x}^{\mathrm{F}}_u - \mathbf{F}_{\mathrm{b}}\mathbf{x}_u)$

where $\overrightarrow{\tau}^{c_{n_a}}_{q,u,s_a} = ((\tau^{c_{n_a}}_{q,u,s_a})^{-1} - (\overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a})^{-1})^{-1}$ and $\overrightarrow{\mathbf{c}}^{n_a}_{q,u,s_a} = \overrightarrow{\tau}^{c_{n_a}}_{q,u,s_a}\left(\hat{\mathbf{c}}^{n_a}_{q,u,s_a}/\tau^{c_{n_a}}_{q,u,s_a} - \overleftarrow{\mathbf{c}}^{n_a}_{q,u,s_a}/\overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a}\right)$. Since non-informative prior is adopted, we update $\Delta^{\mathbf{c}^{n_a}_{q,u,s_a}}_{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}}$ as $\Delta^{\mathbf{c}^{n_a}_{q,u,s_a}}_{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}} = \Delta^{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}}_{\mathbf{c}^{n_a}_{q,u,s_a}}$. Then, the message $\Delta^{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}}_{\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}}$ is obtained as $\Delta^{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}}_{\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}} = \mathcal{CN}(\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a} \mid \overleftarrow{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a}, \overleftarrow{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a}\mathbf{I}_{MN})$ where $\overleftarrow{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a} = \mathbf{F}_{1:L+1}\overleftarrow{\mathbf{c}}^{n_a}_{q,u,s_a}$ and $\overleftarrow{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a} = \overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a}$. Now, we can update the posterior distribution of $\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}$ as

$$\begin{aligned}\Delta^{\mathrm{F}}_{\mathbf{c}_{q,u,n_a,s_a}} &\propto \Delta^{f^{\mathrm{P}}_{q,u,n_a,s_a}}_{\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}}\Delta^{f^{\mathbf{c}_{\mathrm{F}}}_{q,u,n_a,s_a}}_{\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a}}\\ &\propto \mathcal{CN}(\mathbf{c}^{\mathrm{F}}_{q,u,n_a,s_a} \mid \hat{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a}, \mathrm{diag}(\boldsymbol{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a})),\end{aligned} \tag{50}$$

where

$$\boldsymbol{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a} = 1 \oslash ((\overleftarrow{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a})^{-1}\mathbf{1}_{MN} + 1 \oslash \overrightarrow{\boldsymbol{\tau}}^{c_{\mathrm{F}}}_{q,u,n_a,s_a}), \tag{51}$$

$$\begin{aligned}\hat{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a} = \boldsymbol{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a} \odot (&\overleftarrow{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a}/\overleftarrow{\tau}^{c_{\mathrm{F}}}_{q,u,n_a,s_a}\\ &+ \overrightarrow{\mathbf{c}}^{\mathrm{F}}_{q,u,n_a,s_a} \oslash \overrightarrow{\boldsymbol{\tau}}^{c_{\mathrm{F}}}_{q,u,n_a,s_a}).\end{aligned} \tag{52}$$

In the SDR module, by combining the messages from $f^{\mathrm{P}}_{q,u,n_a,s_a}$ to $\mathbf{x}^{\mathrm{F}}_u$, we get the effective likelihood of $\mathbf{x}^{\mathrm{F}}_u$ as

$$\begin{aligned}\Delta^{\mathbf{x}^{\mathrm{F}}_u}_{f^{\mathbf{x}_{\mathrm{F}}}_u} &\propto \prod_{q,n_a,s_a}\Delta^{f^{\mathrm{P}}_{q,u,n_a,s_a}}_{\mathbf{x}^{\mathrm{F}}_u}\\ &= \mathcal{CN}(\mathbf{x}^{\mathrm{F}}_u \mid \overleftarrow{\mathbf{x}}^{\mathrm{F}}_u, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u)),\end{aligned} \tag{53}$$

where $\overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u = 1 \oslash \sum_{q,n_a,s_a}(1 \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a})$ and $\overleftarrow{\mathbf{x}}^{\mathrm{F}}_u = \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u \odot \sum_{q,n_a,s_a}(\overleftarrow{\mathbf{x}}^{\mathrm{F}}_{q,u,n_a,s_a} \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a})$. Similar to (48), with the message $\Delta^{\mathbf{x}_u}_{f^{\mathbf{x}_{\mathrm{F}}}_u} = \mathcal{CN}(\mathbf{x}_u \mid \overrightarrow{\mathbf{x}}_u, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}^{x}_u))$ as prior and (53) as likelihood, the LMMSE estimate of $\mathbf{x}_u$ is given by

$$\hat{\mathbf{x}}_u = \boldsymbol{\Sigma}_{\mathbf{x}_u}(\mathbf{F}^{\mathrm{H}}_{\mathrm{b}}\,\mathrm{diag}^{-1}(\overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u)\overleftarrow{\mathbf{x}}^{\mathrm{F}}_u + \mathrm{diag}^{-1}(\overrightarrow{\boldsymbol{\tau}}^{x}_u)\overrightarrow{\mathbf{x}}_u), \tag{54}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{x}_u} = (\mathbf{F}^{\mathrm{H}}_{\mathrm{b}}\,\mathrm{diag}^{-1}(\overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u)\mathbf{F}_{\mathrm{b}} + \mathrm{diag}^{-1}(\overrightarrow{\boldsymbol{\tau}}^{x}_u))^{-1}. \tag{55}$$

The posterior distribution of $\mathbf{x}_u$ can also be written as $\Delta_{\mathbf{x}_u} = \mathcal{CN}(\mathbf{x}_u \mid \hat{\mathbf{x}}_u, \mathrm{diag}(\boldsymbol{\tau}^x_u))$, where $\boldsymbol{\tau}^x_u = \mathrm{diag}(\boldsymbol{\Sigma}_{\mathbf{x}_u})$. Then, we can get

$$\Delta^{f^{\mathbf{x}_{\mathrm{F}}}_u}_{\mathbf{x}_u} \propto \frac{\Delta_{\mathbf{x}_u}}{\Delta^{\mathbf{x}_u}_{f^{\mathbf{x}_{\mathrm{F}}}_u}} \propto \mathcal{CN}(\mathbf{x}_u \mid \overleftarrow{\mathbf{x}}_u, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}^{x}_u)), \tag{56}$$

where $\overleftarrow{\boldsymbol{\tau}}^{x}_u = 1 \oslash (1 \oslash \boldsymbol{\tau}^x_u - 1 \oslash \overrightarrow{\boldsymbol{\tau}}^{x}_u)$ and $\overleftarrow{\mathbf{x}}_u = \overleftarrow{\boldsymbol{\tau}}^{x}_u \odot (\hat{\mathbf{x}}_u \oslash \boldsymbol{\tau}^x_u - \overrightarrow{\mathbf{x}}_u \oslash \overrightarrow{\boldsymbol{\tau}}^{x}_u)$. With the update of $\Delta^{\mathbf{x}_u}_{f^{\mathbf{x}_{\mathrm{F}}}_u} = \Delta^{f^{\mathbf{x}_{\mathrm{F}}}_u}_{\mathbf{x}_u}$, we can compute $\Delta^{f^{\mathbf{x}_{\mathrm{F}}}_u}_{\mathbf{x}^{\mathrm{F}}_u} = \mathcal{CN}(\mathbf{x}^{\mathrm{F}}_u \mid \overrightarrow{\mathbf{x}}^{\mathrm{F}}_u, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u))$, where $\overrightarrow{\mathbf{x}}^{\mathrm{F}}_u = \mathbf{F}_{\mathrm{b}}\overrightarrow{\mathbf{x}}_u$ and $\overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u = \mathrm{diag}(\mathbf{F}_{\mathrm{b}}\,\mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}^{x}_u)\mathbf{F}^{\mathrm{H}}_{\mathrm{b}})$. Now, the posterior distribution of $\mathbf{x}^{\mathrm{F}}_u$ can be updated as

$$\Delta^{\mathrm{F}}_{\mathbf{x}_u} \propto \Delta^{f^{\mathbf{x}_{\mathrm{F}}}_u}_{\mathbf{x}^{\mathrm{F}}_u}\Delta^{\mathbf{x}^{\mathrm{F}}_u}_{f^{\mathbf{x}_{\mathrm{F}}}_u} \propto \mathcal{CN}(\mathbf{x}^{\mathrm{F}}_u \mid \hat{\mathbf{x}}^{\mathrm{F}}_u, \mathrm{diag}(\boldsymbol{\tau}^{x_{\mathrm{F}}}_u)), \tag{57}$$

where $\boldsymbol{\tau}^{x_{\mathrm{F}}}_u = 1 \oslash (1 \oslash \overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u + 1 \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u)$ and $\hat{\mathbf{x}}^{\mathrm{F}}_u = \boldsymbol{\tau}^{x_{\mathrm{F}}}_u \odot (\overrightarrow{\mathbf{x}}^{\mathrm{F}}_u \oslash \overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u + \overleftarrow{\mathbf{x}}^{\mathrm{F}}_u \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_u)$. Finally, the message $\Delta^{\mathbf{x}^{\mathrm{F}}_u}_{f^{\mathrm{P}}_{q,u,n_a,s_a}}$ can be updated by

$$\begin{aligned}\Delta^{\mathbf{x}^{\mathrm{F}}_u}_{f^{\mathrm{P}}_{q,u,n_a,s_a}} &\propto \frac{\Delta^{\mathrm{F}}_{\mathbf{x}_u}}{\Delta^{f^{\mathrm{P}}_{q,u,n_a,s_a}}_{\mathbf{x}^{\mathrm{F}}_u}}\\ &\propto \mathcal{CN}(\mathbf{x}^{\mathrm{F}}_u \mid \overrightarrow{\mathbf{x}}^{\mathrm{F}}_{q,u,n_a,s_a}, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a})),\end{aligned} \tag{58}$$

where $\overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a} = 1 \oslash (1 \oslash \boldsymbol{\tau}^{x_{\mathrm{F}}}_u - 1 \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a})$ and $\overrightarrow{\mathbf{x}}^{\mathrm{F}}_{q,u,n_a,s_a} = \overrightarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a} \odot (\hat{\mathbf{x}}^{\mathrm{F}}_u \oslash \boldsymbol{\tau}^{x_{\mathrm{F}}}_u - \overleftarrow{\mathbf{x}}^{\mathrm{F}}_{q,u,n_a,s_a} \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\mathrm{F}}}_{q,u,n_a,s_a})$.

Based on the above described procedures, we outline the main steps of the proposed centralized AEP in Algorithm 2. Note that the computation of $\hat{\mathbf{x}}_u$ and $\boldsymbol{\tau}^x_u$ involves costly matrix inversion, which may impose significantly computational burden on the central node. To mitigate this challenge, we employ the covariance-free method (CFM) [19]. It transforms the matrix inversions into solving the linear equations, which can be computed in parallel using efficient algorithms only relying on matrix multiplication, such as the conjugate gradient method utilized in CFM. In addition, we recommend damping the messages that follow the interference cancellation step to prevent the algorithm from diverging. Let message $\Delta^{f^{\mathbf{g}'}_{n_a,s_a}}_{\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a}}$ be computed by (73), and then the new message can be updated as

$$(\Delta^{f^{\mathbf{g}}_{n_a,s_a}}_{\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a}})_{\mathrm{new}} \propto (\Delta^{f^{\mathbf{g}'}_{n_a,s_a}}_{\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a}})^{\eta_\epsilon}((\Delta^{f^{\mathbf{g}}_{n_a,s_a}}_{\mathbf{d}^{\mathrm{B}}_{q,n_a,s_a}})_{\mathrm{old}})^{1-\eta_\epsilon}, \tag{59}$$

where η_ϵ is the damping factor and we set $\eta_\epsilon = 0.5$ in the simulations. The message $\Delta^{f^{\mathrm{W}}_{q,n_a,s_a}}_{\mathbf{d}^{\mathrm{P}}_{q,u,n_a,s_a}}$ can be damped similarly. Besides, the MRF-BGM-AMP and GAMP detector can also be adopted in the end of the centralized AEP for further improving performance.

Next, we describe the complexity of the proposed algorithm which consists of initial phase and refinement phase. In the initial phase, the MRF-BGM-AMP and GAMP detector is adopted, resulting in the complexity of $\mathcal{O}(S_aUM_{\mathrm{p}}N_{\mathrm{p}}(Q_{\mathrm{I}}+1)(L+1)N_zN_y + S_aN_zN_y\,|\mathcal{U}_a|\,(MN)^2)$. Note that due to the BEM modeling, the matrix-vector products involved in the

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

refinement phase can be implemented by fast Fourier transform. Consequently, the complexity is primarily governed by the calculations of $\hat{\mathbf{c}}^{n_a}_{q,u,s_a}$, $\tau^{c_{n_a}}_{q,u,s_a}$, $\hat{\mathbf{x}}_u$, and $\boldsymbol{\tau}^x_u$. Computations for $\hat{\mathbf{c}}^{n_a}_{q,u,s_a}$ and $\tau^{c_{n_a}}_{q,u,s_a}$ involve matrix inversion, resulting in the complexity of $\mathcal{O}(S_a N_z N_y |\mathcal{U}_a| (Q+1)(L+1)^3)$. Typically, the number of channel taps $L+1$ is sufficiently small, e.g., $L=2$ in NTN-TDL-D [28], and hence we can perform the matrix inversion directly. For computing $\hat{\mathbf{x}}_u$, and $\boldsymbol{\tau}^x_u$, we adopt the CFM which only needs the matrix multiplication, and then can also be implemented by FFT. As a result, it induces the complexity of $\mathcal{O}(|\mathcal{U}_a| (K_p+1) MN \log(MN))$, where K_p is the number of probe vectors used in CFM.

Algorithm 2 Centralized AEP

Input: Recieved symbols $\mathbf{y}'_{n_a,s_a}$; the maximum number of iterations T_{out}, T_{c}, T_{x}; the termination threshold η_{c}, η_{x}.
Output: The refined time-varying channel $\hat{h}^{n_a}_{u,s_a}[n,l']$ and data symbols $\hat{\mathbf{x}}_u$.

Get initial CE $\hat{\mathbf{c}}^{\text{I}n_a}_{q,u,s_a}$,$\tau^{\text{c}\text{I}n_a}_{q,u,s_a}$, and $\sigma^2_{s_a}$ by Algorithm 1;
Perform device identification via (37);
Get initial data symbols and variance $\hat{\mathbf{x}}^{\text{I}}_u$ and $\boldsymbol{\tau}^{x_{\text{I}}}_u$ by solving (39);
Initialization: $\forall q,u,n_a,s_a$: $\overleftarrow{\mathbf{c}}^{n_a}_{q,u,s_a} = \hat{\mathbf{c}}^{\text{I}n_a}_{q,u,s_a}$, $\overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a} = \tau^{\text{c}\text{I}n_a}_{q,u,s_a}$, $\overleftarrow{\mathbf{c}}^{\text{F}}_{q,u,n_a,s_a} = \mathbf{F}_{1:L+1}\overleftarrow{\mathbf{c}}^{n_a}_{q,u,s_a}$, $\overleftarrow{\tau}^{c_{\text{F}}}_{q,u,n_a,s_a} = \overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a}$, $\overrightarrow{\mathbf{x}}_u = \hat{\mathbf{x}}^{\text{I}}_u$, $\overrightarrow{\boldsymbol{\tau}}^x_u = \boldsymbol{\tau}^{x_{\text{I}}}_u$, $\overleftarrow{\mathbf{x}}^{\text{F}}_u = \mathbf{0}$, $\overleftarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_u = \boldsymbol{\infty}$, $\overrightarrow{\mathbf{x}}^{\text{F}}_u = \mathbf{F}_{\text{b}}\overrightarrow{\mathbf{x}}_u$, $\overrightarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_u = \text{diag}(\mathbf{F}_{\text{b}}\,\text{diag}(\overrightarrow{\boldsymbol{\tau}}^x_u)\mathbf{F}^{\text{H}}_{\text{b}})$;
for $t_{\text{out}} = 1$ to T_{out} **do**
  $\forall q,u,n_a,s_a$: Compute $\Delta^{\text{F}}_{\mathbf{x}_u}$ and $\Delta^{\mathbf{x}^{\text{F}}_u}_{f^{\text{P}}_{q,u,n_a,s_a}}$ via (57) and (58);
  \\ **Refine channel**
  for $t_{\text{c}} = 1$ to T_{c} **do**
    $\forall q,u,n_a,s_a$: Compute $\Delta^{f^{\text{P}}_{q,u,n_a,s_a}}_{\mathbf{c}^{\text{F}}_{q,u,n_a,s_a}}$ via (72)-(74) and (76);
    $\forall q,u,n_a,s_a$: Compute $\hat{\mathbf{c}}^{n_a}_{q,u,s_a}(t_{\text{c}}+1)$ and $\tau^{c_{n_a}}_{q,u,s_a}$ via (48) and (46);
    $\forall q,u,n_a,s_a$: Update $\Delta^{\mathbf{c}^{n_a}_{q,u,s_a}}_{f^{\text{c}^{\text{F}}}_{q,u,n_a,s_a}}$ via (49);
    $\forall q,u,n_a,s_a$: Update $\overleftarrow{\mathbf{c}}^{\text{F}}_{q,u,n_a,s_a} = \mathbf{F}_{1:L+1}\overleftarrow{\mathbf{c}}^{n_a}_{q,u,s_a}$ and $\overleftarrow{\tau}^{c_{\text{F}}}_{q,u,n_a,s_a} = \overleftarrow{\tau}^{c_{n_a}}_{q,u,s_a}$;
    if $\frac{\sum_{q,u,n_a,s_a}\left\|\hat{\mathbf{c}}^{n_a}_{q,u,s_a}(t_{\text{c}}+1)-\hat{\mathbf{c}}^{n_a}_{q,u,s_a}(t_{\text{c}})\right\|^2}{\sum_{q,u,n_a,s_a}\|\hat{\mathbf{c}}^{n_a}_{q,u,s_a}(t_{\text{c}})\|^2} \le \eta_{\text{c}}$: break; **end if**
  end for
  $\forall q,u,n_a,s_a$: Compute $\Delta^{\text{F}}_{\mathbf{c}_{q,u,n_a,s_a}}$ via (50);
  \\ **Refine data symbols**
  for $t_{\text{x}} = 1$ to T_{x} **do**
    $\forall q,u,n_a,s_a$: Compute $\Delta^{f^{\text{P}}_{q,u,n_a,s_a}}_{\mathbf{x}^{\text{F}}_u}$ via (72)-(74) and (75);
    $\forall u$: Compute $\Delta^{\mathbf{x}^{\text{F}}_u}_{f^{\mathbf{x}_{\text{F}}}_u}$ via (53);
    $\forall u$: Compute $\hat{\mathbf{x}}_u(t_{\text{x}}+1)$ and $\boldsymbol{\tau}^x_u$ via (54) and (55);
    $\forall u$: Update $\Delta^{\mathbf{x}_u}_{f^{\mathbf{x}_{\text{F}}}_u}$ via (56);
    $\forall u$: Update $\overrightarrow{\mathbf{x}}^{\text{F}}_u = \mathbf{F}_{\text{b}}\overrightarrow{\mathbf{x}}_u$ and $\overrightarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_u = \text{diag}(\mathbf{F}_{\text{b}}\,\text{diag}(\overrightarrow{\boldsymbol{\tau}}^x_u)\mathbf{F}^{\text{H}}_{\text{b}})$;
    if $\frac{\sum_u\|\hat{\mathbf{x}}_u(t_{\text{x}}+1)-\hat{\mathbf{x}}_u(t_{\text{x}})\|^2}{\sum_u\|\hat{\mathbf{x}}_u(t_{\text{x}})\|^2} \le \eta_{\text{x}}$: break **end if**
  end for
end for
$\forall n,l',u,n_a,s_a$: Reconstruct $\hat{h}^{n_a}_{u,s_a}[n,l']$ via (6);
$\forall u$: Perform hard decision on $\hat{\mathbf{x}}_u$;

B. Distributed Cooperative Mode

Since the centralized mode imposes the huge computational burden on the central server, we next investigate the distributed method which allows for offloading the computations to edge satellites.

1) Device identification: In the distributed mode, the Algorithm 1 is firstly performed at each satellite independently to get the initial channel estimation $\hat{c}^{\text{I}_{u,s_a}}_{i,j}$. To identify active devices, each satellite computes the channel energy $\sum_{i,j}\left|\hat{c}^{\text{I}_{u,s_a}}_{i,j}\right|^2$, and broadcast it within the constellations. Then, the activity indicator of u-th device can be given by (37).

2) Initial symbol detection: Here, each satellite detect data symbols independently in the initial phase, and then we have the following relationship.

$$\mathbf{y}^{\text{DD}}_{s_a} = \bar{\mathbf{H}}_{s_a}\mathbf{x} + \mathbf{e}_{s_a}, \tag{60}$$

where $\bar{\mathbf{H}}_{s_a} = [\bar{\mathbf{H}}^{\text{T}}_{0,s_a},\ldots,\bar{\mathbf{H}}^{\text{T}}_{N_z N_y-1,s_a}]^{\text{T}}$. The difference from (39) is that the channel matrix $\bar{\mathbf{H}}_{s_a}$ only involves the channels related to the s_a-th satellite. Then, the GAMP with uniform prior can be employed to get the initial guess of data symbols $\hat{\mathbf{x}}^{\text{I}}_{u,s_a}$ for active devices and the corresponding variance $\boldsymbol{\tau}^{x_{\text{I}}}_{u,s_a}$.

3) Jointly refine channel estimation and symbol detection: To exploit diversity provided by the satellite constellations, we propose to share the soft information about data symbols among the S_a satellites during the iteration process of the refinement algorithm. Hence, except for the previous defined auxiliary variables, we introduce $\mathbf{x}_{u,s_a} = \mathbf{x}_u$ and the corresponding factors $f^{\mathbf{x}}_{u,s_a,s'_a} = \delta(\mathbf{x}_{u,s_a} - \mathbf{x}_{u,s'_a})$, $s'_a = \{0,\ldots,S_a-1\}\backslash s_a$. Note that the previous derivations of IC module and CER module for centralized mode can also be applied here. However, there are two significant differences in the SDR module: the one is that the likelihoods from S_a satellites can not be combined directly via (53), since the refinement algorithm is implemented independently at each satellite; the other is that the message from $\mathbf{x}_{u,s_a}$ to $f^{\mathbf{x}_{\text{F}}}_{u,s_a}$ should combine the soft information shared by other satellites.

Firstly, by considering the likelihoods at the s_a-th satellite, we revise the message in (53) as

$$\begin{aligned}\Delta^{\mathbf{x}^{\text{F}}_{u,s_a}}_{f^{\mathbf{x}_{\text{F}}}_{u,s_a}} &\propto \prod_{q,n_a}\Delta^{f^{\text{P}}_{q,u,n_a,s_a}}_{\mathbf{x}^{\text{F}}_{u,s_a}} \\ &\propto \mathcal{CN}(\mathbf{x}^{\text{F}}_{u,s_a} \mid \overleftarrow{\bar{\mathbf{x}}}^{\text{F}}_{u,s_a}, \text{diag}(\overleftarrow{\bar{\boldsymbol{\tau}}}^{x_{\text{F}}}_{u,s_a})),\end{aligned} \tag{61}$$

where $\overleftarrow{\bar{\boldsymbol{\tau}}}^{x_{\text{F}}}_{u,s_a} = 1 \oslash \sum_{q,n_a}(1 \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_{q,u,n_a,s_a})$ and $\overleftarrow{\bar{\mathbf{x}}}^{\text{F}}_{u,s_a} = \overleftarrow{\bar{\boldsymbol{\tau}}}^{x_{\text{F}}}_{u,s_a} \odot \sum_{q,n_a}(\overleftarrow{\mathbf{x}}^{\text{F}}_{q,u,n_a,s_a} \oslash \overleftarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_{q,u,n_a,s_a})$. Besides, with the message shared by other satellites $\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}} = \mathcal{CN}(\mathbf{x}_{u,s_a} \mid \overrightarrow{\bar{\mathbf{x}}}_{u,s_a,s'_a}, \text{diag}(\overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a}))$, the message from $\mathbf{x}_{u,s_a}$ to $f^{\mathbf{x}_{\text{F}}}_{u,s_a}$ can be updated as

$$\begin{aligned}\Delta^{\mathbf{x}_{u,s_a}}_{f^{\mathbf{x}_{\text{F}}}_{u,s_a}} &\propto \Delta^{f^{\mathbf{x}_{\text{F}}}_{u,s_a}}_{\mathbf{x}_{u,s_a}} \prod_{s'_a}\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}} \\ &\propto \mathcal{CN}(\mathbf{x}_{u,s_a} \mid \overrightarrow{\bar{\mathbf{x}}}_{u,s_a}, \text{diag}(\overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a})),\end{aligned} \tag{62}$$

where $\overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} = 1 \oslash (1 \oslash \overleftarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} + \sum_{s'_a}(1 \oslash \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a}))$ and $\overrightarrow{\bar{\mathbf{x}}}_{u,s_a} = \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} \odot (\overleftarrow{\bar{\mathbf{x}}}_{u,s_a} \oslash \overleftarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} + \sum_{s'_a}(\overrightarrow{\bar{\mathbf{x}}}_{u,s_a,s'_a} \oslash \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a}))$. Then, the message transmitted to other satellites can be obtained as

$$\begin{aligned}\Delta^{\mathbf{x}_{u,s_a}}_{f^{\mathbf{x}}_{u,s_a,s'_a}} &\propto \frac{\Delta^{\mathbf{x}_{u,s_a}}_{f^{\mathbf{x}_{\text{F}}}_{u,s_a}}}{\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}}} \\ &\propto \mathcal{CN}(\mathbf{x}_{u,s_a} \mid \overleftarrow{\bar{\mathbf{x}}}_{u,s_a,s'_a}, \text{diag}(\overleftarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a})),\end{aligned} \tag{63}$$

where $\overleftarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a} = 1 \oslash (1 \oslash \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} - 1 \oslash \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a})$ and $\overleftarrow{\bar{\mathbf{x}}}_{u,s_a,s'_a} = \overleftarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a,s'_a} \odot (\overrightarrow{\bar{\mathbf{x}}}_{u,s_a} \oslash \overrightarrow{\bar{\boldsymbol{\tau}}}^x_{u,s_a} - \overrightarrow{\bar{\mathbf{x}}}_{u,s_a,s'_a} \oslash$

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

$\overrightarrow{\boldsymbol{\tau}}^{x}_{u,s_a,s'_a}$). Finally, we can update $\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}}$ as

$$\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}} \propto \int_{\mathbf{x}_{u,s'_a}} f^{\mathbf{x}}_{u,s_a,s'_a} \Delta^{\mathbf{x}_{u,s'_a}}_{f^{\mathbf{x}}_{u,s_a,s'_a}}, \tag{64}$$

where $\overrightarrow{\hat{\mathbf{x}}}_{u,s_a,s'_a} = \overleftarrow{\hat{\mathbf{x}}}_{u,s'_a,s_a}$ and $\overrightarrow{\boldsymbol{\tau}}^{x}_{u,s_a,s'_a} = \overleftarrow{\boldsymbol{\tau}}^{x}_{u,s'_a,s_a}$. Other derivations of SDR module in centralized mode can be reused by adding the index s_a, which represents the computations are implemented at the s_a-th satellite.

The proposed distributed AEP algorithm is summarized in Algorithm 3. In Line 5, T_{ex} represents the number of iterations during which the soft information of data symbols is exchanged. Line 6 indicates that the subsequent computations are individually performed at each satellite, which is executed in parallel for practical deployment. Lines 21 and 24 handle the exchange of soft information about data symbols within the satellite constellations. Totally, the soft information about data symbols is exchanged $T_{\text{ex}} + 1$ times in Algorithm 3. Furthermore, the damping and CFM method can also be adopted here. The complexity analysis for the distributed AEP is similar to that of the centralized AEP, except that the dominant computations are offloaded to edge satellites. Therefore, for a single satellite, the complexities of the initial phase and the refinement phase are in the order of $\mathcal{O}(UM_{\text{p}}N_{\text{p}}(Q_{\text{I}}+1)(L+1)N_zN_y + N_zN_y\,|\mathcal{U}_a|\,(MN)^2)$ and $\mathcal{O}(N_zN_y\,|\mathcal{U}_a|\,(Q+1)(L+1)^3 + |\mathcal{U}_a|\,(K_p+1)MN\log(MN))$, respectively.

Algorithm 3 Distributed AEP

Input: Recieved symbols $\mathbf{y}'_{n_a,s_a}$; the maximum number of iterations T_{ex}, T_{out}, T_{c}, T_{x}; the termination threshold η_{c}, η_{x}.
Output: The refined time-varying channel $\hat{h}^{n_a}_{u,s_a}[n,l']$ and data symbols $\hat{\mathbf{x}}_u$.

1: Get initial CE $\hat{\mathbf{c}}^{\text{r}_{n_a}}_{q,u,s_a}$,$\tau^{\text{cr}_{n_a}}_{q,u,s_a}$, and $\sigma^2_{s_a}$ by Algorithm 1;
2: Perform device identification via (37);
3: Get initial data symbols and variance $\hat{\mathbf{x}}^{\text{I}}_{u,s_a}$ and $\boldsymbol{\tau}^{x_{\text{I}}}_{u,s_a}$ by solving (60);
4: **Initialization**: Similar with line 4 in Algorithm 2 except for $\forall u, s_a, s'_a$: $\overrightarrow{\hat{\mathbf{x}}}_{u,s_a,s'_a} = \mathbf{0}$, $\overrightarrow{\boldsymbol{\tau}}^{x}_{u,s_a,s'_a} = \boldsymbol{\infty}$;
5: **for** $t_{\text{ex}} = 1$ to T_{ex} **do**
6: **for** $s_a = 0$ to $S_a - 1$ **do**
7: **for** $t_{\text{out}} = 1$ to T_{out} **do**
8: $\forall q, u, n_a$: Compute $\Delta^{\text{F}}_{\mathbf{x}_{u,s_a}}$ and $\Delta^{\mathbf{x}^{\text{F}}_{u,s_a}}_{f^{\text{P}}_{q,u,n_a,s_a}}$ via (57) and (58);
9: Refine channel using lines 8-15 of Algorithm 2 with fixed s_a;
10: \\ **Refine data symbols**
11: **for** $t_{\text{x}} = 1$ to T_{x} **do**
12: $\forall q, u, n_a$: Compute $\Delta^{f^{\text{P}}_{q,u,n_a,s_a}}_{\mathbf{x}^{\text{F}}_{u,s_a}}$ via (72)-(74) and (75);
13: $\forall u$: Compute $\Delta^{\mathbf{x}^{\text{F}}_{u,s_a}}_{f^{\mathbf{x}^{\text{F}}_{u,s_a}}}$ via (61);
14: $\forall u$: Compute $\hat{\mathbf{x}}_{u,s_a}(t_{\text{x}}+1)$ and $\boldsymbol{\tau}^{x}_{u,s_a}$ via (54) and (55);
15: $\forall u$: Update $\Delta^{\mathbf{x}_{u,s_a}}_{f^{\mathbf{x}^{\text{F}}_{u,s_a}}}$ via (62);
16: $\forall u$: Update $\overrightarrow{\hat{\mathbf{x}}}^{\text{F}}_{u,s_a} = \mathbf{F}_{\text{b}}\overrightarrow{\mathbf{x}}_{u,s_a}$ and $\overrightarrow{\boldsymbol{\tau}}^{x_{\text{F}}}_{u,s_a} = \text{diag}(\mathbf{F}_{\text{b}}\,\text{diag}(\overrightarrow{\boldsymbol{\tau}}^{x}_{u,s_a})\mathbf{F}^{\text{H}}_{\text{b}})$;
17: **if** $\frac{\sum_u \|\hat{\mathbf{x}}_{u,s_a}(t_{\text{x}}+1) - \hat{\mathbf{x}}_{u,s_a}(t_{\text{x}})\|^2}{\sum_u \|\hat{\mathbf{x}}_{u,s_a}(t_{\text{x}})\|^2} \le \eta_{\text{x}}$: break; **end if**
18: **end for**
19: **end for**
20: **end for**
21: $\forall u, s_a, s'_a$: Compute $\Delta^{\mathbf{x}_{u,s_a}}_{f^{\mathbf{x}}_{u,s_a,s'_a}}$ and $\Delta^{f^{\mathbf{x}}_{u,s_a,s'_a}}_{\mathbf{x}_{u,s_a}}$ via (63) and (64);
22: **end for**
23: $\forall t, l', u, n_a, s_a$: Reconstruct $\hat{h}^{n_a}_{u,s_a}[n,l']$ via (6);
24: $\forall u$: $\hat{\mathbf{x}}_u = 1 \oslash (\sum_{s_a} 1 \oslash \boldsymbol{\tau}^{x}_{u,s_a}) \odot (\sum_{s_a} \hat{\mathbf{x}}_{u,s_a} \oslash \boldsymbol{\tau}^{x}_{u,s_a})$;
25: $\forall u$: Perform hard decision on $\hat{\mathbf{x}}_u$;

Table II: Simulation Parameters

Parameter	Values
Channel Model	NTN-CDL-A
Carrier frequency	2 GHz
Modulation Scheme	QPSK
Orbit altitude	600 km
Spot beam radius	200 km
Satellite velocity	7.5622 km/s
Device velocity	$0 \sim 500$ km/h
Subcarrier spacing Δf	240 kHz
Frame size (M, N)	$(32, 15)$
Number of antennas $N_z \times N_y$	8×8

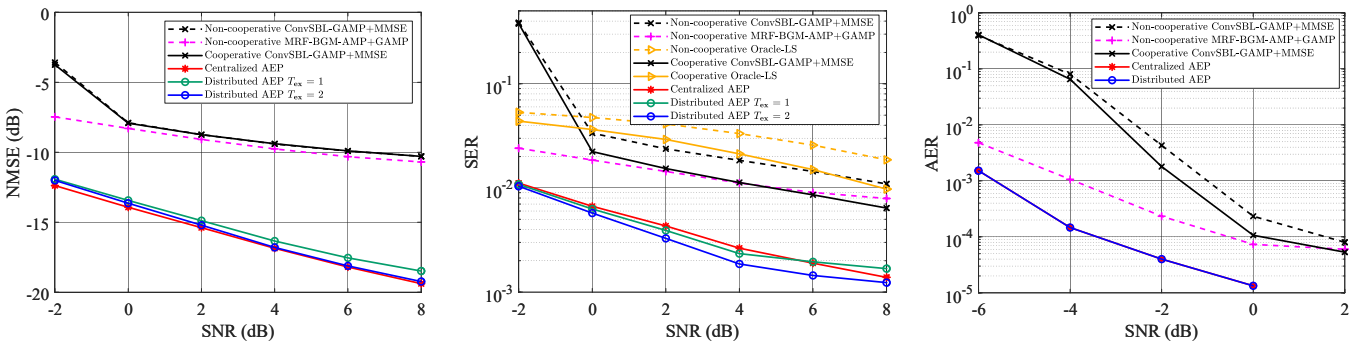
V. Numerical Results

This section presents simulation results to validate the effectiveness of the proposed algorithms, with the NTN-CDL-A channel model [28] recommended by 3GPP for the experiments. We consider a typical random access scenario in the non-terrestrial networks [37], where $U = 50$ potential devices are randomly distributed within the coverage area of the satellites. This area is defined as a circle with the 200 km radius. The satellites themselves are positioned evenly along the circumference of a larger circle with the 400 km radius, at the altitude of 600 km. The main parameters are summarized in Table II. The activity error rate (AER), the normalized mean squared error (NMSE), and the symbol error rate (SER) are adopted as the performance metrics, defined as $\text{AER} = \frac{1}{U}\sum_u \mathbb{I}\{\hat{\lambda}_u \neq \lambda_u\}$, $\text{NMSE} = \frac{\sum_{n,l',n_a,u,s_a}\left|h^{n_a}_{u,s_a}[n,l'] - \hat{h}^{n_a}_{u,s_a}[n,l']\right|^2}{\sum_{n,l',n_a,u,s_a}\left|h^{n_a}_{u,s_a}[n,l']\right|^2}$, and $\text{SER} = \frac{\sum_u \sum_{m\in\mathcal{S}_d}\mathbb{I}\{x_u[m]\neq\hat{x}_u[m]\}}{U|\mathcal{S}_d|}$, where $\mathcal{S}_d$ contains the indexes of the data symbols, assuming the inactive devices transmit zeroes. Then, a symbol is detected correctly only when both the device activity and the symbol itself are judged correctly. In addition, the signal-to-noise ratio is defined as $\text{SNR} = 10\log_{10}\frac{\|\mathbf{Z}_{s_a}\|^2_{\text{F}}}{M_{\text{p}}N_{\text{p}}N_zN_y\sigma^2_{s_a}}$. Two benchmarks are adopted for performance comparison. The first, ConvSBL-GAMP + MMSE, employs the ConvSBL-GAMP [19] for joint device identification and channel estimation, followed by symbol detection using the MMSE equalizer [38]. The second, Oracle-LS, leverages perfect CSI to detect data symbols based on the least squares principle [39]. These benchmarks are evaluated for two modes: (i) centralized cooperative mode where the central server processes all received signals, leveraging the collective information for enhanced performance. (ii) non-cooperative mode where satellites perform the task independently and without signal sharing.

Fig. 4 depicts the NMSE, SER, and AER as the functions of SNR achieved by various algorithms, given the number of satellites $S_a = 2$, device activity $p_\lambda = 0.1$, and pilot overhead $\rho = 0.4$. The BEM orders are set to $Q_I = 4$ and $Q_R = 8$ unless otherwise specified. Additionally, we include results for non-cooperative MRF-BGM-AMP+GAMP, which is the initial phase of the distributed AEP. It is observed that the cooperative schemes achieve superior performance compared to their non-cooperative counterparts, with the proposed cooperative algorithms significantly outperforming the benchmarks. For instance, in Fig. 4(a), when the SNR is 4 dB, the proposed centralized AEP outperforms ConvSBL-GAMP by around 6 dB in terms of NMSE. Similarly, Fig. 4(b) indicates that at the SER of 0.01, both centralized and distributed AEP outperform

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064



(a) Performance of channel estimation. (b) Performance of symbol detection. (c) Performance of device identification.

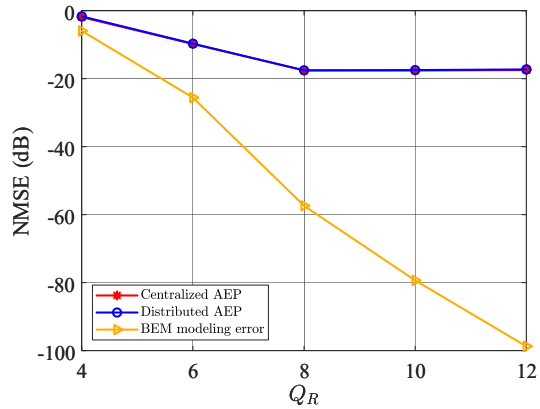
Fig. 4. Performance comparison among different schemes given different SNR values, where $S_a = 2$, $p_\lambda = 0.1$, and $\rho = 0.4$.

the cooperative ConvSBL-GAMP+MMSE and Oracle-LS by approximately 6 dB and 9 dB in terms of SNR, respectively. This is due to the benefit of joint channel estimation and symbol detection design. Fig. 4(c) shows that at the AER of 0.001, both the centralized and distributed AEP outperform the cooperative ConvSBL-GAMP+MMSE by about 4 dB.

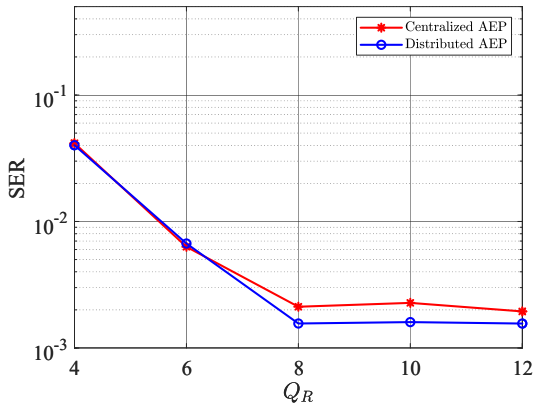
Fig. 4(a) and Fig. 4(b) also illustrate that increasing the iterations of soft information exchange of data symbols in distributed AEP leads to reduced NMSE and SER, and only two exchanges ($T_{\text{ex}} = 1$) are needed for achieving performance comparable to the centralized AEP; it is indicated that distributed AEP can efficiently offload the computational burden while requiring small signaling overhead. Moreover, the distributed AEP achieves a marginal gain over the centralized AEP when $T_{\text{ex}} = 2$, which results from the different ways they exploit the spatial diversity. In the distributed AEP, the soft information related to data symbols are produced by the LMMSE denoiser and then aggregated from different satellites, which is possible to obtain a more precise likelihoods for data symbols. We adopt $T_{\text{ex}} = 2$ for subsequent distributed AEP.

Fig. 5 illustrates how the accuracy of BEM modeling influences the performance of the proposed algorithms by showing NMSE and SER as functions of the BEM order Q_R. The observations reveal that performance enhancements are evident when $Q_R \leq 8$, as higher Q_R values lead to reduced BEM modeling errors. However, for $Q_R > 8$, performance gains plateau due to inter-user and inter-component interference, which limits further NMSE and SER improvements despite ongoing reductions in BEM modeling error. Notably, BEM with $Q_R = 2\lceil 2N\bar{\nu}_{\max}\rceil = 8$ suffices for precisely modeling TSL, achieving an NMSE of approximately -60 dB.

In Fig. 6, we evaluate the performance of different schemes with varying pilot overhead, given the number of satellites $S_a = 2$, device activity $p_\lambda = 0.1$, and SNR $= 5$ dB. Fig. 6(b) reveals that at a pilot overhead of 0.2, Oracle-LS outperforms other methods in SER, benefiting from perfect channel knowledge for symbol detection. In contrast, other methods struggle with accurate channel estimation at such low pilot overhead, as shown in Fig. 6(a). However, when pilot overhead exceeds 0.3, our proposed cooperative algorithms demonstrate remarkable improvements over baselines. This enhancement results from more accurate initial channel estimation, facilitating the initial symbol detection and refinement phase of our algorithms. For instance, to achieve an SER of 0.01, both the centralized and distributed AEP require 0.15 less pilot overhead than the



(a) Performance of channel estimation.



(b) Performance of symbol detection.

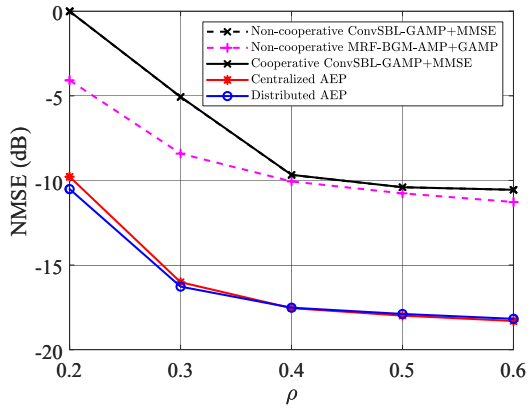
Fig. 5. The impact of BEM modeling accuracy, where SNR = 5 dB, $S_a = 2$, $p_\lambda = 0.1$, and $\rho = 0.4$.

cooperative ConvSBL-GAMP+MMSE, indicating their higher spectral efficiency.

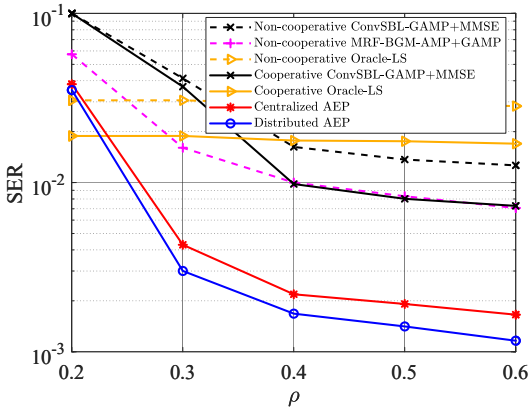
In Fig. 7, we investigate the scalability of the system. Fig. 7(a) plots the SER as a function of the device activity, given the number of satellites $S_a = 2$, pilot overhead $\rho = 0.4$, and SNR $= 5$ dB. It is observed that the proposed centralized AEP and distributed AEP achieves the best performance. Note that the SER of all the schemes deteriorates obviously when p_λ is large. This is because the limited size of the receiving antenna array leads to decreased spatial separation as the number of active devices increases. Then, the inter-user interference and inter-component interference of BEM also become severer in the spatial domain. In Fig. 7(b), we evaluate the SER of different schemes with respect to the number of satellites,

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064



(a) Performance of channel estimation.



(b) Performance of symbol detection.

Fig. 6. Performance comparison among different schemes given different pilot overhead, where $S_a = 2$, $p_\lambda = 0.1$, and SNR $= 5$ dB.

given the device activity $p_\lambda = 0.1$, pilot overhead $\rho = 0.4$ and SNR $= 5$ dB. As expected, the SER for cooperative schemes diminishes with an increase in satellite numbers, whereas the SER for non-cooperative schemes remains relatively unchanged. Besides, only marginal gains are observed when the number of satellites is more than four. This plateau effect likely results from the satellites' circular equidistant arrangement, which shortens inter-satellite distances and leads to highly correlated channels among adjacent satellites. It is worth exploring the optimal satellite positioning to enhance cooperative service to ground devices in the future work. Fig. 7(c) depicts the AER achieved by different schemes as a function of the number of satellites. A similar trend to Fig. 7(b) is observed. The intuitive explanation behind this trend is that when one LEO satellite misjudges the activity of a device due to unfavorable channel conditions, observations from other satellites can greatly correct this error.

Finally, Fig. 8 depicts the NMSE and SER as the functions of the number of antennas achieved by various algorithms, given the number of satellites $S_a = 2$, device activity $p_\lambda = 0.1$, pilot overhead $\rho = 0.4$, and SNR $= 5$ dB. As expected, increasing the number of receive antennas enhances the performance of all schemes. This improvement is due to higher spatial resolution at the receiver, which reduces inter-user and inter-component interference. Additionally, both the proposed centralized and distributed AEP algorithms outperform the benchmarks significantly as the number of antennas increases, highlighting their effectiveness in exploiting the sparsity of the angular domain channel.

VI. Conclusion

This paper proposed a random access scheme tailored for cooperative LEO satellite networks, leveraging OTFS to mitigate the significant Doppler effect in the TSLs. We introduced GCE-BEM to obtain a compact channel representation and then analyzed the input-output relationship of the system. Next, by exploiting the 2D block sparsity of the large antenna array, a message-passing type algorithm was designed for initial channel estimation. Then, based on the EP and CLT, we proposed both a centralized and distributed FFT-based cooperative algorithm for refining channel estimation and symbol detection of active devices, respectively. Simulation results have verified that our proposed schemes outperform existing solutions in terms of AER, NMSE and SER. Furthermore, the distributed method requires only two exchanges of soft information about data symbols within the constellation to achieve performance comparable to the centralized scheme. Future work is worth exploring the optimal satellite positioning to enhance cooperative service to ground devices.

Appendix A
Proof of Proposition 1

Firstly, we denote the (l, n')-th element of $\mathbf{B}_u^{\mathrm{TD}}$ as $B_u^{\mathrm{TD}}[l, n']$, where $l = 0, \ldots, M-1$ and $n' = 0, \ldots, N-1$. Then, substituting (9) into (11), we have

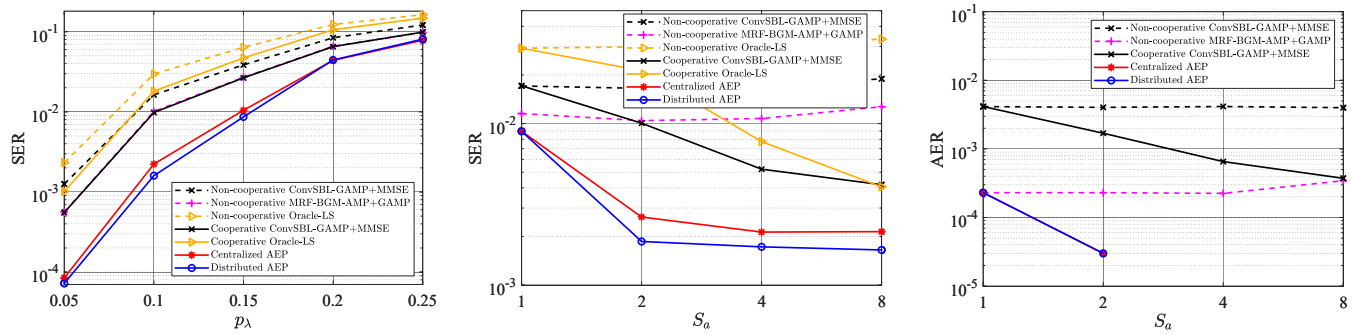
$$\begin{aligned} Y_{n_a,u,s_a}^{\mathrm{DD}}[l,k] = \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} \sum_{l' \leqslant l} h_{u,s_a}^{n_a}\left[M_{\mathrm{cp}} + n'M + l, l'\right] \\ \times B_u^{\mathrm{TD}}\left[l - l', n'\right] e^{-\jmath \frac{2\pi}{N} k n'} \\ + \frac{1}{\sqrt{N}} \sum_{n'=0}^{N-1} \sum_{l' > l} h_{u,s_a}^{n_a}\left[M_{\mathrm{cp}} + n'M + l, l'\right] e^{-\jmath \frac{2\pi}{N} k n'} \\ \times B_u^{\mathrm{TD}}\left[l - l' + M, (n'-1)_N\right]. \end{aligned} \quad (65)$$

It is observed that (65) is the sum of two parts, where the former one is the sum for $l' \leqslant l$ after demodulation and the other represents the sum for $l' > l$ after demodulation. We denote the two parts as $\tilde{S}_{u,s_a}^{n_a}[l,k]$ and $\bar{S}_{u,s_a}^{n_a}[l,k]$, respectively. Then, $Y_{n_a,u,s_a}^{\mathrm{DD}}[l,k]$ can be represented as $Y_{n_a,u,s_a}^{\mathrm{DD}}[l,k] = \tilde{S}_{u,s_a}^{n_a}[l,k] + \bar{S}_{u,s_a}^{n_a}[l,k]$. Substituting (6) and (8) into (65) and given $\omega_q = \frac{2\pi q'}{MNR}$ with $q' = \left(q - \left\lceil \frac{Q}{2} \right\rceil\right)$, we can get

$$\begin{aligned} \tilde{S}_{u,s_a}^{n_a}[l,k] &= \frac{1}{N} \sum_{q=0}^{Q} \sum_{l' \leqslant l} \sum_{n'=0}^{N-1} \sum_{k'=0}^{N-1} c_{q,l',u,s_a}^{n_a} e^{\jmath \frac{2\pi q'}{MNR}\left(M_{\mathrm{cp}} + n'M + l\right)} \\ &\quad \times X_u^{\mathrm{DD}}\left[l - l', k'\right] e^{\jmath \frac{2\pi}{N}(k'-k)n'} \\ &= \frac{1}{N} \sum_{q=0}^{Q} \sum_{l' \leqslant l} \sum_{k'=0}^{N-1} c_{q,l',u,s_a}^{n_a} e^{\jmath \frac{2\pi q'}{MNR}(M_{\mathrm{cp}} + l)} X_u^{\mathrm{DD}}\left[l - l', k'\right] \\ &\quad \times \frac{1 - e^{\jmath 2\pi \left(k' - \left(k - \frac{q'}{R}\right)\right)}}{1 - e^{\jmath \frac{2\pi}{N}\left(k' - \left(k - \frac{q'}{R}\right)\right)}}, \end{aligned} \quad (66)$$

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

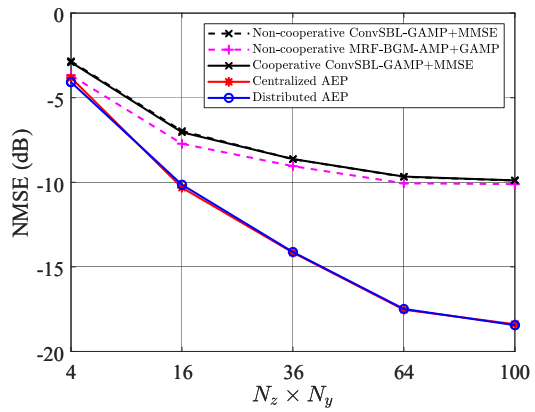


(a) Symbol detection performance for different device acitivities, where SNR = 5 dB.

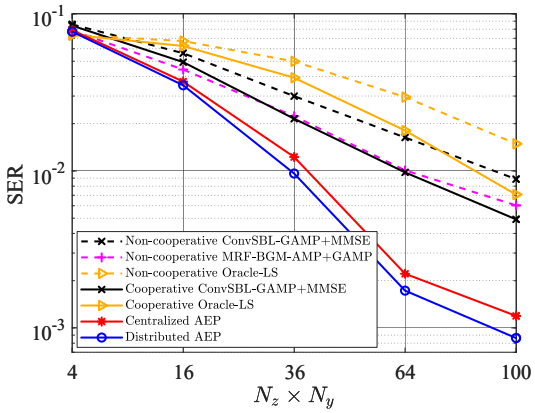
(b) Symbol detection performance for different number of satellites, where SNR = 5 dB.

(c) Device identification performance for different number of satellites, where SNR = −2 dB.

Fig. 7. Performance comparison for the system scalability among different schemes.



(a) Performance of channel estimation.



(b) Performance of symbol detection.

Fig. 8. Performance comparison among different schemes given different number of antennas, where $S_a = 2$, $p_\lambda = 0.1$, $\rho = 0.4$, and SNR = 5 dB.

and

$$\bar{S}^{n_a}_{u,s_a}[l,k] = \frac{1}{N}\sum_{q=0}^{Q}\sum_{l'>l}\sum_{k'=0}^{N-1} c^{n_a}_{q,l',u,s_a} e^{\bar{\jmath}\frac{2\pi q'}{MNR}(M_{\text{cp}}+l)} X^{\text{DD}}_u[l-l'+M,k'] \times \sum_{n'=0}^{N-1} e^{\bar{\jmath}\frac{2\pi q'}{RN}n'} e^{\bar{\jmath}\frac{2\pi}{N}k'(n'-1)_N} e^{-\bar{\jmath}\frac{2\pi}{N}kn'}$$

$$= \frac{1}{N}\sum_{q=0}^{Q}\sum_{l'>l}\sum_{k'=0}^{N-1} c^{n_a}_{q,l',u,s_a} e^{\bar{\jmath}\frac{2\pi q'}{MNR}(M_{\text{cp}}+l)} X^{\text{DD}}_u[l-l'+M,k'] \times e^{-\bar{\jmath}\frac{2\pi}{N}k'} \frac{1-e^{\bar{\jmath}2\pi\left(k'-\left(k-\frac{q'}{R}\right)\right)}}{1-e^{\bar{\jmath}\frac{2\pi}{N}\left(k'-\left(k-\frac{q'}{R}\right)\right)}}. \quad (67)$$

Finally, by absorbing the constant $e^{\bar{\jmath}\frac{2\pi q'}{MNR}M_{\text{cp}}}$ into $c^{n_a}_{q,l',u,s_a}$ and substituting $\frac{2\pi q'}{MNR}$ as ω_q, we get the result in Proposition 1.

Appendix B
Message Scheduling in the MRF

With the inputs $\Delta^{f^{c^{u,s_a}}_{i,j}}_{s^{u,s_a}_{i,j}}$, we are now ready to describe the messages involved in the MRF. Recalling that the satellite adopts the 2D UPA, $\mathcal{D}_j$ can be given as $\mathcal{D}_j = \{j+N_z, j-N_z, j+1, j-1\}$. To clearly characterize the relative position, the left, right, top, and bottom neighbors of $s^{u,s_a}_{i,j}$ are reindexed by $\{s^{u,s_a}_{i,j_{\text{L}}}, s^{u,s_a}_{i,j_{\text{R}}}, s^{u,s_a}_{i,j_{\text{T}}}, s^{u,s_a}_{i,j_{\text{B}}}\}$ corresponding to $\mathcal{D}_j$. The left, right, top, and bottom input messages of $s^{u,s_a}_{i,j}$ are denoted as $\Omega^{\text{L}^{u,s_a}}_{i,j}$, $\Omega^{\text{R}^{u,s_a}}_{i,j}$, $\Omega^{\text{T}^{u,s_a}}_{i,j}$, and $\Omega^{\text{B}^{u,s_a}}_{i,j}$. Then, the input message of $s^{u,s_a}_{i,j}$ from the left is given by

$$\Omega^{\text{L}^{u,s_a}}_{i,j} \propto \int_{\sim s^{u,s_a}_{i,j}} \Delta^{f^{c^{u,s_a}}_{i,j}}_{s^{u,s_a}_{i,j}} \Gamma(s^{u,s_a}_{i,j}, s^{u,s_a}_{i,j_{\text{L}}})\Psi(s^{u,s_a}_{i,j_{\text{L}}}) \prod_{\text{p}\in\{\text{L,T,B}\}} \Omega^{\text{p}^{u,s_a}}_{i,j_{\text{L}}} = \xi^{\text{L}^{u,s_a}}_{i,j}\delta\left(s^{u,s_a}_{i,j}-1\right) + (1-\xi^{\text{L}^{u,s_a}}_{i,j})\delta\left(s^{u,s_a}_{i,j}+1\right), \quad (68)$$

where $\xi^{\text{L}^{u,s_a}}_{i,j}$ is given by (69). The input messages of $s^{u,s_a}_{i,j}$ from right, top, and bottom have a similar form to $\Omega^{\text{L}^{u,s_a}}_{i,j}$. Then, the output message for $s^{u,s_a}_{i,j}$ is given by

$$\Delta^{s^{u,s_a}_{i,j}}_{f^{c^{u,s_a}}_{i,j}} \propto \Psi(s^{u,s_a}_{i,j_{\text{L}}}) \prod_{\text{p}\in\{\text{L,R,T,B}\}} \Omega^{\text{p}^{u,s_a}}_{i,j} = \zeta^{u,s_a}_{i,j}\delta\left(s^{u,l'}_{i,j}-1\right) + (1-\zeta^{u,s_a}_{i,j})\delta\left(s^{u,l'}_{i,j}+1\right), \quad (70)$$

where

$$\zeta^{u,s_a}_{i,j} = \frac{e^{-\gamma}\prod_{\text{p}\in\{\text{L,R,T,B}\}}\xi^{\text{p}^{u,s_a}}_{i,j}}{e^{-\gamma}\prod_{\text{p}\in\{\text{L,R,T,B}\}}\xi^{\text{p}^{u,s_a}}_{i,j} + e^{\gamma}\prod_{\text{p}\in\{\text{L,R,T,B}\}}(1-\xi^{\text{p}^{u,s_a}}_{i,j})}. \quad (71)$$

Appendix C
Message Scheduling in the IC Module

In this appendix, we investigate the message passing in the IC module. With the input message $\Delta^{\mathbf{x}^{\text{F}}_u}_{f^{\text{P}}_{q,u,n_a,s_a}} =$

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

$$\xi_{i,j}^{\mathrm{L},u,s_a} = \frac{e^{-\gamma+\beta}\eta_{i,j_{\mathrm{L}}}^{u,s_a}\prod_{\mathrm{p}\in\{\mathrm{L,T,B}\}}\xi_{i,j_{\mathrm{L}}}^{\mathrm{p},u,s_a} + e^{\gamma-\beta}(1-\eta_{i,j_{\mathrm{L}}}^{u,s_a})\prod_{\mathrm{p}\in\{\mathrm{L,T,B}\}}(1-\xi_{i,j_{\mathrm{L}}}^{\mathrm{p},u,s_a})}{(e^{\beta}+e^{-\beta})\left(e^{-\gamma}\eta_{i,j_{\mathrm{L}}}^{u,s_a}\prod_{\mathrm{p}\in\{\mathrm{L,T,B}\}}\xi_{i,j_{\mathrm{L}}}^{\mathrm{p},u,s_a} + e^{\gamma}(1-\eta_{i,j_{\mathrm{L}}}^{u,s_a})\prod_{\mathrm{p}\in\{\mathrm{L,T,B}\}}(1-\xi_{i,j_{\mathrm{L}}}^{\mathrm{p},u,s_a})\right)}. \tag{69}$$

$\mathcal{CN}(\mathbf{x}_u^{\mathrm{F}} \mid \overrightarrow{\mathbf{x}}_{q,u,n_a,s_a}^{\mathrm{F}}, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{x_{\mathrm{F}}}))$ and $\Delta_{f_{q,u,n_a,s_a}^{\mathrm{P}}}^{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}} = \mathcal{CN}(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \overleftarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}}, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}}))$, the message $\Delta_{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}}$ passed into the IC module can be approximated as a Gaussian distribution, i.e., $\Delta_{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}} = \mathcal{CN}(\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}} \mid \overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}}, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}}))$, where $\overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}} = \overrightarrow{\mathbf{x}}_u^{\mathrm{F}} \odot \overleftarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}}$ and $\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}} = \overrightarrow{\boldsymbol{\tau}}_u^{x_{\mathrm{F}}} \odot \overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}} + \overrightarrow{\mathbf{x}}_u^{\mathrm{F}} \odot \overrightarrow{\mathbf{x}}_u^{\mathrm{F}^*} \odot \overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}} + \overleftarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} \odot \overleftarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}^*} \odot \overrightarrow{\boldsymbol{\tau}}_u^{x_{\mathrm{F}}}$. Since $\Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}} = \Delta_{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}}$, the message from factor node $f_{q,n_a,s_a}^{\mathrm{W}}$ to the variable $\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}$ is given as

$$\begin{aligned}
&\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}}^{f_{q,n_a,s_a}^{\mathrm{W}}} \\
&\propto \frac{\mathrm{Proj}\left[\Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}} \int_{\sim\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}} f_{q,n_a,s_a}^{\mathrm{W}} \prod_u \Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}\right]}{\Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}}} \\
&\overset{(a)}{\propto} \mathcal{CN}(\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{W}} \mid \overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{W}}, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{W}}})),
\end{aligned} \tag{72}$$

where (a) is due to CLT, $\overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{W}} = \sum_u \overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}}$, and $\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{W}}} = \sum_u \overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}}$. Since $\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}$ is transformed to $\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}$ by the DFT matrix and we have $\Delta_{f_{q,n_a,s_a}^{\mathrm{F}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}} = \Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}}^{f_{q,n_a,s_a}^{\mathrm{W}}}$, the message $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}}^{f_{q,n_a,s_a}^{\mathbf{d}_{\mathrm{F}}}}$ can be obtained as $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}}^{f_{q,n_a,s_a}^{\mathbf{d}_{\mathrm{F}}}} = \mathcal{CN}(\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}} \mid \overleftarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}}, \overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}}\mathbf{I}_{MN})$, where $\overleftarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}} = \mathbf{F}_{MN}^{\mathrm{H}}\overleftarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{W}}$, $\overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}} = < \overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{W}}} >$, and $< \cdot >$ is the empirical averaging operation. Recalling that $\mathbf{b}_q$ represents the q-th basis function of BEM, and hence $\mathrm{diag}(\mathbf{b}_q)$ is a unitary matrix. The message $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}}^{f_{q,n_a,s_a}^{\mathrm{B}}}$ can be computed as $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}}^{f_{q,n_a,s_a}^{\mathrm{B}}} = \mathcal{CN}(\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}} \mid \overleftarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{B}}, \overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{B}}}\mathbf{I}_{MN})$, where $\overleftarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{B}} = \mathbf{b}_q \odot \overleftarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}}$ and $\overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{B}}} = \overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}}$. Then, the message keep passing from right to left in the factor graph and we have $\Delta_{f_{n_a,s_a}^{\mathbf{g}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}} = \Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}}^{f_{q,n_a,s_a}^{\mathrm{B}}}$. At this point, the message start to pass in the opposite direction. According to the CLT, the output message of $f_{n_a,s_a}^{\mathbf{g}}$ is derived as

$$\begin{aligned}
&\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}}^{f_{n_a,s_a}^{\mathbf{g}}} \\
&\propto \frac{\mathrm{Proj}\left[\Delta_{f_{n_a,s_a}^{\mathbf{g}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}} \int_{\sim\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}} f_{n_a,s_a}^{\mathbf{g}} \Delta_{f_{n_a,s_a}^{\mathbf{g}}}^{\mathbf{g}_{n_a,s_a}} \prod_{\bar{q}\neq q} \Delta_{\mathbf{d}_{\bar{q},n_a,s_a}^{\mathrm{B}}}^{f_{\bar{q},n_a,s_a}^{\mathrm{B}}}\right]}{\Delta_{f_{n_a,s_a}^{\mathbf{g}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}}}} \\
&\propto \mathcal{CN}(\mathbf{d}_{q,n_a,s_a}^{\mathrm{B}} \mid \overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{B}}, \overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{B}}}\mathbf{I}_{MN}),
\end{aligned} \tag{73}$$

where $\overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{B}} = \mathbf{y}'_{n_a,s_a} - \sum_{\bar{q}\neq q}\overleftarrow{\mathbf{d}}_{\bar{q},n_a,s_a}^{\mathrm{B}}$ and $\overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{B}}} = \sigma_{s_a}^2 + \sum_{\bar{q}\neq q}\overleftarrow{\tau}_{\bar{q},n_a,s_a}^{d_{\mathrm{B}}}$. It can be observed that this step is similar with the SIC and eliminates the inter-component interference in BEM using the soft estimates. By the symmetry, we can follow the similar steps to derive the message in the opposite direction, and the message $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}}}^{f_{q,n_a,s_a}^{\mathrm{B}}} = \mathcal{CN}(\mathbf{d}_{q,n_a,s_a}^{\mathrm{F}} \mid \overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}}, \overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}}\mathbf{I}_{MN})$ with $\overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}} = \mathbf{b}_q^* \odot \overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{B}}$ and $\overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}} = \overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{B}}}$; the message $\Delta_{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}}^{f_{q,n_a,s_a}^{\mathbf{d}_{\mathrm{F}}}} = \mathcal{CN}(\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}} \mid \overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{W}}, \overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{W}}}\mathbf{I}_{MN})$, where $\overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{W}} = \mathbf{F}_{MN}\overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{F}}$ and $\overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{W}}} = \overrightarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{F}}}$. Similar to (73), the output message of $f_{q,n_a,s_a}^{\mathrm{W}}$ is obtained by combining relevant messages, i.e.,

$$\begin{aligned}
&\Delta_{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}^{f_{q,n_a,s_a}^{\mathrm{W}}} \propto \\
&\frac{\mathrm{Proj}\left[\Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}} \int_{\sim\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}} f_{q,n_a,s_a}^{\mathrm{W}} \Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,n_a,s_a}^{\mathrm{W}}} \prod_{\bar{u}\neq u} \Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,\bar{u},n_a,s_a}^{\mathrm{P}}}\right]}{\Delta_{f_{q,n_a,s_a}^{\mathrm{W}}}^{\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}}}} \\
&\propto \mathcal{CN}(\mathbf{d}_{q,u,n_a,s_a}^{\mathrm{P}} \mid \overrightarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}}, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}})),
\end{aligned} \tag{74}$$

where $\overrightarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}} = \overrightarrow{\mathbf{d}}_{q,n_a,s_a}^{\mathrm{W}} - \sum_{\bar{u}\neq u}\overleftarrow{\mathbf{d}}_{q,\bar{u},n_a,s_a}^{\mathrm{P}}$ and $\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}} = \overleftarrow{\tau}_{q,n_a,s_a}^{d_{\mathrm{W}}}\mathbf{1}_{MN} + \sum_{\bar{u}\neq u}\overleftarrow{\boldsymbol{\tau}}_{q,\bar{u},n_a,s_a}^{d_{\mathrm{P}}}$. Note that the inter-user interference is canceled in this step. Similar to [40], we next resort to the mean-filed approximation to provide the likelihoods for $\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}$ and $\mathbf{x}_u^{\mathrm{F}}$. Given the posterior distribution $\Delta_{\mathbf{x}_u}^{\mathrm{F}} = \mathcal{CN}(\mathbf{x}_u^{\mathrm{F}} \mid \hat{\mathbf{x}}_u^{\mathrm{F}}, \mathrm{diag}(\boldsymbol{\tau}_u^{x_{\mathrm{F}}}))$ and $\Delta_{\mathbf{c}_{q,u,n_a,s_a}}^{\mathrm{F}} = \mathcal{CN}(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}}, \mathrm{diag}(\boldsymbol{\tau}_{q,u,n_a,s_a}^{c_{\mathrm{F}}}))$, respectively, we can get

$$\begin{aligned}
&\Delta_{\mathbf{x}_u^{\mathrm{F}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}} \propto \exp\int_{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}} \Delta_{\mathbf{c}_{q,u,n_a,s_a}}^{\mathrm{F}} \times \\
&\log\mathcal{CN}(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \odot \mathbf{x}_u^{\mathrm{F}} \mid \overrightarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}}, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}})) \\
&\propto \mathcal{CN}(\mathbf{x}_u^{\mathrm{F}} \mid \overleftarrow{\mathbf{x}}_{q,u,n_a,s_a}^{\mathrm{F}}, \mathrm{diag}(\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{x_{\mathrm{F}}})),
\end{aligned} \tag{75}$$

where $\overleftarrow{\mathbf{x}}_{q,u,n_a,s_a}^{\mathrm{F}} = \frac{\overrightarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}} \odot \hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}^*}}{\hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} \odot \hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}^*} + \boldsymbol{\tau}_{q,u,n_a,s_a}^{c_{\mathrm{F}}}}$ and $\overleftarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{x_{\mathrm{F}}} = \frac{\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}}}{\hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} \odot \hat{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}^*} + \boldsymbol{\tau}_{q,u,n_a,s_a}^{c_{\mathrm{F}}}}$. By the symmetry, we also have

$$\Delta_{\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}}}^{f_{q,u,n_a,s_a}^{\mathrm{P}}} = \mathcal{CN}(\mathbf{c}_{q,u,n_a,s_a}^{\mathrm{F}} \mid \overrightarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}}, \mathrm{diag}(\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}})), \tag{76}$$

where $\overrightarrow{\mathbf{c}}_{q,u,n_a,s_a}^{\mathrm{F}} = \frac{\overrightarrow{\mathbf{d}}_{q,u,n_a,s_a}^{\mathrm{P}} \odot \hat{\mathbf{x}}_u^{\mathrm{F}^*}}{\hat{\mathbf{x}}_u^{\mathrm{F}} \odot \hat{\mathbf{x}}_u^{\mathrm{F}^*} + \boldsymbol{\tau}_u^{x_{\mathrm{F}}}}$ and $\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{c_{\mathrm{F}}} = \frac{\overrightarrow{\boldsymbol{\tau}}_{q,u,n_a,s_a}^{d_{\mathrm{P}}}}{\hat{\mathbf{x}}_u^{\mathrm{F}} \odot \hat{\mathbf{x}}_u^{\mathrm{F}^*} + \boldsymbol{\tau}_u^{x_{\mathrm{F}}}}$.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

REFERENCES

[1] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-things-based smart cities: Recent advances and challenges," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 16–24, Sep. 2017.

[2] 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, Oct. 2016.

[3] X. Lin, S. Cioni, G. Charbit, N. Chuberre, S. Hellsten, and J.-F. Boutillon, "On the path to 6G: Embracing the next wave of low earth orbit satellite access," *IEEE Commun. Mag.*, vol. 59, no. 12, pp. 36–42, Dec. 2021.

[4] J. Jiao, S. Wu, R. Lu, and Q. Zhang, "Massive access in space-based internet of things: Challenges, opportunities, and future directions," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 118–125, Oct. 2021.

[5] Z. Gao, X. Zhou, J. Zhao, J. Li, C. Zhu, C. Hu, P. Xiao, S. Chatzinotas, D. W. K. Ng, and B. Ottersten, "Grant-free NOMA-OTFS paradigm: Enabling efficient ubiquitous access for LEO satellite internet-of-things," *IEEE Network*, vol. 37, no. 1, pp. 18–26, Apr. 2023.

[6] S. Liu, Z. Gao, Y. Wu, D. W. Kwan Ng, X. Gao, K.-K. Wong, S. Chatzinotas, and B. Ottersten, "LEO satellite constellations for 5G and beyond: How will they reshape vertical domains?" *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 30–36, Jul. 2021.

[7] J. Choi, J. Ding, N.-P. Le, and Z. Ding, "Grant-free random access in machine-type communication: Approaches and challenges," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 151–158, Feb. 2022.

[8] 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.

[9] Z. Gao, M. Ke, Y. Mei, L. Qiao, S. Chen, D. W. K. Ng, and H. V. Poor, "Compressive-sensing-based grant-free massive access for 6G massive communication," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7411–7435, Mar. 2024.

[10] 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, Mar. 2018.

[11] 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, May 2021.

[12] 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, Jan. 2020.

[13] G. Sun, Y. Li, X. Yi, W. Wang, X. Gao, L. Wang, F. Wei, and Y. Chen, "Massive grant-free OFDMA with timing and frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3365–3380, May 2022.

[14] 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.

[15] Z. Zhang, Y. Li, C. Huang, Q. Guo, L. Liu, C. Yuen, and Y. L. Guan, "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.

[16] Y. Zuo, M. Yue, M. Zhang, S. Li, S. Ni, and X. Yuan, "OFDM-based massive connectivity for LEO satellite internet of things," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8244–8258, Nov. 2023.

[17] Y. Li, S. Chen, and W. Meng, "Asynchronous grant-free massive access with Doppler frequency offset in NTN," *IEEE Wireless Commun. Lett.*, vol. 13, no. 2, pp. 367–371, Feb. 2024.

[18] Y. Ma, G. Ma, N. Wang, Z. Zhong, J. Yuan, and B. Ai, "Enabling OTFS-TSMA for smart railways mMTC over LEO satellite: A differential Doppler shift perspective," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 4799–4814, Mar. 2023.

[19] B. Shen, Y. Wu, J. An, C. Xing, L. Zhao, and W. Zhang, "Random access with massive MIMO-OTFS in LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2865–2881, Oct. 2022.

[20] X. Zhou, K. Ying, Z. Gao, Y. Wu, Z. Xiao, S. Chatzinotas, J. Yuan, and B. Ottersten, "Active terminal identification, channel estimation, and signal detection for grant-free NOMA-OTFS in LEO satellite internet-of-things," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2847–2866, Apr. 2023.

[21] X. Wang, W. Shen, C. Xing, J. An, and L. Hanzo, "Joint Bayesian channel estimation and data detection for OTFS systems in LEO satellite communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4386–4399, Jul. 2022.

[22] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, May 2017, pp. 1–6.

[23] 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.

[24] I. Leyva-Mayorga, B. Soret, and P. Popovski, "Inter-plane inter-satellite connectivity in dense LEO constellations," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3430–3443, Jan. 2021.

[25] M. Y. Abdelsadek, G. Karabulut-Kurt, H. Yanikomeroglu, P. Hu, G. Lamontagne, and K. Ahmed, "Broadband connectivity for handheld devices via LEO satellites: Is distributed massive MIMO the answer?" *IEEE Open J. Commun. Soc.*, vol. 4, pp. 713–726, Mar. 2023.

[26] K. Ying, Z. Gao, S. Chen, M. Zhou, D. Zheng, S. Chatzinotas, B. Ottersten, and H. V. Poor, "Quasi-synchronous random access for massive MIMO-based LEO satellite constellations," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1702–1722, May 2023.

[27] G. Leus, "On the estimation of rapidly time-varying channels," in *Proc. IEEE Euro. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 2227–2230.

[28] 3rd Generation Partnership Project, "3rd generation partnership project; technical specification group radio access network; study on new radio (NR) to support non-terrestrial networks (Release 15)," *3GPP TR 38.811 V15.4.0*, Sep. 2020.

[29] A. G. Kanatas and A. D. Panagopoulos, *Radio Wave Propagation and Channel Modeling for Earth-Space Systems*. CRC Press, 2016.

[30] M. K. Tsatsanis and G. B. Giannakis, "Modelling and equalization of rapidly fading channels," *Int. J. Adapt. Control Signal Process.*, vol. 10, no. 2-3, pp. 159–176, Mar. 1996.

[31] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2226–2238, May 2007.

[32] Y. Liu, Y. L. Guan, and D. G. G., "Near-optimal BEM OTFS receiver with low pilot overhead for high-mobility communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3392–3406, 2022.

[33] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, Nov. 2021.

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

[35] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2001.

[36] J. P. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[37] S. Bhandari, T. X. Vu, and S. Chatzinotas, "User-centric flexible resource management framework for LEO satellites with fully regenerative payload," *IEEE J. Sel. Areas Commun.*, pp. 1–1, Feb. 2024 (Early Access).

[38] G. D. Surabhi and A. Chockalingam, "Low-complexity linear equalization for OTFS modulation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 330–334, Feb. 2020.

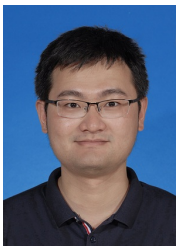
[39] S. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Pearson, 1993.

[40] F. Huang, Q. Guo, Y. Zhang, and Y. Zakharov, "Message passing-based joint channel estimation and signal detection for OTFS with superimposed pilots," Sep. 2023. [Online]. Available: https://arxiv.org/abs/2309.08177.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal on Selected Areas in Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2024.3460064

Boxiao Shen (Student Member, IEEE) received the B.E. degree in communication engineering from 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. From 2019 to 2021, he was a visiting scholar at the Future Network of Intelligence Institute (FNii), the Chinese University of Hong Kong, Shenzhen, China. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering in the Shanghai Jiao Tong University (SJTU). His research interests include satellite communications, massive machine-type communications, information theory, and the application of artificial intelligence on wireless communications.



Yongpeng Wu (S'08–M'13–SM'17) received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in July 2007, the Ph.D. degree in communication and signal processing with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in November 2013. Dr. Wu is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was senior research fellow with Institute for Communications Engineering, Technical University of Munich, Germany and the Humboldt research fellow and the senior research fellow with Institute for Digital Communications, University Erlangen-Nürnberg, Germany. During his doctoral studies, he conducted cooperative research at the Department of Electrical Engineering, Missouri University of Science and Technology, 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 for IEEE International Conference on Communications (ICC) 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants for IEEE Communication Theory Workshop 2016, the Excellent Doctoral Thesis Awards of China Communications Society 2016, the Exemplary Editor Award of IEEE Communication Letters 2017, Young Elite Scientist Sponsorship Program by CAST 2017, and Excellent Youth Science Fund Project of National Natural Science Foundation of China 2021. He has been the lead guest editor of IEEE Journal of Selected Topics in Signal Processing, IEEE Journal on Selected Areas in Communications and IEEE Wireless Communications. He is currently an editor of IEEE Wireless Communications, and was an Editor of IEEE Transactions on Communications and IEEE Communications Letters. He has been symposium chairs of various conferences, including Globecom, ICC, VTC, and PIMRC, etc.

Shiqi Gong (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Beijing Institute of Technology, China, in 2014 and 2020, respectively. From 2021 to 2022, she was a postdoctoral fellow with the State Key Laboratory of Internet of Things for Smart City, University of Macau, China. She is currently an associate professor with the School of Cyberspace Science and Technology, Beijing Institute of Technology. Her research interests are in the area of intelligent reflecting surface, physical layer security, resource allocation, and convex optimization. She was a recipient of the Best Ph.D. Thesis Award of Beijing Institute of Technology in 2020.

Heng Liu (Member, IEEE) received the B.E. degree and the Ph.D. degree from the Beijing Institute of Technology, China, in 2017 and 2022, respectively. From 2022 to 2024, he was a postdoctoral fellow with the School of Cyberspace Science and Technology, Beijing Institute of Technology. He is currently an assistant professor with the School of Information and Electronics, Beijing Institute of Technology. His main research interests include signal processing, reconfigurable intelligent surface and convex optimization.

Björn Ottersten (S'87–M'89–SM'99–F'04) received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Ottersten has been Head of the Department for Signals, Sensors, and Systems, KTH, and Dean of the School of Electrical Engineering, KTH. He is the founding Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. He is a recipient of the IEEE Signal Processing Society Technical Achievement Award, the EURASIP Group Technical Achievement Award, and the European Research Council (ERC) advanced research grant twice. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and 9 IEEE conference papers best paper awards. He has been a board member of IEEE Signal Processing Society, the Swedish Research Council and currently serves of the boards of EURASIP and the Swedish Foundation for Strategic Research as well as on the ERC Scientific Council. Dr. Ottersten has served as Editor in Chief of EURASIP Signal Processing, and acted on the editorial boards of IEEE Transactions on Signal Processing, IEEE Signal Processing Magazine, IEEE Open Journal for Signal Processing, EURASIP Journal of Advances in Signal Processing and Foundations and Trends in Signal Processing. He is a fellow of IEEE, EURASIP, and AAIA.

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.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on November 15,2024 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.