Tensor-Based Unsourced Random Access for LEO Satellite Internet of Things

Ziqi Kang, Dongxuan He, *Member, IEEE*, Hua Wang, *Member, IEEE*, Weijie Yuan, *Member, IEEE*, and Tony Q.S. Quek, *Fellow, IEEE*

***Abstract*—With the rapid expansion of Internet of Things (IoT) applications, the demand of wide coverage and massive connectivity is inevitable. In this context, this paper investigates massive unsourced random access (URA) paradigm for low earth orbit (LEO) satellite IoT applications, focusing on device separation and signal detection. By exploiting the structured Grassmannian constellation to generate the codebook, a tensor-based URA transmission scheme is provided, which models the separation and detection problem as a general canonical polyadic (CP) decomposition. Then, to evaluate the access capability of our considered URA scheme, a comprehensive uniqueness analysis considering both sufficient conditions and necessary conditions is presented. Accordingly, an efficient generalized line-search-accelerated alternating least squares (GLSA-ALS) method is proposed to conduct the device separation and signal detection, which can avoid a large number of inverse computations for large-scale matrices. To be specific, with the help of the relaxation factors during the iteration, our proposed method can converge at a fast speed with negligible performance loss, which facilitates a better trade-off between the detection accuracy and computational complexity. Furthermore, depending on the demand of a specific application scenario, the flexible selection of relaxation factors enables the proposed method to be compatible to the classical ALS method, which can enhance the performance at the cost of additional complexity. Finally, relying on the maximum likelihood (ML)-based detection approach, the message list transmitted by active devices from one common codebook can be recovered. Simulation results demonstrate that the proposed GLSA-ALS method outperforms the state-of-the-art methods for practical LEO satellite IoT applications.**

***Index Terms*—Internet of Things, LEO satellite, unsourced random access, tensor decomposition, ALS, line search.**

I. Introduction

INTERNET of Things (IoT) has been regarded as a typical massive machine type communication (mMTC) application, which is playing a key role in various industrial fields [1]-[3]. Compared to the conventional human-type communication, the wide coverage and massive connectivity are the main distinctive features of mMTC. To be specific, IoT has been projected to serve in remote areas, such as oceans, deserts, and forests. However, restricted by the costs and geographical limitations, the wireless coverage is unsatisfactory through the current terrestrial networks. Therefore, by perceiving the wide coverage effect of low earth orbit (LEO) satellites, LEO satellite IoT has been a popular trend in IoT development, where the IoT devices are supported the direct access of satellite. On the other hand, the number of IoT devices is predicted to reach 500 billion by 2030, which requires more efficient access techniques to support the applications [4],[5]. Therefore, the efficient and reliable access of LEO satellite IoT devices is urgently needed.

In response to this demand, grant-free random access (GF-RA) has been attracting widespread attention, which has been regarded as a promising access technique [6]. Compared to the grant-based random access (GB-RA), which requires multiple handshakes and excessive signaling overhead, GF-RA successively improves the access efficiency by avoiding the complex access request especially in large-scale access scenarios [7]. Specifically, depending on whether it is necessary to detect the identity (ID) of active devices, GF-RA can be further divided into two categories, namely, sourced random access (SRA) and unsourced random access (URA) [8]. Among them, SRA can realize active device detection (ADD) based on non-orthogonal pilot sequence and channel estimation (CE), where the ADD problem is modeled as a sparse matrix reconstruction problem with the help of compressed sensing (CS) [9],[10]. For instance, orthogonal matching pursuit (OMP) was utilized to conduct CE efficiently for the subsequent ADD, which shows extremely fast convergence speed [9]. To enhance the reconstruction accuracy of OMP, the message passing algorithm was introduced into CS, facilitating a novel approximate message passing (AMP) algorithm with improved CE and ADD performance [10]. Accordingly, CS-based methodology has been proven to be still effective for SRA systems in the LEO satellite IoT scenarios [11],[12]. By combining the orthogonal time frequency space (OTFS) modulation with CS-based methods, the authors in [11] and [12] proposed effective joint CE and ADD which can adapt to the fast time-varying channel in the LEO satellite communication. However, restricted by the complexity of CS-based methods and unique pilot sequence allocation demand on each device, the performance of SRA is still limited.

Indeed, due to the fact that the pilot overhead is reduced and the corresponding ID detection through the ADD is avoided, URA can enable lower access latency and larger number of device access [13]. The critical issue of URA is to recover the message list based on one common codebook shared by all active devices, where an unknown permutation of the

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFE0200404; in part by the National Research Foundation, Singapore and Infocomm Media Development Authority under its Future Communications Research & Development Programme. (*Corresponding author: Dongxuan He*)

Ziqi. Kang, Dongxuan. He and Hua. Wang are with Beijing Institute of Technology, Beijing 100081, China (e-mails: ziqi_kang@bit.edu.cn; dongxuan_he@bit.edu.cn; wanghua@bit.edu.cn)

Weijie Yuan is with the School of Automation and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: yuanwj@sustech.edu.cn)

T. Q. S. Quek is with the Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg)

TABLE I
A COMPARISON OF THE RELATED LITERATURE WITH OUR WORK

Reference	System Model	Category	Fundamental	Proposed algorithm	Benchmark
[11]	MIMO-OTFS + LEO satellite IoT	SRA	CS	OMP-based	oracle-LS
[12]	MIMO-OTFS + LEO satellite IoT	SRA	CS	ConvSBL-GAMP	SBL-GAMP / GMMV-AMP
[14]	Massive MIMO	SRA + URA	CS	BiG-AMP + SIC	Coherent detection
[17]	Single antenna	URA	CS	CCS + tree-based	Random coding / SIC
[18]	Single antenna + quasi-static fading	URA	MAP	Alternating BP	ALOHA
[19]	Slotted ALOHA + asynchronous fading	URA	CS + MAP	OMP + SIC	SIC (synchronous)
[20]	OFDMA + LEO satellite IoT	URA	CS	PD + JD	DOT / DOCT
[25]	Block fading	URA	Tensor decomposition	GN-based + ML	Random coding
[26]	RIS-aided + blocked direct links	URA	Tensor decomposition + SBL	CTAD	ALS + CVX
[27]	Block fading + high-mobility	URA	Tensor decomposition	BTD + GN-based	TBM
Our work	LEO satellite IoT	URA	Tensor decomposition	GLSA-ALS	GN-based / ALS

estimated messages is inevitable [14]. On the basis of this fundamental structure, the main challenge in URA scheme stems from the huge size of codebook, which grows exponentially as the length of payload data increases [15]. To tackle this problem, the authors in [16] designed a low-complexity coding scheme for URA, which provided an innovative idea of segmenting the transmission period into multiple sub-blocks to transmit the codewords. To enhance the efficiency, a coded CS (CCS)-based approach was proposed to further reduce the computational complexity of decoding process by leveraging the advances of CS theory [17]. However, despite these studies contributing significantly with respect of codebook design, the considered communication scenario is limited to the Gaussian multiple access channel, thus hindering its practical implementation. Therefore, the authors in [18] derived an approximate performance bound under the fading channels with the help of ideal coding and decoding framework. Similarly, the authors in [19] preliminarily analyzed the impact of block misalignment under the asynchronous and fading channels. Furthermore, for the more complicated LEO satellite IoT scenario, the authors in [20] proposed an OFDMA-based CCS approach, where the designed inner decoder fairly controls its computational complexity and accelerates the convergence speed by the information interaction with the outer decoder. In summary, these studies demonstrate the significance of a competitive design of codebook and decoder based on realistic channel conditions, which plays an important role in facilitating efficient device access in practical applications.

In recent years, by comprehensively utilizing the inherent-structured characteristics of channels or signals, the attractive methods based on tensor decomposition have been proven to be capable of effectively handling variant estimation and detection problems [21]-[24]. In accordance with the existing outstanding tensor-based studies, the research on novel codebooks for URA based on tensor modeling was also investigated, where the device separation and signal detection can be conducted without the pilot sequence [25]. Besides, the proposed paradigm can be well applied to fading channels and take advantage of the spatial diversity by accommodating multiple antenna receivers. At the receiver side, as one of the most representative methods of non-linear least square approaches, the Gauss-Newton (GN)-based method was uti-

This article has been accepted for publication in IEEE Transactions on Wireless 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/TWC.2025.3612505

lized to compute the canonical polyadic (CP) decomposition, which can return relatively accurate estimation of factor matrices for the detection. Inspired by this, the authors in [26] proposed a reconfigurable intelligent surface (RIS)-aided URA approach, where the detection performance can be enhanced under the complex RIS channel condition by combining the tensor modeling and sparse Bayesian learning (SBL). Moreover, to fit for the time-varying channels in high-mobility scenarios, a modified tensor-based scheme on the basis of block term decomposition (BTD) was designed, where the structured tensor paradigm effectively improved the detection performance in shorter coherence blocks [27]. However, the high computational complexity during the detection has not been resolved yet, which limits the practicality of tensor-based URA scheme, thus making it impractical to LEO satellite IoT applications due to the low signal to noise ratio (SNR) brought by the long-distance transmission. For clarity, the comparison of the aforementioned related works is summarized in TABLE I.

Against this background, this paper intends to design an efficient tensor-based URA scheme via the CP decomposition for LEO satellite IoT applications. The contributions of this paper are summarized as follows:

1) A structured URA scheme based on tensor format is established for the LEO satellite IoT scenario. By fully exploiting the Grassmannian constellation for modulation, the device separation and signal detection are modeled as a general CP decomposition under higher-order cases, where the overhead of pilot allocation can be well saved.
2) The uniqueness analysis of the proposed tensor-based model is elaborated to provide the theoretical support for the access capability of the URA system. Particularly, the conditions for unique decomposition are discussed from both sufficient and necessary aspects, demonstrating the boundary for the number of supported devices in detail, which remains open in the studies on LEO satellite IoT scenario.
3) An effective generalized line-search-accelerated alternating least squares (GLSA-ALS) method is proposed to solve the CP decomposition, where the efficiently recovered factor matrices are utilized to facilitate the device separation and signal detection. Specifically, the designed acceleration optimizes the iterative update of factor matrices estimation to reach faster convergence, which can be compatible to the classical ALS method with the help of relaxation factor selection to satisfy a demand of more accuracy.

The rest of this paper is organized as follows. In Section II, we introduce the considered URA transmission scheme in LEO satellite IoT scenarios. Section III provides the CP-based tensor modeling of the received signal, as well as the uniqueness analysis of the corresponding decomposition problem. After that, by analyzing the existing conventional solutions of the proposed scheme, a GLSA-ALS method is designed to obtain the estimation of factor matrices for the subsequent signal detection in section IV. Simulation results are presented in Section V, followed by conclusions in Section VI.

Fig. 1. An illustration of URA system in LEO satellite IoT scenarios.

Notations: Vectors and matrices are denoted by lowercase and uppercase boldface letters, respectively. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the l_2-norm and Frobenius norm, respectively. The operations of transpose, conjugate, conjugate transpose and pseudo inverse are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^\dagger$, respectively. $\mathbf{I}_M$ is the identity matrix with dimension $M \times M$. $[\![\cdot]\!]$ denotes the Kruskal operator. $[\mathbf{x}]_l, [\mathbf{X}]_{l,m}, [\mathcal{X}]_{l,m,n}, \mathbf{X}_l, [\mathbf{X}]_{:,m:n}$ denote the l-th entry of $\mathbf{x}$, element (l,m) of $\mathbf{M}$, element (l,m,n) of third-order tensor $\mathcal{X}$, the l-th column of $\mathbf{X}$ and submatrix of $\mathbf{X}$ from the m-th to the n-th columns, respectively. $\lceil\cdot\rceil$ and $\max(\cdot)$ denote the operator of taking the ceiling and maximization, respectively. The real part of a complex value is denoted by $\mathrm{Re}(\cdot)$, and $j = \sqrt{-1}$. $|\mathbb{A}|_c$ is the cardinal number of the set $\mathbb{A}$. $\mathbf{\Lambda}(\mathbf{x})$ denotes the diagonal matrix formed by $\mathbf{x}$. $\otimes, \odot, *$, and $\circ$ denote Kronecker, Khatri-Rao, Hadamard, and outer product, respectively. Moreover, given N matrices $\mathbf{X}_i \in \mathbb{C}^{L_i \times R}$, we consider the following products:

$$\bigodot_{i=N}^{1} \mathbf{X}_i = \mathbf{X}_N \odot \cdots \odot \mathbf{X}_i \odot \cdots \odot \mathbf{X}_1, \tag{1}$$

$$\bigodot_{i=1}^{N} \mathbf{X}_i = \mathbf{X}_1 \odot \cdots \odot \mathbf{X}_i \odot \cdots \odot \mathbf{X}_N, \tag{2}$$

$$\bigotimes_{i=1}^{N} \mathbf{X}_i = \mathbf{X}_1 \otimes \cdots \otimes \mathbf{X}_i \otimes \cdots \otimes \mathbf{X}_N. \tag{3}$$

II. System Model

A. Channel Model in LEO Satellite IoT Scenarios

As shown in Fig. 1, we consider a LEO satellite IoT system including U_a IoT devices with single antenna and a LEO satellite with M antennas [28]-[31]. The variables are detailed in TABLE II. According to the transmission characteristics of LEO satellite communication, the channel between the u-th

TABLE II
VARIABLE LIST OF CHANNEL MODEL

Notation	Definition
α_u	Satellite receive antenna gain
c	Speed of light
f_c	Carrier frequency
d_s	Diameter of circular antenna array
θ_u	Off-axis of the satellite boresight
β_u	Large-scale fading factor
d_0	Propagation distance
G_u	Device transmit antenna gain
W	Carrier bandwidth
T_0	Received noise temperature
κ	Boltzmann's constant
r_u	Rain attenuation coefficient
γ_u	Rician factor

IoT device and the LEO satellite, $u \in \{1, \cdots, U_a\}$, can be written by [32]

$$\mathbf{h}_u = \alpha_u \left(\sqrt{\frac{\gamma_u \beta_u}{\gamma_u + 1}} \mathbf{h}_u^{\text{LOS}} + \sqrt{\frac{\beta_u}{\gamma_u + 1}} \mathbf{h}_u^{\text{NLOS}} \right), \tag{4}$$

where $\mathbf{h}_u^{\text{LOS}}$ and $\mathbf{h}_u^{\text{NLOS}}$ represent the line-of-sight (LOS) and non-line-of-sight (NLOS) exponent in the channel model, respectively, and the satellite receive antenna gain α_u is given by [33]

$$\alpha_u = \frac{J_1(\phi_u)}{2\phi_u} + 36\frac{J_3(\phi_u)}{\phi_u^3}, \tag{5}$$

with

$$\phi_u = \frac{\pi d_s f_c}{c} \sin(\theta_u). \tag{6}$$

Besides, the large-scale fading factor β_u is given by [34],[35]

$$\beta_u = (\frac{c}{4\pi f_c d_0})^2 \cdot \frac{G_u}{\kappa W T_0} \cdot \frac{1}{r_u}, \tag{7}$$

where $(\frac{c}{4\pi f_c d_0})^2$ is the free space loss, and the power gain of rain attenuation effect in dB commonly follows log-normal random distribution $\ln(r_{\text{dB}}) \sim \mathcal{CN}(\mu_r, \sigma_r^2)$ with mean μ_r and variance σ_r^2. Moreover, since the transmit elevation angles to the LEO satellite can be regarded as invariant, the LOS component $\mathbf{h}_u^{\text{LOS}} \in \mathbb{C}^{M\times 1}$ is considered as a constant over a relatively long time. For the NLOS component of the channel, $\mathbf{h}_u^{\text{NLOS}} \in \mathbb{C}^{M\times 1}$ commonly follows a independent and identically distributed (i.i.d.) complex Gaussian distribution $\mathbf{h}_u^{\text{NLOS}} \sim \mathcal{CN}(\mathbf{0}, (\sigma_u^{\text{NLOS}})^2 \mathbf{I}_M)$. Hence, as the combination of its LOS component and NLOS component, the LEO satellite channel $\mathbf{h}_u \in \mathbb{C}^{M\times 1}$ follows the complex Gaussian distribution $\mathbf{h}_u \sim \mathcal{CN}(\alpha_u \sqrt{\frac{\gamma_u \beta_u}{\gamma_u+1}} \mathbf{h}_u^{\text{LOS}}, \frac{\alpha_u^2 \beta_u}{\gamma_u+1} (\sigma_u^{\text{NLOS}})^2 \mathbf{I}_M)$ [36].

B. Transmission Scheme of Unsourced Random Access

Following the URA paradigm [15], the active IoT devices transmit their payload information simultaneously by sharing the common constellation. Therefore, according to the channel and signal propagation characteristics of LEO satellite IoT scenarios, the received signal $\mathbf{Y} \in \mathbb{C}^{L\times M}$ at LEO satellite can be expressed as

$$\mathbf{Y} = \sum_{u=1}^{U_a} \sqrt{P_u} \mathbf{s}_u \mathbf{h}_u^H + \mathbf{Z}, \tag{8}$$

where P_u is the transmit power of the u-th device, $\mathbf{s}_u \in \mathscr{C} \subset \mathbb{C}^L$ denotes the sequence of complex baseband symbols transmitted by the u-th device over L channel uses and selected from the same constellation $\mathscr{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{2^B}\}$, B is the information bits, and $\mathbf{Z} \in \mathbb{C}^{L\times M}$ is the additive white Gaussian noise (AWGN) with mean zero and variance σ_n^2 [25]. Due to the property of Kronecker product, the received signal in matrix form can be vectorized as

$$\mathbf{y} = vec(\mathbf{Y}) = \sum_{u=1}^{U_a} \sqrt{P_u} \mathbf{s}_u \otimes \mathbf{h}_u + \mathbf{z}, \tag{9}$$

where $\mathbf{z} = vec(\mathbf{Z})$ is the column vectorization of $\mathbf{Z}$, and $\mathbf{y}, \mathbf{z} \in \mathbb{C}^{LM\times 1}$. Moreover, the outperformed multi-dimensional constellation structure is constructed on the basis of tensor-based modulation format, which can be decomposed into multiple sub-constellations [25]-[27]. Specifically, let $\mathscr{C}_n \subset \mathbb{C}^{L_n}$ denote the n-th sub-constellation of $\mathscr{C}$ over L_n channel use, where L_n is the n-th part of the factorized total channel use, i.e., $L = \prod_{n=1}^{N} L_n$, $N \geq 2$, $L_n \geq 2$ [27]. Then, the original constellation can be regarded as being comprised of all possible combinations of elements selected from each $\mathscr{C}_n$, given by

$$\mathscr{C} = \{\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_N, \mathbf{x}_1 \in \mathscr{C}_1, \ldots, \mathbf{x}_N \in \mathscr{C}_N\}. \tag{10}$$

Therefore, $\mathbf{s}_u$ can be accordingly factorized as follows

$$\mathbf{s}_u = \mathbf{x}_{1,u} \otimes \mathbf{x}_{2,u} \otimes \cdots \otimes \mathbf{x}_{N,u}, \tag{11}$$

where $\mathbf{x}_{n,u} \in \mathscr{C}_n \subset \mathbb{C}^{L_n}$ represents the transmitted sub-constellation points of the u-th device. By substituting (11) into (9), the received signal can be rewritten as

$$\mathbf{y} = \underbrace{\sum_{u=1}^{U_a} \sqrt{P_u} \mathbf{x}_{1,u} \otimes \mathbf{x}_{2,u} \otimes \cdots \otimes \mathbf{x}_{N,u} \otimes \mathbf{h}_u}_{\triangleq \mathbf{y}_0} + \mathbf{z}, \tag{12}$$

where $\mathbf{y}_0$ denotes the noise-free part of the received signal.

III. TENSOR MODELING AND UNIQUENESS ANALYSIS

In this section, we firstly derive the CP-based tensor form of the received signal. Then, the conditions that guarantee the uniqueness of the CP decomposition are analyzed.

A. Tensor Modeling of The Receiver

With the help of the property of outer product and vectorization, the signal model presented in (12) can be equivalently reformulated in a rank-one tensor form, given by [26],[32]

$$\mathcal{Y} = \mathcal{Y}_0 + \mathcal{Z} = \sum_{u=1}^{U_a} \mathbf{x}_{1,u} \circ \mathbf{x}_{2,u} \circ \cdots \circ \mathbf{x}_{N,u} \circ \mathbf{h'}_u + \mathcal{Z}, \quad (13)$$

where $\mathcal{Y}, \mathcal{Z} \in \mathbb{C}^{L_1 \times L_2 \times \cdots \times L_N \times M}$ are the received signal and the AWGN in tensor space, respectively, $\mathcal{Y}$ satisfies $\mathbf{y} = vec(\mathcal{Y})$, $\mathcal{Y}_0$ is the noise-free part, $\mathbf{h'}_u = \sqrt{P_u}\mathbf{h}_u$ denotes the equivalent channel vector. For convenience, we will drop the prime of $\mathbf{h'}_u$ in the following discussion. Consequently, given the estimated $\hat{\mathbf{h}}_u$, the multi-user detection can be formulated as

$$\{\hat{\mathbf{x}}_{n,u}\} = \underset{\mathbf{x}_{n,u} \in \mathscr{C}_n}{\operatorname{argmin}} \left\| \mathcal{Y} - \sum_{u=1}^{U_a} \mathbf{x}_{1,u} \circ \mathbf{x}_{2,u} \circ \cdots \circ \mathbf{x}_{N,u} \circ \hat{\mathbf{h}}_u \right\|_F^2, \quad (14)$$

where the symbols are mapped from Grassmannian constellations and can be segmented as (10). Particularly, benefiting from the specific decoding methodology, $\hat{\mathbf{x}}_{n,u}$ can be detected without an accurate CE beforehand. Therefore, the multi-user detection problem can be equivalently relaxed to simultaneously obtain $\{\hat{\mathbf{x}}_{n,u}, \hat{\mathbf{h}}_u\}$, which satisfies

$$\begin{aligned} \min_{\hat{\mathbf{x}}_{n,u} \in \mathscr{C}_n, \hat{\mathbf{h}}_u \in \mathbb{C}^M} & \left\| \hat{\mathbf{x}}_{1,u} \circ \hat{\mathbf{x}}_{2,u} \circ \cdots \circ \hat{\mathbf{x}}_{N,u} \circ \hat{\mathbf{h}}_u \right. \\ & \left. - \mathbf{x}_{1,u} \circ \mathbf{x}_{2,u} \circ \cdots \circ \mathbf{x}_{N,u} \circ \mathbf{h}_u \right\|_F^2. \end{aligned} \quad (15)$$

In fact, the device separation and signal detection problem modeling in (14)-(15) can be handled by a typical CP decomposition, where $\hat{\mathbf{x}}_{n,u}$ can be obtained by the accurate reconstruction of factor matrices. Accordingly, (13) can be equivalently expressed by the composition of factor matrices, given by

$$\begin{aligned} \mathcal{Y} &= \sum_{u=1}^{U_a} \mathbf{x}_{1,u} \circ \mathbf{x}_{2,u} \circ \cdots \circ \mathbf{x}_{N,u} \circ \mathbf{h}_u + \mathcal{Z} \\ &= [\![\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N, \mathbf{H}]\!] + \mathcal{Z}, \end{aligned} \quad (16)$$

where the factor matrices are denoted by

$$\mathbf{X}_n \triangleq [\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,U_a}] \in \mathbb{C}^{L_n \times U_a}, \quad (17)$$

$$\mathbf{H} \triangleq [\mathbf{h}_1, \ldots, \mathbf{h}_{U_a}] \in \mathbb{C}^{M \times U_a}. \quad (18)$$

B. Uniqueness Analysis of N-Order Tensors

Considering the impact of the number of active devices on signal detection performance, the conditions of essential uniqueness should be clarified and satisfied, which is mostly related to the constraints of tensor rank. In addition, despite the fact that most existing works only provides the sufficient conditions limited to the third-order tensors, the necessary conditions are equally important to present for the selection of system parameters. Moreover, due to the different levels of tightness, the conditions that are derived to satisfy the uniqueness of CP decomposition are not unique. Therefore, based on different theoretical basis, a group of sufficient and necessary conditions for the uniqueness are given as follows.

1) Sufficient Condition: The most widely adopted sufficient condition is the Kruskal's condition derived from the specific third-order case, which is exploited here by extending the condition to $(N+1)$-order cases, given by [37],[38]

$$\sum_{n=1}^{N} rank(\mathbf{X}_n) + rank(\mathbf{H}) \geq 2U_a + N. \quad (19)$$

For the sake of simplicity, we denote $\mathbf{X}_{N+1} \triangleq \mathbf{H}$ and rearrange (19) to obtain the corresponding rank constraint, given by

$$U_a \leq \frac{\sum_{n=1}^{N+1} rank(\mathbf{X}_n) - N}{2}. \quad (20)$$

As a sufficient condition, the constraint can be reasonably relaxed to provide a higher upper bound. Therefore, an alternative sufficient condition is considered based on the geometrical concept of tangential weak defectivity, given by [39]

$$\begin{aligned} U_a &\leq \left\lceil \frac{\Pi_{n=1}^{N} L_n M}{1 + \Sigma_{n=1}^{N}(L_n - 1) + (M-1)} \right\rceil - 1 \\ &= \left\lceil \frac{LM}{M + \Sigma_{n=1}^{N}(L_n - 1)} \right\rceil - 1, \end{aligned} \quad (21)$$

where a more relaxed rank constraint is obtained to guarantee the tensor identifiability in the cases of larger tensor space.

2) Necessary Condition: Considering the mode-n unfolding of $\mathcal{Y}_0$, we have

$$\begin{aligned} \mathbf{Y}_{(n)} &= \mathbf{X}_n \left(\mathbf{X}_{N+1} \odot \cdots \odot \mathbf{X}_{n+1} \odot \mathbf{X}_{n-1} \odot \cdots \odot \mathbf{X}_1 \right)^T \\ &= \mathbf{X}_n \left(\bigodot_{i=N+1, i \neq n}^{1} \mathbf{X}_i \right)^T. \end{aligned} \quad (22)$$

Then, on the basis of Liu's condition [40], a general necessary condition for N-order cases can be given by

$$U_a = \min_{n=1,\ldots,N+1} rank \left(\bigodot_{i=1, i \neq n}^{N+1} \mathbf{X}_i \right). \quad (23)$$

By exploiting the property of Khatri-Rao product, i.e., $rank(\mathbf{X}_1 \odot \mathbf{X}_2) \leq rank(\mathbf{X}_1) \cdot rank(\mathbf{X}_2)$, the upper bound of U_a can be given by

$$U_a \leq \min_{n=1,\ldots,N+1} \left(\prod_{i=1, i \neq n}^{N+1} rank(\mathbf{X}_i) \right), \quad (24)$$

where the rank constraint sets a limit from the perspective of necessary condition to avoid non-unique decomposition. In summary, by comprehensively considering the sufficient and necessary conditions, the credible number range of active devices that can be identified by the tensor modeling can be obtained.

IV. Low Complexity Device Separation and Signal Detection via CP Decomposition

In this section, we discuss the solutions based on GN and ALS, respectively. Then, an efficient GLSA-ALS method is designed to handle the CP decomposition. Based on the results of GLSA-ALS, the maximum likelihood (ML)-based signal detection is presented. Besides, the computational complexity analysis is provided.

A. Proposed GLSA-ALS Method

As the solution of the existing works about tensor-based URA, GN method has shown outstanding performance in terms of the accuracy of tensor decomposition [25],[27]. By computing the Jacobian and Hessian matrix of the objective residual tensor, GN method can be viewed as a refinement of ALS methods, where the performance improvement is more obvious as the swamps arise [41]. Specifically, the swamps caused by the collinear components of factor matrices will significantly slow down the convergence speed, which is more likely to occur in large-scale cases. However, due to the inherent inverse operation of large-scale Hessian matrices for iterative updating in GN method, the detection accuracy may be deteriorated because of the near singular feature, where the total computational complexity of factor matrices estimation is also unendurable for practical applications [42]. In contrast, the conventional ALS has the lowest computational complexity among the methods of CP decomposition. However, ALS method is not robust to the swamps, which will make the convergence hard to be guaranteed. Therefore, we propose a GLSA-ALS method for the device separation and signal detection problem of the tensor-based URA to achieve better trade-off between performance and convergence.

In the proposed GLSA-ALS method, we focus on accelerating the convergence to reduce the number of iterations. Specifically, GLSA-ALS can be divided into three stages, i.e., initialization, acceleration and estimation, which will be discussed in detail as follows.

1) Iterative Updates Initialization: As one of the most typical nonlinear least squares methods, GN method allows the initialization of factor matrices to be randomly initialized or all zero values. Comparatively, ALS-based methods require specific initialization to ensure the convergence speed. Besides, in contrast to the original ALS method, two initial values are demanded for starting the proposed GLSA-ALS defined as $\hat{\mathbf{X}}_n^{(\eta-1)}$ and $\hat{\mathbf{X}}_n^{(\eta-2)}$, where η denotes the current iteration, $n = 1, \ldots, N+1$. For this purpose, we design to perform $2(N+1)$ iterations of updates to obtain the above two initial values, where the updating rule is given by

$$\begin{aligned}\hat{\mathbf{X}}_n^{(\eta)} &= \mathbf{Y}_{(n)} \left(\left(\hat{\mathbf{X}}_{N+1}^{(\eta-1)} \odot \cdots \odot \hat{\mathbf{X}}_{n+1}^{(\eta-1)} \odot \hat{\mathbf{X}}_{n-1}^{(\eta)} \odot \cdots \odot \hat{\mathbf{X}}_1^{(\eta)} \right)^{\dagger} \right)^T \\ &= \mathbf{Y}_{(n)} \left(\left(\left(\bigodot_{i=N+1}^{n+1} \hat{\mathbf{X}}_i^{(\eta-1)} \right) \odot \left(\bigodot_{j=n-1}^{1} \hat{\mathbf{X}}_j^{(\eta)} \right) \right)^{\dagger} \right)^T . \end{aligned} \tag{25}$$

It is noted that the factor matrices should be updated in order from $\hat{\mathbf{X}}_1$ to $\hat{\mathbf{X}}_{N+1}$.

In addition, for the first time of the updating calculation, the value of $\hat{\mathbf{X}}_n^{(0)}$ is obtained with the help of the singular value decomposition (SVD) of the corresponding mode-n unfolding, given by

$$\mathbf{Y}_{(n)} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^H, \tag{26}$$

where $\mathbf{U}_n$ represents the left singular matrix. Assume that $\mathbf{U}_n$ has the size of $Q_{n,r} \times Q_{n,c}$, where $Q_{n,c}$ will not be always larger than U_a. The value of $\hat{\mathbf{X}}_n^{(0)}$ is selected as

$$\hat{\mathbf{X}}_n^{(0)} = \begin{cases} [\mathbf{U}_n]_{1:U_a}, & Q_{n,c} \geq U_a, \\ [\mathbf{U}_n \, \mathbf{G}_n], & Q_{n,c} < U_a. \end{cases} \tag{27}$$

where $\mathbf{G}_n \triangleq [\mathbf{g}_{n,1}, \ldots, \mathbf{g}_{n,(U_a - Q_{n,c})}] \in \mathbb{C}^{Q_{n,r} \times (U_a - Q_{n,c})}$ is exploited to supplement the number of columns, $\mathbf{g}_{n,q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{Q_{n,r}})$, $q = 1, \ldots, (U_a - Q_{n,c})$.

After finishing the $2(N+1)$ iterations of updates, each factor matrix determines its initial values, i.e., $\hat{\mathbf{X}}_n^{(\eta-1)}$ and $\hat{\mathbf{X}}_n^{(\eta-2)}$, by the results of the last two updates, respectively.

2) Line Search Acceleration: According to the important fact revealed by [43],[44], there exist cycles of convergence defined by a unique direction when the whole convergence speed is slow. For one given cycle, the factor matrices will be updated in the same direction until reaching convergence or the maximum number of iteration. To be specific, the direction of the cycle can be represented by $(\hat{\mathbf{X}}_n^{(\eta-1)} - \hat{\mathbf{X}}_n^{(\eta-2)})$.

To effectively guarantee the convergence of ALS, the iteration numbers of a given cycle should be reasonably conserved, which can be precisely implemented by computing the following linear regression with the help of the given direction, given by

$$\hat{\mathbf{X}}_n^{(LS)} = \hat{\mathbf{X}}_n^{(\eta-2)} + \mu_{LS}(\hat{\mathbf{X}}_n^{(\eta-1)} - \hat{\mathbf{X}}_n^{(\eta-2)}), \tag{28}$$

where $\hat{\mathbf{X}}_n^{(LS)}$ denotes the alternative factor matrix against $\hat{\mathbf{X}}_n^{(\eta-1)}$ for updating $\hat{\mathbf{X}}_n^{(\eta)}$, μ_{LS} is the relaxation factor to regulate the savings of iterations. To obtain a pronounced acceleration effect on convergence performance without introducing remarkable computational complexity, the relaxation factor is considered to be set as $\mu_{LS} = \eta^{1/(N+1)}$ instead of fixed value or other adaptive values, which is capable of balancing the efficiency and stability well [47].

Subsequently, we need to judge whether $\hat{\mathbf{X}}_n^{(LS)}$ is superior to the corresponding $\hat{\mathbf{X}}_n^{(\eta-1)}$ during each iteration, which requires the comparison of residuals between $\mathcal{Y}$ and the tensors formed by these two group of factor matrices, given by

$$\mathcal{E}^{(LS)} = \| \mathcal{Y} - [\![\hat{\mathbf{X}}_1^{(LS)}, \hat{\mathbf{X}}_2^{(LS)}, \ldots, \hat{\mathbf{X}}_{N+1}^{(LS)}]\!] \|_F^2, \tag{29}$$

$$\mathcal{E}^{(\eta-1)} = \| \mathcal{Y} - [\![\hat{\mathbf{X}}_1^{(\eta-1)}, \hat{\mathbf{X}}_2^{(\eta-1)}, \ldots, \hat{\mathbf{X}}_{N+1}^{(\eta-1)}]\!] \|_F^2, \tag{30}$$

where $\mathcal{E}^{(LS)}$ and $\mathcal{E}^{(\eta)}$ respectively denotes the above residuals. To be specific, the expected result of $\mathcal{E}^{(LS)} < \mathcal{E}^{(\eta-1)}$ demonstrates that the alternative $\hat{\mathbf{X}}_n^{(LS)}$ is closer to the real factor matrix than $\hat{\mathbf{X}}_n^{(\eta-1)}$. Therefore, $\mathcal{E}^{(LS)} < \mathcal{E}^{(\eta-1)}$ validates a successful acceleration, which indicates that we have equivalently jumped a few iterations and improved the convergence speed [47]. On the contrary, the undesirable results of $\mathcal{E}^{(LS)} \geq \mathcal{E}^{(\eta-1)}$ demonstrates that the value of μ_{LS} is considered to be inappropriate in this iteration, where the acceleration cannot be guaranteed and $\hat{\mathbf{X}}_n^{(\eta-1)}$ remains unchanged. In summary, the acceleration results of line search can be expressed as

$$\hat{\mathbf{X}}_n^{(\eta-1)} = \begin{cases} \hat{\mathbf{X}}_n^{(LS)}, & \mathcal{E}^{(LS)} < \mathcal{E}^{(\eta-1)}, \\ \hat{\mathbf{X}}_n^{(\eta-1)}, & \mathcal{E}^{(LS)} \geq \mathcal{E}^{(\eta-1)}. \end{cases} \tag{31}$$

Algorithm 1 Tensor-based GLSA-ALS Algorithm

Input: Received signal tensor $\mathcal{Y}$, relaxation factor μ_{LS}, threshold ξ, and maximum number of iterations I_{iter}

Output: $\hat{\mathbf{X}}_n, n = 1, \ldots, N+1$

Stage 1: Initialization:

Compute $\hat{\mathbf{X}}_n^{(0)}$ by (27);
for $\eta = 1 : 2(N+1)$ **do**
  for $n = 1 : N+1$ **do**
    Update $\hat{\mathbf{X}}_n^{(\eta)}$ according to (25);
  end for
end for

Stage 2: Acceleration:

Take $\hat{\mathbf{X}}_n^{(\eta)}$ and $\hat{\mathbf{X}}_n^{(\eta-1)}$ as the initial values;
Let $\eta = \eta + 1$, compute $\hat{\mathbf{X}}_n^{(LS)}$ by (28);
Reconstruct $\mathcal{E}^{(LS)}$ and $\mathcal{E}^{(\eta-1)}$ by (29) and (30);
if $\mathcal{E}^{(LS)} < \mathcal{E}^{(\eta-1)}$ **then**
  $\hat{\mathbf{X}}_n^{(\eta-1)} = \hat{\mathbf{X}}_n^{(LS)}$;
else
  $\hat{\mathbf{X}}_n^{(\eta-1)} = \hat{\mathbf{X}}_n^{(\eta-1)}$.
end if

Stage 3: Estimation:

if $\eta \leq I_{iter}$ **then**
  for $n = 1 : N+1$ **do**
    Update $\hat{\mathbf{X}}_n^{(\eta)}$ according to (25);
  end for
  Reconstruct $\mathcal{E}^{(\eta)}$ in the way of (30);
  if $\left|\mathcal{E}^{(\eta)} - \mathcal{E}^{(\eta-1)}\right| > \xi$ **then**
    do 7;
  else
    $\hat{\mathbf{X}}_n = \hat{\mathbf{X}}_n^{(\eta)}$, **break**.
  end if
end if

3) Factor Matrices Estimation: To accurately and efficiently obtain the estimation of the factor matrices $\hat{\mathbf{X}}_n$, two steps are required. Firstly, based on acceleration shown in (31), the iterative updates are continuously conducted by (25). Similarly, $\hat{\mathbf{X}}_n^{(\eta)}$ should be updated in order from $n = 1$ to $n = N+1$. Then, by exploiting the preseted threshold denoted by ξ, we will determine whether the results of iteration are acceptable for further signal detection. If $\left|\mathcal{E}^{(\eta)} - \mathcal{E}^{(\eta-1)}\right| > \xi$, the update of $\hat{\mathbf{X}}_n$ will continue by repeating the initialization and acceleration. Specifically, the value of $\hat{\mathbf{X}}_n^{(\eta)}$ and $\hat{\mathbf{X}}_n^{(\eta-1)}$ will be assigned to $\hat{\mathbf{X}}_n^{(\eta-1)}$ and $\hat{\mathbf{X}}_n^{(\eta-2)}$, which are subsequently exploited to calculate new $\hat{\mathbf{X}}_n^{(LS)}$ and $\mathcal{E}^{(LS)}$ in the next iteration. Conversely, if $\left|\mathcal{E}^{(\eta)} - \mathcal{E}^{(\eta-1)}\right| \leq \xi$ or the number of iteration has already reached the maximum value, denoted by I_{iter}, the update of $\hat{\mathbf{X}}_n$ stops, where $\hat{\mathbf{X}}_n = \hat{\mathbf{X}}_n^{(\eta)}$ will be utilized for signal detection. In summary, the proposed GLSA-ALS method based on tensor modeling for URA in LEO satellite IoT scenarios can be detailed as **Algorithm 1**.

B. Computational Complexity Analysis for GLSA-ALS

The computational complexity of the proposed GLSA-ALS algorithm is elaborated as follows, which is mainly contributed by the complex multiplications during the factor matrices estimation and LS acceleration. For simplicity, the corresponding analysis is presented with the aspect of the n-th factor matrix, i.e., $\hat{\mathbf{X}}_n \in \mathbb{C}^{L_n \times U_a}$, recalling that $L = \prod_{n=1}^{N} L_n$. In terms of factor matrices estimation, the computational complexity is mainly consist of SVD and matrices updating. The SVD takes $\mathcal{O}(L_n M^2 (L/L_n)^2 + M^3 (L/L_n)^3)$, which can be approximated as $\mathcal{O}(\max(L_1, \ldots, L_N, M)^3)$ when the magnitude of $(L_1, \ldots, L_N, M)$ is similar. The matrices updating requires $\mathcal{O}(I_{GLSA}(U_a ML(1+1/L_n) + U_a^2 L_n + U_a L_n))$, where I_{GLSA} denotes the total iteration number of GLSA-ALS and contains $2(N+1)$ times iterations for LS preparation, $U_a ML/L_n$ is the cost of successive Khatri-Rao product in (25). In terms of the other part, the computational complexity is mainly consist of the relaxation and additional judgment in (29), which requires $\mathcal{O}(U_a L_n)$ and $\mathcal{O}(L)$, respectively.

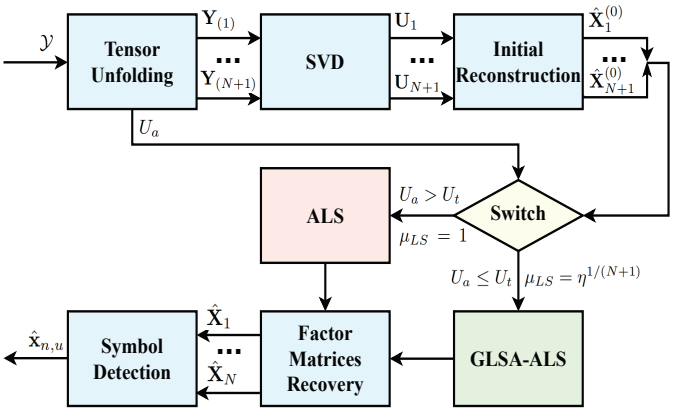

Fig. 2. The block diagram of the proposed detection method for the tensor-based URA scheme.

C. ML-based Signal Detection

Based on the results of tensor decomposition, the relationship between the estimated factor matrices and the true factor matrices can be presented as

$$\hat{\mathbf{X}}_n = \mathbf{X}_n \boldsymbol{\Delta}_n \boldsymbol{\Pi} + \mathbf{E}_n, \quad \forall n, \tag{32}$$

where $\boldsymbol{\Delta}_n$ is the unknown diagonal matrix satisfying $\prod_{n=1}^{N+1} \boldsymbol{\Delta}_n = \mathbf{I}_{U_a}$, $\boldsymbol{\Pi} \in \mathbb{C}^{U_a \times U_a}$ is the unknown permutation matrix, $\mathbf{E}_n$ is the corresponding estimation errors. Therefore, the signal detection should be able to handle the ambiguity of scaling and permutation.

In fact, benefited from the proper selection of transmitted constellations, the scalar indeterminacy can be resolved with the help of the Grassmannian structure design, which is insensitive to the collinearity [45]. Moreover, due to the characteristic of CP decomposition, $\boldsymbol{\Pi}$ is identical to all $\hat{\mathbf{X}}_n$, which indicates that the sub-constellation symbols of all devices are column aligned. Hence, $\hat{\mathbf{x}}_{n,u}$ can be separated and detected from $\hat{\mathbf{X}}_n$.

Recalling that we have denoted $\mathbf{X}_{N+1} \triangleq \mathbf{H}$, accompanied with the information-bearing factor matrices, the CSI matrix is simultaneously updated, which is coupled with $\hat{\mathbf{X}}_n$ and cannot be extracted without pilot sequence. Nevertheless, we mainly focus on the signal detection instead of CE, and the impact of $\mathbf{H}$ can be ignored since the ML-based method can be used

to accurately recover the signal without the aid of CE [46]. Hence, the detection is conducted in a non-coherent way as most URA studies. To be specific, motivated by the concept of ALS, we firstly fix all $\hat{\mathbf{x}}_{n,u}$ to solve (15) with regard to $\mathbf{h}_u$, where the property in (12)-(13) is utilized in reverse, given by

$$\begin{aligned}\mathbf{h}_u^* &= \frac{(\mathbf{x}_{1,u}\otimes\cdots\otimes\mathbf{x}_{N,u})^H(\hat{\mathbf{x}}_{1,u}\otimes\cdots\otimes\hat{\mathbf{x}}_{N,k})}{\|\mathbf{x}_{1,u}\otimes\cdots\otimes\mathbf{x}_{N,u}\|_2^2}\hat{\mathbf{h}}_u \\ &= \frac{(\bigotimes_{n=1}^N\mathbf{x}_{n,u})^H(\bigotimes_{n=1}^N\hat{\mathbf{x}}_{n,u})}{\|\bigotimes_{n=1}^N\mathbf{x}_{n,u}\|_2^2}\hat{\mathbf{h}}_u.\end{aligned} \tag{33}$$

Then, we have the relationship between $\mathbf{h}_u$ and $\hat{\mathbf{h}}_u$, which can be substituted into (15) to simplify the objective function with regard to $\mathbf{x}_{n,u}$, given by

$$\begin{aligned}&\left\|\hat{\mathbf{x}}_{1,u}\otimes\cdots\otimes\hat{\mathbf{x}}_{N,u}\otimes\hat{\mathbf{h}}_u-\mathbf{x}_{1,u}\otimes\cdots\otimes\mathbf{x}_{N,u}\otimes\mathbf{h}_u^*\right\|_2^2 \\ &=\left\|vec\left(\bigotimes_{n=1}^N\hat{\mathbf{x}}_{n,u}\cdot\hat{\mathbf{h}}_u^T\right)-vec\left(\bigotimes_{n=1}^N\mathbf{x}_{n,u}\cdot(\mathbf{h}_u^*)^T\right)\right\|_2^2 \\ &=\left\|\left(\mathbf{I}_L-\frac{(\bigotimes_{n=1}^N\mathbf{x}_{n,u})(\bigotimes_{n=1}^N\mathbf{x}_{n,u})^H}{\|\bigotimes_{n=1}^N\mathbf{x}_{n,u}\|^2}\right)\left(\bigotimes_{n=1}^N\hat{\mathbf{x}}_{n,u}\cdot\hat{\mathbf{h}}_u^T\right)\right\|_2^2,\end{aligned} \tag{34}$$

where $\left(\mathbf{I}_L-\left((\bigotimes_{n=1}^N\mathbf{x}_{n,u})(\bigotimes_{n=1}^N\mathbf{x}_{n,u})^H\right)/\|\bigotimes_{n=1}^N\mathbf{x}_{n,u}\|^2\right)$ denotes the corresponding projection matrix. Based on the properties of 2-norm and projection, (15) can be equivalently transformed as follows

$$\max_{\mathbf{x}_{n,u}\in\mathscr{C}_n}\frac{\left\|(\bigotimes_{n=1}^N\mathbf{x}_{n,u})(\bigotimes_{n=1}^N\mathbf{x}_{n,u})^H(\bigotimes_{n=1}^N\hat{\mathbf{x}}_{n,u})\right\|_2^2}{\|\bigotimes_{n=1}^N\mathbf{x}_{n,u}\|_2^4}. \tag{35}$$

Eventually, according to the property of Kronecker product that $(\mathbf{a}\otimes\mathbf{b})^H(\mathbf{c}\otimes\mathbf{d})=(\mathbf{a}^H\mathbf{c})(\mathbf{b}^H\mathbf{d})$, the detection problem in (15) can be expressed as

$$\max_{\hat{\mathbf{x}}_{n,u}\in\mathscr{C}_n}\prod_{n=1}^N\frac{\left|\mathbf{x}_{n,u}^H\hat{\mathbf{x}}_{n,u}\right|}{\|\hat{\mathbf{x}}_{n,u}\|_2\|\mathbf{x}_{n,u}\|_2}, \tag{36}$$

which confirms the discussion about the robustness of detection to $\mathbf{h}_u$.

D. Compatibility with Classical ALS

Although the proposed GLSA-ALS method is basically competent enough to recover the factor matrices efficiently for ML-based detection, a slight loss in estimation accuracy is still inevitable due to the iteration acceleration especially under a larger number of active devices. Fortunately, the selection of the relaxation factor of the desgined method is flexible, where $\mu_{LS}=1$ will make the GLSA-ALS degrade into the ALS to pursue a better accuracy performance under the premise of convergence. Therefore, the proposed GLSA-ALS can be reduced to the classical ALS for variant demands of accuracy or complexity. Accordingly, the block diagram of the proposed detection method is depicted in Fig. 2, where U_t denotes the empirical switch position under a particular application scenario.

TABLE III
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency f_c	30GHz
Propagation distance d_0	1000km
Satellite receive antenna gain α_u	20dBi
Transmit gain to noise temperature $\frac{G_u}{T_0}$	34dB/K
Carrier bandwidth W	25MHz
Rain attenuation mean μ_r	-2.6dB
Rain attenuation variance σ_r^2	1.63dB
Rician factor γ_u	8
Boltzmann's constant κ	1.38×10^{-23} J/K
LOS component $\left\|\mathbf{h}_u^{\mathrm{LOS}}\right\|_2^2$	$\mathcal{U}[0.6,0.7]$
NLOS variance $(\sigma_u^{\mathrm{NLOS}})^2$	$\mathcal{CN}[0.2,0.25]$
Maximum number of iterations I_{iter}	100
Length of information bits B	140, 112
Size of factorized channel use N	4, 5
Number of antennas M	2, 48
Number of active devices U_a	20∼100
Termination threshold ξ	10^{-6}

V. SIMULATION RESULTS

In this section, we conduct numerical simulations to verify the effectiveness of the proposed tensor-based unsourced random access scheme for LEO satellite IoT. According to 3GPP technical report (TR) 38.811, TR 38.821 and some existing related works, the simulation parameters are set in TABLE III unless otherwise specified [32],[36],[48]. Generally, the average per user probability of error (PUPE) is considered to measure the detection performance, which can be defined as [14],[25]

$$\mathrm{PUPE}=\mathbb{E}\left\{\frac{|\mathcal{L}-\hat{\mathcal{L}}|_c}{|\mathcal{L}|_c}\right\}, \tag{37}$$

where $\mathcal{L}$ and $\hat{\mathcal{L}}$ denote the detected message list and the transmitted message list, respectively. In addition, we use the average number of iterations and the variance of iteration numbers to assess the convergence performance. To ensure the reliability of performance analysis, the simulation results are obtained by averaging over 10000 independent channel realizations. For the signal in tensor form, the SNR is defined as

$$\mathrm{SNR}=\frac{\|\mathcal{Y}-\mathcal{Z}\|_F^2}{\|\mathcal{Z}\|_F^2}. \tag{38}$$

A. Impact of Different Relaxation Factors

In this part, we comprehensively present the effectiveness of the proposed GLSA-ALS method in terms of convergence and detection performance under different μ_{LS}. Recalling that

This article has been accepted for publication in IEEE Transactions on Wireless 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/TWC.2025.3612505

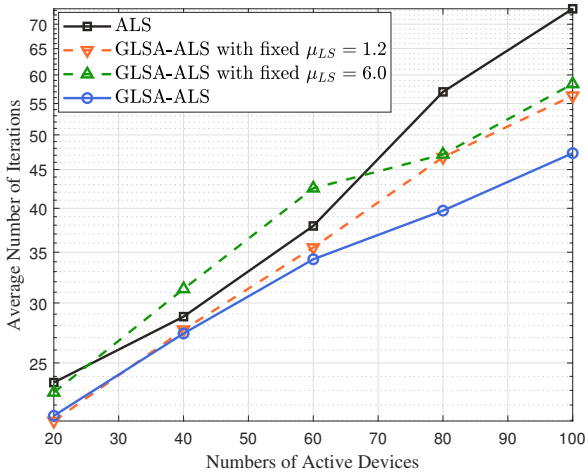

Fig. 3. Required number of iterations under various numbers of active devices.

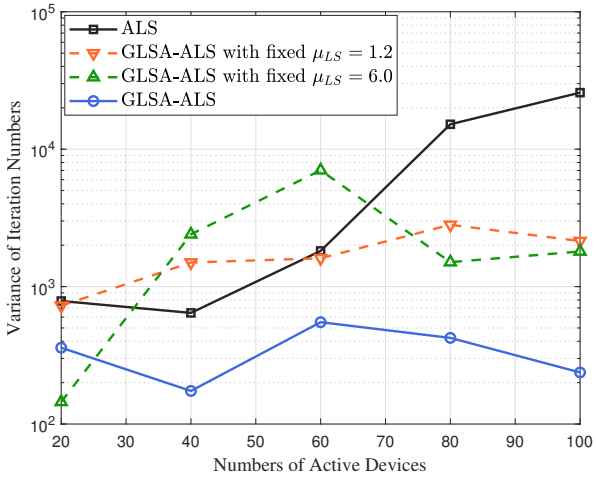

Fig. 4. Variance of iteration numbers under various numbers of active devices.

the relaxation factor of GLSA-ALS is considered to be set as $\mu_{LS} = \eta^{1/(N+1)}$ and the conventional ALS method can be regarded as $\mu_{LS} = 1$. Therefore, for illustration, we set two additional fixed-valued relaxation factors, i.e., $\mu_{LS} = 1.2$ and $\mu_{LS} = 6.0$, to reveal the acceleration effect in detail. Without loss of generality, the simulation parameters are selected as SNR = 5dB, $N = 4$, $L_n = (10, 10, 8, 8)$ and $M = 48$ for the convenience of comparison.

In Fig. 3, we compare the convergence performance of the conventional ALS method and the proposed GLSA-ALS method with different μ_{LS} in terms of the total number of iterations. It is noted that, given a specific value of N, μ_{LS} will gradually increase along with the iteration, which indicates its automatic adjustment feature in the line search process. Thus, $\mu_{LS} = 1.2$ and $\mu_{LS} = 6.0$ roughly represent the smallest and largest value during the iteration, which can help illustrate the impact of fixed relaxation factors on line search. Fig. 3 shows that, as the number of active devices increases, the total numbers of iterations required for all schemes to converge increase approximately linearly. However, due to the line search acceleration, the proposed GLSA-ALS method always require less iterations to converge than the conventional ALS method, where the gap significantly increases when $U_a > 60$. The gap of iteration numbers demonstrates that convergence speed of GLSA-ALS is consistently faster than ALS, especially under the condition of large number of active devices, where the iteration numbers of ALS is more than 150% of GLSA-ALS at $U_a = 100$. In addition, although the convergence speeds under fixed μ_{LS} are faster than ALS when $U_a > 80$, they are still evidently slower than $\mu_{LS} = \eta^{1/(N+1)}$, where the performance of $\mu_{LS} = 6.0$ is even worse than ALS before $U_a \leq 60$. Therefore, the reasonable selection of relaxation factor guarantees the convergence speed improvement of the proposed GLSA-ALS method.

To further illustrate the convergence performance from the aspect of stability, we provide the corresponding variance of iteration numbers required to guarantee the convergence in Fig. 4. Overall, the larger variance indicates the poorer stability, i.e., the convergence speed will significantly slow down in some certain transmissions, which is far from the average performance. Fig. 4 shows that, as the number of active devices increases, the variance of ALS generally exhibits an increasing trend, while the variance of GLSA-ALS remains relatively stable within a certain range. To be specific, the corresponding variance gap reaches approximately 1-2 orders of magnitude, which widens significantly when $U_a > 60$. In fact, the size of factor matrices will be enlarged along with the increase of active device numbers, where the greater numerical uncertainty affects the iteration efficiency remarkably. Therefore, without considering the limitation of iteration numbers, ALS method can converge in each transmission, but the required computational complexity also significantly increases. In addition, although the performance of $\mu_{LS} = 1.2$ is more stable than ALS, it consistently remains worse than GLSA-ALS with $\mu_{LS} = \eta^{1/(N+1)}$. Besides, in spite of the smallest variance under $U_a = 20$, the stability of GLSA-ALS with $\mu_{LS} = 6.0$ is the worst, which confirms the importance of appropriate relaxation factor selection.

Fig. 5 exhibits the PUPE performance under the same conditions in Fig. 3 and 4, which indicates the influence of iteration acceleration on detection accuracy. It can be observed that, the PUPE performance of ALS and GLSA-ALS is very close when $U_a \leq 60$ regardless of the relaxation factor selection, which can achieve on the order of 10^{-5}. However, under $U_a > 60$, the performance gap is enlarged due to the detection accuracy deterioration of GLSA-ALS. Based on the results of Fig. 3 and 4, in fact, when $U_a > 60$, the required iteration numbers of ALS increase significantly as compared with the effectively accelerated GLSA-ALS, which enables ALS to achieve better performance at the cost of additional computational complexity. However, if the computational complexity of the detector is limited in some specific application scenarios, when the ALS method converges slowly due to the possible swamps, it would fail to detect correctly within finite number of iterations. To be specific, the reduction of computational complexity between GLSA-

This article has been accepted for publication in IEEE Transactions on Wireless 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/TWC.2025.3612505

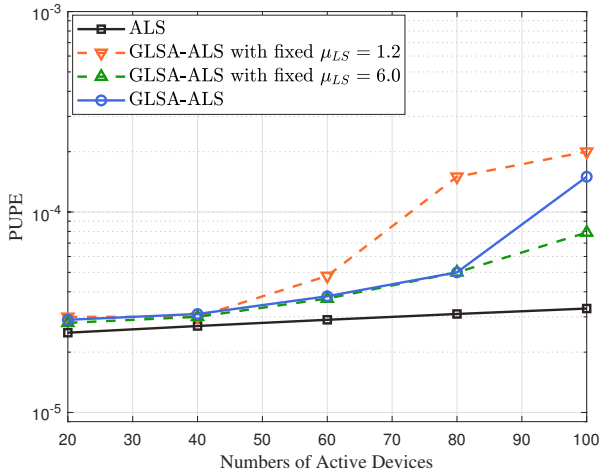

Fig. 5. PUPE performance under various numbers of active devices.

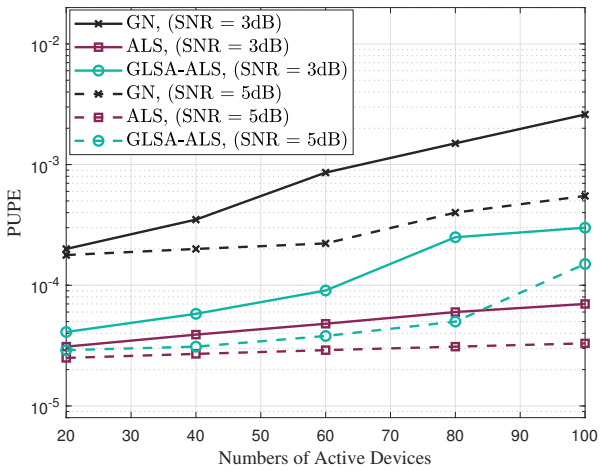

Fig. 6. PUPE performance under various SNRs, $N = 4$, $L_n = (10, 10, 8, 8)$, $M = 48$.

ALS and ALS mainly lies in the matrices updating, taking $\mathcal{O}((I_{ALS} - I_{GLSA})(U_a ML(1 + 1/L_n) + U_a^2 L_n + U_a L_n))$, where I_{ALS} denotes the total iteration number of ALS. Despite a small amount of additional relaxation and judgement operation of GLSA-ALS taking $\mathcal{O}((I_{GLSA} - 2(N+1))(U_a L_n + L))$, the total computational complexity of GLSA-ALS compared to ALS is significantly reduced when $U_a > 60$. Therefore, the proposed GLSA-ALS method is more efficient, where the flexible selection of μ_{LS} enables the adaptive switch of the detector between GLSA-ALS and conventional ALS to meet different requirements of accuracy and complexity. Besides, although the PUPE performance of $\mu_{LS} = 1.2$ and $\mu_{LS} = 6.0$ is similar to that of $\mu_{LS} = \eta^{1/(N+1)}$ in GLSA-ALS, the more required iteration numbers, as shown in Fig. 3, make them an undesired selection for possibly switching between ALS.

B. Performance Comparison of Different Detection Methods

To validate the effectiveness of the proposed method under variant simulation parameters, Fig. 6, Fig. 7, Fig. 8, and Fig. 9 have been presented to show the PUPE performance of different detection methods, where the GN-based method [25],[27],[41],[42] is selected for comparison. Specifically, due to the superior capability of handling the collinearity challenges, the damped GN algorithm is chosen as the benchmark [42](cf. Sec. 5).

In Fig. 6, we compare the PUPE performance under different SNRs. It can be observed that the performance of all methods is sensibly enhanced with the increase of SNR, where the proposed efficient GLSA-ALS method exhibits evidently better performance than the GN method. In fact, during the estimation of factor matrices in GN method, the calculation of matrix inversion is required in each iterative update, which involves the Hessian matrix of the objective residual tensor. After certain times iteration, the term that needs to calculate the inversion approaches a singular matrix, which deteriorates the accuracy of recovered factor matrices, thus leading to a worse PUPE performance. This phenomenon mainly stems from the specific factor matrices with the entries came from

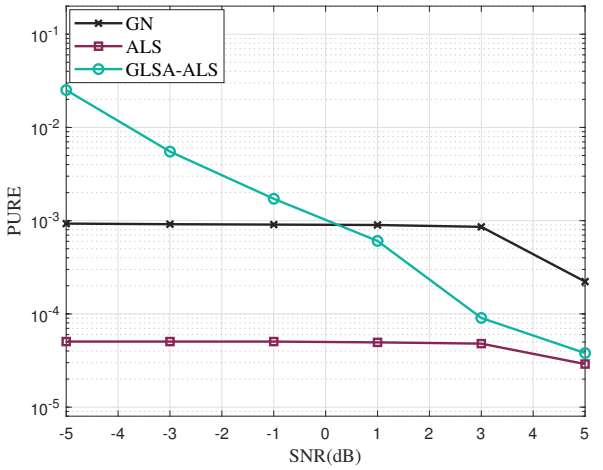

Fig. 7. PUPE performance under various SNRs with a certain number of active devices, $N = 4$, $L_n = (10, 10, 8, 8)$, $M = 48$, $U_a = 60$.

practical communication systems, whose numerical feature is close to ill-condition to some extent. Therefore, the proposed GLSA-ALS method generally outperforms the GN method by more than one order of magnitude. In addition, performance gaps between GLSA-ALS and conventional ALS are similar under different SNRs, which is consistent with the analysis of Fig. 5 and provides an recommendation of adaptive switch with $U_t = 60$ for a better trade-off between the detection accuracy and computational complexity.

Subsequently, Fig. 7 further shows the PUPE performance under various SNRs with a certain number of active devices. It can be observed that under the low SNR conditions, the performance of both GN and conventional ALS is relatively stable, where ALS performs better and maintains at the order of 10^{-5}. In addition, although GLSA-ALS method performs worse than GN and ALS when SNR is less than 0dB, its detection performance can significantly improve with the increase of SNR, which verifies the effectiveness of the proposed method. Besides, benefiting from the compatibility with ALS, the proposed GLSA-ALS can reasonably adjust the switch

This article has been accepted for publication in IEEE Transactions on Wireless 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/TWC.2025.3612505

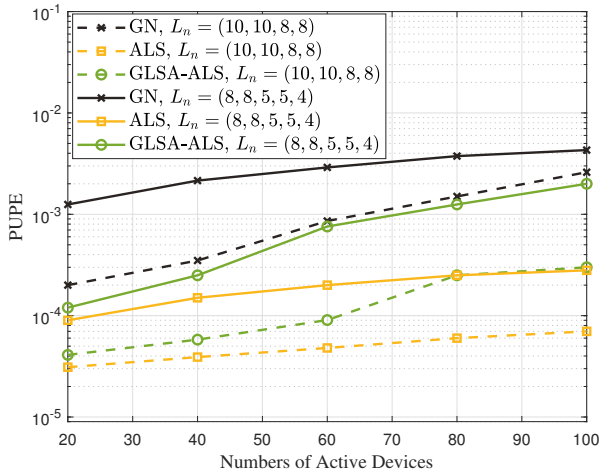

Fig. 8. PUPE performance under various factorization of channel use with SNR = 3dB and $M = 48$.

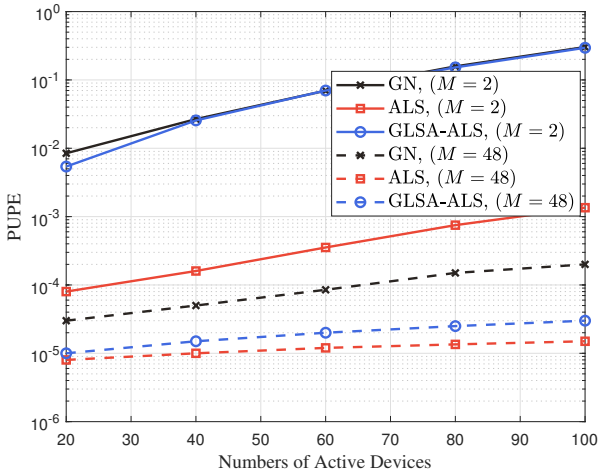

Fig. 9. PUPE performance under various numbers of antennas, SNR = 10dB, $N = 4$, $L_n = (10, 10, 8, 8)$.

timing according to the specific level of detection accuracy requirements, which guarantees the performance under low SNR conditions.

To illustrate the detection accuracy with respect to varying tensor sizes in the URA system we investigated, Fig. 8 plots the PUPE performance under different L_n. To be specific, the factorization is set as $N = 4, 5$ and $L_n = (10, 10, 8, 8), (8, 8, 5, 5, 4)$ respectively, where a larger N leads to a greater number of smaller-sized factor matrices to be estimated. It can be observed that the PUPE performance of all methods under $N = 4$ is better than that of $N = 5$, where the performance gap between the different methods is consistent with the results in Fig. 6. In fact, the effect of different factorization on PUPE performance can be explained through the number of degrees of freedom (DoF) per active device, which would achieve its upper bound by the cooperation of devices [25]. Particularly, considering a variable in Grassmannian of lines in dimension L_n has $L_n - 1$ DoF [45], the sum-DoF of all active devices based on the rank-U_a tensor-form model can be defined as $\text{DoF}(U_a) = U_a \sum_{n=1}^{N}(L_n - 1)$. Therefore, given a certain number of active devices, the available DoF of $N = 4$ is higher than that of $N = 5$, which facilitates a better detection for the proposed tensor-based URA system. On the contrary, the larger-sized factor matrices caused by smaller N will slightly increase the computational complexity of all methods, which directly complicates the matrix inversion in GN and iteration update in ALS-based methods. Thus, this is also a selection of trade-off based on the practical demands.

Fig. 9 shows the PUPE performance under various numbers of receiving antennas. It can be observed that the detection accuracy of same method under $M = 48$ is significantly superior than the performance under $M = 2$, where the PUPE of GN and GLSA-ALS is worse than 10^{-1} under $M = 2$ as $U_a \geq 80$. This phenomenon demonstrates that the proposed tensor-based URA scheme can benefit from the spatial diversity brought by a large number of antennas, which can also be explained from the aspect of identifiability in the uniqueness analysis part, i.e., (15)-(20). In addition, the PUPE of GLSA-ALS is almost the same as GN under $M = 2$, while it is worse than ALS about two orders of magnitude. The corresponding performance gap indicates the importance of implementing a large number of antennas for the URA system, which guarantees an acceptable performance trade-off brought by the proposed detection method. Besides, Fig. 9 also presents a negligible performance gap between GLSA-ALS and ALS under $M = 48$ and SNR = 10dB, which confirms the effectiveness of GLSA-ALS at high SNR regime. Therefore, for the sake of efficiency, the proposed accelerated detection method can be continuously exploited from $U_a = 20$ to $U_a = 100$ without switching in this situation.

VI. Conclusion

In this paper, we investigated an effective unsourced random access scheme for LEO satellite Internet of Things scenarios. By leveraging the Grassmannian constellation for modulation, a tensor-based URA scheme was proposed, which can be adopted to handle the device separation and signal detection by general CP decomposition. In addition, to demonstrate the boundary for the number of supported active devices, a comprehensive uniqueness analysis from both sufficient and necessary conditions was presented, which also provides a theoretical guarantee for the detection. Then, by exploiting the acceleration effect of line search, an effective GLSA-ALS method was proposed, which is able to converge at a fast speed and recover the factor matrices efficiently to facilitate the subsequent detection. Besides, benefiting from the flexibility of relaxation factor selection, the proposed GLSA-ALS method can be compatible to the classical ALS method, which enables the designed URA scheme to satisfy the variant demands of computational complexity and detection accuracy in practical scenarios. Simulation results verified that the proposed method outperforms the state-of-the-art solutions in the tensor-based URA schemes for LEO satellite IoT scenarios, showcasing the superiority of our proposed methodology.

REFERENCES

[1] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[2] A. Zanella, N. Bui, et al., "Internet of things for smart cities," *IEEE Internet Things J.*, 2014, vol. 1, no. 1, pp. 22–32.

[3] N. Ahmed, D. De, and I. Hussain, "Internet of Things (IoT) for smart precision agriculture and farming in rural areas," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4890–4899, Dec. 2018.

[4] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[5] D. C. Nguyen et al., "6G Internet-of-Things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[6] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805-1838, 3rd Quart. 2020.

[7] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 148–156, Aug. 2020.

[8] X. Shao et al., "Cooperative activity detection: Sourced and unsourced massive random access paradigms," *IEEE Trans. Signal Process.*, vol. 68, pp. 6578–6593, Dec. 2020.

[9] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2320–2323, Nov. 2016.

[10] 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.

[11] X. Zhou et al., "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.

[12] 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.

[13] Y. Li et al., "Unsourced multiple access for 6G massive machine type communications," *China commun.*, vol. 19, no. 3, pp. 70-87, March 2022.

[14] M. Ke, Z. Gao, M. Zhou, D. Zheng, D. W. K. Ng and H. V. Poor, "Next-generation URLLC with massive devices: A unified semi-blind detection framework for sourced and unsourced random access," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 7, pp. 2223-2244, Jul. 2023.

[15] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2523–2527.

[16] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2528–2532.

[17] V. K. Amalladinne, J. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, Oct. 2020.

[18] S. S. Kowshik, K. Andreev, A. Frolov and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, 2019, pp. 2768-2772.

[19] K. Andreev, S. S. Kowshik, A. Frolov and Y. Polyanskiy, "Low complexity energy efficient random access scheme for the asynchronous fading MAC," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Honolulu, HI, USA, 2019, pp. 1-5.

[20] J. Fang, G. Sun, W. Wang, L. You and R. Ding, "OFDMA-based unsourced random access in LEO satellite Internet of Things," *China Commun.*, vol. 21, no. 1, pp. 13-23, Jan. 2024.

[21] P. Chen and L. Cheng, "Estimating channels with hundreds of sub-paths for MU-MIMO uplink: A structured high-rank tensor approach," *IEEE Signal Process. Lett.*, vol. 31, pp. 2320-2324, 2024.

[22] R. Zhang et al., "Integrated sensing and communication with massive MIMO: A unified tensor approach for channel and target parameter estimation," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8571-8587, Aug. 2024.

[23] L. Cheng et al., "Towards flexible sparsity-aware modeling: Automatic tensor rank learning using the generalized hyperbolic prior," *IEEE Trans. Signal Process.*, vol. 70, pp. 1834-1849, 2022.

[24] T. Yang et al., "A unifed tensor-based joint AUD and ISAC parameter estimation with large-scale user access," *IEEE Trans. Cognit. Commun. Networking*, early access, Feb. 25, 2025, doi:10.1109/TCCN.2025.3545690.

[25] A. Decurninge, I. Land, and M. Guillaud, "Tensor-based modulation for unsourced massive random access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 552–556, Mar. 2020.

[26] X. Shao, L. Cheng, X. Chen, C. Huang and D. W. K. Ng, "Reconfigurable intelligent surface-aided 6G massive access: Coupled tensor modeling and sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10145-10161, Dec. 2022.

[27] Z. Luan, Y. Wu, S. Liang, W. Han, B. Bai and L. Zhang, "Modulation for massive unsourced random access based on tensor block term decomposition," in *Proc. IEEE Global Commun. Conf. Workshops (GC Wkshps)*, Rio de Janeiro, Brazil, 2022, pp. 637-643

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

[29] Z. Gao, K. Ying, C. He, Z. Xiao, D. Zheng, and J. Zhang, "Grant-free random access in massive MIMO based LEO satellite Internet of Things," in *Proc. IEEE/CIC Interfaces Conf. Commun. China (ICCC)*, Aug. 2021, pp. 700–705.

[30] 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.

[31] Z. Lin, M. Lin, J.-B. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 3, pp. 657–670, Jun. 2019.

[32] M. Ying et al., "Exploiting tensor-based bayesian learning for massive grant-free random access in LEO satellite Internet of Things," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1141-1152, Feb. 2023.

[33] J. Chu and X. Chen, "Robust design for integrated satellite–terrestrial Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9072–9083, Jun. 2021.

[34] M. R. Bhatnagar, "Making two-way satellite relaying feasible: A differential modulation based approach," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2836–2847, Aug. 2015.

[35] A. Gharanjik, M. R. B. Shankar, P. D. Arapoglou, M. Bengtsson, and B. Ottersten, "Robust precoding design for multibeam downlink satellite channel with phase uncertainty," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 3083–3087.

[36] Non-Terrestrial Networks Channel Models, document TR 38.811 v15.4.0, 3GPP, Sep. 2020.

[37] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[38] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays", *J. Chemometrics*, 14 (2000), pp. 229–239.

[39] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM J. Matrix Anal. App.*, vol. 35, no. 4, pp. 1265–1287, 2014.

[40] X. Liu and N. Sidiropoulos, "Cram´er-Rao lower bounds for low-rank decomposition of multi-dimensional arrays," *IEEE Trans. Signal Process.*, 49 (2001), pp. 2074–2086.

[41] L. Sorber, M. V. Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(Lr , Lr , 1) terms, and a new generalization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 695–720, 2013.

[42] Anh-Huy Phan, Petr Tichavský, and Andrzej Cichocki, "Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC," *SIAM J. Matrix Anal. App.*, vol. 354, no. 1, pp. 126–147, 2013.

[43] R. Bro, "Multi-way analysis in the food industry: models, algorithms, and applications," Ph.D. thesis, University of Amsterdam, Amsterdam, The Netherlands, 1998.

[44] R. A. Harshman, "Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, 16 (1970), pp. 1–84.

[45] K.-H. Ngo, A. Decurninge, M. Guillaud, and S. Yang, "Cube-split: A structured Grassmannian constellation for non-coherent SIMO communications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1948–1964, Mar. 2020.

[46] A. Decurninge, I. Land, and M. Guillaud, "Tensor-based modulation for unsourced massive random access," 2020. [Online]. Available: https://arxiv.org/abs/2006.06797.

[47] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced line search: a novel method to accelerate PARAFAC," *SIAM J. Matrix Anal. Appl.*, 30 (2008), pp. 1128–1147.

[48] Solutions for NR to Support Non-Terrestrial Networks (NTN), document TR 38.821 v16.2.0, 3GPP, Apr. 2023.