

# Tensor-based Unified Joint Channel Estimation and Active Device Detection Scheme for High-Mobility Grant-Free Random Access Scenarios

Ziqi Kang, Dongxuan He, *Member, IEEE*, Hua Wang, *Member, IEEE*,

Zhaocheng Wang, *Fellow, IEEE*, and Zhu Han, *Fellow, IEEE*

**Abstract**—With the rapid development of Internet of Things (IoT), efficient and reliable massive IoT device connections need to be widely supported in the upcoming next-generation communication networks, especially for emerging high-mobility scenarios. In this context, this paper investigates massive grant-free random access (GF-RA) in high mobility scenarios, focusing on active device detection (ADD) and channel estimation (CE) under fast time-varying channels. By exploiting the inherent low-rank structure of the observed pilot-signal-tensor, a tensor-based GF-RA transmission scheme is provided. On this basis, we propose a joint ADD and CE method based on the canonical polyadic (CP) model for both sourced and unsourced RA frameworks. More specifically, by remodelling the observation signal as a third-order tensor, the channel parameters can be grouped in the factor matrices of the CP model. However, the excessive number of potential device connections in massive GF-RA scenarios lead to excessively large dimensions of the factor matrices, thus resulting in severe ill-condition. To solve this problem, the Vandermonde structure of factor matrices is developed, which enables the effective exploitation of the tensor subspace for CP decomposition. Then, by utilizing the pre-allocated training precoders, an effective two-dimensional search method is proposed to jointly detect active devices and initialize the iterative estimation of channel parameters. Finally, due to the grouping situation, independent and coupled channel parameters are estimated by appropriate methods based on maximum likelihood (ML) and iterative updating, respectively. Moreover, the pre-allocation of training precoders can be unified to the unsourced RA scenarios, where the joint ADD and CE can be regard as a simple degenerate method compared to sourced RA. Simulation results demonstrate that the proposed tensor-based GF-RA framework outperforms the state-of-the-art schemes in terms of both ADD and CE performance.

**Index Terms**—grant-free random access, tensor decomposition, high mobility, correlation-based, channel estimation, active device detection.

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFE0200404; in part by the NSF ECCS-2302469, CMMI-2222810, Toyota, Amazon and Japan Science and Technology Agency (JST) Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) JPMJAP2326. (*Corresponding author: Dongxuan He; Hua Wang*)

Z. Kang, D. He and H. 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 )

Z. Wang is with Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (E-mail: zcwang@tsinghua.edu.cn)

Z. Han is with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea, 446-701 (E-mail: hanzhu22@gmail.com)

Copyright (c) 2025 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

## I. INTRODUCTION

MASSIVE machine type communication (mMTC) has been regarded as one major application scenario of the fifth generation (5G). However, the efficient access of massive Internet of Things (IoT) devices is still very challenging. By 2030, the number of IoT devices worldwide will reach 500 billion, where the conventional 4G and 5G mobile communication networks would hardly support such massive device access requirements of future IoT applications effectively. As such, for the upcoming beyond fifth generation (B5G) and the sixth generation (6G) mobile communication networks, efficient and reliable IoT device access is still a key issue for massive communication in academic and industry [1][2]. On the one hand, although the development of many advanced multiple access technologies, such as space division multiple access (SDMA) [3]-[5] and rate-splitting multiple access (RSMA) [6], have been further adopted in IoT, the number of supported access devices is still very limited. On the other hand, the widely used authorized random access (RA) protocol requires multiple handshakes to accurately schedule device access, where the excessive signaling overhead and interaction will make it inefficient when a large number of device accesses simultaneously [7].

In recent years, grant-free random access (GF-RA) technology has received widespread attention, which allows IoT terminals to directly transmit their non-orthogonal preamble sequences through the uplink transmission, thus avoiding the complex access request for resource scheduling in large-scale access scenarios [8]. Depending on whether the device's identity (ID) is considered, GF-RA can be divided into two categories, namely, sourced random access and unsourced random access [9]. The unsourced RA base station (BS) only needs to detect which devices are activated to transmit messages, while sourced RA BS identifies the specific IDs of these devices. In fact, regardless of the type of RA, the received non-orthogonal pilot sequence based active device detection (ADD) is one of the critical issues of GF-RA. By fully utilizing the inherent feature of sporadic access caused by the bursty characteristics of IoT applications, one popular solution is to model this problem as a sparse signal reconstruction problem based on compressed sensing (CS) [11]-[14]. For example, the active device detection methodology based on the results of sparse channel estimation (CE) was proposed in [11] and [12], where orthogonal matching pursuit (OMP) is adopted. Benefiting from its fast convergence, OMP can efficiently reconstruct

TABLE I  
A COMPARISON OF THE RELATED LITERATURE WITH OUR WORK

Reference	Scenario	Modeling	Transmission mode	Target issue	Proposed method	Benchmark
[15]	LEO + IoT	MIMO-OTFS + sparse recovery	uplink	CE + ADD	OMP	oracle-LS
[16]	LEO + IoT	MIMO-OTFS + sparse recovery	uplink	CE + ADD	ConvSBL-GAMP	SBL-GAMP / GMMV-AMP
[17]	LEO + IoT	single carrier + tensor decomposition	uplink	CE + ADD	tensor-based Bayesian learning	OMP / AMP
[18]	mmWave	MIMO-OFDM + tensor decomposition	uplink	CE	CP-ALS	OMP
[19]	mmWave	MIMO-OFDM + tensor decomposition	downlink	CE	CP-ALS	OMP
[20]	mmWave	MIMO-OFDM + tensor decomposition	downlink	CE	SCPD	ALS / OMP
[21]	mmWave + high-mobility	MIMO-OFDM + tensor decomposition	downlink	CE	CP-TVFS + CRAF	ALS / A-BOMP
[22]	mmWave + high-mobility	MIMO-OFDM + tensor decomposition	uplink	CE	subspace-based	ALS / OMP
[23]	mmWave + high-mobility	MIMO-OTFS + tensor decomposition	uplink	CE	tensor-based OMP	OMP / S-OMP
Our work	IoT + high-mobility	MIMO-OFDM + tensor decomposition	uplink	CE + ADD	subspace-based + correlation-based	ALS / OMP

the target sparse signal but its reconstruction accuracy is limited. By introducing the message passing algorithm into CS, the approximate message passing (AMP) algorithm based receivers was designed in [13] and [14], which can improve both channel estimation and device detection performance simultaneously. However, due to the limitations in system and algorithm design, most studies can only deal with traditional low-mobility ground scenarios, while the widespread high-mobility scenarios in practice fail to be sufficiently supported.

The main challenge of high-mobility scenarios is the severe Doppler effect, which leads to fast time-varying channel and in turn limits the performance of CE and ADD. Taking low-orbit satellites (LEO) as an example, the authors in [15] revealed that the performance of the framework and algorithms designed based on traditional low-mobility scenarios suffers severe performance loss in high-mobility scenarios. Besides,

the generalized AMP (GAMP) was combined with sparse Bayesian learning (SBL) to conduct joint ADD and CE in [16], which confirmed the adverse impact of high mobility on conventional GF-RA schemes. By combining CS-based methods with new modulation scheme, i.e., orthogonal time frequency space (OTFS) modulation, these works can demonstrate superior performance even under fast time-varying channel. To further mitigate the effect of high mobility, the device's preamble sequence was divided into segments to apply the tensor model in the form of rank-1 decomposition, where the Bayesian learning was also exploited to solve the joint ADD and CE [17]. In fact, the authors in [17] demonstrated that the methods based on tensor decomposition have better performance compared to conventional CS-based methods such as OMP and AMP in high-mobility scenarios. However, little works have considered how to design the framework of

GF-RA based on tensor model.

Indeed, the tensor decomposition model has been widely applied to complete CE in conventional scenarios such as millimeter wave (mmWave) massive MIMO since it can fully exploit the inherent structure of the channel and the received pilot signal [18]-[20]. For instance, by strategically decomposing and combining the received pilot signals, the downlink single-user MIMO-CE can be modeled as a third-order canonical polyadic (CP) decomposition and reconstruction problem, where the method based on alternating least squares (ALS) optimization was proposed to recover the factor matrices containing channel parameters [18][19]. In addition, these parameters are estimated by simple correlation-based methods whose performance can be guaranteed by grid-based one-dimensional search. The authors in [19] revealed that the proposed CP-ALS-correlation framework outperforms OMP in both complexity and performance, which made it wildly followed by other works. For example, a tensor model of spatial frequency channels for the dual-bandwidth effect in millimeter-wave massive MIMO scenarios was designed in [20], where the proposed spatial smoothing method for CE reached the better performance compared to the traditional simultaneous weighted-OMP (SW-OMP). However, despite the inherent extensibility of these tensor-based channel estimation methods in dealing with high-mobility scenarios, the majority of research in this area still primarily focuses on traditional low-mobility scenarios.

Recently, by comprehensively considering the channel characteristics in high-mobility scenarios, a CE method based on the tensor CP model was proposed in [21], which characterizes the impact of fast time-varying channels as an additional parameter to be estimated. Nevertheless, such modeling requires reliable estimation of coupled channel parameters, where the conventional correlation-based methods fail to work. Therefore, an iteratively joint estimation algorithm was also proposed, effectively solving the CE problem brought by extra coupled parameter. Although the authors in [21] showed the application potential of tensor-based methods in high-mobility scenarios, the estimation performance is still limited to the simple downlink single-user transmission. In this regard, the impact of multiple users on the tensor model was characterized as the accumulation of the total number of transmission paths in [22]. Since the time delay was not considered in the channel modeling, the challenge of multi-parameter estimation which is supposed to be faced from one factor matrix was evidently avoided. To fully consider the time delay and Doppler shift in the uplink multi-user transmission, a distinct tensor model instead of the CP was developed to explore and utilize the inherent characteristics of the fast time-varying channel model [23]. Nevertheless, though the tensor-based method in [23] still outperforms OMP and S-OMP, it will significantly deteriorate when the number of devices is large. Moreover, the proposed scheme relies heavily on the unique pilot design pattern, which is very challenging in GF-RA with a large number of device accesses. These studies motivate us to develop efficient GF-RA schemes with the help of accurate CE in high-mobility IoT scenarios, which benefits from the excellent tensor-based methods. For clarity, the comparison of the aforementioned

related works is summarized in TABLE I.

Against the background, this paper intends to design an effective unified GF-RA framework based on the CP model of tensor decomposition for uplink high-mobility scenarios. The contributions of this paper are summarized as follows:

- 1) A tensor-based GF-RA framework for high-mobility scenarios is proposed. By strategically performing identity transformation to rearrange the tensor CP model, the Vandermonde structure of factor matrix is exploited to facilitate a dimension reduction method, which can solve the ill-conditioned problem of obtaining the factor matrices, thus helping the following CE and ADD. For the scenario specifically targeting massive device access, i.e., for large-sized factor matrices to be estimated in the proposed framework, the accuracy and convergence of tensor decomposition can be well guaranteed.
- 2) An effective correlation-based ADD method independent of complete channel state information (CSI) is proposed. Particularly, considering whether the channel parameters are coupled or independent, two appropriate parameter estimation schemes are proposed respectively. On this basis, by exploring and utilizing the device activity information contained in the factor matrix where certain channel parameters are located, a simple and effective detection method is further designed. In contrast to the state-of-the-art complete CSI-aided ADD methods, the proposed method is executed more efficiently, which can be processed in parallel with the estimated channel parameters.
- 3) The extensibility of the proposed framework for ADD in unsourced RA application scenarios has been derived and analyzed. Specifically, due to the common IDs of active devices, the proposed detection method can be directly degraded into a more concise and effective form, which can support further investigation on subsequent data detection in the future. The flexibility and adaptability of the proposed ADD method facilitates our tensor-based unified framework to meet different GF-RA requirements.

The rest of this paper is organized as follows. In Section II, we introduce the considered GF-RA transmission scheme in high-mobility scenarios. Section III provides the CP-based tensor model of the observation signal, and then conducts the tensor decomposition based on conventional ALS and tensor subspace, respectively. After that, we propose a correlation-based detection method for active devices and diverse estimation methods for channel parameters, and analyze the potential expansion and computation complexity of the proposed scheme in section IV. Simulation results are presented in Section V, followed by conclusions in Section VI.

*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  $M \times M$ .  $\llbracket \cdot \rrbracket$  denotes the Kruskal operator.  $[\mathbf{x}]_l$ ,  $[\mathbf{X}]_{l,m}$ ,  $[\mathcal{X}]_{l,m,n}$ ,  $\mathbf{X}_l$ ,  $[\mathbf{X}]_{:,m:n}$

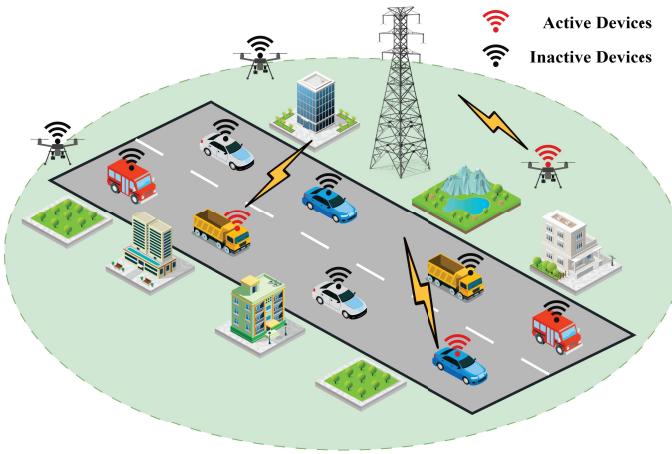


Fig. 1. An illustration of Grant-free random access system in high-mobility scenarios.

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.  $\otimes, \odot, *,$  and  $\circ$  denote Kronecker, Khatri-Rao, Hadamard, and outer product, respectively.  $\Lambda(\mathbf{x})$  denotes the diagonal matrix formed by  $\mathbf{x}$ . The real part of a complex value is denoted by  $\text{Re}(\cdot)$ , and  $j = \sqrt{-1}$ .  $|\mathbb{A}|_c$  is the cardinal number of the set  $\mathbb{A}$ .

## II. SYSTEM MODEL

### A. Time-Varying Channel in High Mobility Scenarios

As shown in Fig. 1, we consider a massive GF-RA system including multiple mobile devices with  $N_T$  antennas and a BS with  $N_R$  antennas, where  $K_0$  orthogonal subcarriers are employed to send  $N_s$  independent data streams. A hybrid MIMO architecture is utilized in our considered system, where the devices and BS are equipped with  $N_{\text{RF}}^T$  and  $N_{\text{RF}}^R$  chains, respectively. Let  $\mathbf{s}_{k,u}(n) \in \mathbb{C}^{N_s \times 1}$  be the transmitted symbol sequence of  $u$ -th device at the  $k$ -th subcarrier of the  $n$ -th frame,  $\forall k \in \{1, \dots, K_0\}$ ,  $\forall u \in \{1, \dots, U\}$ ,  $n = 1, 2, \dots, N$ , the transmitted precoded signal can be written by

$$\mathbf{p}_{k,u}(n) = \mathbf{P}_{\text{RF},u} \mathbf{P}_{\text{BB},k,u}(n) \mathbf{s}_{k,u}(n), \quad (1)$$

where  $\mathbf{P}_{\text{RF}} \in \mathbb{C}^{N_T \times N_{\text{RF}}^T}$  and  $\mathbf{P}_{\text{BB},k}(n) \in \mathbb{C}^{N_{\text{RF}}^T \times N_s}$  are the analog radio frequency (RF) precoder and digital baseband precoder, respectively. The equivalent combiner at the BS is given by

$$\mathbf{W}_k = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB},k}, \quad (2)$$

where  $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_R \times N_{\text{RF}}^R}$  and  $\mathbf{W}_{\text{BB},k} \in \mathbb{C}^{N_{\text{RF}}^R \times N_s}$  are the analog RF combiner and digital baseband combiner, respectively.

Due to the severe Doppler effect in high mobility scenarios, the time-varying massive MIMO channel matrix  $\mathbf{H}_{k,u}(n)$  in the frequency domain can be expressed as [25]

$$\mathbf{H}_{k,u}(n) = \sum_{l_u=1}^{L_u} \alpha_{l_u} e^{-j2\pi\tau_{l_u} f_s \frac{k}{K_0}} \mathbf{a}_R(\theta_{l_u}) \mathbf{a}_T^T(\phi_{l_u}) e^{j\nu_{l_u} n}, \quad (3)$$

where  $\alpha_{l_u}, \tau_{l_u}, \theta_{l_u}, \phi_{l_u}$ , and  $\nu_{l_u}$  denote the path gain, time delay, angles of arrival (AoA), angles of departure (AoD), and normalized Doppler shift of the  $l_u$ -th path, respectively,  $L_u$  is the number of paths at the  $u$ -th device, and  $f_s$  is the sampling rate. Assume the uniform linear array (ULA) is deployed, the two array steering vectors are given by [26]

$$\mathbf{a}_R(\theta_{l_u}) = \frac{1}{\sqrt{N_R}} \left[ 1, e^{j\frac{2\pi}{\lambda_c} d \sin \theta_{l_u}}, \dots, e^{j\frac{2\pi}{\lambda_c} (N_R-1) d \sin \theta_{l_u}} \right]^T, \quad (4)$$

$$\mathbf{a}_T(\phi_{l_u}) = \frac{1}{\sqrt{N_T}} \left[ 1, e^{j\frac{2\pi}{\lambda_c} d \sin \phi_{l_u}}, \dots, e^{j\frac{2\pi}{\lambda_c} (N_T-1) d \sin \phi_{l_u}} \right]^T, \quad (5)$$

where  $d$  and  $\lambda_c$  denote the antenna spacing and wavelength of the carrier frequency, respectively.

### B. Transmission Scheme of GF-RA

Assume that the activity possibility is  $\epsilon$ , the activity indicator  $a_u$  of the  $u$ -th device is given by

$$a_u = \begin{cases} 1, & p = \epsilon, \\ 0, & p = 1 - \epsilon. \end{cases} \quad (6)$$

To facilitate the channel estimation,  $K$  out of  $K_0$  subcarriers are exploited for training, where the frequency-flat beam training scheme is employed [21]. In addition, we assume that the precoders and combiners satisfy  $\mathbf{p}_{k,u}(n) = \mathbf{p}_u(n)$  and  $\mathbf{W}_k = \mathbf{W}$ ,  $\forall k \in \{1, \dots, K\}$  to formulate an efficient tensor-based CE method. Therefore, the received training signal at the BS can be expressed as

$$\begin{aligned} \mathbf{y}_k(n) &= \sum_{u=1}^U a_u \mathbf{W}^H \mathbf{H}_{k,u}(n) \mathbf{p}_u(n) + \mathbf{z}_k(n) \\ &= \sum_{u=1}^U \sum_{l_u=1}^{L_u} a_u \alpha_{l_u} e^{-j2\pi\tau_{l_u} f_s \frac{k}{K_0}} \mathbf{W}^H \mathbf{a}_R(\theta_{l_u}) \mathbf{a}_T^T(\phi_{l_u}) \\ &\quad \times \mathbf{p}_u(n) e^{j\nu_{l_u} n} + \mathbf{z}_k(n) \\ &= \sum_{u=1}^U \sum_{l_u=1}^{L_u} h_{l_u} e^{-j2\pi\tau_{l_u} f_s \frac{k}{K_0}} \tilde{\mathbf{a}}_R(\theta_{l_u}) \mathbf{a}_T^T(\phi_{l_u}) \\ &\quad \times \mathbf{p}_u(n) e^{j\nu_{l_u} n} + \mathbf{z}_k(n) \\ &= \sum_{l=1}^L h_l e^{-j2\pi\tau_l f_s \frac{k}{K_0}} \tilde{\mathbf{a}}_R(\theta_l) \mathbf{a}_T^T(\phi_l) \bar{\mathbf{p}}_{l,u}(n) e^{j\nu_l n} + \mathbf{z}_k(n), \end{aligned} \quad (7)$$

where  $L = UL_u$  collects the paths of all devices,  $h_{l_u} = a_u \alpha_{l_u}$  denotes the equivalent path gain,  $\bar{\mathbf{p}}_{l,u}(n)$  represents  $\mathbf{p}_u(n)$  when the  $l$ -th path belongs to the  $u$ -th device, subject to  $l = \sum_{j=1}^{u-1} L_j + l_u$ , and  $\mathbf{z}_k(n) \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$  is the additive white Gaussian noise with mean zero and variance  $\sigma_n^2$ .

## III. TENSOR MODEL AND DECOMPOSITION

In this section, we firstly derive the CP-based tensor form of the received training signal. Then, the conventional ALS-based tensor decomposition and subspace-based tensor decomposition are presented to handle the CP model.

### A. CP-Based Tensor Model

By collecting the received training signals  $\mathbf{y}_k(n)$  from all  $N$  frames, (7) can be rewritten as

$$\begin{aligned}\mathbf{Y}_k &= \sum_{l=1}^L h_l e^{-j2\pi\tau_l f_s \frac{k}{K_0}} \tilde{\mathbf{a}}_R(\theta_l) \mathbf{a}_T^T(\phi_l) \bar{\mathbf{P}}_{l,u} \mathbf{\Xi}(\nu_l) + \mathbf{Z}_k \\ &= \sum_{l=1}^L h_l e^{-j2\pi\tau_l f_s \frac{k}{K_0}} \tilde{\mathbf{a}}_R(\theta_l) \tilde{\mathbf{a}}_T^T(\phi_l) \mathbf{\Xi}(\nu_l) + \mathbf{Z}_k,\end{aligned}\quad (8)$$

where  $\mathbf{Y}_k = [\mathbf{y}_k(1), \dots, \mathbf{y}_k(N)] \in \mathbb{C}^{N_{RF}^R \times N}$  contains the overall received measurements at the  $k$ -th subcarrier,  $\forall k = \{1, \dots, K\}$ ,  $\bar{\mathbf{P}}_{l,u} = [\bar{\mathbf{p}}_{l,u}(1), \dots, \bar{\mathbf{p}}_{l,u}(N)] \in \mathbb{C}^{N_T \times N}$  denotes the training precoder,  $\tilde{\mathbf{a}}_T(\phi_l) \triangleq \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l) \in \mathbb{C}^{N_T \times N}$  is the equivalent steering vector after precoding,  $\mathbf{\Xi}(\nu_l) = \Lambda([e^{j\nu_l}, \dots, e^{j\nu_l N}])$  collects Doppler shift terms  $e^{j\nu_l n}$  as its diagonal elements,  $\mathbf{Z}_k = [\mathbf{z}_k(1), \mathbf{z}_k(2), \dots, \mathbf{z}_k(N)] \in \mathbb{C}^{N_{RF}^R \times N}$  denotes the noise matrix.

Consequently, by concatenating the received signal matrices in  $K$  subcarriers, a low-rank structure tensor  $\mathcal{Y}$  enabling the CP decomposition can be obtained according to the following proposition.

*Proposition 1:* Given the slices of  $\mathcal{Y}$  as  $\mathbf{Y}_k$  in (8), the received low-rank tensor signal  $\mathcal{Y}$  can be expressed as

$$\mathcal{Y} = [\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}] + \mathcal{Z} = \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ \tilde{\mathbf{c}}(\tau_l) + \mathcal{Z}, \quad (9)$$

where  $\mathcal{Z}$  is the noise tensor,  $\tilde{\mathbf{b}}_T(\nu_l, \phi_l) = \mathbf{\Xi}(\nu_l) \tilde{\mathbf{a}}_T(\phi_l)$  is the equivalent steering vector containing the normalized Doppler shift  $\nu_l$ ,  $\tilde{\mathbf{c}}(\tau_l) = h_l \mathbf{c}(\tau_l) \in \mathbb{C}^{K \times 1}$ , where  $\mathbf{c}(\tau_l)$  is defined by

$$\mathbf{c}(\tau_l) = [e^{-j2\pi\tau_l f_s \frac{1}{K_0}}, \dots, e^{-j2\pi\tau_l f_s \frac{K}{K_0}}]^T. \quad (10)$$

Note that, (10) can be viewed as a frequency-domain steering vector with respect to the time delay  $\tau_l$ .

*Proof:* See Appendix A.

Leveraging the structured CP model in proposition 1, the unknown channel parameters  $\{\theta_l, \nu_l, \phi_l, h_l, \tau_l\}$  can be estimated from the three factor matrices obtained by decomposing the third-order observation tensor  $\mathcal{Y}$ , given by

$$\tilde{\mathbf{A}} \triangleq [\tilde{\mathbf{a}}_R(\theta_1), \dots, \tilde{\mathbf{a}}_R(\theta_L)] \in \mathbb{C}^{N_{RF}^R \times L}, \quad (11)$$

$$\tilde{\mathbf{B}} \triangleq [\tilde{\mathbf{b}}_T(\nu_1, \phi_1), \dots, \tilde{\mathbf{b}}_T(\nu_L, \phi_L)] \in \mathbb{C}^{N_T \times L}, \quad (12)$$

$$\tilde{\mathbf{C}} \triangleq [h_1 \mathbf{c}(\tau_1), \dots, h_L \mathbf{c}(\tau_L)] \in \mathbb{C}^{K \times L}, \quad (13)$$

which satisfy the following requirement, given by

$$\underset{\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}}{\text{minimize}} \left\| \mathcal{Y} - \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\phi_l, \nu_l) \circ \tilde{\mathbf{c}}(\tau_l) \right\|_F^2. \quad (14)$$

### B. Conventional ALS-Based Tensor Decomposition

To solve the tensor-decomposition problem shown in (14), ALS turns to be an efficient method, which alternatively updates one factor matrix by fixing the other two factor

matrices [27]. More specifically, the updating process in the  $i$ -th iterations can be expressed as

$$\tilde{\mathbf{A}}^{(i)} = \arg \min_{\tilde{\mathbf{A}}} \left\| \mathbf{Y}_{(1)}^T - (\tilde{\mathbf{C}}^{(i-1)} \odot \tilde{\mathbf{B}}^{(i-1)}) \tilde{\mathbf{A}}^T \right\|_F^2, \quad (15)$$

$$\tilde{\mathbf{B}}^{(i)} = \arg \min_{\tilde{\mathbf{B}}} \left\| \mathbf{Y}_{(2)}^T - (\tilde{\mathbf{C}}^{(i-1)} \odot \tilde{\mathbf{A}}^{(i)}) \tilde{\mathbf{B}}^T \right\|_F^2, \quad (16)$$

$$\tilde{\mathbf{C}}^{(i)} = \arg \min_{\tilde{\mathbf{C}}} \left\| \mathbf{Y}_{(3)}^T - (\tilde{\mathbf{B}}^{(i)} \odot \tilde{\mathbf{A}}^{(i)}) \tilde{\mathbf{C}}^T \right\|_F^2. \quad (17)$$

By updating  $\tilde{\mathbf{A}}$ ,  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$  iteratively in an alternating fashion, ALS method can converge to a stationary solution of (14) under the mild condition. Once convergent, the ALS method provides the factor matrix estimates as  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$ , which are subject to two types of ambiguities, namely, scaling and permutation. To be specific, the estimated factor matrices  $\{\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}\}$  can be presented as

$$\hat{\mathbf{A}} = \tilde{\mathbf{A}} \Delta_{\tilde{\mathbf{A}}} \Pi + \mathbf{E}_{\tilde{\mathbf{A}}}, \quad (18)$$

$$\hat{\mathbf{B}} = \tilde{\mathbf{B}} \Delta_{\tilde{\mathbf{B}}} \Pi + \mathbf{E}_{\tilde{\mathbf{B}}}, \quad (19)$$

$$\hat{\mathbf{C}} = \tilde{\mathbf{C}} \Delta_{\tilde{\mathbf{C}}} \Pi + \mathbf{E}_{\tilde{\mathbf{C}}}, \quad (20)$$

where  $\Delta_{\tilde{\mathbf{A}}}$ ,  $\Delta_{\tilde{\mathbf{B}}}$  and  $\Delta_{\tilde{\mathbf{C}}}$  are unknown diagonal matrices satisfying  $\Delta_{\tilde{\mathbf{A}}} \Delta_{\tilde{\mathbf{B}}} \Delta_{\tilde{\mathbf{C}}} = \mathbf{I}_L$ ,  $\Pi \in \mathbb{C}^{L \times L}$  is the unknown permutation matrix,  $\mathbf{E}_{\tilde{\mathbf{A}}}$ ,  $\mathbf{E}_{\tilde{\mathbf{B}}}$  and  $\mathbf{E}_{\tilde{\mathbf{C}}}$  represent the corresponding estimation errors.

### C. Subspace-Based Tensor Decomposition

The traditional ALS enjoys desirable performance in estimating the factor matrix [28]. However, as the number of potential devices increases, the number of columns in the factor matrix increases significantly, which will in turn affect the ALS optimization obviously. Firstly, the convergence of ALS may become uncertain, which cannot be guaranteed even if the number of iterations is increased [29]. Secondly, even if ALS converges after adequate iterations, the error between the estimated factor matrix and the actual factor matrix will still be significant, i.e.,  $\mathbf{E}_{\tilde{\mathbf{A}}}$ ,  $\mathbf{E}_{\tilde{\mathbf{B}}}$  and  $\mathbf{E}_{\tilde{\mathbf{C}}}$  are too large to enable the channel parameters be reliably estimated from the factor matrix<sup>1</sup>. Therefore, the tensor decomposition is badly in need of restraining the numerical instability to tackle this issue.

*1) Equivalent Transformation of Tensor Model:* To solve this ill-conditioned problem, we fully utilize the basic characteristic of CP model to conduct a strategic equivalent transformation to the original estimation problem in the following proposition.

*Proposition 2:* Given the observation tensor  $\mathcal{Y}$  in (9), the decomposition can be handled on the linear subspace to facilitate the subsequent dimensionality reduction, which should be restructured by

$$\mathcal{Y} = [\mathbf{A}, \mathbf{B}, \mathbf{C}] + \mathcal{Z} = \sum_{l=1}^L (\bar{h}_l \tilde{\mathbf{a}}_R(\theta_l)) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ \tilde{\mathbf{c}}(\tau_l) + \mathcal{Z}, \quad (21)$$

<sup>1</sup>The elements of factor matrices are obtained by practical problems, whose inherent errors are amplified by the significantly increased columns, thus deteriorating the accuracy of ALS optimization even under the convergence condition.

where the equivalent path gain  $\bar{h}_l = h_l e^{-j2\pi\tau_l f_s \frac{1}{K_0}}$  contains the information of delay  $\tau_l$ , and the steering vector with respect to  $\tau_l$  in (10) can be rewritten as

$$\bar{\mathbf{c}}(\tau_l) = [1, e^{-j2\pi\tau_l f_s \frac{1}{K_0}}, \dots, e^{-j2\pi\tau_l f_s \frac{K-1}{K_0}}]^T. \quad (22)$$

*Proof:* See Appendix B.

According to the proposition 2, the rearranged factor matrices are defined as

$$\mathbf{A} \triangleq [\bar{h}_1 \tilde{\mathbf{a}}_{\mathbf{R}}(\theta_1), \dots, \bar{h}_L \tilde{\mathbf{a}}_{\mathbf{R}}(\theta_L)] \in \mathbb{C}^{N_{\text{RF}}^{\text{R}} \times L}, \quad (23)$$

$$\mathbf{B} \triangleq [\tilde{\mathbf{b}}_{\mathbf{T}}(\nu_1, \phi_1), \dots, \tilde{\mathbf{b}}_{\mathbf{T}}(\nu_L, \phi_L)] \in \mathbb{C}^{N \times L}, \quad (24)$$

$$\mathbf{C} \triangleq [\bar{\mathbf{c}}(\tau_1), \dots, \bar{\mathbf{c}}(\tau_L)] \in \mathbb{C}^{K \times L}, \quad (25)$$

It is evident that, new coupling relationships of parameters are introduced, which allows the calculation of factor matrices more effectively in the low-dimensional subspace compared to the conventional ALS method.

2) *Vandermonde Structure-Based Subspace Processing:* To solve the CP decomposition in subspace, the structural characteristics of the factor matrix is utilized, i.e., the Vandermonde structure [22]. Specifically, the entries of each column in  $\mathbf{C}$  constitute a geometric series, which can be formulated by a set of distinct generators  $\{g_l = e^{-j2\pi\tau_l f_s \frac{1}{K_0}}\}_{l=1}^L$ . Therefore, by combining the Vandermonde structure of  $\mathbf{C}$  and the fundamental linear algebra [30], the CP decomposition in (21) can be solved in the matrix subspace. Firstly, a group of key intermediate variables  $(\Omega_1, \Omega_2)$  subjected to  $\Omega_1 + \Omega_2 = K+1$  are chosen to define a cyclic selection matrix, given by

$$\boldsymbol{\Omega}_l = [\mathbf{0}_{\Omega_1 \times (l-1)} \quad \mathbf{I}_{\Omega_1} \quad \mathbf{0}_{\Omega_1 \times (\Omega_2-l)}] \in \mathbb{C}^{\Omega_1 \times T}, \quad (26)$$

where  $\Omega_1 = \lfloor (K+1)/2 \rfloor$  and  $\Omega_2 = K+1 - \Omega_1$  are selected for simplicity. Then, based on the basic rules of tensor unfolding, the slice of  $\mathcal{Y}$  in the first dimension can be written as [27]

$$\mathbf{Y}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T \in \mathbb{C}^{N_{\text{RF}}^{\text{R}} \times NK}. \quad (27)$$

Left multiplying  $\boldsymbol{\Omega}_l$  by  $\mathbf{C}$  can select its submatrix from the  $l$ -th to the  $(l + \Omega_1)$ -th rows, i.e.,  $[\mathbf{C}]_{l:l+\Omega_1}$ . Since  $\mathbf{C}$  is a Vandermonde matrix, we have

$$\boldsymbol{\Omega}_l \mathbf{C} = \boldsymbol{\Omega}_1 \mathbf{C} \boldsymbol{\Lambda} ([\mathbf{C}]_{l,:}). \quad (28)$$

As a result, we can expand the dimension of  $\mathbf{Y}_{(1)}$  as

$$\begin{aligned} \mathbf{Y}_{(1)}^{\Omega} &= [(\boldsymbol{\Omega}_1 \otimes \mathbf{I}_N) \mathbf{Y}_{(1)}^T, \dots, (\boldsymbol{\Omega}_{\Omega_2} \otimes \mathbf{I}_N) \mathbf{Y}_{(1)}^T] \\ &= [((\boldsymbol{\Omega}_1 \mathbf{C}) \odot \mathbf{B}) \mathbf{A}^T, \dots, ((\boldsymbol{\Omega}_{\Omega_2} \mathbf{C}) \odot \mathbf{B}) \mathbf{A}^T] \\ &= [((\boldsymbol{\Omega}_1 \mathbf{C} \boldsymbol{\Lambda} ([\mathbf{C}]_{1,:})) \odot \mathbf{B}) \mathbf{A}^T, \dots, \\ &\quad ((\boldsymbol{\Omega}_1 \mathbf{C} \boldsymbol{\Lambda} ([\mathbf{C}]_{\Omega_2,:})) \odot \mathbf{B}) \mathbf{A}^T] \\ &= [((\boldsymbol{\Omega}_1 \mathbf{C}) \odot \mathbf{B} \boldsymbol{\Lambda} ([\mathbf{C}]_{1,:})) \mathbf{A}^T, \dots, \\ &\quad ((\boldsymbol{\Omega}_1 \mathbf{C}) \odot \mathbf{B} \boldsymbol{\Lambda} ([\mathbf{C}]_{\Omega_2,:})) \mathbf{A}^T] \\ &= ([\mathbf{C}]_{1:\Omega_1,:} \odot \mathbf{B}) [\boldsymbol{\Lambda}([\mathbf{C}]_{1,:}) \mathbf{A}^T, \dots, \boldsymbol{\Lambda}([\mathbf{C}]_{\Omega_2,:}) \mathbf{A}^T] \\ &= ([\mathbf{C}]_{1:\Omega_1,:} \odot \mathbf{B}) ([\mathbf{C}]_{1:\Omega_2,:} \odot \mathbf{A})^T, \end{aligned} \quad (29)$$

where  $([\mathbf{C}]_{1:\Omega_1,:} \odot \mathbf{B})$  and  $([\mathbf{C}]_{1:\Omega_2,:} \odot \mathbf{A})$  are of full column rank and row rank, respectively. Let  $\mathbf{Y}_{(1)}^{\Omega} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H$  denote the truncated singular value decomposition (SVD) of  $\mathbf{Y}_{(1)}^{\Omega}$ . Since the columns of  $\mathbf{U}$  span the same subspace as the columns of  $\mathbf{Y}_{(1)}^{\Omega}$ , there exists a nonsingular matrix  $\mathbf{M} \in \mathbb{C}^{L \times L}$  satisfying

$$\mathbf{U} \mathbf{M} = [\mathbf{C}]_{1:\Omega_1,:} \odot \mathbf{B}, \quad (30)$$

$$(\mathbf{M}^{-1} \boldsymbol{\Sigma} \mathbf{V}^H)^T = [\mathbf{C}]_{1:\Omega_2,:} \odot \mathbf{A}. \quad (31)$$

Similarly, by leveraging the Vandermonde structure of  $\mathbf{C}$ , we can define two partially overlapping matrices related to its submatrix  $[\mathbf{C}]_{1:\Omega_1,:}$ , given by

$$\mathbf{C}_{1,\Omega_1} \triangleq [[\mathbf{C}]_{1:\Omega_1,:}]_{1:\Omega_1-1,:} \in \mathbb{C}^{(\Omega_1-1) \times L}, \quad (32)$$

$$\mathbf{C}_{2,\Omega_1} \triangleq [[\mathbf{C}]_{1:\Omega_1,:}]_{2:\Omega_1,:} \in \mathbb{C}^{(\Omega_1-1) \times L}. \quad (33)$$

Here,  $\mathbf{C}_{1,\Omega_1}$  and  $\mathbf{C}_{2,\Omega_1}$  are apparently submatrices of  $\mathbf{C}$  whose entries subjected to the characteristic of geometric series. Hence, by introducing the generating matrix  $\mathbf{G} = \boldsymbol{\Lambda}([g_1, g_2, \dots, g_L])$  composed by the generators of  $\mathbf{C}$ , the relationship of  $\mathbf{C}_{1,\Omega_1}$  and  $\mathbf{C}_{2,\Omega_1}$  can be expressed as

$$\mathbf{C}_{2,\Omega_1} = \mathbf{C}_{1,\Omega_1} \mathbf{G}. \quad (34)$$

3) *Estimation of Factor Matrices:* Considering the size of these two submatrices and combining (30), (31) and (34), we divide the unitary matrix  $\mathbf{U}$  into two submatrices to facilitate the estimation of factor matrices, which can be expressed as

$$\mathbf{U}_1 \triangleq [\mathbf{U}]_{1:(\Omega_1-1)N,:} \in \mathbb{C}^{(\Omega_1-1)N \times L}, \quad (35)$$

$$\mathbf{U}_2 \triangleq [\mathbf{U}]_{N+1:\Omega_1 N,:} \in \mathbb{C}^{(\Omega_1-1)N \times L}. \quad (36)$$

where  $\mathbf{U}_1$  and  $\mathbf{U}_2$  satisfy

$$\mathbf{U}_2 \mathbf{M} = \mathbf{U}_1 \mathbf{M} \mathbf{G}. \quad (37)$$

Clearly, (37) can be directly rewritten as

$$\mathbf{U}_1^\dagger \mathbf{U}_2 = \mathbf{M} \mathbf{G} \mathbf{M}^{-1}, \quad (38)$$

which indicates a typical similarity transformation that  $\mathbf{G} \sim \mathbf{U}_1^\dagger \mathbf{U}_2$ . For convenience, we define the similarity matrix of  $\mathbf{G}$  as  $\tilde{\mathbf{G}} = \mathbf{U}_1^\dagger \mathbf{U}_2$ . Recalling the Vandermonde structure of  $\mathbf{C}$  and the definition of  $\mathbf{G}$ , it can be noticed that the information of delay  $\tau_l$  can be estimated from the eigenvalues of  $\tilde{\mathbf{G}}$ . Let  $\tilde{\mathbf{G}} = \mathbf{Q} \mathbf{D} \mathbf{Q}^{-1}$  denote the eigenvalue decomposition (EVD) of  $\tilde{\mathbf{G}}$ , we can obtain

$$\boldsymbol{\lambda}(\mathbf{D}) = \boldsymbol{\lambda}(\mathbf{G}) \boldsymbol{\Pi}, \quad (39)$$

$$\mathbf{Q} = \mathbf{M} \boldsymbol{\Delta} \boldsymbol{\Pi}, \quad (40)$$

where  $\boldsymbol{\Delta}$  and  $\boldsymbol{\Pi}$  represent the diagonal scaling matrix and permutation matrix, respectively. Thus, the eigenvalues of  $\tilde{\mathbf{G}}$  are the diagonal elements of  $\mathbf{G}$  with permutation ambiguity. By exploiting the estimated generators, the Vandermonde matrix  $\hat{\mathbf{C}}' = [\hat{\mathbf{C}}'_1, \dots, \hat{\mathbf{C}}'_L]$  can be recovered, given by

$$\hat{\mathbf{C}}' = \mathbf{C} \boldsymbol{\Pi}. \quad (41)$$

For convenience, we rewrite the  $l$ -th column of the submatrix of  $\mathbf{C}$  in (30) and (31), which can be expressed as

$$\left[ \left[ \hat{\mathbf{C}}' \right]_{1:\Omega_1,:} \right]_l \triangleq \hat{\mathbf{C}}'_{\Omega_1,l}, \quad (42)$$

$$\left[ \left[ \hat{\mathbf{C}}' \right]_{1:\Omega_2,:} \right]_l \triangleq \hat{\mathbf{C}}'_{\Omega_2,l}. \quad (43)$$

Consequently, the estimate of the rest factor matrices  $\mathbf{B}$  and  $\mathbf{A}$  can be obtained with the help of  $\hat{\mathbf{C}}'$  as the following proposition shows.

*Proposition 3:* Given the result of SVD in (30)-(31) and EVD in (39)-(40),  $\hat{\mathbf{B}}'$  and  $\hat{\mathbf{A}}'$  can be recovered column by column based on the mixed product property of Kronecker product, the  $l$ -th column of which are respectively given by

$$\hat{\mathbf{B}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_1,l}}{\Omega_1} \otimes \mathbf{I}_N \right) \mathbf{U} \mathbf{Q}_l, \quad (44)$$

$$\hat{\mathbf{A}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_2,l}}{\Omega_2} \otimes \mathbf{I}_{N_{RF}^R} \right) \mathbf{V}^* \boldsymbol{\Sigma} [\mathbf{Q}^{-T}]_{:,l}. \quad (45)$$

*Proof:* See Appendix C.

In summary, considering the estimation error, the recovered factor matrices and the true factor matrices are related as

$$\hat{\mathbf{A}}' = \mathbf{A} \Delta_{\mathbf{A}} \Pi + \mathbf{E}_{\mathbf{A}} = \mathbf{A} \Delta^{-1} \Pi + \mathbf{E}_{\mathbf{A}}, \quad (46)$$

$$\hat{\mathbf{B}}' = \mathbf{B} \Delta_{\mathbf{B}} \Pi + \mathbf{E}_{\mathbf{B}} = \mathbf{B} \Delta \Pi + \mathbf{E}_{\mathbf{B}}, \quad (47)$$

$$\hat{\mathbf{C}}' = \mathbf{C} \Delta_{\mathbf{C}} \Pi + \mathbf{E}_{\mathbf{C}} = \mathbf{C} \Pi + \mathbf{E}_{\mathbf{C}}, \quad (48)$$

For the sake of notation simplification, the prime of  $\hat{\mathbf{A}}'$ ,  $\hat{\mathbf{B}}'$  and  $\hat{\mathbf{C}}'$  will be dropped in the sequel.

#### IV. JOINT ACTIVE DEVICES DETECTION AND CHANNEL ESTIMATION

In this section, based on the above analysis, a joint ADD and CE method is designed. Besides, the computational complexity of the proposed method is analyzed.

##### A. Overview of The Proposed Framework

Based on the proposed tensor-based GF-RA framework, an efficient joint ADD and CE scheme relying on the training precoders is designed, which can be divided into the following four parts:

1) *Factor Matrices Acquisition:* The channel matrix to be estimated is divided into three sets of parameters, which can be estimated by three factor matrices, regardless of whether they are coupled or independent. Therefore, effective estimation of the factor matrices is an important foundation for completing ADD and CE. This step has been accurately completed through Section III, i.e., the recovered  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{C}}$ .

2) *Activity Detection and Coupled Parameter Estimation:* Recalling the rank-one decomposition of the observed tensor, the training precoder  $\tilde{\mathbf{P}}_{l,u}$  containing the active device ID can be merged into the transmitting steering vector  $\tilde{\mathbf{a}}_T(\phi_l)$ , which is further combined with the diagonal matrix  $\boldsymbol{\Xi}(\nu_l)$  containing Doppler frequency shift to form a rank-one tensor  $\tilde{\mathbf{b}}_T(\nu_l, \phi_l)$ . Consequently, ADD and the estimation of coupled  $\phi$  and  $\nu$  can be conducted simultaneously by the recovered  $\hat{\mathbf{B}}$ . In fact, the parameter estimation based on  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{C}}$  are parallel to each other, which indicates that device activity can be detected before the complete CSI is estimated.

3) *Remaining Parameter Estimations:* Due to the grouping of channel parameters in (23)-(25), this part involves two types of methods to estimate the rest parameters  $\{\tau_l, \theta_l, h_l\}$ . Firstly, the existing popular correlation-based method with one-dimensional search is adopted to obtain the estimation of independent  $\tau_l$  and  $\theta_l$ . Then, the final estimation of  $h_l$  is calculated based on the outcomes of other parameters, and ultimately, the CSI matrix  $\mathbf{H}_{k,u}$  is recovered.

4) *Potential Expansion:* For unsourced RA, the BS allocate identical preamble sequences for each device previously, where the CE would even be unnecessary for some particular applications. Thus, the device detection should be conducted without relying on unique preambles and complete CSI, which is exactly the advantage of our proposed framework and will be further proved in the following.

##### B. Activity Detection and Coupled Parameter Estimation

For at least two parameters that are coupled with each other and need to be estimated from one factor matrix, the conventional correlation-based method cannot be applied to effectively estimate the parameters. Inspired by the ALS algorithm, the joint estimation of coupled parameters can be conducted by an iterative alternating detection method, which requires appropriate initialization and maximum number of iterations.

Specifically, to estimate the coupled AoD  $\phi$  and Doppler frequency shift  $\nu$  based on (8) and (45), the joint parameter estimation problem on the  $l$ -th path can be formulated as

$$\{\hat{\delta}_l, \hat{\nu}_l, \hat{\phi}_l\} = \arg \min_{\delta_l, \nu_l, \phi_l} \left\| \hat{\mathbf{B}}_l - \delta_l \boldsymbol{\Xi}(\nu_l) \tilde{\mathbf{a}}_T(\phi_l) \right\|_2^2, \quad (49)$$

where  $\hat{\delta}_l$  is the  $l$ -th diagonal element of  $\Delta$  instead of a channel parameter. Yet, as an important intermediate variable,  $\hat{\delta}_l$  is required by the estimation of both coupled parameters and the path gain finally. Therefore, based on a reasonable initialization, the idea of ALS replaced in the tensor decomposition part is adopted here, i.e.,  $\delta_l, \nu_l$  and  $\phi_l$  are alternately estimated and updated through a certain number of iterations. Firstly, taking the order of estimation as  $\hat{\delta}_l, \hat{\nu}_l$  and  $\hat{\phi}_l$  in (49), the initialization of the latter two parameters is given by [21]

$$\hat{\nu}_l^{(0)} = 0, \quad (50)$$

$$\hat{\phi}_l^{(0)} = \arg \max_{\phi_l \in [-\frac{\pi}{2}, \frac{\pi}{2}]} \frac{\left| \tilde{\mathbf{a}}_T^H(\phi_l) \hat{\mathbf{B}}_l \right|^2}{\|\tilde{\mathbf{a}}_T(\phi_l)\|_2^2}, \quad (51)$$

respectively. Recalling the definition of  $\tilde{\mathbf{a}}_T(\phi_l)$ , we have  $\tilde{\mathbf{a}}_T(\phi_l) \triangleq \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l) \in \mathbb{C}^{N_T \times N}$ , where  $\bar{\mathbf{P}}_{l,u}$  represents the  $l$ -th path belongs to  $u$ -th device. Then, (51) can be derived as

$$\begin{aligned} \{\hat{\mathbf{P}}_{l,u}, \hat{\phi}_l^{(0)}\} &= \arg \max_{\bar{\mathbf{P}}_{l,u}, \phi_l} \frac{\left|\mathbf{a}_T^H(\phi_l) \bar{\mathbf{P}}_{l,u}^* \hat{\mathbf{B}}_l\right|^2}{\left\|\bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l)\right\|_2^2} \\ &= \arg \max_{\bar{\mathbf{P}}_{l,u}, \phi_l} f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l), \end{aligned} \quad (52)$$

where  $f_{la}(\cdot)$  denotes the correlation function to determine the activity of the  $l$ -th path. Due to the activity indicator  $a_u$ , the maximum values of  $f_{la}(\cdot)$  on the paths belongs to active devices, i.e., active paths, are far larger than the value on inactive paths. Moreover, benefiting from the training precoder  $\bar{\mathbf{P}}_{l,u}$  that has been known at transceiver, the maximum value of  $f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l)$  can be obtained by a simple two-dimensional search, where the initialization of  $\hat{\phi}_l$  can be determined by using dense search grids.

Therefore, based on the empirical threshold  $\xi$ , we can detect the active paths through the correlation function. Specifically, if  $f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l) \geq \xi$ , the path  $l$  is active and device  $u$  which it belongs to is active as well, i.e.,  $l \in \{la\}$  and  $\hat{a}_u = 1$ . On the contrary, if  $f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l) < \xi$ , the opposite conclusion will be obtained. The active detection methodology can be summarized as

$$\begin{cases} \hat{a}_u = 1, & \{l \in \{la\} | f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l) \geq \xi\}, \\ \hat{a}_u = 0, & \{l \notin \{la\} | f_{la}(\bar{\mathbf{P}}_{l,u}, \phi_l) < \xi\}. \end{cases} \quad (53)$$

Then, based on the preset initialization (50) and (51), the parameters are alternately estimated and updated by fixing the other two parameters [33]. Specifically, through simple algebraic derivation, the updating rules for iteration can be given in the following steps, i.e.,

$$\begin{aligned} \hat{\delta}_l^{(i)} &= \arg \min_{\delta_l} \left\| \hat{\mathbf{B}}_l - \delta_l \Xi(\hat{\nu}_l^{(i-1)}) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)}) \right\|_2^2 \\ &= \frac{\mathbf{a}_T^H(\hat{\phi}_l^{(i-1)}) \bar{\mathbf{P}}_{l,u}^* \Xi^{-1}(\nu_l^{(i-1)}) \hat{\mathbf{B}}_l}{\|\bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)})\|_2^2}, \end{aligned} \quad (54)$$

$$\begin{aligned} \hat{\nu}_l^{(i)} &= \arg \min_{\nu_l} \left\| \hat{\mathbf{B}}_l - \hat{\delta}_l^{(i)} \Xi(\nu_l) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)}) \right\|_2^2 \\ &= \arg \min_{\nu_l} \left( \hat{\mathbf{B}}_l^H \hat{\mathbf{B}}_l - \hat{\delta}_l^{(i)} \hat{\mathbf{B}}_l^H \Xi(\nu_l) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)}) \right. \\ &\quad \left. - (\hat{\delta}_l^{(i)} \hat{\mathbf{B}}_l^H \Xi(\nu_l) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)}))^H \right. \\ &\quad \left. + |\hat{\delta}_l^{(i)}|^2 \mathbf{a}_T^H(\hat{\phi}_l^{(i-1)}) \bar{\mathbf{P}}_{l,u}^* \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)}) \right) \\ &= \arg \max_{\nu_l} \operatorname{Re}(\hat{\delta}_l^{(i)} \hat{\mathbf{B}}_l^H \Xi(\nu_l) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l^{(i-1)})), \end{aligned} \quad (55)$$

$$\begin{aligned} \hat{\phi}_l^{(i)} &= \arg \min_{\phi_l} \left\| \hat{\mathbf{B}}_l - \hat{\delta}_l^{(i)} \Xi(\hat{\nu}_l^{(i)}) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l) \right\|_2^2 \\ &= \arg \min_{\phi_l} \left\| \left( \Xi(\hat{\nu}_l^{(i)}) \right)^{-1} \hat{\mathbf{B}}_l - \hat{\delta}_l^{(i)} \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l) \right\|_2^2 \\ &= \arg \max_{\phi_l} \frac{\left| \mathbf{a}_T^H(\phi_l) \bar{\mathbf{P}}_{l,u}^* \left( \Xi(\hat{\nu}_l^{(i)}) \right)^{-1} \hat{\mathbf{B}}_l \right|^2}{\|\bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l)\|_2^2}. \end{aligned} \quad (56)$$

### C. Remaining Parameter Estimations

As the remaining parameters, the transmission delay  $\tau_1$ , the arrival angle  $\theta_l$ , and the path gain  $h_l$  are required to be estimated from the factor matrices  $\hat{\mathbf{C}}$  and  $\hat{\mathbf{A}}$ , respectively.

1) *Independent Parameter Estimations*: Based on the proposed framework and existing popular solutions,  $\tau_1$  can be estimated as an independent parameter from the factor matrix  $\hat{\mathbf{C}}$ , where the simple correlation-based one-dimensional search scheme is still an effective methodology, given by [19]

$$\hat{\tau}_l = \arg \max_{\tau_l} \frac{\left| \hat{\mathbf{C}}_l^H \tilde{\mathbf{c}}(\tau_l) \right|^2}{\left\| \hat{\mathbf{C}}_l \right\|_2^2 \left\| \tilde{\mathbf{c}}(\tau_l) \right\|_2^2}. \quad (57)$$

Although  $\hat{\mathbf{A}}$  contains two channel parameters to be estimated as  $\hat{\mathbf{B}}$ ,  $\theta_l$  can still be estimated in the same way as  $\tau_l$ , given by

$$\begin{aligned} \hat{\theta}_l &= \arg \max_{\theta_l} \frac{\left| \hat{\mathbf{A}}_l^H(\bar{h}_l \tilde{\mathbf{a}}_R(\theta_l)) \right|^2}{\left\| \hat{\mathbf{A}}_l \right\|_2^2 \left\| \bar{h}_l \tilde{\mathbf{a}}_R(\theta_l) \right\|_2^2} \\ &= \arg \max_{\theta_l} \frac{\left| \hat{\mathbf{A}}_l^H \tilde{\mathbf{a}}_R(\theta_l) \right|^2}{\left\| \hat{\mathbf{A}}_l \right\|_2^2 \left\| \tilde{\mathbf{a}}_R(\theta_l) \right\|_2^2}. \end{aligned} \quad (58)$$

Clearly, the path gain will not affect the estimation of AoA in a correlation-based scheme, which can be estimated separately.

2) *Path Gain Estimation*: Regardless of the order in which the previous parameters are estimated, the estimation of the path gain can only be performed at the end. According to the intrinsic relationship between the estimated factor matrices and the real ones,  $\hat{h}_l$  can be estimated through the following steps, given by [35]

$$[\Delta_B]_{l,l} = (\Xi(\hat{\nu}_l) \bar{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\hat{\phi}_l))^\dagger \hat{\mathbf{B}}_l = \hat{\delta}_l, \quad (59)$$

$$\Delta_A = \Delta_B^{-1}, \quad (60)$$

$$\hat{h}_l = [\Delta_A]_{l,l}^{-1} \tilde{\mathbf{a}}_R^\dagger(\hat{\theta}_l) \hat{\mathbf{A}}_l. \quad (61)$$

Recalling the equivalent transformation from (21) to (22), the estimation of path gain can be expressed as

$$\begin{aligned} \hat{h}_l &= \hat{h}_l / e^{-j2\pi\hat{\tau}_l f_s \frac{1}{K_0}} \\ &= e^{j2\pi\hat{\tau}_l f_s \frac{1}{K_0}} [\Delta_A]_{l,l}^{-1} \tilde{\mathbf{a}}_R^\dagger(\hat{\theta}_l) \hat{\mathbf{A}}_l. \end{aligned} \quad (62)$$

### D. Potential Expansion

For unsourced scenarios, the active devices are deployed without unique IDs, which simplifies the design of training precoders to  $\bar{\mathbf{P}}_{l,u} = \bar{\mathbf{P}}_l$ . Yet, the detection of activity is based on a correlation scheme, where the impact on the maximum value of correlation function  $f_{la}(\cdot)$  comes from the activity indicator  $a_u$  rather than the unique training precoder. Thus, the proposed method of activity detection can be expressed by an effective and simpler form, given by

$$\begin{aligned} \hat{\phi}_l^{(0)} &= \arg \max_{\phi_l \in [-\frac{\pi}{2}, \frac{\pi}{2}]} \frac{\left| \tilde{\mathbf{a}}_T^H(\phi_l) \hat{\mathbf{B}}_l \right|^2}{\left\| \tilde{\mathbf{a}}_T(\phi_l) \right\|_2^2} \\ &= \arg \max_{\phi_l \in [-\frac{\pi}{2}, \frac{\pi}{2}]} f_{la}(\phi_l), \end{aligned} \quad (63)$$

### Algorithm 1 Tensor-based Joint ADD and CE Algorithm

**Input:** Received observation tensor  $\mathcal{Y}$ , training precoder  $\bar{\mathbf{P}}_{l,u}$ , combiner  $\mathbf{W}$ , and total number of iterations  $I_{iter}$

**Output:** Estimated activity indicator  $\hat{a}_u$ ,  $\forall u$ , the channel parameters  $\{\hat{\theta}_l, \hat{\nu}_l, \hat{\phi}_l, \hat{h}_l, \hat{\tau}_l\}$ ,  $l \in \{la\}$ , and the CSI  $\hat{\mathbf{H}}_{k,u}(n)$ ,  $l_u \in \{la\}$ ,  $\forall k$

- 1: Recover the Vandermonde factor matrix  $\hat{\mathbf{C}}$  by the eigenvalues of its generating matrix  $\mathbf{G}$  according to (39);
- 2: Recover the rest factor matrices  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{A}}$  by (44) and (45), respectively;
- 3: **for**  $l = 1 : L$  **do**
- 4:   **Initialization:**  $\hat{\nu}_l^{(0)}$  and  $\hat{\phi}_l^{(0)}$  are initialized as (50) and (51), respectively;
- 5:   **if** for sourced RA scenarios, compute the estimated activity indicator  $\hat{a}_u$  by (52) and (53); **else**, compute the estimated activity indicator  $\hat{a}_u$  by (63) and (64).
- 6:   **end for**
- 7:   **for** each  $l \in \{la\}$  **do**
- 8:      $i = 1$ ;
- 9:     **while**  $i \leq I_{iter}$  **do**
- 10:       Update  $\hat{\delta}_l^{(i)}$  according to (54);
- 11:       Update  $\hat{\nu}_l^{(i)}$  according to (55);
- 12:       Update  $\hat{\phi}_l^{(i)}$  according to (56);
- 13:        $i = i + 1$ .
- 14:     **end while**
- 15:     Compute the estimate of  $\hat{\tau}_l$  by (57);
- 16:     Compute the estimate of  $\hat{\theta}_l$  by (58);
- 17:     Compute the estimate of  $\hat{h}_l$  by (62);
- 18:   **end for**

where the acquisition of the maximum value of  $f_{la}(\phi_l)$  is simplified from a two-dimensional search to a one-dimensional search, and the activity of devices can still be detected by the preset empirical threshold  $\xi$ , given by

$$\begin{cases} \hat{a}_u = 1, & \{l \in \{la\} | f_{la}(\phi_l) \geq \xi\}, \\ \hat{a}_u = 0, & \{l \notin \{la\} | f_{la}(\phi_l) < \xi\}. \end{cases} \quad (64)$$

Thereby, the flexibility of the detection method enables the proposed framework extend to be a unified one. In summary, the joint ADD and CE method based on CP model for a unified sourced and unsourced GF-RA framework in high-mobility scenarios can be described as **Algorithm 1**.

*Proposition 4: Due to the insensitivity of the initialization to the training precoder as shown in (51), the robustness of the proposed **Algorithm 1** is guaranteed, i.e., the effectiveness of joint ADD and coupled parameter estimation method can be promised regardless of the training precoder design.*

*Proof:* See Appendix D.

### E. Computational Complexity Analysis

The computational complexity of the proposed framework is elaborated as follows, which is mainly contributed by the complex multiplications of tensor decomposition and joint active devices detection and channel estimation. For simplicity, the complexity level of the intermediate variables  $\Omega_1$  and  $\Omega_2$  are considered to be the same as  $\mathcal{O}(K)$ [22]. In terms of tensor decomposition, the expansion  $\mathbf{Y}_{(1)}^{\Omega_2}$  and its SVD operation

TABLE II  
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency $f_c$	60GHz
Number of device antennas $N_T$	32
Number of BS antennas $N_R$	64
AoD and AoA $\phi, \theta$	$(-\pi/2, \pi/2)$
Number of RF chains at BS $N_{RF}^R$	40
Sampling rate $f_s$	100MHz
Total number of subcarriers $K_0$	256
Number of training subcarriers $K$	40~120
Time delay $\tau$	(0, 320) ns
Velocity of device $v$	(0, 120) km/h
Number of device paths $L_u$	2~3
Path gain $\alpha$	$\mathcal{CN}(0, 1)$
Number of potential devices $U$	100
Activity probability $\epsilon$	0.1~0.3
Number of frames $N$	50
Transmit SNR	0~30 dB

require  $\mathcal{O}(K^3 N^2 N_{RF}^R)$  and  $\mathcal{O}(K^3 N N_{RF}^R)$  respectively. The EVD operation of  $\mathbf{G}$  takes  $\mathcal{O}(L^3)$ , and the estimation of factor matrices has the complexity  $\mathcal{O}((N^2 + (N_{RF}^R)^2)KL^2)$ . In terms of the other part, given the number of grid points for search as  $I_{grid}$ , the complexity of joint activity detection and initialization is  $\mathcal{O}(N_T N U I_{grid} L)$ . With the number of iterations  $I_{iter}$ , the estimation of coupled parameters costs  $\mathcal{O}(I_{iter}(N_T N + N)I_{grid} L)$ ; the estimation of independent parameters costs  $\mathcal{O}((N_R N_{RF}^R + K)I_{grid} L)$ ; and finally, the calculation of path gains based on previous estimations requires  $\mathcal{O}((N_{RF}^R + K)L)$ .

### V. SIMULATION RESULTS

In this section, we conduct numerical simulations to verify the effectiveness of the proposed tensor-based GF-RA framework. The simulation parameters are set in TABLE II unless otherwise specified<sup>2</sup>. Generally, the false alarm probability, missed detection probability, and activity error rate (AER) are considered to measure the detection performance. In addition, we use the normalized mean square error (NMSE) of the channel matrix and MSE of the channel parameters to assess the CE performance. To ensure the reliability of performance analysis, the simulation results are obtained by averaging over  $10^3$  independent channel realizations. For the tensor framework, the signal to noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\|\mathcal{Y} - \mathcal{N}\|_F^2}{\|\mathcal{N}\|_F^2}. \quad (65)$$

#### A. Performance under Different Number of Active Devices

In Fig. 2, we compare the detection performance of the conventional ALS-based method and the proposed method in

<sup>2</sup>To satisfy the increasing requirement of high data rate in the upcoming 6G era, millimeter-wave band (30GHz-300GHz) has been widely investigated in GF-RA, which motivates us to consider a system operating in 60GHz [36]-[37]. However, it should be noticed that the CE performance of GF-RA operating in 60 GHz is deteriorated when compared to its counterpart in sub-6GHz, which is a topic for future consideration.

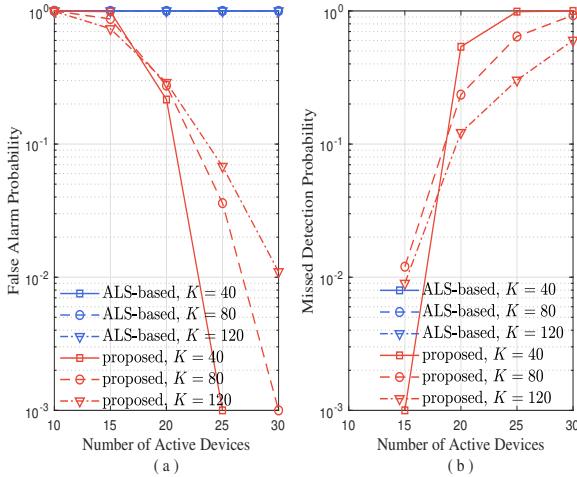


Fig. 2. False Alarm and Missed Detection performance under various numbers of active devices, SNR = 15dB.

terms of false alarm probability and miss detection probability, which are generally used to characterize the performance of detection errors in the number of active devices [24]. Typically, detecting a device that is originally inactive as active is considered as false alarm, and vice versa is considered as miss detection. For our proposed tensor-based GF-RA framework, the activity of a device is determined by the activity of its paths, rather than directly detecting  $a_u$  itself. In the setting of our simulation conditions, the active devices are assumed to have multiple transmission paths, i.e.,  $L_u > 1$ . Thus, the false alarm probability and miss detection probability of active paths  $\{la\}$  from all devices are presented, which makes the numerical results more rigorous. Fig. 2 shows that, as the number of active devices increases, the false alarm probability of the proposed scheme decreases, while the miss detection probability increases. In fact, the result of multiple active paths would match the same ID of one active device as the proposed ADD method designed, which indicates that the miss detection of an active path does not necessarily lead to the increase of AER. On the contrary, the false alarm would have a severe impact on AER, which should maintain a low level for a large number of devices. Therefore, in the case of a large number of active devices, the false alarm probability and miss detection probability of the proposed framework are acceptable.

In addition, the intersection of the curves at near SNR=17dB with different  $K$ s indicates that the increase in the number of training subcarriers does not have a constant effect on the false alarm and miss detection performance, which will be significantly affected by the device interference resulting from accessing more active devices. For the conventional ALS-based scheme, the simulation results demonstrate that its false alarm and miss detection probabilities will maintain 1 and 0 with different simulation conditions, respectively. Due to the previous analysis about false alarm, the detection performance will inevitably lead to an unacceptable AER, which will be verified in the following subsections.

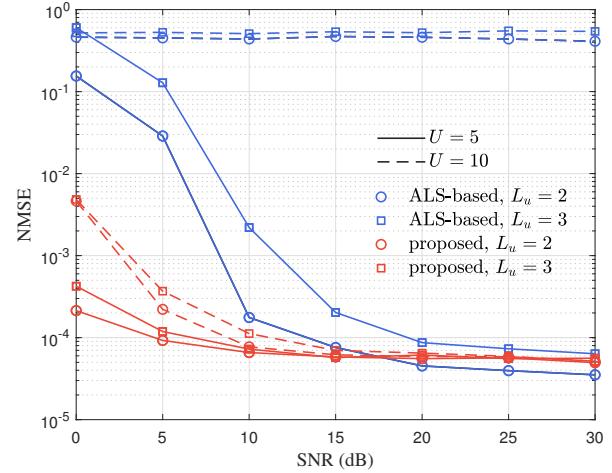


Fig. 3. Channel estimation performance under various numbers of potential devices,  $K = 60$ ,  $\epsilon = 0.2$ .

### B. Performance under Different Number of Potential Devices

Taking the performance of CE as an example, we provide a comparison under different total number of potential device  $U$  and transmission path numbers  $L_u$  to illustrate the effectiveness change of the two schemes. Specifically, the NMSE of CE is given by

$$\text{NMSE}(\mathbf{H}) = \sum_u^K \frac{\|\mathbf{H}_{k,u}(n) - \hat{\mathbf{H}}_{k,u}(n)\|_F^2}{\|\mathbf{H}_{k,u}(n)\|_F^2}. \quad (66)$$

According to the proposed framework, the column number of factor matrices  $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$  under the simulation conditions in Fig. 3 ranges from 10 to 30. The numerical results demonstrate that the effectiveness of the conventional ALS-based scheme can be guaranteed as  $U$  stays small. In addition, the performance of ALS is better than the proposed scheme at  $U = 5$  and  $L_u = 2$  under high SNRs, which stems from the convergent iteration process in ALS for small size factor matrices. However, as these sizes grows, ALS's performance drops sharply. Indeed, the effectiveness change of ALS confirms the analysis in Section III about the ill-conditioned problem brought by large  $U$  in the system, which can also explain the reason why the performance of ALS remains unchanged in Fig. 2.

For the proposed scheme, despite the performance of CE also slightly drops as  $U$  grows, its NMSE under high SNRs is basically consistent. Above all, the effectiveness will not be influenced with large size factor matrices. Therefore, in the following comparison of ADD and CE under  $U = 100$ , the performance of ALS-based scheme would be only provided under the best simulation conditions, i.e., the minimum number of active devices and the maximum number of training subcarriers. In other words, we are more concerned about the performance changes of the proposed scheme under different simulation conditions. To further confirm the necessity and effectiveness of our proposed scheme, the performance of the representative CS-based method, i.e., OMP [34] are provided.

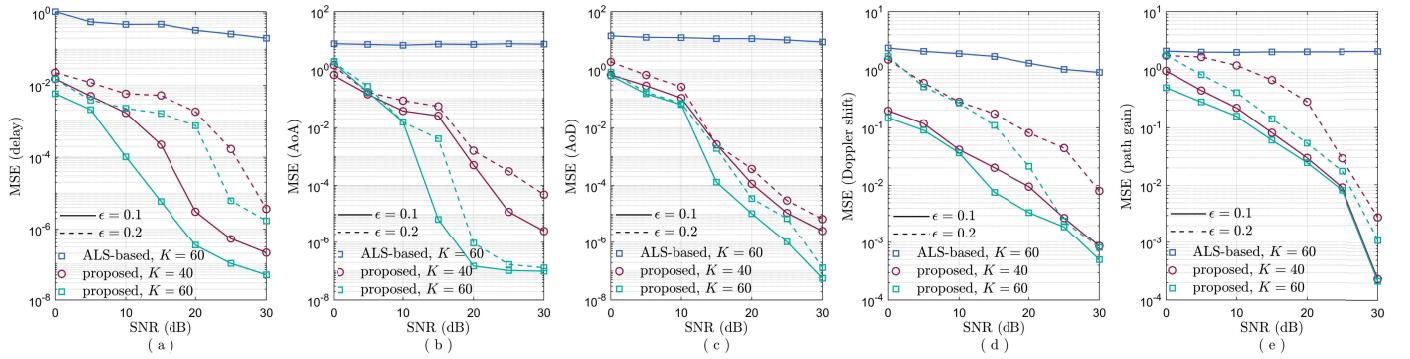


Fig. 4. MSE of channel parameters under various SNRs.

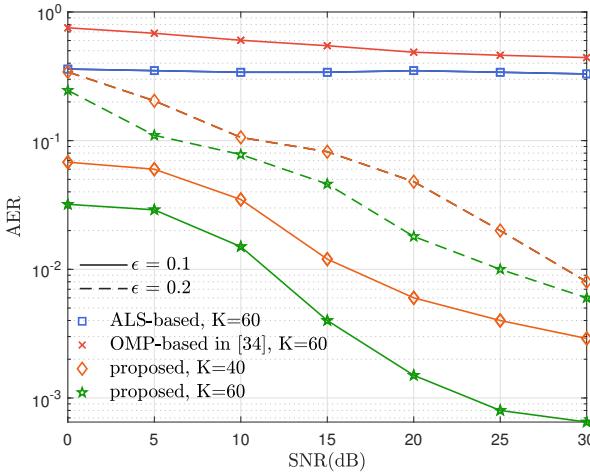


Fig. 5. AER performance under various SNRs.

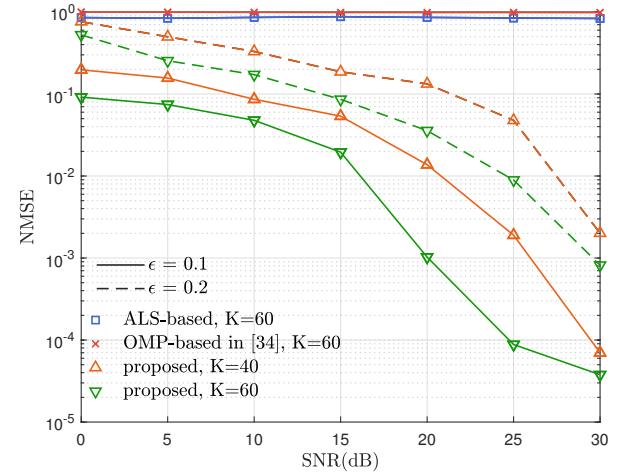


Fig. 6. Channel estimation performance under various SNRs.

### C. Performance under Different Signal-to-Noise Ratio

To validate the effectiveness of the proposed scheme under large number of devices, Fig. 4 shows the mean square error (MSE) of all channel parameter estimations  $\{\hat{\theta}_l, \hat{\nu}_l, \hat{\phi}_l, \hat{h}_l, \hat{\tau}_l\}$ . Taking the delay  $\hat{\tau}_l$  as an example, the MSE is defined as

$$\text{MSE}(\tau) = \sum_{l=1}^L |\tau_l - \hat{\tau}_l|^2. \quad (67)$$

Particularly, for the iteratively joint estimated AoD  $\hat{\phi}_l$  and Doppler shift  $\hat{\nu}_l$ , the maximum number of iterations is set to  $I_{iter} = 40$ . For the proposed scheme, it can be observed that the performance of all parameter estimations deteriorate with more active devices, i.e., the larger  $\epsilon$ , which is attributed to the increase of interference brought by devices. On the other hand, the accuracy of estimation can be significantly improved by increasing the number of training subcarriers  $K$ . Overall, the MSE of channel parameters will reach an order ranging from  $10^{-4}$  to  $10^{-7}$  at high SNR regime, where the difference in performance mainly comes from the distinct estimation methods for coupled parameters and independent parameters. The estimation of path gain  $\hat{h}_l$  has to be obtained based on the other parameters, thus its MSE would be slightly increased

especially under low SNR regime. For the conventional ALS-based scheme, the MSE of all channel parameter estimations remains at a high level, which cannot be improved with the increase in SNR, thus significantly affecting the performance of ADD as designed in our proposed framework.

Fig. 5 exhibits the AER performance of ADD under different SNRs. Depending on the detection scheme designed as (53) or (64), we further define the detected active device set  $\widehat{\mathbb{A}}$  to collect the information of unique IDs from the devices with  $\hat{a}_u = 1$ . Besides, the real active device set is defined as  $\mathbb{A}$ . Therefore, the AER is given by [24]

$$\text{AER} = \frac{|\mathbb{A} - \widehat{\mathbb{A}}|_c + |\widehat{\mathbb{A}} - \mathbb{A}|_c}{U}. \quad (68)$$

In addition, similar to the conventional CS methods [15], the necessary threshold is set to  $\xi = 0.1$  as an empirical value to minimize the detection error. It can be observed that the proposed scheme can conduct the ADD effectively in contrast to the ALS-based scheme, which can reach an order of  $10^{-4}$  at high SNR regime. Moreover, the better performance with  $K = 60$  compared to  $K = 40$  under the same  $\epsilon$  demonstrates that the increase of training subcarriers evidently contributes to the improvement of ADD performance. On the other hand,

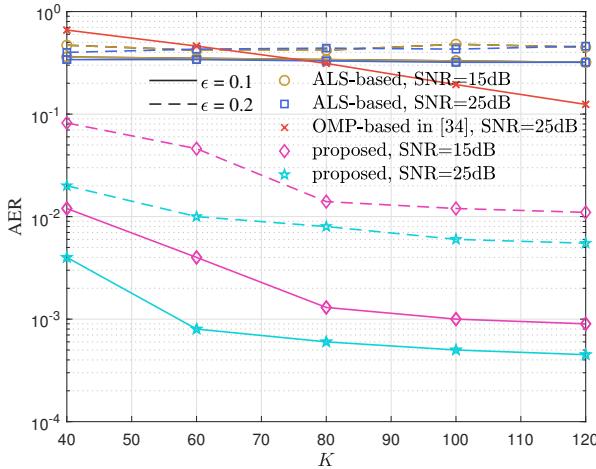


Fig. 7. AER performance under various numbers of training subcarriers.

the performance of AER deteriorating at  $\epsilon = 0.2$  confirms the impact of more device interference as Fig. 4 shows.

To further demonstrate the performance of CE in our proposed framework, Fig. 6 provides the NMSE in terms of CSI based on the estimated channel parameters. It can be observed that the impact of changes in  $K$  and  $\epsilon$  on CE performance is clearly consistent with the numerical results of parameter estimation in Fig. 4. Compared to the unacceptable performance of ALS-based scheme, the NMSE of the proposed scheme can achieve nearly  $10^{-5}$  at high SNR regime. Besides, for the proposed joint ADD and CE method under the same SNRs, Figs. 5 and 6 also demonstrate that the performance degradation caused by more active devices has a greater impact than the performance improvement brought by more training subcarriers.

#### D. Performance under Different Number of Training Subcarriers

Although Figs. 4, 5 and 6 illustrate the performance enhancement brought by increasing the training overhead, the comparisons under  $K = 40$  and  $K = 60$  are not sufficient to fully display its contribution to the ADD and CE. In addition, the performance improvement caused by larger  $K$  would not be consistently remarkable. Thus, in Fig. 7 and Fig. 8, we present the AER and NMSE performance under different  $K$ , respectively.

It can be observed in Fig. 7 that the conventional ALS-based scheme cannot achieve satisfactory performance in terms of ADD even as the number of training subcarriers increases to  $K = 120$ . In contrast, the detection performance of the proposed scheme improves significantly with the increase of  $K$ , where the acceptable AER on the order of  $10^{-4}$  can be obtained under  $\epsilon = 0.1$  as  $K$  achieves  $K = 80$ . It is worth noting that although the OMP-based scheme has better AER performance than the conventional ALS-based scheme after  $K > 80$ , it is still visibly inferior to the proposed scheme. Moreover, blindly increasing training overhead is not practical. On the other hand, the similar overall trend

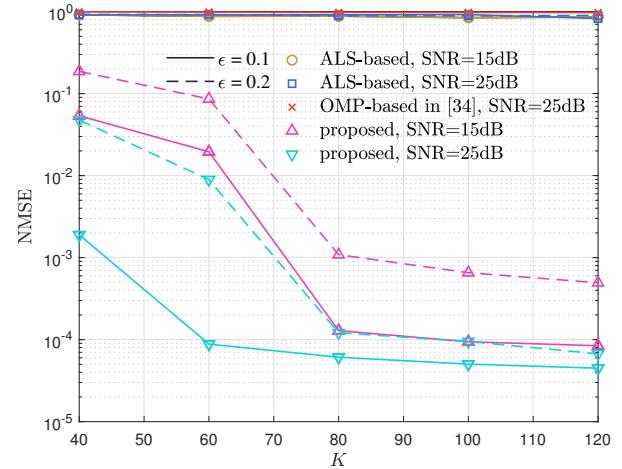


Fig. 8. Channel estimation performance under various numbers of training subcarriers.

of CE performance is exhibited in Fig. 8, where the NMSE comparison under distinct  $\epsilon$  and same SNR illustrates the contribution of increasing the training subcarriers to alleviating the deterioration caused by more active devices. Yet, the insignificant improvement of both AER and NMSE curves after  $K > 80$  demonstrates that blindly increasing the training overhead to obtain a better performance is meaningless in our proposed framework. Therefore, the numerical results show that a better trade-off between the overhead and performance would be achieved by reasonably saving the number of training subcarriers.

## VI. CONCLUSION

In this paper, we investigated a unified joint channel estimation and active device detection scheme for high-mobility GF-RA scenarios. By leveraging CP decomposition, a tensor-based GF-RA was proposed, which can be adopted to complete the joint ADD and CE. More specifically, to solve the ill-condition generated in large-size factor matrices, Vandermonde structure was utilized, where the considered model was strategically transformed into an equivalent form. Then, depending on whether the channel parameters are coupled or independent in factor matrices, two estimation methods were designed separately. By fully exploring the active devices information bearing in the coupled parameters, an effective correlation-based ADD method was designed, which can work efficiently even without complete CSI. Besides, to satisfy the requirement of ADD in both sourced and unsourced RA, the designed method was efficiently extended to the devices equipped with common preambles. Simulation results verified that the proposed tensor-based GF-RA framework outperforms the existing conventional ALS-based method particularly in high-mobility scenarios, showcasing the superiority of our proposed framework.

## APPENDIX A PROOF OF PROPOSITION 1

*Proof:* Recalling that  $\tilde{\mathbf{a}}_R(\theta_l) \in \mathbb{C}^{N_{RF}^R \times 1}$  and  $\tilde{\mathbf{a}}_T^T(\phi_l) \in \mathbb{C}^{N \times 1}$ , the entry located in the  $n_{RF}^R$ -th row and  $n$ -th column of  $\mathbf{Y}_k$  in (8) can be written as

$$\begin{aligned} [\mathbf{Y}_k]_{n_{RF}^R, n} &= \sum_{l=1}^L h_l e^{-j2\pi\tau_l f_s \frac{k}{K_0}} [\tilde{\mathbf{a}}_R(\theta_l)]_{n_{RF}^R} [\tilde{\mathbf{a}}_T^T(\phi_l)]_n e^{j\nu_l n} \\ &\quad + [\mathbf{Z}_k]_{n_{RF}^R, n} \\ &= \sum_{l=1}^L h_l [\tilde{\mathbf{a}}_R(\theta_l)]_{n_{RF}^R} [\Xi(\nu_l) \tilde{\mathbf{a}}_T(\phi_l)]_n [\mathbf{c}(\tau_l)]_k \\ &\quad + [\mathbf{Z}_k]_{n_{RF}^R, n} \\ &= \sum_{l=1}^L [\tilde{\mathbf{a}}_R(\theta_l)]_{n_{RF}^R} [\tilde{\mathbf{b}}_T(\nu_l, \phi_l)]_n [\tilde{\mathbf{c}}(\tau_l)]_k \\ &\quad + [\mathbf{Z}_k]_{n_{RF}^R, n}, \end{aligned} \quad (69)$$

where  $\tilde{\mathbf{b}}_T(\nu_l, \phi_l) = \Xi(\nu_l) \tilde{\mathbf{a}}_T(\phi_l) \in \mathbb{C}^{N \times 1}$ ,  $\tilde{\mathbf{c}}(\tau_l) = h_l \mathbf{c}(\tau_l) \in \mathbb{C}^{K \times 1}$  and  $\mathbf{c}(\tau_l) = [e^{-j2\pi\tau_l f_s \frac{1}{K_0}}, \dots, e^{-j2\pi\tau_l f_s \frac{K-1}{K_0}}]^T$ . According to the definition of a third-order tensor in [27, Section 2],  $[\mathbf{Y}_k]_{n_{RF}^R, n}$  denotes the  $(n_{RF}^R, n, k)$ -th entry of the third-order tensor  $\mathcal{Y} \in \mathbb{C}^{N_{RF}^R \times N \times K}$ . Then, by comparing (69) to the definition of CP decomposition in [27, Section 3, (3.1)],  $\mathcal{Y}$  can be obtained, given by

$$\mathcal{Y} = [\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}] + \mathcal{Z} = \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ \tilde{\mathbf{c}}(\tau_l) + \mathcal{Z}. \quad (70)$$

The proof is completed.

## APPENDIX B PROOF OF PROPOSITION 2

*Proof:* Recalling the original expression of  $\mathcal{Y}$  in (9), the format of CP decomposition can be equivalently transformed by rearranging  $\tilde{\mathbf{c}}(\tau_l)$  according to the property of CP model [27], given by

$$\begin{aligned} \mathcal{Y} &= \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ \tilde{\mathbf{c}}(\tau_l) + \mathcal{Z} \\ &= \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ (h_l \mathbf{c}(\tau_l)) + \mathcal{Z} \\ &= \sum_{l=1}^L \tilde{\mathbf{a}}_R(\theta_l) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ (h_l e^{-j2\pi\tau_l f_s \frac{1}{K_0}} \bar{\mathbf{c}}(\tau_l)) + \mathcal{Z} \\ &= \sum_{l=1}^L (\bar{h}_l \tilde{\mathbf{a}}_R(\theta_l)) \circ \tilde{\mathbf{b}}_T(\nu_l, \phi_l) \circ \bar{\mathbf{c}}(\tau_l) + \mathcal{Z} \\ &= [\mathbf{A}, \mathbf{B}, \mathbf{C}] + \mathcal{Z}, \end{aligned} \quad (71)$$

where  $\bar{\mathbf{c}}(\tau_l) = [1, e^{-j2\pi\tau_l f_s \frac{1}{K_0}}, \dots, e^{-j2\pi\tau_l f_s \frac{K-1}{K_0}}]^T$  and  $\bar{h}_l = h_l e^{-j2\pi\tau_l f_s \frac{1}{K_0}}$ . The proof is completed.

## APPENDIX C PROOF OF PROPOSITION 3

*Proof:* To estimate the rest factor matrices  $\mathbf{A}$  and  $\mathbf{B}$  with the help of  $\hat{\mathbf{C}}'$ , the mixed product property of Kronecker product is fully exploited here, whose general definition is given by

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}), \quad (72)$$

where  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\mathbf{B} \in \mathbb{C}^{p \times q}$ ,  $\mathbf{C} \in \mathbb{C}^{n \times s}$ , and  $\mathbf{D} \in \mathbb{C}^{q \times t}$ .

By utilizing (72), we can firstly perform an identity transformation of the  $l$ -th column of  $\mathbf{B} \in \mathbb{C}^{N \times L}$ , given by

$$\hat{\mathbf{B}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_1, l}}{\hat{\mathbf{C}}'^H_{\Omega_1, l} \hat{\mathbf{C}}'_{\Omega_1, l}} \otimes \mathbf{I}_N \right) (\hat{\mathbf{C}}'_{\Omega_1, l} \otimes \hat{\mathbf{B}}'_l). \quad (73)$$

Recalling that  $\hat{\mathbf{C}}'$  is generated by  $[\mathbf{D}]_{l, l}$  according to (39), we have  $\hat{\mathbf{C}}'^H_{\Omega_1, l} \hat{\mathbf{C}}'_{\Omega_1, l} = \Omega_1$ . Moreover, by combining (30), (40) and (41), we can obtain

$$\hat{\mathbf{C}}'_{\Omega_1, l} \otimes \hat{\mathbf{B}}'_l = \mathbf{U} \mathbf{Q}_l. \quad (74)$$

Therefore, the second factor matrix  $\hat{\mathbf{B}}' = [\hat{\mathbf{B}}'_1, \dots, \hat{\mathbf{B}}'_L]$  up to permutation and scaling ambiguity can be recovered column by column, given by

$$\hat{\mathbf{B}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_1, l}}{\Omega_1} \otimes \mathbf{I}_N \right) \mathbf{U} \mathbf{Q}_l. \quad (44)$$

Finally, the estimation procedure of  $\hat{\mathbf{A}}' = [\hat{\mathbf{A}}'_1, \dots, \hat{\mathbf{A}}'_L]$  is consistent with that of  $\hat{\mathbf{B}}'$ . Based on (72), the identity transformation of  $\hat{\mathbf{A}}'_l$  can be expressed as

$$\hat{\mathbf{A}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_2, l}}{\hat{\mathbf{C}}'^H_{\Omega_2, l} \hat{\mathbf{C}}'_{\Omega_2, l}} \otimes \mathbf{I}_{N_{RF}^R} \right) (\hat{\mathbf{C}}'_{\Omega_2, l} \otimes \hat{\mathbf{A}}'_l). \quad (75)$$

Similarly, we have  $\hat{\mathbf{C}}'^H_{\Omega_2, l} \hat{\mathbf{C}}'_{\Omega_2, l} = \Omega_2$  and the following equivalent substitution by combining (31), (40) and (41), given by

$$\hat{\mathbf{C}}'_{\Omega_2, l} \otimes \hat{\mathbf{A}}'_l = \mathbf{V}^* \Sigma [\mathbf{Q}^{-1}]_{:, l}^T. \quad (76)$$

Consequently, the last factor matrix can be recovered by

$$\hat{\mathbf{A}}'_l = \left( \frac{\hat{\mathbf{C}}'^H_{\Omega_2, l}}{\Omega_2} \otimes \mathbf{I}_{N_{RF}^R} \right) \mathbf{V}^* \Sigma [\mathbf{Q}^{-T}]_{:, l}. \quad (45)$$

The proof is completed.

## APPENDIX D PROOF OF PROPOSITION 4

*Proof:* Based on (49),  $\hat{\mathbf{B}}_l$  can be expressed as

$$\hat{\mathbf{B}}_l = \delta_l \Xi(\nu_l) \tilde{\mathbf{a}}_T(\phi_l) + \mathbf{e}_l, \quad (77)$$

where  $\mathbf{e}_l \sim \mathcal{CN}(0, \sigma_e^2)$  is the  $l$ -th column of  $\mathbf{E}_B$ . By introducing the discrete Fourier transform (DFT) basis matrix  $\mathbf{F} \in \mathbb{C}^{N_T \times N_T}$ , (77) can be rewritten as

$$\begin{aligned} \hat{\mathbf{B}}_l &= \Xi(\nu_l)(\delta_l \tilde{\mathbf{a}}_T(\phi_l)) + \mathbf{e}_l \\ &= \Xi(\nu_l)(\delta_l \tilde{\mathbf{P}}_{l,u}^T \mathbf{a}_T(\phi_l)) + \mathbf{e}_l \\ &= \Xi(\nu_l) \tilde{\mathbf{P}}_{l,u}^T \mathbf{F} \mathbf{x}_l + \mathbf{e}_l, \end{aligned} \quad (78)$$

where  $\mathbf{x}_l \in \mathbb{C}^{N_T \times 1}$  is a sparse vector whose largest non-zero element is  $\delta_l$  and the corresponding index points to the quantized  $\phi_l$ . Consequently, (49) can be reformulated as

$$\hat{\mathbf{x}}_l = \arg \min_{\mathbf{x}_l} \left\| \left| \hat{\mathbf{B}}_l \right| - \left| \bar{\mathbf{P}}_{l,u}^T \mathbf{F} \mathbf{x}_l \right| \right\|_2^2, \quad \text{s.t.} \|\mathbf{x}_l\|_0 = 1, \quad (79)$$

$$\hat{\nu}_l = \arg \min_{\nu_l} \left\| \left[ \hat{\mathbf{B}}_l \right]_n - e^{jn\nu_l} \bar{\mathbf{P}}_{l,u}^T(n) \mathbf{F} \hat{\mathbf{x}}_l \right\|_2^2. \quad (80)$$

To effectively conduct the coupled parameter estimation, the initial parameters are selected as follows [21] [32]

$$\eta_i = \frac{|\Psi_i^H \hat{\mathbf{B}}_l|^2}{N \|\Psi_i\|_2^2}, \quad i = 1, 2, \dots, N_T, \quad (81)$$

$$\hat{\mathcal{S}} = \left\{ i_S | \eta_{iS} > \eta_{i1} > \eta_{i2} > \dots > \eta_{iN_T} \right\}, \quad (82)$$

which is equivalent to (51), and  $\Psi = \bar{\mathbf{P}}_{l,u}^T \mathbf{F}$  represents the equivalent training precoder matrix,  $\hat{\mathcal{S}}$  denotes the support set. Specifically, the maximum index in the test statistics  $\{\eta_i\}_{i=1}^{N_T}$  is detected and exploited as the initial guess of  $\hat{\mathcal{S}}$ . Assuming that each entry of  $\Psi$  is zero-mean and i.i.d, the analytical expression of the test statistics  $\mathbb{E}\{\eta_i\}$  can be expressed as

$$\begin{aligned} \mathbb{E}\{\eta_i\} &= \|\mathbf{x}\|_2^2 \mathbb{E}^2\{|\varphi|^2\} + \mathbb{E}\{|\varphi|^2\} \sigma_e^2 + |\mathbf{x}_i|^2 \mathbb{E}\{|\varphi|^4\} \\ &\quad + |\mathbf{x}_i|^2 \mathbb{E}^2\{|\varphi|^2\} \left( \frac{2}{N} \sum_{n=1}^{N-1} (N-n) \cos(\nu_l n) - 1 \right), \end{aligned} \quad (83)$$

where  $\varphi = [\Psi]_{n,i}$ ,  $\mathbb{E}\{|\varphi|^2\}$  and  $\mathbb{E}\{|\varphi|^4\}$  represent the second-order moment and fourth-order moment of  $\varphi$ , respectively, and the terms related to  $\mathbf{x}_i$  contribute to the gaps between the statistics corresponding to  $i \in \mathcal{S}$  and  $i \notin \mathcal{S}$ . In fact, the zero-value gap will result in same values returned by test statistics whether the index falls into true  $\mathcal{S}$  or not, which in turn leads to the incorrect support detection. Subsequently, the unreliable initialization caused by the incorrect support detection will hinder the proposed correlation-based ADD method in (52). It is noted that, the value of the gap is mainly determined by the distribution of  $\varphi$  as described in (83), which essentially is the statistical characteristic of  $\bar{\mathbf{P}}_{l,u}$ . Therefore, a strictly positive gap is required to be guaranteed to ensure the proposed algorithm to be robust to the training precoder design.

For clarity, the gap function as follows is rewritten as

$$\begin{aligned} g_S(\mathbf{x}_i, \varphi | \eta_i) &= |\mathbf{x}_i|^2 (\mathbb{E}\{|\varphi|^4\} - \mathbb{E}^2\{|\varphi|^2\}) \\ &\quad + \frac{2}{N} |\mathbf{x}_i|^2 \mathbb{E}^2\{|\varphi|^2\} \sum_{n=1}^{N-1} (N-n) \cos(\nu_l n), \end{aligned} \quad (84)$$

where  $|\mathbf{x}_i|^2 (\mathbb{E}\{|\varphi|^4\} - \mathbb{E}^2\{|\varphi|^2\}) \geq 0$ . Besides, the second term on the right side of (84) returns a strictly positive value, where the normalized Doppler shift satisfies  $\cos(\nu_l(N-1)) > 0$  [21]. Therefore, by defining the probability distribution function (PDF) of  $\varphi$  as  $F(\varphi)$ , we can obtain

$$g_S(\mathbf{x}_i, \varphi | \eta_i) > 0, \quad \forall F(\varphi), \quad (85)$$

which demonstrates the strictly positive gap value can be guaranteed regardless of the statistical characteristic of the training precoders. In summary, the insensitivity of the initialization to

the training precoders facilitates the following joint ADD and CE process, which guarantees the robustness of the **Algorithm 1**. The proof is completed.

## REFERENCES

- [1] X. Chen et al., "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.
- [2] 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.
- [3] Z. Lin et al., "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717-3724, Aug. 2022.
- [4] Z. Lin et al., "Secrecy-energy efficient hybrid beamforming for satellite-terrestrial integrated networks", *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6345-6360, Sep. 2021.
- [5] K. An et al., "Exploiting multi-layer refracting RIS-assisted receiver for HAP-SWIPT networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 12638-12657, Oct. 2024.
- [6] Z. Lin et al., "Supporting IoT with rate-splitting multiple access in satellite and aerial-integrated networks", *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11123-11134, Jul. 2021.
- [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] M. B. Shahab, R. Abbas, M. Shirvanimoghadam 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.
- [9] 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.
- [10] Z. Zhang, X. Wang, Y. Zhang, and Y. Chen, "Grant-free rateless multiple access: A novel massive access scheme for Internet of Things," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2019-2022, Oct. 2016.
- [11] 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.
- [12] Y. Du et al., "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812-2828, Dec. 2017.
- [13] L. Liu and W. Yu, "Massive connectivity with massive MIMO – Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.
- [14] M. Ke et al., "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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] Z. Zhou et al., "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.
- [19] Z. Zhou et al., "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524-1538, Jul. 2017.
- [20] Y. Lin, S. Jin, M. Matthaiou and X. You, "Tensor-based channel estimation for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4218-4232, Jul. 2020.
- [21] R. Zhang et al., "Tensor decomposition-based channel estimation for hybrid mmWave massive MIMO in high-mobility scenarios," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6325-6340, Sep. 2022.
- [22] J. Wang et al., "Time-varying channel estimation scheme for uplink MU-MIMO in 6G systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11820-11831, Nov. 2022.
- [23] X. Wu, S. Ma and X. Yang, "Tensor-based low-complexity channel estimation for mmWave massive MIMO-OTFS systems," *J. Commun. Inf. Netw.*, vol. 5, no. 3, pp. 324-334, Sep. 2020.

- [24] 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.
- [25] A. Liao et al., "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329-8345, Dec. 2019.
- [26] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, Jr, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946-2960, May. 2018.
- [27] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455-500, Aug. 2009.
- [28] N. D. Sidiropoulos et al., "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551-3582, Jul. 2017.
- [29] N. Li, "Variants of ALS on tensor decompositions and applications," *Diss. Thesis Gradworks*, vol. 51, no. 3, pp. 455-500, 2013.
- [30] M. Sørensen et al., "Blind signal separation via tensor decomposition with vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507-5519, Nov. 2013
- [31] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772-781, Sep. 1978.
- [32] L. Zhang, G. Wang, G. B. Giannakis, and J. Chen, "Compressive phase retrieval via reweighted amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5029-5040, Oct. 2018.
- [33] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, Fac. Grad. School, Univ. Minnesota, Minneapolis, MN, May. 2014.
- [34] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [35] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [36] X. Shao, X. Chen, C. Zhong and Z. Zhang, "Exploiting Simultaneous Low-Rank and Sparsity in Delay-Angular Domain for Millimeter-Wave/Terahertz Wideband Massive Access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2336-2351, April 2022.
- [37] L. You, X. Gao, G. Y. Li, X.-G. Xia, and N. Ma, "BDMA for millimeter-wave/terahertz massive MIMO transmission with per-beam synchronization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1550-1563, Jul. 2017.