



**QUEEN'S  
UNIVERSITY  
BELFAST**

## Tensor-Based Channel Estimation for Millimeter Wave MIMO-OFDM with Dual-Wideband Effects

Lin, Y., Jin, S., Matthaiou, M., & You, X. (2020). Tensor-Based Channel Estimation for Millimeter Wave MIMO-OFDM with Dual-Wideband Effects. *IEEE Transactions on Communications*. Advance online publication. <https://doi.org/10.1109/TCOMM.2020.2983673>

**Published in:**  
IEEE Transactions on Communications

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
© 2020 IEEE.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

**General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

**Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

# Tensor-Based Channel Estimation for Millimeter Wave MIMO-OFDM with Dual-Wideband Effects

Yuxing Lin, *Student Member, IEEE*, Shi Jin, *Senior Member, IEEE*, Michail  
Matthaiou, *Senior Member, IEEE*, and Xiaohu You, *Fellow, IEEE*

## Abstract

We consider the channel estimation problem in millimeter wave (mmWave) multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems with hybrid analog-digital architectures. Leveraging the spatial- and frequency-wideband (dual-wideband) effects in massive MIMO scenarios, we derive a spatial-frequency channel model with dual-wideband effects that incorporates the multipath parameters, i.e., time delay, complex gain, angle of departure/arrival. We adopt a successive beam training scheme and formulate the training OFDM signal as a third-order low-rank tensor fitting a canonical polyadic (CP) model with factor matrices containing the channel parameters. Exploiting the Vandermonde nature of factor matrices, we propose a structured CP decomposition-based channel estimation strategy aided by the spatial smoothing method, where two dedicated algorithms with particular tensor modeling and parameter recovery operations are developed. The proposed scheme leverages standard linear algebra, and, hence, avoids the random initialization problem and iterative procedure. An analysis of the uniqueness condition of CP decomposition is also pursued. Simulation results indicate that the proposed strategy achieves enhanced estimation performance, which outperforms the traditional approaches in terms of accuracy, robustness and complexity.

## Index Terms

Y. Lin, S. Jin, and X. You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, P. R. China, e-mail: {yxlin; jinshi; xhyu}@seu.edu.cn.

M. Matthaiou is with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, Belfast, BT3 9DT, U.K., e-mail: m.matthaiou@qub.ac.uk.

Manuscript received Sept. 28, 2019; revised Jan. 23, 2020. Part of this paper has appeared in [1].

Dual-wideband effects, mmWave channel estimation, MIMO-OFDM, structured CP decomposition, tensor signal processing.

## I. INTRODUCTION

Millimeter wave (mmWave) transmission technologies have been investigated as a promising candidate against the increasing data traffic and frequency spectrum shortage as we experience the development of next-generation communications [2]. Gigabit-per-second data rates can be delivered by mmWave communications in various environments and applications [3]. As a feasible alternative to microwave systems, mmWave systems work at extremely high frequency bands (30–300 GHz) with mm-level carrier wavelength, which supports the miniaturized implementation of large antenna arrays [4]. Thus, massive multiple-input multiple-output (MIMO) architectures can be efficiently integrated into mmWave systems, which guarantee a large beamforming gain to compensate for the exacerbated free-space and penetration losses [5]. Yet, in order to achieve considerable directional beamforming performance and transmission reliability, accurate channel state information (CSI) should be acquired via channel estimation.

In recent years, the fundamental problem of mmWave channel estimation has been widely investigated. By leveraging the sparse nature of mmWave channels, the estimation issue can be equivalently formulated as a sparse signal recovery problem, which can be solved by compressed sensing (CS) tools, e.g., orthogonal matching pursuit (OMP) [6]. Hierarchical multi-resolution codebooks, as well as, adaptive beam training schemes have been developed to improve the efficiency of channel estimation [7]. The work of [8] developed a generalized approximate message passing algorithm to perform channel estimation with one-bit analog-to-digital converters. The work of [9] proposed an Arnoldi iteration-based method to estimate the singular subspaces of channels. A two-dimensional ESPRIT-based scheme was proposed to explore the angle domain information of channels in [10]. A beamspace channel estimation scheme based on cospase image reconstruction principles was proposed in [11]. A denoising-aided deep learning network was introduced to implement the CSI acquirement in [12]. Moreover, tensor-based signal processing schemes have been exploited for the design of channel estimators [13]–[17].

As indicated in [18], [19], there exist non-negligible time delays across the array aperture for the same data symbol in massive MIMO, leading to the so-called *spatial-wideband* effect [20], [21]. Moreover, large bandwidths of frequency-selective systems induce to the *frequency-wideband* effect. However, most of the existing works [6]–[17] have not jointly considered

the spatial- and frequency-wideband (dual-wideband) effects. In practice, there exist numerous challenges caused by dual-wideband effects in wideband MIMO systems. For example, channel estimation or user scheduling schemes need to track the changes of the effective channel parameters. Hybrid beamforming or sidelobe control schemes should update the codebook design against the beam squint effect. Decoding with dual-wideband effects under low-resolution analog-to-digital converters is also an open issue [19]–[21]. For these reasons, we believe that in order to optimize the transmission performance, the dual-wideband effects from both the spatial and frequency domains should be taken into account for the estimation design of mmWave massive MIMO systems.

In this paper, we precisely address the channel estimation problem of a wideband mmWave orthogonal frequency division multiplexing (OFDM) system. The main contributions of the paper are summarized as follows:

- By leveraging the channel sparsity in angle and delay domains, we transform the massive MIMO channel with dual-wideband effects into a *spatial-frequency wideband (SFW)* model, which is represented by limited multipath parameters, i.e., time delay, complex gain, angle of departure (AoD) and angle of arrival (AoA).
- For the case with negligible dual-wideband effects, we implement a successive beam training strategy within a single time slot, where the base station (BS) transmits precoded pilots in continuous subframes and the mobile station (MS) combines the received signals with parallel measurement streams. We formulate the training signal as a trilinear tensor fitting the canonical polyadic (or CANDECOMP/PARAFAC, CP) model, where the factor matrices contain the multipath parameters [22], [23]. Expanding our prior work [1], we utilize the Vandermonde factor matrix and the spatial smoothing method to develop a **Structured CP Decomposition**-based channel estimation algorithm, abbreviated as *SCPD* [24], [25].
- For the case with significant dual-wideband effects, we perform the beam training procedure across sequential time slots. By introducing a new factor matrix containing time-varying path gains, we reformulate the trilinear tensor model. Furthermore, we employ specially designed hybrid beamforming schemes to exploit the structural information of factor matrices. With the updated CP decomposition procedure and multipath parameter recovery operation, we develop a **Dual-Wideband Effects-oriented Structured CP Decomposition**-based estimation algorithm, abbreviated as *DWE-SCPD*.

- We analyze the uniqueness condition of CP decomposition, which determines the feasible application field of the proposed algorithms. We also evaluate the accuracy of multipath parameter recovery, and compare it with the Cramér-Rao bounds (CRB).

Simulation results indicate that the proposed schemes outperform the traditional methods, e.g., trilinear alternating least squares (TALS) [26] and complex parallel factor analysis (COMFAC) [27], in terms of accuracy, robustness and complexity.

The rest of the paper is organized as follows. Section II introduces preliminaries about the tensor theory and spatial smoothing method. Section III presents the channel model with dual-wideband effects and the hybrid analog-digital architecture for mmWave MIMO-OFDM systems. Section IV introduces the dedicated SCPD algorithm for the case with negligible dual-wideband effects. Section V develops the DWE-SCPD algorithms for the case with significant dual-wideband effects. Section VI presents the numerical results of the channel estimation performance derived by the proposed schemes. Section VII draws the most important conclusions.

*Notations:*  $\mathbf{a}$ ,  $\mathbf{A}$  and  $\mathcal{A}$  denote a vector, a matrix and a tensor, respectively;  $(\cdot)^T$ ,  $(\cdot)^H$ ,  $(\cdot)^{-1}$ ,  $(\cdot)^{-T}$  and  $(\cdot)^\dagger$  denote the transpose, conjugate transpose, inverse, transpose-inverse and pseudo-inverse, respectively;  $[\mathbf{a}]_m$ ,  $[\mathbf{A}]_{m,n}$ ,  $\mathbf{a}_m$ ,  $[\mathbf{A}]_{m:n,:}$  and  $[\mathbf{A}]_{:,m:n}$  denote the  $m$ th entry of  $\mathbf{a}$ , the  $(m,n)$ th entry of  $\mathbf{A}$ , the  $m$ th column of  $\mathbf{A}$ , the submatrix of  $\mathbf{A}$  from the  $m$ th to the  $n$ th rows, and the submatrix of  $\mathbf{A}$  from the  $m$ th to the  $n$ th columns, respectively;  $\|\cdot\|$  and  $\|\cdot\|_F$  denote the 2-norm and Frobenius norm, respectively;  $\otimes$ ,  $\odot$ ,  $*$  and  $\circ$  denote the Kronecker, Khatri-Rao, Hadamard and outer products, respectively;  $\text{r}(\mathbf{A})$  and  $\text{kr}(\mathbf{A})$  denote the rank and Kruskal-rank of  $\mathbf{A}$ , respectively;  $\mathbf{d}(\mathbf{A})$  and  $\mathbf{D}(\mathbf{a})$  denote the diagonal element vector of  $\mathbf{A}$  and the diagonal matrix formed by  $\mathbf{a}$ , respectively;  $\mathcal{I}(n)$  denotes the index set  $\{1, 2, \dots, n\}$ ;  $\mathbf{I}_n$ ,  $\mathbf{1}_{m \times n}$  and  $\mathbf{0}_{m \times n}$  denote an  $n \times n$  identity matrix, a  $m \times n$  all-ones matrix and a  $m \times n$  all-zeros matrix, respectively.  $\mathcal{U}(a, b)$  and  $\mathcal{CN}(\mu, \sigma^2)$  denote a uniform distribution within range  $(a, b)$  and a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ , respectively.

## II. PRELIMINARIES

### A. Concepts of Tensor Theory

We first review some basic concepts of tensor algebra. A tensor stores data in its entries as a multi-dimensional array. Vectors and matrices are one-dimensional and two-dimensional tensors, respectively. Fibers and slices are vectors and matrices derived by fixing all the tensor element indices except those along one dimension and two dimensions, respectively.

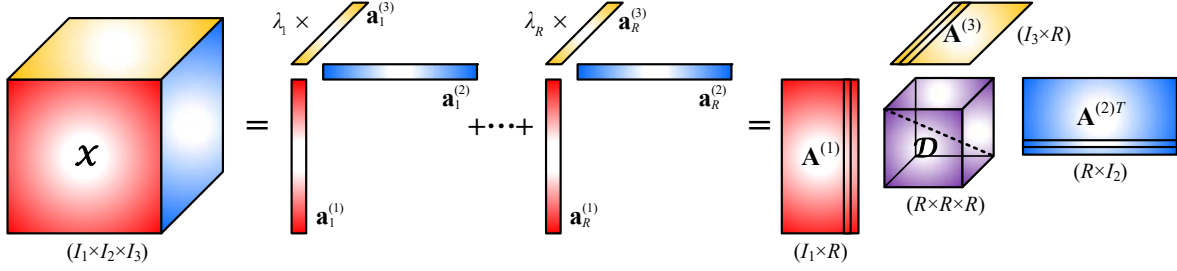


Fig. 1. Diagram of CP decomposition of a third-order tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ .

**Definition 1.**  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  is a  $N$ th-order tensor with the  $(i_1, i_2, \dots, i_N)$ th entry  $\mathcal{X}_{i_1, i_2, \dots, i_N}$ .

**Definition 2 (Tensor Unfolding).**  $\mathbf{X}_{(n)} \in \mathbb{C}^{I_n \times \prod_{k \neq n} I_k}$  is the mode- $n$  unfolding of  $\mathcal{X}$  with the element  $[\mathbf{X}_{(n)}]_{i_n, j}$  mapped from  $\mathcal{X}_{i_1, i_2, \dots, i_N}$ , which satisfies

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^N (i_k - 1) J_{k,n}, J_{k,n} \triangleq \begin{cases} \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m, \mathcal{I}(k-1) \setminus n \neq \emptyset, \\ 1, \text{ otherwise.} \end{cases} \quad (1)$$

**Definition 3 (CP Decomposition).** The CP decomposition factorizes a tensor into a (weighted) sum of a set of rank-one outer products, i.e.,

$$\mathcal{X} = \sum_{r=1}^R \lambda_r \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \dots \circ \mathbf{a}_r^{(N)}, \quad (2)$$

where  $R$  is the tensor rank;  $\{\lambda_r\}_{r=1}^R$  are the weighting coefficients;  $\mathbf{A}^{(n)} \triangleq [\mathbf{a}_1^{(n)}, \dots, \mathbf{a}_R^{(n)}] \in \mathbb{C}^{I_n \times R}$  is the mode- $n$  factor matrix. The mode- $n$  unfolding  $\mathbf{X}_{(n)}$  is equally represented as

$$\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{\Lambda} \left( \mathbf{A}^{(N)} \circ \dots \circ \mathbf{A}^{(n+1)} \circ \mathbf{A}^{(n-1)} \circ \dots \circ \mathbf{A}^{(1)} \right)^T, \quad (3)$$

where  $\mathbf{\Lambda} \triangleq \mathbf{D}([\lambda_1, \dots, \lambda_R]^T)$ . The CP decomposition of a third-order  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  is shown in Fig. 1, where  $\mathcal{D} \in \mathbb{C}^{R \times R \times R}$  is a diagonal tensor containing  $\{\lambda_r\}_{r=1}^R$  in  $\{\mathcal{D}_{r,r,r}\}_{r=1}^R$ .

**Definition 4.** For a tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times \dots \times I_N}$  with rank  $R$ , we introduce the following *matricization* representation [24]

$$\begin{aligned} \mathbf{X}^{[P]} &\triangleq \begin{bmatrix} \mathcal{X}_{1, \dots, 1, 1, \dots, 1} & \mathcal{X}_{1, \dots, 1, 1, \dots, 2} & \dots & \mathcal{X}_{1, \dots, 1, I_{P+1}, \dots, I_N} \\ \mathcal{X}_{1, \dots, 2, 1, \dots, 1} & \mathcal{X}_{1, \dots, 2, 1, \dots, 2} & \dots & \mathcal{X}_{1, \dots, 2, I_{P+1}, \dots, I_N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{I_1, \dots, I_P, 1, \dots, 1} & \mathcal{X}_{I_1, \dots, I_P, 1, \dots, 2} & \dots & \mathcal{X}_{I_1, \dots, I_P, I_{P+1}, \dots, I_N} \end{bmatrix} \\ &= (\mathbf{A}^{(1)} \circ \dots \circ \mathbf{A}^{(P)}) \mathbf{\Lambda} (\mathbf{A}^{(P+1)} \circ \dots \circ \mathbf{A}^{(N)})^T. \end{aligned} \quad (4)$$

## B. Spatial Smoothing

*Spatial smoothing* is a technique commonly applied in sensor array processing to overcome the potential problems caused by rank-deficient matrices [25].

*Definition 5 (Vandermonde matrix).* The factor matrix  $\mathbf{A}^{(n)}$  is said to be Vandermonde if

$$\mathbf{a}_r^{(n)} = [1, z_{n,r}, z_{n,r}^2, \dots, z_{n,r}^{I_n-1}]^T, \quad (5)$$

where  $\{z_{n,r}\}_{r=1}^R$  are called the *generators* of  $\mathbf{A}_n$ .  $\mathbf{A}^{(K_n, n)} \triangleq [\mathbf{A}^{(n)}]_{1:K_n, :}$  denotes the submatrix containing the first  $K_n$  rows of  $\mathbf{A}^{(n)}$ .

We assume that  $\{\mathbf{A}^{(p)}\}_{p=1}^P$ ,  $P < N$  are Vandermonde matrices. By choosing  $K_p + L_p = I_p + 1$ , spatial smoothing maps  $\mathbf{A}^{(p)}$  to  $\mathbf{A}^{(K_p, p)} \mathbf{A}^{(L_p, p)T} \in \mathbb{C}^{K_p \times L_p}$ . Then, one can map  $\mathcal{X}$  to a new tensor  $\mathcal{Y} \in \mathbb{C}^{(K_1 \times \dots \times K_P) \times (I_{P+1} \times \dots \times I_Q) \times (L_1 \times \dots \times L_P) \times (I_{Q+1} \times \dots \times I_N)}$  as

$$\begin{aligned} \mathcal{Y}_{k_1, \dots, k_P, i_{P+1}, \dots, i_Q, l_1, \dots, l_P, i_{Q+1}, \dots, i_N} &= \mathcal{X}_{l_1+k_1-1, \dots, l_P+k_P-1, i_{P+1}, \dots, i_N} \\ &= \sum_{r=1}^R \lambda_r \prod_{p=1}^P z_{p,r}^{l_p+k_p-2} \prod_{q=P+1}^N [\mathbf{a}_r^{(q)}]_{i_q} \\ &= \sum_{r=1}^R \lambda_r \prod_{s=1}^P z_{s,r}^{l_s-1} \prod_{t=1}^P z_{t,r}^{k_t-1} \prod_{q=P+1}^N [\mathbf{a}_r^{(q)}]_{i_q}, \end{aligned} \quad (6)$$

where  $P+1 \leq Q \leq N$ ,  $k_p \in \mathcal{I}(K_p)$ ,  $l_p \in \mathcal{I}(L_p)$ . With (4), the tensor (6) can be matricized to a matrix  $\mathbf{Y}^{[P]} \in \mathbb{C}^{\prod_{p=1}^P K_p \prod_{n=P+1}^Q I_n \times \prod_{p=1}^P L_p \prod_{n=Q+1}^N I_n}$  as

$$\begin{aligned} \mathbf{Y}^{[P]} &= (\mathbf{A}^{(K_1, 1)} \odot \dots \odot \mathbf{A}^{(K_P, P)} \odot \mathbf{A}^{(P+1)} \odot \dots \odot \mathbf{A}^{(Q)}) \\ &\quad \times \Lambda (\mathbf{A}^{(L_1, 1)} \odot \dots \odot \mathbf{A}^{(L_P, P)} \odot \mathbf{A}^{(Q+1)} \odot \dots \odot \mathbf{A}^{(N)})^T. \end{aligned} \quad (7)$$

More specifically, for a third-order tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ , we define two smoothing transformations for the cases of the  $P = 1, 2$  Vandermonde matrices respectively as [28]

*Definition 6.* The smoothing transformation with  $P = 1$  is defined as

$$\mathcal{S}_1(\mathbf{X}) \triangleq [\mathbf{J}_1 \mathbf{X} \quad \mathbf{J}_2 \mathbf{X} \quad \dots \quad \mathbf{J}_{L_1} \mathbf{X}], \quad (8)$$

where  $\mathbf{J}_{l_1} \triangleq [\mathbf{0}_{K_1 \times (l_1-1)} \quad \mathbf{I}_{K_1} \quad \mathbf{0}_{K_1 \times (L_1-l_1)}] \otimes \mathbf{I}_{I_2}$ . The smoothing transformation with  $P = 2$  is defined as

$$\mathcal{S}_2(\mathbf{X}) \triangleq [\mathbf{J}_{1,1} \mathbf{X} \quad \dots \quad \mathbf{J}_{1,L_2} \mathbf{X} \quad \mathbf{J}_{2,1} \mathbf{X} \quad \dots \quad \mathbf{J}_{2,L_2} \mathbf{X} \quad \dots \quad \mathbf{J}_{L_1,L_2} \mathbf{X}], \quad (9)$$

where  $\mathbf{J}_{l_1, l_2} \triangleq \begin{bmatrix} \mathbf{0}_{K_1 \times (l_1-1)} & \mathbf{I}_{K_1} & \mathbf{0}_{K_1 \times (L_1-l_1)} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{0}_{K_2 \times (l_2-1)} & \mathbf{I}_{K_2} & \mathbf{0}_{K_2 \times (L_2-l_2)} \end{bmatrix}$ .

The matrix-oriented operators  $\mathcal{S}_1(\cdot)$ ,  $\mathcal{S}_2(\cdot)$  can achieve the same smoothing results as the tensor-oriented operation (6) as follows [29]:

*Lemma 1.* Let  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  be a tensor with factor matrices  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$ ,  $n \in \{1, 2, 3\}$ . Let  $\mathbf{A}^{(1)}$  and  $\mathbf{A}^{(2)}$  be Vandermonde matrices. Define  $\mathbf{X}^{[2]} = (\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}) \mathbf{\Lambda} \mathbf{A}^{(3)T}$  as a matricization of  $\mathcal{X}$  following (4), then

$$\begin{aligned} \mathcal{S}_1(\mathbf{X}^{[2]}) &= (\mathbf{A}^{(K_1,1)} \odot \mathbf{A}^{(2)}) \mathbf{\Lambda} (\mathbf{A}^{(L_1,1)} \odot \mathbf{A}^{(3)})^T, \\ \mathcal{S}_2(\mathbf{X}^{[2]}) &= (\mathbf{A}^{(K_1,1)} \odot \mathbf{A}^{(K_2,2)}) \mathbf{\Lambda} (\mathbf{A}^{(L_1,1)} \odot \mathbf{A}^{(L_2,2)} \odot \mathbf{A}^{(3)})^T. \end{aligned} \quad (10)$$

### III. CHANNEL AND SYSTEM MODEL

#### A. Channel Model with Dual-Wideband Effects

We consider a MIMO-OFDM system consisting of one BS with an  $N_{\text{bs}}$ -antenna uniform linear array (ULA) and several MSs, each with an  $N_{\text{ms}}$ -antenna ULA. The total number of transmitting subcarriers is denoted by  $K_0$ . In this paper, we focus on the downlink transmission with a downlink carrier frequency  $f_c$  and bandwidth  $f_s$  (symbol duration  $T_s = 1/f_s$ ), where the MSs individually perform their channel estimation procedure.<sup>1</sup>

Exploiting the sparsity nature of mmWave channels, we adopt the extended Saleh-Valenzuela model, which consists of multiple individual scattering paths [30]. The physical AoD and AoA of the  $l$ th path are denoted by  $\varphi_l$  and  $\vartheta_l$ , respectively. For an antenna array with adjacent spacing  $d$ , the corresponding spatial AoD and AoA are defined as  $\phi_l \triangleq \frac{d}{\lambda_c} \sin \varphi_l$  and  $\theta_l \triangleq \frac{d}{\lambda_c} \sin \vartheta_l$ , respectively, where  $\lambda_c$  is the carrier wavelength. We denote the time delay of the  $l$ th path from the  $n$ th BS antenna to the  $m$ th MS antenna by  $\tau_{l,m,n}$ <sup>2</sup>, and the time-varying complex gain of the  $l$ th path at the  $t$ th time slot by  $\bar{\alpha}_{t,l}$ . For a MIMO system with ULA configurations,  $\tau_{l,m,n}$  is represented as

$$\tau_{l,m,n} \triangleq \tau_l + (m-1) \frac{\theta_l}{f_c} + (n-1) \frac{\phi_l}{f_c}, \quad (11)$$

<sup>1</sup>For the uplink scenario, our proposed scheme permits the MSs to simultaneously perform channel estimation through non-overlapping subcarrier bands.

<sup>2</sup>For simplicity, we assume that the maximal time delay is within the duration  $\tau_0$  of a single time slot.



where  $\tau_l$  is the free-space propagation delay of the  $l$ th path. Then, given a baseband signal  $x_n(\tau)$  transmitted from the  $n$ th BS antenna, the received baseband signal at the  $m$ th MS antenna during the  $t$ th slot can be expressed as

$$\begin{aligned} y_{t,m}(\tau) &= \sum_{l=1}^L \bar{\alpha}_{t,l} x_n(\tau - \tau_{l,m,n}) e^{-j2\pi f_c \tau_{l,m,n}} \\ &= \sum_{l=1}^L \bar{\alpha}_{t,l} e^{-j2\pi f_c \tau_l} x_n(\tau - \tau_{l,m,n}) e^{-j2\pi f_c (m-1) \frac{\theta_l}{f_c}} e^{-j2\pi f_c (n-1) \frac{\phi_l}{f_c}} \\ &= \sum_{l=1}^L \alpha_{t,l} x_n(\tau - \tau_{l,m,n}) e^{-j2\pi (m-1) \theta_l} e^{-j2\pi (n-1) \phi_l}, \end{aligned} \quad (12)$$

where  $L$  is the total number of signal paths;  $\alpha_{t,l} \triangleq \bar{\alpha}_{t,l} e^{-j2\pi f_c \tau_l}$  is the equivalent path gain. Then, the downlink spatial-time channel response at the  $m$ th MS antenna from the  $n$ th BS antenna during the  $t$ th time slot is represented as

$$[\mathbf{H}_t(\tau)]_{m,n} = \sum_{l=1}^L \alpha_{t,l} [\mathbf{a}_{\text{ms}}(\theta_l)]_m [\mathbf{a}_{\text{bs}}(\phi_l)]_n \delta(\tau - \tau_{l,m,n}), \quad (13)$$

where

$$\mathbf{a}_{\text{ms}}(\theta) \triangleq [1, e^{-j2\pi\theta}, \dots, e^{-j2\pi(N_{\text{ms}}-1)\theta}]^T, \quad (14a)$$

$$\mathbf{a}_{\text{bs}}(\phi) \triangleq [1, e^{-j2\pi\phi}, \dots, e^{-j2\pi(N_{\text{bs}}-1)\phi}]^T, \quad (14b)$$

are the well-known *spatial* steering vectors of the MS and BS antenna arrays, respectively.

Performing the continuous time Fourier transform of (13), we can derive the *spatial-frequency* channel response at the  $k$ th subcarrier during the  $t$ th time slot as

$$\begin{aligned} [\mathbf{H}_{t,k}]_{m,n} &= \int_0^{\tau_0} [\mathbf{H}_t(\tau)]_{m,n} e^{-j2\pi f_k \tau} d\tau \\ &= \sum_{l=1}^L \alpha_{t,l} [\mathbf{a}_{\text{ms}}(\theta_l)]_m [\mathbf{a}_{\text{bs}}(\phi_l)]_n e^{-j2\pi f_k \tau_{l,m,n}} \\ &= \sum_{l=1}^L \alpha_{t,l} e^{-j2\pi f_k \tau_l} [\mathbf{a}_{\text{ms},k}(\theta_l)]_m [\mathbf{a}_{\text{bs},k}(\phi_l)]_n, \end{aligned} \quad (15)$$

where  $f_k \triangleq \frac{k f_s}{K_0}$  is the frequency shift of the  $k$ th subcarrier; and

$$\mathbf{a}_{\text{ms},k}(\theta) \triangleq \left[ 1, e^{-j2\pi \left(1 + \frac{f_k}{f_c}\right) \theta}, \dots, e^{-j2\pi (N_{\text{ms}}-1) \left(1 + \frac{f_k}{f_c}\right) \theta} \right]^T, \quad (16a)$$

$$\mathbf{a}_{\text{bs},k}(\phi) \triangleq \left[ 1, e^{-j2\pi \left(1 + \frac{f_k}{f_c}\right) \phi}, \dots, e^{-j2\pi (N_{\text{bs}}-1) \left(1 + \frac{f_k}{f_c}\right) \phi} \right]^T, \quad (16b)$$

are the so-called *spatial-frequency* steering vectors at the  $k$ th subcarrier. Specifically, the array-dependent terms  $\{(m-1)\theta\}_{m \in \mathcal{I}(N_{\text{ms}})}$  come from the spatial-wideband effect, and the additional phase shift  $f_k \theta / f_c$  results from the frequency-wideband effect. Finally, the channel matrix at the  $k$ th subcarrier during the  $t$ th time slot is derived as

$$\begin{aligned} \mathbf{H}_{t,k} &= \sum_{l=1}^L \alpha_{t,l} e^{-j2\pi f_k \tau_l} \mathbf{a}_{\text{ms},k}(\theta_l) \mathbf{a}_{\text{bs},k}^T(\phi_l) \\ &= \sum_{l=1}^L \alpha_{t,l} e^{-j2\pi f_k \tau_l} (\mathbf{a}_{\text{ms}}(\theta_l) \mathbf{a}_{\text{bs}}^T(\phi_l)) * \Psi_k(\theta_l, \phi_l). \end{aligned} \quad (17)$$

where  $[\Psi_k(\theta, \phi)]_{m,n} \triangleq e^{-j2\pi \frac{f_k}{f_c} ((m-1)\theta + (n-1)\phi)}$ ,  $m \in \mathcal{I}(N_{\text{ms}})$ ,  $n \in \mathcal{I}(N_{\text{bs}})$ . Clearly, (17) provides a more accurate channel model for massive MIMO by considering the dual-wideband effects, which is referred to as the *SFW* channel.

In small-scale antennas or in large-scale antennas but with very narrow bandwidth,  $\Psi_k$  approximates a nearly all-ones matrix. For wideband massive MIMO systems, however,  $\Psi_k$  will fundamentally influence the system performance. For example, for a typical  $32 \times 64$ -ULA system with  $d = \lambda_c/2$ ,  $f_s = 1$  GHz at  $f_c = 60$  GHz, the maximum phase shift of  $\{\Psi_k(\theta, \phi)\}_{k \in \mathcal{I}(K_0)}$ , i.e.,  $\frac{f_s}{f_c} \frac{d}{\lambda_c} (N_{\text{ms}} + N_{\text{bs}})$ , is approximated to  $0.25\pi$ . Concretely, the spatial-wideband effect yields a maximum delay of  $0.80T_s$  across the array aperture; the frequency-wideband effect yields a maximum deviation of  $10.39^\circ$  from the physical angle  $79.61^\circ$ .<sup>3</sup>

## B. Hybrid Transceiver Model

To facilitate efficient hardware implementation, a hybrid analog-digital architecture with  $N_d$  data streams is considered. The system supports  $M_{\text{bs}}$  and  $M_{\text{ms}}$  RF chains at the BS and MS respectively, satisfying  $N_d \leq M_{\text{bs}} \leq N_{\text{bs}}$ ,  $N_d \leq M_{\text{ms}} \leq N_{\text{ms}}$  (see Fig. 2).

At the  $k$ th subcarrier, the BS employs an analog RF precoder  $\mathbf{F}_A \in \mathbb{C}^{N_{\text{bs}} \times M_{\text{bs}}}$  and a digital baseband precoder  $\mathbf{F}_{D,k} \in \mathbb{C}^{M_{\text{bs}} \times N_d}$ , while the MS employs an RF combiner  $\mathbf{W}_A \in \mathbb{C}^{N_{\text{ms}} \times M_{\text{ms}}}$  and a baseband combiner  $\mathbf{W}_{D,k} \in \mathbb{C}^{M_{\text{ms}} \times N_d}$ . Since the analog circuits are implemented by phase shifters, a constant-magnitude hardware constraint on the elements of analog beamformers is introduced, i.e.,  $|\mathbf{F}_A]_{i,j}| = |\mathbf{W}_A]_{i,j}| = 1, \forall i, j$ .

<sup>3</sup>When the physical angle exceeds  $79.61^\circ$ , a phase wrapping happens, which may increase the maximum deviation to  $-169.52^\circ$ .

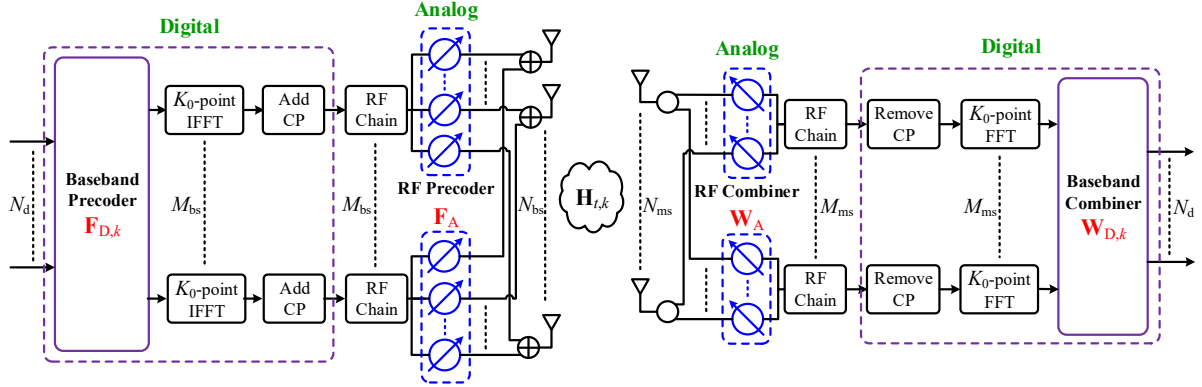


Fig. 2. Hardware block diagram of mmWave MIMO-OFDM system with hybrid beamforming structure.

At the BS side, the baseband signal  $\mathbf{s}_{t,k} \in \mathbb{C}^{N_d}$  at the  $k$ th subcarrier within the  $t$ th time slot is precoded by  $\mathbf{F}_{D,k}$ , and transformed to the time-domain via a  $K_0$ -point inverse discrete Fourier transform. After a cyclic prefix addition,  $\mathbf{F}_A$  is applied to all subcarriers. Then, the precoded signal at the  $k$ th subcarrier during the  $t$ th time slot is represented as

$$\mathbf{x}_{t,k} = \mathbf{F}_A \mathbf{F}_{D,k} \mathbf{s}_{t,k} = \mathbf{F}_k \mathbf{s}_{t,k}, \quad (18)$$

where  $\mathbf{F}_k \triangleq \mathbf{F}_A \mathbf{F}_{D,k}$  is the equivalent precoding matrix.

At the MS side, the received signal is combined by  $\mathbf{W}_A$ . Then the cyclic prefix is removed and the signal is transformed back to the frequency-domain via a discrete Fourier transform. The baseband symbols are processed by  $\mathbf{W}_{D,k}$  for each subcarrier. Finally, under the perfect synchronization assumption, the received signal can be given by<sup>4</sup>

$$\begin{aligned} \mathbf{y}_{t,k} &= \mathbf{W}_{D,k}^T \mathbf{W}_A^T (\mathbf{H}_{t,k} \mathbf{x}_{t,k} + \mathbf{n}_{t,k}) \\ &= \mathbf{W}_k^T \mathbf{H}_{t,k} \mathbf{F}_k \mathbf{s}_{t,k} + \mathbf{W}_k^T \mathbf{n}_{t,k}, \end{aligned} \quad (19)$$

where  $\mathbf{W}_k \triangleq \mathbf{W}_A \mathbf{W}_{D,k}$  is the equivalent combining matrix;  $\mathbf{n}_{t,k} \in \mathbb{C}^{N_{ms}}$  is the additive noise following independent and identically distributed (i.i.d.)  $\mathcal{CN}(0, \sigma_n^2)$ .

<sup>4</sup>Imperfect hardware or synchronization yields carrier frequency offsets (CFO), which result in unknown phase errors in the received signal that linearly increase with time. Since these errors can be regarded as unidentifiable phase shifts of the time-varying path gains  $\{\alpha_{t,l}\}$ , we assume perfect synchronization without CFO in this paper.

#### IV. CHANNEL ESTIMATION WITHOUT WIDEBAND EFFECTS

We first consider the special case with negligible wideband-effects and apply a successive beam training strategy to perform the channel estimation procedure.<sup>5</sup> When the phase shifts inside  $\{\Phi_k(\theta_l, \phi_l)\}$  in (17) approximate to ones, i.e., the array size or system bandwidth is not large enough, the steering vectors in (16) approximate those in (14), and the channel matrix (17) at the  $k$ th subcarrier within an arbitrary time slot simplifies to

$$\mathbf{H}_k = \sum_{l=1}^L \alpha_l e^{-j2\pi f_k \tau_l} \mathbf{a}_{\text{ms}}(\theta_l) \mathbf{a}_{\text{bs}}^T(\phi_l). \quad (20)$$

##### A. Beam Training Scheme

We employ  $K$  subcarriers for beam training.<sup>6</sup> Without prior CSI information, we apply a frequency-flat training beamforming scheme, which will subsequently help us formulate the third-order signal tensor model. Considering that one time slot is divided into  $P$  subframes, the BS transmits a precoded pilot signal  $\mathbf{f}_p = \mathbf{F}_A \mathbf{F}_{D,p} \mathbf{s}_p \in \mathbb{C}^{N_{\text{bs}}}$  with  $\mathbf{F}_A \in \mathbb{C}^{N_{\text{bs}} \times M_{\text{bs}}}$ ,  $\mathbf{F}_{D,p} \in \mathbb{C}^{M_{\text{bs}} \times N_d}$ ,  $\mathbf{s}_p \in \mathbb{C}^{N_d}$  at the  $p$ th subframe.<sup>7</sup> The MS employs a measurement vector  $\mathbf{w}_q = \mathbf{W}_A \mathbf{w}_{D,q} \in \mathbb{C}^{N_{\text{ms}}}$  with  $\mathbf{W}_A \in \mathbb{C}^{N_{\text{ms}} \times M_{\text{ms}}}$ ,  $\mathbf{w}_{D,q} \in \mathbb{C}^{M_{\text{ms}}}$  at the  $q$ th stream to combine the received signal, yielding

$$y_{k,q,p} = \mathbf{w}_q^T \mathbf{H}_k \mathbf{f}_p + \mathbf{w}_q^T \mathbf{n}_{k,q,p}, \quad (21)$$

where  $\mathbf{n}_{k,q,p} \in \mathbb{C}^{N_{\text{ms}}}$  is the noise vector.

We assume that the MS activates  $Q \leq M_{\text{ms}}$  parallel streams to simultaneously combine the received training signal, yielding

$$\mathbf{y}_{k,p} = \mathbf{W}^T \mathbf{H}_k \mathbf{f}_p + \mathbf{d}(\mathbf{W}^T \mathbf{N}_{k,p}), \quad (22)$$

where  $\mathbf{y}_{k,p} = [y_{k,1,p}, \dots, y_{k,Q,p}]^T \in \mathbb{C}^Q$ ;  $\mathbf{W} \triangleq [\mathbf{w}_1, \dots, \mathbf{w}_Q] \in \mathbb{C}^{N_{\text{ms}} \times Q}$  contains the measurement combiners, and  $\mathbf{N}_{k,p} = [\mathbf{n}_{k,1,p}, \dots, \mathbf{n}_{k,Q,p}] \in \mathbb{C}^{N_{\text{ms}} \times Q}$ .

<sup>5</sup>In this section, both the beam training and signal processing are performed within a single time slot. Hence, the subscript  $t$  of the channel matrix or path gain is omitted for notation simplicity.

<sup>6</sup>In this work, we adopt a comb-type pilot arrangement, where successive subcarriers with indices  $k \in \mathcal{I}(K)$  are selected for beam training.

<sup>7</sup>If  $\mathbf{F}_{D,p}$  is frame-invariant and  $P \geq N_d$ , one can utilize frequency-dependent orthogonal pilots  $\mathbf{S}_k \triangleq [\mathbf{s}_{k,1}, \dots, \mathbf{s}_{k,P}] \in \mathbb{C}^{N_d \times P}$  at the  $k$ th subcarrier to better control the peak-to-average power ratio, which can be removed at the MS end via multiplying the received signal matrix by  $\mathbf{S}_k^H$ .

We assume that the BS successively switches the precoded pilots  $\mathbf{f}_p$  at  $P$  subframes within a single time slot, leading to the received training signals as

$$\mathbf{Y}_k = \mathbf{W}^T \mathbf{H}_k \mathbf{F} + \mathbf{N}_k, \quad (23)$$

where  $\mathbf{Y}_k = [\mathbf{y}_{k,1}, \dots, \mathbf{y}_{k,P}] \in \mathbb{C}^{Q \times P}$  denotes the training signal matrix at the  $k$ th subcarrier;  $\mathbf{F} \triangleq [\mathbf{f}_1, \dots, \mathbf{f}_P] \in \mathbb{C}^{N_{\text{bs}} \times P}$  contains the precoded pilot signals, and  $\mathbf{N}_k \in \mathbb{C}^{Q \times P}$  is the equivalent combined noise.

By substituting the channel coefficients of (20) into (23), we rewrite the received signal as

$$\begin{aligned} \mathbf{Y}_k &= \sum_{l=1}^L \alpha_l e^{-j2\pi f_k \tau_l} \mathbf{W}^T \mathbf{a}_{\text{ms}}(\theta_l) \mathbf{a}_{\text{bs}}^T(\phi_l) \mathbf{F} + \mathbf{N}_k \\ &= \sum_{l=1}^L \alpha_l e^{-j2\pi \frac{k f_s}{K_0} \tau_l} \tilde{\mathbf{a}}_{\text{ms}}(\theta_l) \tilde{\mathbf{a}}_{\text{bs}}^T(\phi_l) + \mathbf{N}_k, \end{aligned} \quad (24)$$

where  $\tilde{\mathbf{a}}_{\text{ms}}(\theta) \triangleq \mathbf{W}^T \mathbf{a}_{\text{ms}}(\theta) \in \mathbb{C}^Q$  and  $\tilde{\mathbf{a}}_{\text{bs}}(\phi) \triangleq \mathbf{F}^T \mathbf{a}_{\text{bs}}(\phi) \in \mathbb{C}^P$  are the equivalent steering vectors after combining and precoding, respectively.

By concatenating the received data at  $K$  subcarriers, we can derive a third-order tensor  $\mathcal{Y} \in \mathbb{C}^{Q \times P \times K}$  with its  $(q, p, k)$ th entry given by  $[\mathbf{Y}_k]_{q,p}$ , whose three modes stand for the combining data streams, the precoding subframes and the training subcarriers, respectively. According to (24), we note that each slice  $\mathbf{Y}_k$  of tensor  $\mathcal{Y}$  can be viewed as a weighted sum of a set of rank-one outer products. Hence, the tensor  $\mathcal{Y}$  admits the CP model, which can be expressed as

$$\mathcal{Y} = \sum_{l=1}^L \tilde{\mathbf{a}}_{\text{ms}}(\theta_l) \circ \tilde{\mathbf{a}}_{\text{bs}}(\phi_l) \circ (\alpha_l \mathbf{g}(\tau_l)) + \mathcal{N}, \quad (25)$$

where  $\mathcal{N} \in \mathbb{C}^{Q \times P \times K}$  is the tensor form of noise; and

$$\mathbf{g}(\tau) \triangleq \left[ e^{-j2\pi \frac{f_s}{K_0} \tau}, \dots, e^{-j2\pi \frac{K f_s}{K_0} \tau} \right]^T, \quad (26)$$

which can be seen as a frequency-domain steering vector pointing towards the time delay  $\tau$ .

Due to the sparse scattering nature of mmWave channels, the number of paths  $L$  is commonly small relative to the dimensions of tensor. Therefore,  $\mathcal{Y}$  has an inherent low-rank characteristic, which ensures that the CP decomposition is unique up to scaling and permutation ambiguities. An estimation of channel parameters  $\{\theta_l, \phi_l, \alpha_l, \tau_l\}$  can be obtained by analyzing the factor matrices

of (25), which are defined as

$$\mathbf{B}^{(1)} \triangleq [\tilde{\mathbf{a}}_{\text{ms}}(\theta_1), \dots, \tilde{\mathbf{a}}_{\text{ms}}(\theta_L)] \in \mathbb{C}^{Q \times L}, \quad (27a)$$

$$\mathbf{B}^{(2)} \triangleq [\tilde{\mathbf{a}}_{\text{bs}}(\phi_1), \dots, \tilde{\mathbf{a}}_{\text{bs}}(\phi_L)] \in \mathbb{C}^{P \times L}, \quad (27b)$$

$$\mathbf{B}^{(3)} \triangleq [\alpha_1 \mathbf{g}(\tau_1), \dots, \alpha_L \mathbf{g}(\tau_L)] \in \mathbb{C}^{K \times L}. \quad (27c)$$

### B. Uniqueness Condition

We now elaborate on the uniqueness condition of CP decomposition, which guarantees that the derived factor matrices contain the accurate information of channel parameters. We first present a well-known sufficient uniqueness condition as follows [31].

*Lemma 2.* Let  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  be a tensor with factor matrices  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$ ,  $n \in \{1, 2, 3\}$ .<sup>8</sup> If

$$\text{kr}(\mathbf{A}^{(1)}) + \text{kr}(\mathbf{A}^{(2)}) + \text{kr}(\mathbf{A}^{(3)}) \geq 2R + 2, \quad (28)$$

then the rank of  $\mathcal{X}$  is  $R$ , and the CP decomposition is unique. In the generic case,<sup>9</sup> condition (28) becomes

$$\min(I_1, R) + \min(I_2, R) + \min(I_3, R) \geq 2R + 2. \quad (29)$$

Lemma 2 indicates that each factor matrix must have a Kruskal-rank greater than one, which means that there exist no linear dependent columns. By ingeniously designing the beamforming matrices  $\mathbf{W}$  and  $\mathbf{F}$ , this uniqueness condition can be satisfied.

Leveraging the structural characteristic inside the tensor contributes to the relaxation of uniqueness condition [24]. In our case, we find that the factor matrix  $\mathbf{B}^{(3)}$  has a Vandermonde nature due to the shift-invariant vectors  $\{\mathbf{g}(\tau_l)\}_{l=1}^L$ , where the generators are  $\{z_l \triangleq e^{-j2\pi f_s \tau_l / K_0}\}_{l=1}^L$ .

By utilizing the spatial smoothing, we can obtain a relaxed uniqueness condition of CP decomposition as follows [14], [15], [24]

*Theorem 1.* Let  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  be a tensor with factor matrices  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$ ,  $n \in \{1, 2, 3\}$ .

Let  $\mathbf{A}^{(1)}$  be Vandermonde matrix with distinct generators  $\{z_r\}_{r=1}^R$ . If

$$\begin{cases} \text{r}(\mathbf{A}^{(K_1-1,1)} \odot \mathbf{A}^{(2)}) = R, \\ \text{r}(\mathbf{A}^{(L_1,1)} \odot \mathbf{A}^{(3)}) = R, \end{cases} \quad (30)$$

<sup>8</sup>Since the scaling weights  $\mathbf{\Lambda}$  can be incorporated into the factor matrices, we assume that  $\mathbf{\Lambda} = \mathbf{I}_R$  without loss of generality.

<sup>9</sup>A generic property means that it holds with probability one when the entries of factor matrices are drawn from absolutely continuous probability density functions.

with  $K_1 + L_1 = I_1 + 1$ , then the rank of  $\mathcal{X}$  is  $R$  and the CP decomposition is unique. In the generic case, condition (30) becomes

$$\min_{K_1+L_1=I_1+1} ((K_1 - 1)I_2, L_1I_3) \geq R. \quad (31)$$

Note that Theorem 1 does not prevent  $\text{kr}(\mathbf{A}^{(2)}) = 1$  and/or  $\text{kr}(\mathbf{A}^{(3)}) = 1$ , and even allows  $I_2 = 1$  or  $I_3 = 1$ . In our case, the uniqueness holds even for the circumstances that some channel paths share the same AoD and/or AoA. Furthermore, Theorem 1 allows us to deal with CP decomposition problem with  $\min_{m \neq n} I_m I_n < R$ . In contrast,  $\min_{m \neq n} I_m I_n \geq R$  is necessary for unique decomposition of common CP models without structural constraints.

### C. Proposed Algorithm

If the tensor rank is unknown, we need a preliminary step to determine it for the realization of CP decomposition. Assuming that the beamformed noise  $\mathcal{N}$  in (25) follows an i.i.d. Gaussian distribution and is independent of  $\mathcal{Y}$ , we adopt a minimum description length (MDL) method to estimate the number of signal path components  $L$  [32], [33].

We perform mode-1 unfolding on  $\mathcal{Y}$  by (1), and compute the eigenvalue decomposition (EVD) of the sample covariance matrix as

$$\hat{\Sigma}_{\mathbf{Y}_{(1)}} \triangleq \frac{1}{PK} \mathbf{Y}_{(1)} \mathbf{Y}_{(1)}^H = \mathbf{V}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^H, \quad (32)$$

where  $\mathbf{\Lambda}_1 \triangleq \mathbf{D}([\lambda_{1,1}, \dots, \lambda_{1,Q}]^T)$  contains the eigenvalues in descending order, and  $\mathbf{V}_1 \in \mathbb{C}^{Q \times Q}$  contains the corresponding eigenvectors. The matrix rank of  $\mathbf{Y}_{(1)} \in \mathbb{C}^{Q \times PK}$  can be estimated by the MDL criterion as

$$\hat{L}_1 = \arg \min_{\ell} \frac{\ell}{2} (2Q - \ell) \log(PK) - \log \left( \frac{\prod_{i=\ell+1}^Q \lambda_{1,i}^{\frac{1}{Q-\ell}}}{\frac{1}{Q-\ell} \sum_{i=\ell+1}^Q \lambda_{1,i}} \right)^{PK(Q-\ell)}. \quad (33)$$

The ranks  $\hat{L}_2, \hat{L}_3$  can be obtained by similarly applying (32), (33) to the mode-2, 3 unfolding  $\mathbf{Y}_{(2)}, \mathbf{Y}_{(3)}$  respectively. The tensor rank can be estimated as  $\hat{L} = \min\{\hat{L}_n\}_{n=1}^3$ . Moreover, one can incrementally utilize the eigenvalues  $\{\mathbf{\Lambda}_n\}_{n=1}^3$  to develop a sequential detection MDL procedure [32].

We turn to the realization of CP decomposition (25), which can be accomplished by solving

$$\min_{\hat{\mathbf{B}}^{(1)}, \hat{\mathbf{B}}^{(2)}, \hat{\mathbf{B}}^{(3)}} \left\| \left\| \mathcal{Y} - \sum_{l=1}^{\hat{L}} \hat{\mathbf{b}}_l^{(1)} \circ \hat{\mathbf{b}}_l^{(2)} \circ \hat{\mathbf{b}}_l^{(3)} \right\|_F \right\|^2, \quad (34)$$

where  $\hat{\mathbf{b}}_l^{(n)}$  is the column of the estimated factor matrices  $\hat{\mathbf{B}}^{(n)}$ . The most common algorithm of CP decomposition, i.e., TALS [13], [22], sequentially updates one of the factor matrices to its least squares (LS) solution, fixing the other two matrices until convergence as

$$\hat{\mathbf{B}}_{[i+1]}^{(1)} = \arg \min_{\hat{\mathbf{B}}^{(1)}} \left\| \mathbf{Y}_{(1)}^T - \left( \hat{\mathbf{B}}_{[i]}^{(3)} \odot \hat{\mathbf{B}}_{[i]}^{(2)} \right) \hat{\mathbf{B}}^{(1)T} \right\|_F^2, \quad (35a)$$

$$\hat{\mathbf{B}}_{[i+1]}^{(2)} = \arg \min_{\hat{\mathbf{B}}^{(2)}} \left\| \mathbf{Y}_{(2)}^T - \left( \hat{\mathbf{B}}_{[i]}^{(3)} \odot \hat{\mathbf{B}}_{[i+1]}^{(1)} \right) \hat{\mathbf{B}}^{(2)T} \right\|_F^2, \quad (35b)$$

$$\hat{\mathbf{B}}_{[i+1]}^{(3)} = \arg \min_{\hat{\mathbf{B}}^{(3)}} \left\| \mathbf{Y}_{(3)}^T - \left( \hat{\mathbf{B}}_{[i+1]}^{(2)} \odot \hat{\mathbf{B}}_{[i+1]}^{(1)} \right) \hat{\mathbf{B}}^{(3)T} \right\|_F^2, \quad (35c)$$

where  $\hat{\mathbf{B}}_{[i]}^{(n)}$  is the updated version of  $\hat{\mathbf{B}}^{(n)}$  during the  $i$ th iteration. The main drawbacks of the TALS and its modified versions [34], [35] are: (i) the convergence speed of this iterative method is relatively slow, which incurs high burden of computational complexity; (ii) the alternating result heavily depends on the initialization state, which may converge to a biased local optimal solution. Moreover, the TALS method does not consider the intrinsic structure of factor matrices, which are usually *a priori* known to the system [13]–[17]. By exploiting the Vandermonde nature of  $\mathbf{B}^{(3)}$ , we develop an improved channel parameter estimation scheme aided by the SCPD algorithm according to Theorem 1.

We define an integer pair  $(K_3, L_3)$  subject to  $K_3 + L_3 = K + 1$ , and modify the selection matrices in (8) as  $\mathbf{J}_{l_3} \triangleq [\mathbf{0}_{K_3 \times (l_3-1)} \mathbf{I}_{K_3} \mathbf{0}_{K_3 \times (L_3-l_3)}] \otimes \mathbf{I}_P, l_3 \in \mathcal{I}(L_3)$ . Then, we sequentially perform mode-1 unfolding and spatial smoothing on  $\mathcal{Y}$  by (1) and (8) respectively, yielding

$$\mathbf{Y}_{\mathcal{S}} \triangleq \mathcal{S}_1(\mathbf{Y}_{(1)}^T) = (\mathbf{B}^{(K_3,3)} \odot \mathbf{B}^{(2)}) (\mathbf{B}^{(L_3,3)} \odot \mathbf{B}^{(1)})^T + \mathbf{N}_{\mathcal{S}}, \quad (36)$$

where  $\mathbf{N}_{\mathcal{S}} \in \mathbb{C}^{K_3 P \times L_3 Q}$  is the noise matrix. We then follow [14], [15], [24] and employ an ESPRIT-like approach to recover the factor matrices. The relationship between the estimated factor matrix  $\hat{\mathbf{B}}^{(n)}$  and the real one is represented as ( $\hat{L} = L$ )

$$\hat{\mathbf{B}}^{(n)} = \mathbf{B}^{(n)} \mathbf{\Lambda}_n \mathbf{\Pi}_L + \mathbf{E}_n, n \in \{1, 2, 3\} \quad (37)$$

where  $\mathbf{\Lambda}_n \in \mathbb{C}^{L \times L}$  is the unknown nonsingular diagonal scaling ambiguity matrix, which satisfy  $\prod_{n=1}^3 \mathbf{\Lambda}_n = \mathbf{I}_L$ ;  $\mathbf{\Pi}_L \in \mathbb{C}^{L \times L}$  is an unknown permutation matrix that can be ignored since it is common to all factor matrices;  $\mathbf{E}_n \in \mathbb{C}^{L \times L}$  is the estimation error matrix.

After completing the tensor decomposition, we turn to the estimation of multipath parameters. First, the time delays  $\{\hat{\tau}_l\}$  can be directly obtained from the estimated generators  $\{\hat{z}_l\}$  as

$$\hat{\tau}_l = \frac{jK_0}{2\pi f_s} \log \hat{z}_l = -\frac{K_0}{2\pi f_s} \angle \hat{z}_l, \quad (38)$$



where  $\angle$  denotes the phase angle extraction operator.

Then, the estimated AoA/AoDs can be derived by a correlation-based strategy as follows:<sup>10</sup>

$$\hat{\theta}_l = \arg \max_{\theta} \frac{|\hat{\mathbf{b}}_l^{(1)H} \tilde{\mathbf{a}}_{\text{ms}}(\theta)|^2}{\|\hat{\mathbf{b}}_l^{(1)}\|^2 \|\tilde{\mathbf{a}}_{\text{ms}}(\theta)\|^2}, \quad (39a)$$

$$\hat{\phi}_l = \arg \max_{\phi} \frac{|\hat{\mathbf{b}}_l^{(2)H} \tilde{\mathbf{a}}_{\text{bs}}(\phi)|^2}{\|\hat{\mathbf{b}}_l^{(2)}\|^2 \|\tilde{\mathbf{a}}_{\text{bs}}(\phi)\|^2}, \quad (39b)$$

which can be solved by the one-dimensional searching method. Interestingly, considering the ULA geometries, the solution can also be derived by a polynomial method [23]. Denoting the steering vector  $\mathbf{a}_{\text{ms}}(\theta)$  as  $\mathbf{z}(h) = [1, h, \dots, h^{N_{\text{ms}}-1}]^T$  with  $h \triangleq e^{-j2\pi\theta}$ , we rewrite (39a) as

$$\begin{aligned} \hat{h}_l &= \arg \min_h \left( 1 - \frac{|\hat{\mathbf{b}}_l^{(1)H} \mathbf{W}^T \mathbf{z}(h)|^2}{\|\hat{\mathbf{b}}_l^{(1)}\|^2 \|\mathbf{W}^T \mathbf{z}(h)\|^2} \right) \\ &= \arg \min_h \left( \frac{\mathbf{z}^H(h) \mathbf{W}^* \left( \|\hat{\mathbf{b}}_l^{(1)}\|^2 \mathbf{I} - \hat{\mathbf{b}}_l^{(1)} \hat{\mathbf{b}}_l^{(1)H} \right) \mathbf{W}^T \mathbf{z}(h)}{\|\hat{\mathbf{b}}_l^{(1)}\|^2 \mathbf{z}^H(h) \mathbf{W}^* \mathbf{W}^T \mathbf{z}(h)} \right). \end{aligned} \quad (40)$$

Let  $\mathbf{T}_l = \mathbf{W}^* \left( \|\hat{\mathbf{b}}_l^{(1)}\|^2 \mathbf{I} - \hat{\mathbf{b}}_l^{(1)} \hat{\mathbf{b}}_l^{(1)H} \right) \mathbf{W}^T$ , the numerator in (40) can be further expressed as a polynomial with respect to  $h$  as

$$\mathbf{z}^H(h) \mathbf{T}_l \mathbf{z}(h) = \sum_{k=-N_{\text{ms}}+1}^{N_{\text{ms}}-1} \left( \sum_{j-i=k} [\mathbf{T}_l]_{i,j} \right) h^k. \quad (41)$$

Since all the  $(N_{\text{ms}} - 1)$  pairs of roots for (41) are solutions of the minimization problem (40), we can estimate the AoAs  $\{\hat{\theta}_l\}$  by searching over the normalized  $(N_{\text{ms}} - 1)$  roots with unit amplitudes denoted by  $\{\nu_n\}_{n \in \mathcal{I}(N_{\text{ms}}-1)}$ . The solution of AoDs  $\{\hat{\phi}_l\}$  can be similarly derived. This approach introduces an extra polynomial root calculation to discretize the searching range.

Finally, we substitute the estimated  $\hat{\theta}_l$  and  $\hat{\phi}_l$  back to  $\tilde{\mathbf{a}}_{\text{ms}}(\cdot)$  and  $\tilde{\mathbf{a}}_{\text{bs}}(\cdot)$ , respectively. Combining the definitions (27) and relationship (37), we can estimate the scaling ambiguities as well as the complex gains  $\{\hat{\alpha}_l\}$  with a LS criterion as

$$\begin{aligned} [\mathbf{\Lambda}_1]_{l,l} &= \tilde{\mathbf{a}}_{\text{ms}}^\dagger(\hat{\theta}_l) \hat{\mathbf{b}}_l^{(1)}, [\mathbf{\Lambda}_2]_{l,l} = \tilde{\mathbf{a}}_{\text{bs}}^\dagger(\hat{\phi}_l) \hat{\mathbf{b}}_l^{(2)}, \mathbf{\Lambda}_3 = \mathbf{\Lambda}_1^{-1} \mathbf{\Lambda}_2^{-1}, \\ \hat{\alpha}_l &= [\mathbf{\Lambda}_3]_{l,l}^{-1} \mathbf{g}(\pi_l)^\dagger \hat{\mathbf{b}}_l^{(3)}. \end{aligned} \quad (42)$$

<sup>10</sup>Assuming that the estimation error of factor matrices follows an i.i.d. symmetric Gaussian distribution, this scheme is proved to be a maximum likelihood estimator [22].

---

**Algorithm 1** SCPD Algorithm
 

---

**Require:** observation signal tensor  $\mathbf{Y} \in \mathbb{C}^{Q \times P \times K}$ .

- 1: Estimate the tensor rank  $\hat{L}$  by the MDL criterion (32), (33).
  - 2: Define  $(K_3, L_3)$  subject to  $K_3 + L_3 = K + 1$ . Derive the spatial smoothing  $\mathbf{Y}_S \in \mathbb{C}^{K_3 P \times L_3 Q}$  by (36).
  - 3: Compute singular value decomposition (SVD) as  $\mathbf{Y}_S = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ .
  - 4: Build submatrices of  $\mathbf{U}$  as  $\mathbf{U}_1 = [\mathbf{U}]_{1:(K_3-1)P,:}$ ,  $\mathbf{U}_2 = [\mathbf{U}]_{P+1:K_3 P,:}$ .
  - 5: Compute EVD as  $\mathbf{U}_1^\dagger \mathbf{U}_2 = \mathbf{M} \mathbf{Z} \mathbf{M}^{-1}$ .
  - 6: Estimate the normalized generators as  $\hat{z}_l = [\mathbf{Z}]_{l,l}$ ,  $\hat{z}_l \leftarrow \hat{z}_l / |\hat{z}_l|$ .
  - 7: Reconstruct  $\hat{\mathbf{B}}^{(3)}$  as  $\hat{\mathbf{b}}_l^{(3)} = [\hat{z}_l, \hat{z}_l^2, \dots, \hat{z}_l^K]^T$ .
  - 8: Reconstruct  $\hat{\mathbf{B}}^{(2)}$  as  $\hat{\mathbf{b}}_l^{(2)} = \left( \hat{\mathbf{b}}_l^{(K_3,3)H} \otimes \mathbf{I}_P \right) \mathbf{U} \mathbf{m}_l$ .
  - 9: Compute  $\mathbf{T} = \mathbf{M}^{-T}$ , and reconstruct  $\hat{\mathbf{B}}^{(1)}$  as  $\hat{\mathbf{b}}_l^{(1)} = \left( \hat{\mathbf{b}}_l^{(L_3,3)H} \otimes \mathbf{I}_Q \right) \mathbf{V}^* \mathbf{\Sigma} \mathbf{t}_l$ .
  - 10: Estimate the path time delays  $\{\tau_l\}$ , AoAs  $\{\hat{\theta}_l\}$  and AoDs  $\{\hat{\phi}_l\}$  by (38), (39a) and (39b), respectively.
  - 11: Estimate the scaling ambiguities  $\{\Lambda_n\}$  and path complex gains  $\{\hat{\alpha}_l\}$  by (42).
  - 12: **return** estimated multipath parameters  $\{\hat{\theta}_l, \hat{\phi}_l, \hat{\alpha}_l, \hat{\tau}_l\}$  and channel matrices  $\{\hat{\mathbf{H}}_k\}$ .
- 

We summarize the tensor decomposition-based channel estimation scheme, i.e., SCPD, as Algorithm 1, which only harnesses standard linear algebra and avoids the iterative runs and unstable random initialization. Moreover, the SCPD algorithm is guaranteed to return the correct solution for the noiseless case.

One may choose the pair  $(K_3, L_3)$  as defined in Algorithm 1 such that the dimensions of the matrices  $\mathbf{B}^{(K_3,3)} \odot \mathbf{B}^{(2)}$  and  $\mathbf{B}^{(L_3,3)} \odot \mathbf{B}^{(1)}$  are similar, with the inequalities  $K_3 - 1 \geq \lceil L/P \rceil$ ,  $L_3 \geq \lceil L/Q \rceil$  satisfied. At the cost of computational complexity, one can process the  $\mathbf{Y}$  with multiple pairs of  $(K_3, L_3)$  and calculate the average values of recovered parameters to reduce the variance of estimation error. When the number of training subcarriers  $K$  is of moderate size, one can even try all possible pairs, i.e., let  $K_3$  vary from 2 to  $K$ , to process the received signal tensor, which is still more efficient than an optimization-based method, such as TALS [26].

## V. CHANNEL ESTIMATION WITH DUAL-WIDEBAND EFFECTS

For the case with significant dual-wideband effects, the *spatial-frequency* steering vectors  $\mathbf{a}_{\text{ms},k}(\theta)$  and  $\mathbf{a}_{\text{bs},k}(\phi)$  in (16) are frequency-dependent. Unfortunately, the multi-carrier beam training data of (17) cannot be decoupled into three independent dimensions, and, hence, does not admit the CP decomposition model. In this section, we design a channel estimation scheme across multiple time slots to formulate the CP model.

### A. Beam Training Scheme

Similar as the beam training (21)–(23), the BS switches  $P$  precoding vectors within a single time slot, and the MS combines the received signal with  $Q$  measurement streams. This procedure is duplicated across consecutive  $T$  time slots, yielding the received signal at the  $k$ th subcarrier within the  $t$ th time slot as

$$\begin{aligned} \mathbf{Y}_{t,k} &= \mathbf{W}^T \mathbf{H}_{t,k} \mathbf{F} + \mathbf{N}_{t,k} \\ &= \sum_{l=1}^L \alpha_{t,l} e^{-j2\pi f_k \tau_l} \tilde{\mathbf{a}}_{\text{ms},k}(\theta_l) \tilde{\mathbf{a}}_{\text{bs},k}(\phi_l) + \mathbf{N}_{t,k}, \end{aligned} \quad (43)$$

where  $\tilde{\mathbf{a}}_{\text{ms},k}(\theta) \triangleq \mathbf{W}^T \mathbf{a}_{\text{ms},k}(\theta) \in \mathbb{C}^Q$  and  $\tilde{\mathbf{a}}_{\text{bs},k}(\phi) \triangleq \mathbf{F}^T \mathbf{a}_{\text{bs},k}(\phi) \in \mathbb{C}^P$  are the combined and precoded steering vectors at the  $k$ th subcarrier, respectively;  $\mathbf{N}_{t,k} \in \mathbb{C}^{Q \times P}$  is the combined noise.

Concatenating the received data  $\{\mathbf{Y}_{t,k}\}_{t=1}^T$ , we derive a third-order tensor  $\mathcal{Y}_k \in \mathbb{C}^{Q \times P \times T}$  as

$$\mathcal{Y}_k = \sum_{l=1}^L \tilde{\mathbf{a}}_{\text{ms},k}(\theta_l) \circ \tilde{\mathbf{a}}_{\text{bs},k}(\phi_l) \circ (e^{-j2\pi f_k \tau_l} \mathbf{r}_l) + \mathcal{N}_k, \quad (44)$$

where  $\mathbf{r}_l \triangleq [\alpha_{1,l}, \dots, \alpha_{T,l}]^T \in \mathbb{C}^T$  is a time slot-domain vector containing the time-varying gains of the  $l$ th path. Then, the factor matrices for the three dimensions are denoted by

$$\mathbf{C}_k^{(1)} \triangleq [\tilde{\mathbf{a}}_{\text{ms},k}(\theta_1), \dots, \tilde{\mathbf{a}}_{\text{ms},k}(\theta_L)] \in \mathbb{C}^{Q \times L}, \quad (45a)$$

$$\mathbf{C}_k^{(2)} \triangleq [\tilde{\mathbf{a}}_{\text{bs},k}(\phi_1), \dots, \tilde{\mathbf{a}}_{\text{bs},k}(\phi_L)] \in \mathbb{C}^{P \times L}, \quad (45b)$$

$$\mathbf{C}_k^{(3)} \triangleq [e^{-j2\pi f_k \tau_1} \mathbf{r}_1, \dots, e^{-j2\pi f_k \tau_L} \mathbf{r}_L] \in \mathbb{C}^{T \times L}. \quad (45c)$$

### B. Training Beamforming Design

Typically, we adopt a random beamforming method  $\mathcal{M}_0$  as follows

$$\mathcal{M}_0: [\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N_{\text{bs}}}} e^{j\eta_{m,n}}, [\mathbf{W}]_{m,n} = \frac{1}{\sqrt{N_{\text{ms}}}} e^{j\xi_{m,n}}, \eta_{m,n}, \xi_{m,n} \sim \mathcal{U}(0, 2\pi), \quad (46)$$

which can be implemented by analog phase shifter networks. The benefits of  $\mathcal{M}_0$  are: (i) the elements of  $\mathbf{C}_k^{(1)}, \mathbf{C}_k^{(2)}$  follow i.i.d.  $\mathcal{CN}(0, 1)$  with sufficiently large  $N_{\text{ms}}, N_{\text{bs}}$ , generically leading to  $\text{kr}(\mathbf{C}_k^{(1)}) = \min(Q, L)$ ,  $\text{kr}(\mathbf{C}_k^{(2)}) = \min(P, L)$  in Lemma 2 [22]; (ii) the elements of beamformed noise  $\mathcal{N}_k$  follow i.i.d.  $\mathcal{CN}(0, \sigma_n^2)$ , contributing to the MDL operation (33). However,  $\mathbf{C}_k^{(1)}, \mathbf{C}_k^{(2)}$  lose the shift-invariance feature of  $\mathbf{a}_{\text{ms},k}(\theta), \mathbf{a}_{\text{bs},k}(\phi)$ . Since  $\mathbf{C}_k^{(3)}$  is generically not Vandermonde, Algorithm 1 cannot be directly applied to  $\mathcal{Y}_k$  now.

We can exploit the shift-invariance property of  $\mathbf{a}_{\text{ms},k}(\theta), \mathbf{a}_{\text{bs},k}(\phi)$  by carefully designing the training beamformers. At the MS end, for example, we can apply the following beamforming method  $\mathcal{M}_1$ :

$$\begin{aligned} \mathcal{M}_1: \mathbf{W} &\triangleq \begin{bmatrix} \overline{\mathbf{W}} & \underline{\mathbf{W}} \end{bmatrix} \\ \overline{\mathbf{w}}_q &\triangleq [w_{1,q}, \dots, w_{N_{\text{ms}}-1,q}, 0]^T, \\ \underline{\mathbf{w}}_q &\triangleq [0, w_{1,q}, \dots, w_{N_{\text{ms}}-1,q}]^T, q \in \mathcal{I}(Q/2), \end{aligned} \quad (47)$$

which leverages the shift-invariance property of  $\mathbf{a}_{\text{ms},k}(\theta)$  as

$$\underline{\mathbf{W}}^T \mathbf{a}_{\text{ms},k}(\theta) = \overline{\mathbf{W}}^T \mathbf{a}_{\text{ms},k}(\theta) e^{-j2\pi(1+\frac{f_k}{f_c})\theta}. \quad (48)$$

With this structure, we can derive a new uniqueness condition of CP decomposition (See Section V. C). To realize  $\mathcal{M}_1$  with hybrid beamformers  $\mathbf{W}_A \in \mathbb{C}^{N_{\text{ms}} \times M_{\text{ms}}}$ ,  $\mathbf{W}_D \in \mathbb{C}^{M_{\text{ms}} \times Q}$ , we have

$$\begin{cases} \mathbf{w}_{A,\{N_{\text{ms}}\}}^T \mathbf{w}_{D,q} = 0, q \in \mathcal{I}(Q/2) \\ \mathbf{w}_{A,\{1\}}^T \mathbf{w}_{D,q+Q/2} = 0, q \in \mathcal{I}(Q/2) \\ \mathbf{w}_{A,\{n\}}^T \mathbf{w}_{D,q} = \mathbf{w}_{A,\{n+1\}}^T \mathbf{w}_{D,q+Q/2}, n \in \mathcal{I}(N_{\text{ms}} - 1), q \in \mathcal{I}(Q/2), \end{cases} \quad (49)$$

where  $\mathbf{w}_{A,\{n\}}^T$  denotes the  $n$ th row of  $\mathbf{W}_A$ . Given  $\mathbf{W}_A$  with unit-norm elements, (49) is equivalent to a set of homogeneous linear equations with  $(N_{\text{ms}} + 1)Q/2$  equations and  $M_{\text{ms}}Q$  unknowns. Hence, we can derive a solution of (49) if and only if  $M_{\text{ms}} > (N_{\text{ms}} + 1)/2$ .

Since  $\mathcal{M}_1$  cannot be combined with the spatial smoothing technique, we can employ another beamforming method  $\mathcal{M}_2$  as follows

$$\mathcal{M}_2: \mathbf{W} \triangleq \begin{bmatrix} \mathbf{I}_Q & \mathbf{0}_{Q \times (N_{\text{ms}}-Q)} \end{bmatrix}^T. \quad (50)$$

Clearly,  $\mathcal{M}_2$  extracts the first  $Q$  elements and discards the rest, i.e.,  $\mathbf{W}^T \mathbf{a}_{\text{ms},k}(\theta) = [\mathbf{a}_{\text{ms},k}(\theta)]_{1:Q}$ . It preserves a complete Vandermonde structure with generators  $\{z_{1,k,l} \triangleq e^{-j2\pi(1+f_k/f_c)\theta_l}\}$  at the

expense of losing a portion of training signal. The hybrid beamforming can be realized as [10]

$$\begin{aligned}\mathbf{W}_D &= \frac{1}{\sqrt{M_{\text{ms}}}} \begin{bmatrix} \bar{\mathbf{u}}_1 & \dots & \bar{\mathbf{u}}_Q \end{bmatrix}, \\ \mathbf{W}_A &= \begin{bmatrix} \mathbf{W}_D & \bar{\mathbf{u}}_{M_{\text{ms}}} \mathbf{1}_{1 \times (N_{\text{ms}} - Q)} \end{bmatrix}^H,\end{aligned}\quad (51)$$

where  $Q = M_{\text{ms}} - 1$ ;  $\bar{\mathbf{u}}_n$  is the  $n$ th column of a unitary DFT matrix  $\bar{\mathbf{U}} \in \mathbb{C}^{M_{\text{ms}} \times M_{\text{ms}}}$ . Similarly,  $\mathcal{M}_2$  can be applied at the BS end with  $P = M_{\text{bs}} - 1$ , yielding a Vandermonde  $\mathbf{C}_k^{(2)}$  with generators  $\{z_{2,k,l} \triangleq e^{-j2\pi(1+f_k/f_c)\phi_l}\}$ .

### C. Uniqueness Condition

If the MS employs the beamforming  $\mathcal{M}_1$ , we propose the following uniqueness condition.

*Proposition 1.* Let  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  be a tensor with factor matrices  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$ ,  $n \in \{1, 2, 3\}$ . Let  $\mathbf{A}^{(1)} = \mathbf{W}^T \mathbf{B}$ , where  $\mathbf{B} \in \mathbb{C}^{J_1 \times R}$  is a Vandermonde matrix with distinct generators  $\{z_r\}_{r=1}^R$  and  $\mathbf{W} \in \mathbb{C}^{J_1 \times I_1}$ ,  $J_1 \geq I_1$  is a weighting matrix with a form of (47). If

$$\begin{cases} \text{r}(\mathbf{A}^{(I_1/2,1)} \odot \mathbf{A}^{(2)}) = R, \\ \text{r}(\mathbf{A}^{(3)}) = R, \end{cases}\quad (52)$$

then the rank of  $\mathcal{X}$  is  $R$  and the CP decomposition is unique. In the generic case, condition (52) becomes

$$\min(I_1 I_2 / 2, I_3) \geq R. \quad (53)$$

*Proof:* The proof of Proposition 1 similarly follows that of [24, Proposition III.2]. The only difference is the submatrix definition of singular space of  $\mathbf{X}^{[2]}$ , i.e.,  $\mathbf{X}^{[2]} = \mathbf{U}\Sigma\mathbf{V}^H$ ,  $\mathbf{U}_1 = [\mathbf{U}]_{1:I_1 I_2/2, :}$ ,  $\mathbf{U}_2 = [\mathbf{U}]_{I_1 I_2/2+1:I_1 I_2, :}$ . ■

If both the MS and BS employ method  $\mathcal{M}_2$ , factorizing  $\mathcal{Y}_k$  is equivalent to solving a two-dimensional harmonic retrieval problem with multiple snapshots [36], [37]. Therefore, by leveraging the concept of forward-backward smoothing [28], [29], [36], we obtain the following uniqueness condition.

*Theorem 2.* Let  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$  be a tensor with factor matrices  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$ ,  $n \in \{1, 2, 3\}$ . Let  $\mathbf{A}^{(1)}$  and  $\mathbf{A}^{(2)}$  be Vandermonde matrices with generators  $\{z_{1,r} \triangleq e^{j\omega_{1,r}}\}_{r=1}^R$  and  $\{z_{2,r} \triangleq e^{j\omega_{2,r}}\}_{r=1}^R$ , respectively. Define

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{A}^{(K_1,1)} \odot \mathbf{A}^{(K_2,2)} \odot \mathbf{A}^{(3)} \\ \mathbf{A}^{(L_1,1)} \odot \mathbf{A}^{(L_2,2)} \odot \mathbf{A}^{(3)*} \mathbf{D}(\boldsymbol{\omega}) \end{bmatrix}, \quad (54)$$

where  $[\boldsymbol{\omega}]_r \triangleq e^{-j((I_1-1)\omega_{1,r}+(I_2-1)\omega_{2,r})}$ ,  $r \in \mathcal{I}(R)$ . If

$$\begin{cases} z_{1,r_1} \neq z_{1,r_2}, \forall r_1 \neq r_2, \\ \text{r}(\mathbf{A}^{(K_1-1,1)} \odot \mathbf{A}^{(K_2,2)}) = R, \\ \text{r}(\mathbf{B}) = R, \end{cases} \quad (55)$$

with  $K_n + L_n = I_n + 1$ ,  $n \in \{1, 2\}$ , then the rank of  $\boldsymbol{\mathcal{X}}$  is  $R$  and the CP decomposition is unique. In the generic case, condition (55) becomes

$$\min_{\substack{K_n + L_n = I_n + 1 \\ n \in \{1, 2\}}} ((K_1 - 1)K_2, 2L_1L_2I_3) \geq R. \quad (56)$$

Theorem 2 does not prevent  $\text{kr}(\mathbf{A}^{(1)}) = 1$  or  $\text{kr}(\mathbf{A}^{(2)}) = 1$  and/or  $\text{kr}(\mathbf{A}^{(3)}) = 1$ . In our case, the uniqueness holds even for the circumstances that the path complex gains are time-invariant.

#### D. Proposed Algorithm

If the MS employs beamforming  $\mathcal{M}_1$  or  $\mathcal{M}_2$ , one can similarly follow [1] or Algorithm 1 respectively to solve the channel estimation problem. Here we consider the case that both the MS and BS employ  $\mathcal{M}_2$ . We define  $(K_1, L_1), (K_2, L_2)$  subject to  $K_1 + L_1 = Q + 1, K_2 + L_2 = P + 1$ . Then, we obtain the matricization  $\mathbf{Y}_k^{[2]}$  from  $\mathbf{Y}$  by (4), and define  $\tilde{\mathbf{Y}}_k^{[2]}$  as follows

$$\begin{aligned} \mathbf{Y}_k^{[2]} &\triangleq \left( \mathbf{C}_k^{(1)} \odot \mathbf{C}_k^{(2)} \right) \mathbf{C}_k^{(3)T} + \mathbf{N}_k^{[2]}, \\ \tilde{\mathbf{Y}}_k^{[2]} &\triangleq \boldsymbol{\Pi} \mathbf{Y}_k^{[2]*} = \left( \mathbf{C}_k^{(1)} \odot \mathbf{C}_k^{(2)} \right) \mathbf{D}(\mathbf{z}_k) \mathbf{C}_k^{(3)H} + \tilde{\mathbf{N}}_k^{[2]}, \end{aligned} \quad (57)$$

where  $[\mathbf{z}_k]_l \triangleq (z_{1,k,l}^*)^{Q-1} (z_{2,k,l}^*)^{P-1}$ ,  $l \in \mathcal{I}(L)$ ;  $\boldsymbol{\Pi} \in \mathbb{C}^{QP \times QP}$  is a permutation matrix with ones on its antidiagonal;  $\mathbf{N}_k^{[2]}, \tilde{\mathbf{N}}_k^{[2]} \in \mathbb{C}^{QP \times T}$  are the corresponding noise matrices. We perform spatial smoothing (9) on  $\mathbf{Y}_k^{[2]}$  and  $\tilde{\mathbf{Y}}_k^{[2]}$  respectively, yielding

$$\begin{aligned} \mathbf{Y}_{S,k} &\triangleq \mathcal{S}_2 \left( \mathbf{Y}_k^{[2]} \right) = \left( \mathbf{C}_k^{(K_1,1)} \odot \mathbf{C}_k^{(K_2,2)} \right) \left( \mathbf{C}_k^{(L_1,1)} \odot \mathbf{C}_k^{(L_2,2)} \odot \mathbf{C}_k^{(3)} \right)^T + \mathbf{N}_{S,k}, \\ \tilde{\mathbf{Y}}_{S,k} &\triangleq \mathcal{S}_2 \left( \tilde{\mathbf{Y}}_k^{[2]} \right) = \left( \mathbf{C}_k^{(K_1,1)} \odot \mathbf{C}_k^{(K_2,2)} \right) \left( \mathbf{C}_k^{(L_1,1)} \odot \mathbf{C}_k^{(L_2,2)} \odot \tilde{\mathbf{C}}_k^{(3)} \right)^T + \tilde{\mathbf{N}}_{S,k}, \end{aligned} \quad (58)$$

where  $\tilde{\mathbf{C}}_k^{(3)} \triangleq \mathbf{C}^{(3)*} \mathbf{D}(\mathbf{z}_k)$ ;  $\mathbf{N}_{S,k}, \tilde{\mathbf{N}}_{S,k} \in \mathbb{C}^{K_1 K_2 \times L_1 L_2 T}$  are the corresponding noise matrices. By concatenating the forward and backward smoothing results in (58), we obtain

$$\mathbf{G}_{S,k} \triangleq \begin{bmatrix} \mathbf{Y}_{S,k} & \tilde{\mathbf{Y}}_{S,k} \end{bmatrix} = \left( \mathbf{C}_k^{(K_1,1)} \odot \mathbf{C}_k^{(K_2,2)} \right) \begin{bmatrix} \mathbf{C}_k^{(L_1,1)} \odot \mathbf{C}_k^{(L_2,2)} \odot \mathbf{C}_k^{(3)} \\ \mathbf{C}_k^{(L_1,1)} \odot \mathbf{C}_k^{(L_2,2)} \odot \tilde{\mathbf{C}}_k^{(3)} \end{bmatrix}^T + \begin{bmatrix} \mathbf{N}_{S,k} & \tilde{\mathbf{N}}_{S,k} \end{bmatrix}. \quad (59)$$

Then, we can follow [28], [29] and employ an improved multidimensional folding (IMDF)-like approach to estimate  $\{\mathbf{C}_k^{(n)}\}_{n \in \{1,2,3\}}$ . After the factorization of (44), we turn to the parameter recovery. If the dedicated hybrid precoding and/or combining designs are applied, then the estimation of AoAs  $\{\hat{\theta}_l\}$  and/or AoDs  $\{\hat{\phi}_l\}$  can be directly derived from the generators of  $\mathbf{C}_k^{(1)}$  and  $\mathbf{C}_k^{(2)}$  respectively as

$$\hat{\theta}_l = -\frac{f_k}{2\pi(f_c + f_k)} \angle \hat{z}_{1,k,l}, \quad (60a)$$

$$\hat{\phi}_l = -\frac{f_k}{2\pi(f_c + f_k)} \angle \hat{z}_{2,k,l}. \quad (60b)$$

Following a similar operation as (42), the scaling ambiguity of  $\mathbf{C}_k^{(n)}$ , i.e.,  $\mathbf{\Lambda}_{n,k} \in \mathbb{C}^{L \times L}$ , as well as the equivalent path gains  $\hat{\mathbf{r}}_l e^{-j2\pi f_k \hat{\tau}_l}$  can be respectively computed as

$$\begin{aligned} [\mathbf{\Lambda}_{1,k}]_{l,l} &= \tilde{\mathbf{a}}_{\text{ms},k}^\dagger(\hat{\theta}_l) \hat{\mathbf{c}}_{k,l}^{(1)}, [\mathbf{\Lambda}_{2,k}]_{l,l} = \tilde{\mathbf{a}}_{\text{bs},k}^\dagger(\hat{\phi}_l) \hat{\mathbf{c}}_{k,l}^{(2)}, \mathbf{\Lambda}_{3,k} = \mathbf{\Lambda}_{1,k}^{-1} \mathbf{\Lambda}_{2,k}^{-1}, \\ e^{-j2\pi f_k \hat{\tau}_l} \hat{\mathbf{r}}_l &= [\mathbf{\Lambda}_{3,k}]_{l,l}^{-1} \hat{\mathbf{c}}_{k,l}^{(3)}. \end{aligned} \quad (61)$$

If the MS or BS employs the special hybrid beamforming scheme, then its corresponding scaling ambiguity is exactly an identity matrix. Note that carrying out the beam training (44) is sufficient for the estimation of channel matrix  $\mathbf{H}_k$ . However, if the exact  $\{\hat{\tau}_l\}$  and  $\{\hat{\alpha}_{t,l}\}$  need to be estimated respectively, it is necessary to perform a parallel beam training on at least two different subcarriers.

With the estimated mode-3 factor matrix  $\hat{\mathbf{C}}_k^{(3)}$  and its scaling ambiguity  $\mathbf{\Lambda}_{3,k}$  at arbitrary two different subcarriers, i.e.,  $f_{k_1}$  and  $f_{k_2}$ , we can estimate the path time delays  $\{\hat{\tau}_l\}$  and the time-varying path gains  $\{\hat{\alpha}_{t,l}\}$  of  $T$  time slots as

$$\hat{\tau}_l = -\frac{1}{2\pi(f_{k_2} - f_{k_1})} \angle \left[ \left( [\mathbf{\Lambda}_{3,k_1}]_{l,l}^{-1} \hat{\mathbf{c}}_{k_1,l}^{(3)} \right)^\dagger \left( [\mathbf{\Lambda}_{3,k_2}]_{l,l}^{-1} \hat{\mathbf{c}}_{k_2,l}^{(3)} \right) \right], \quad (62a)$$

$$\hat{\alpha}_{t,l} = \left[ e^{j2\pi f_{k_i} \hat{\tau}_l} [\mathbf{\Lambda}_{3,k_i}]_{l,l}^{-1} \hat{\mathbf{c}}_{k_i,l}^{(3)} \right]_t, \quad t \in \mathcal{I}(T), i \in \{1, 2\}. \quad (62b)$$

We can further combine the tensor decomposition results along multiple subcarriers to derive an averaged estimation of  $\{\hat{\tau}_l\}$  and  $\{\hat{\alpha}_{t,l}\}$ . In this estimation scheme, the indices of training subcarriers can be flexibly selected, which are not constrained as  $\mathcal{I}(K)$ .

For the case that the MS employs  $\mathcal{M}_1$ , we call the channel estimation scheme *DWE-SCPD-0*; for the case that only the MS employs  $\mathcal{M}_2$ , we call the corresponding scheme *DWE-SCPD-1*; for the case that both the MS and BS employ  $\mathcal{M}_2$ , we develop a variant algorithm version called

**Algorithm 2** DWE-SCPD-2 Algorithm**Require:** observation signal tensors  $\{\mathbf{Y}_k\} \in \mathbb{C}^{Q \times P \times T}$ .

- 1: Define pairs  $(K_1, L_1)$  and  $(K_2, L_2)$  subject to  $K_1 + L_1 = Q + 1$  and  $K_2 + L_2 = P + 1$ .
- 2: **for**  $k = 1$  to  $K$  **do**
- 3: Estimate the tensor rank  $\hat{L}$  by the MDL criterion (32), (33).
- 4: Derive the forward-backward spatial smoothing  $\mathbf{G}_{S,k} \in \mathbb{C}^{K_1 K_2 \times 2L_1 L_2 T}$  by (57)–(59).
- 5: Compute SVD as  $\mathbf{G}_{S,k} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H$ .
- 6: Build submatrices of  $\mathbf{U}_k$  as  $\mathbf{U}_{k,1} = [\mathbf{U}_k]_{1:(K_1-1)K_2,:}$ ,  $\mathbf{U}_{k,2} = [\mathbf{U}_k]_{K_2+1:K_1 K_2,:}$ .
- 7: Perform EVD  $\begin{bmatrix} \mathbf{U}_{k,1}^H \\ \mathbf{U}_{k,2}^H \end{bmatrix} [\mathbf{U}_{k,1} \ \mathbf{U}_{k,2}] = \begin{bmatrix} \mathbf{V}_{k,11} & \mathbf{V}_{k,12} \\ \mathbf{V}_{k,21} & \mathbf{V}_{k,22} \end{bmatrix} \mathbf{\Lambda}_U \begin{bmatrix} \mathbf{V}_{k,11} & \mathbf{V}_{k,12} \\ \mathbf{V}_{k,21} & \mathbf{V}_{k,22} \end{bmatrix}^H$ , where  $\mathbf{\Lambda}_U \in \mathbb{C}^{2\hat{L} \times 2\hat{L}}$  contains the eigenvalues [38], [39].
- 8: Perform EVD  $-\mathbf{V}_{k,12} \mathbf{V}_{k,22}^{-1} = \mathbf{M}_k \mathbf{Z}_k \mathbf{M}_k^{-1}$  with submatrices  $\mathbf{V}_{k,12}, \mathbf{V}_{k,22} \in \mathbb{C}^{\hat{L} \times \hat{L}}$ .
- 9: Estimate the normalized generators of  $\mathbf{C}_k^{(1)}$  as  $\hat{z}_{1,k,l} = [\mathbf{Z}_k]_{l,l}$ ,  $\hat{z}_{1,k,l} \leftarrow \hat{z}_{1,k,l} / |\hat{z}_{1,k,l}|$ .
- 10: Reconstruct  $\mathbf{C}_k^{(1)}$  as  $\hat{\mathbf{c}}_{k,l}^{(1)} = [1, \hat{z}_{1,k,l}, \dots, \hat{z}_{1,k,l}^{Q-1}]^T$ .
- 11: Reconstruct  $\hat{\mathbf{C}}_k^{(K_2,2)}$  as  $\hat{\mathbf{c}}_{k,l}^{(K_2,2)} = \left( \hat{\mathbf{c}}_{k,l}^{(K_1,1)H} \otimes \mathbf{I}_{K_2} \right) \mathbf{U}_k \mathbf{m}_{k,l}$ .
- 12: Estimate the normalized generators of  $\mathbf{C}_k^{(2)}$  as  $\hat{z}_{2,k,l} = [\mathbf{c}_{k,l}^{(L_2,2)}]_{1:L_2-1}^\dagger [\mathbf{c}_{k,l}^{(L_2,2)}]_{2:L_2}$ ,  $\hat{z}_{2,k,l} \leftarrow \hat{z}_{2,k,l} / |\hat{z}_{2,k,l}|$ .
- 13: Reconstruct  $\mathbf{C}_k^{(2)}$  as  $\hat{\mathbf{c}}_{k,l}^{(2)} = [1, \hat{z}_{2,k,l}, \dots, \hat{z}_{2,k,l}^{P-1}]^T$ .
- 14: Compute  $\mathbf{T}_k = \mathbf{M}_k^{-T}$ , and reconstruct  $\hat{\mathbf{C}}_k^{(3)}$  as  $\hat{\mathbf{c}}_{k,l}^{(3)} = \left( \hat{\mathbf{c}}_{k,l}^{(L_1,1)H} \otimes \hat{\mathbf{c}}_{k,l}^{(L_2,2)H} \otimes \mathbf{I}_T \right) \mathbf{V}_k^* \mathbf{\Sigma}_k \mathbf{t}_{k,l}$ .
- 15: Estimate the AoAs  $\{\hat{\theta}_l\}$  and AoDs  $\{\hat{\phi}_l\}$  by (60a) and (60b), respectively.
- 16: Estimate the scaling ambiguities  $\{\mathbf{\Lambda}_{n,k}\}$  by (61).
- 17: **end for**
- 18: Estimate the path time delays  $\{\hat{\tau}_l\}$  and complex gains  $\{\hat{\alpha}_{t,l}\}$  by (62).
- 19: **return** estimated multipath parameters  $\{\hat{\theta}_l, \hat{\phi}_l, \hat{\alpha}_{t,l}, \hat{\tau}_l\}$  and channel matrices  $\{\hat{\mathbf{H}}_{t,k}\}$ .

DWE-SCPD-2 following Theorem 2, which is summarized as Algorithm 2. The suffix “-0/-1/-2” corresponds to the number of Vandermonde factor matrices of  $\mathbf{Y}_k$  in each scheme.

## VI. NUMERICAL RESULTS

### A. Computational Complexity

The tensor rank determination (32), (33) is common for all the channel estimation algorithms, which has a computational complexity  $\mathcal{O}(QP K^2)$ . The complexity of subsequent steps of



TABLE I  
COMPUTATIONAL COMPLEXITY OF ALGORITHM 1

Main Steps	Computational Complexity
3: Compute SVD of $\mathbf{Y}_S$	$\mathcal{O}(K_3 L_3 Q P L)$
5: Compute EVD of $\mathbf{U}_1^\dagger \mathbf{U}_2$	$\mathcal{O}(K_3 P L^2)$
7: Reconstruct $\hat{\mathbf{B}}^{(3)}$	$\mathcal{O}(K L)$
8: Reconstruct $\hat{\mathbf{B}}^{(2)}$	$\mathcal{O}(K_3 P L^2)$
9: Reconstruct $\hat{\mathbf{B}}^{(1)}$	$\mathcal{O}(L_3 Q L^2)$
10: Estimate $\{\hat{\tau}_l, \hat{\theta}_l, \hat{\phi}_l\}$	$\mathcal{O}((Q^3 + P^3)L)$
11: Estimate $\{\alpha_l\}$	$\mathcal{O}((Q + P + K)L)$
<b>Total</b>	$\mathcal{O}(K_3 L_3 Q P L)$

Algorithm 1 is listed in Table I, where step 3 has a dominating complexity  $\mathcal{O}(K_3 L_3 Q P L)$ .<sup>11</sup> The total complexity of Algorithm 2 is dominated by  $\mathcal{O}(K_1 K_2 L_1 L_2 T L K)$ , which approximates to  $\mathcal{O}(Q P T L K)$ . As a comparison, the complexity of TALS algorithm [26] for the case with dual-wideband effects is  $\mathcal{O}(N_{\text{iter}} Q P T L K)$  with a commonly large number of iterations  $N_{\text{iter}}$ . The complexity of simultaneous weighted-OMP (SW-OMP) method [6] is of order  $\mathcal{O}(Q P T K (N_1 + N_2))$  with commonly large numbers of AoA/AoD grids  $N_1/N_2$ . Therefore, our proposed algebraic algorithms can be employed at mobile terminals for fast channel estimation.

### B. Simulation Results

In this section, we present simulation results to examine the estimation performance of the proposed schemes. Typical values are set as follows: The antenna array sizes for data transmission are set as  $N_t = 64$ ,  $N_r = 32$ ; the number of RF chains are set as  $M_{\text{ms}} \geq Q$ ,  $M_{\text{bs}} \geq P$ ; the total number of subcarriers is set as  $K_0 = 128$ ; the number of channel paths is  $L = 4$ ;<sup>12</sup> the delay spreads follow i.i.d.  $\mathcal{U}(0, 10^{-7}$  sec).

<sup>11</sup>With different values of  $(K_3, L_3)$ , the complexity varies from  $\mathcal{O}(Q P K L)$  to  $\mathcal{O}(Q P K^2 L/4)$ .

<sup>12</sup>In order to better evaluate the channel parameter estimation performance, the tensor rank  $L$  is assumed to be known or perfectly estimated *a priori* in this simulation.

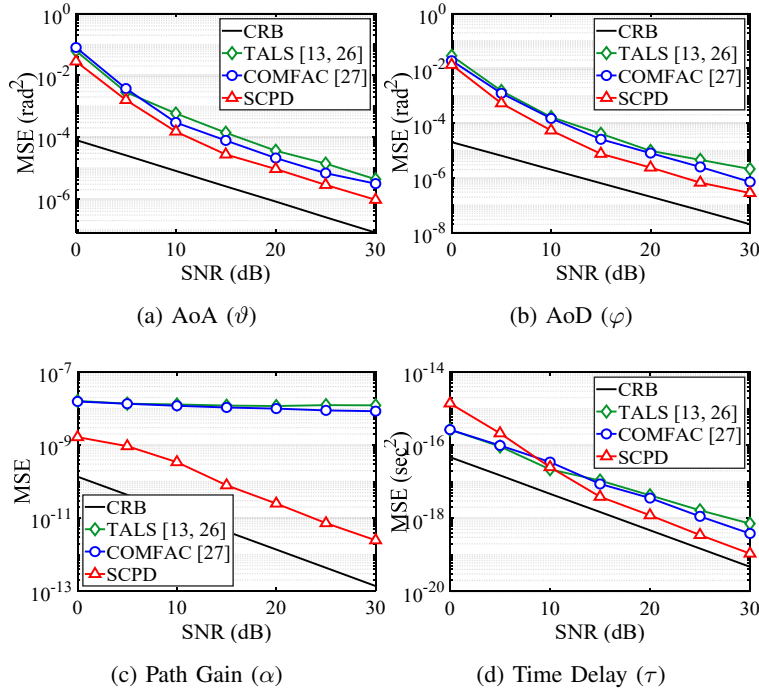


Fig. 3. MSE performance of multipath parameters vs. system SNR,  $Q = P = K = 6$ ,  $f_s = 0.32$  GHz.

1) *Performance Without Dual-Wideband Effects:* We first consider the case with negligible dual-wideband effects, where the carrier frequency is set as  $f_c = 60$  GHz.

We evaluate the estimation accuracy of the multipath parameters  $\{\vartheta_l, \varphi_l, \alpha_l, \tau_l\}$  measured by the mean square error (MSE). To provide a benchmark, we derive the CRB of each parameter as a lower bound of unbiased estimators [40]. Fig. 3 plots the MSE and CRB curves versus the receiving signal-to-noise ratio (SNR)  $\|\mathcal{Y} - \mathcal{N}\|_F^2 / \|\mathcal{N}\|_F^2$ , where the proposed SCPD is compared with TALS [26] and COMFAC [27]. It indicates that the performance of SCPD is exponentially improved against the increasing SNR. Specifically, the MSE of time delay is relatively close to its CRB, while the gap between the MSE and CRB of path gain is relatively wider due to the error accumulation effect. Moreover, the proposed SCPD consistently outperforms the counterparts, especially in terms of the estimation accuracy of path gain.

Then, we focus on the overall estimation performance of channel matrices measured by the normalized MSE (NMSE)  $\sum_{k=1}^K \|\hat{\mathbf{H}}_k - \mathbf{H}_k\|_F^2 / \sum_{k=1}^K \|\mathbf{H}_k\|_F^2$ . Fig. 4 depicts the NMSE curves versus the system SNR obtained by the SCPD, TALS and COMFAC methods with  $Q = P = K = 6$ . Moreover, the SW-OMP [6] and the hierarchical codebook method [7] with

256 × 256 grids are also simulated. The tensor-OMP [17] is a tensorial variant of SW-OMP, while the tensor-minimum MSE [17] does not return accurate estimation results under our simulation configuration. It indicates that the SCPD achieves improved performance against the increasing SNR, which systematically outperforms the counterparts. Moreover, as  $f_s$  increases from 0.32 GHz to 1.28 GHz, the tensor-based strategies advance their performance by nearly one order of magnitude. This can be explained as follows: the increasing  $f_s$  or decreasing  $K_0$  enlarges the subcarrier interval as well as the statistical phase distances of generators  $\{z_l = e^{-j2\pi\frac{f_s}{K_0}\tau_l}\}$ . This phenomenon enhances the resolution across the time delay domain, which we call a *phase rarefaction effect*. We also infer that as the distribution of  $\{2\pi f_s \tau_l / K_0\}$  broadens to  $\mathcal{U}(0, 2\pi)$ , the parameter-based schemes achieve optimized performance, which has been validated by simulations not presented here due to the space constraints.

Fig. 5 plots the NMSE curves versus the number of training subcarriers  $K$  with  $Q = P = 6$  and SNR = 20 dB. It illustrates that as  $K$  increases, the tensor-based methods achieve better performance, while the CS-based and codebook-based methods improve little. With higher sampling rate  $f_s$ , the SCPD scheme returns lower NMSE, which corroborates the inference from Fig. 4. It also shows that with large values of  $f_s$ , increasing  $K$  leads to limited advancement to the performance of TALS and COMFAC.

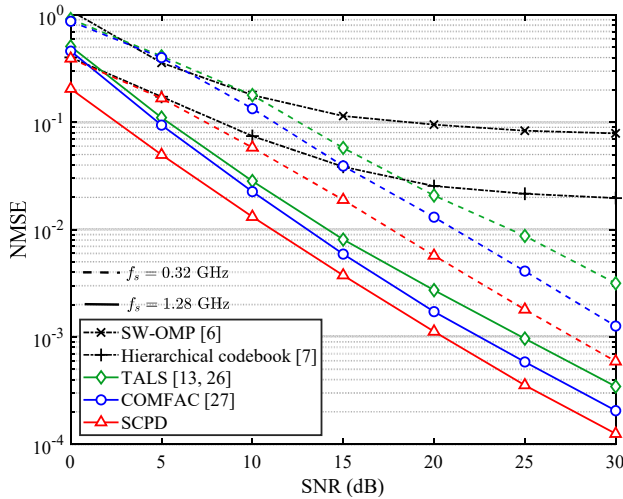


Fig. 4. NMSE performance of channel matrices vs. system SNR,  $Q = P = K = 6$ .

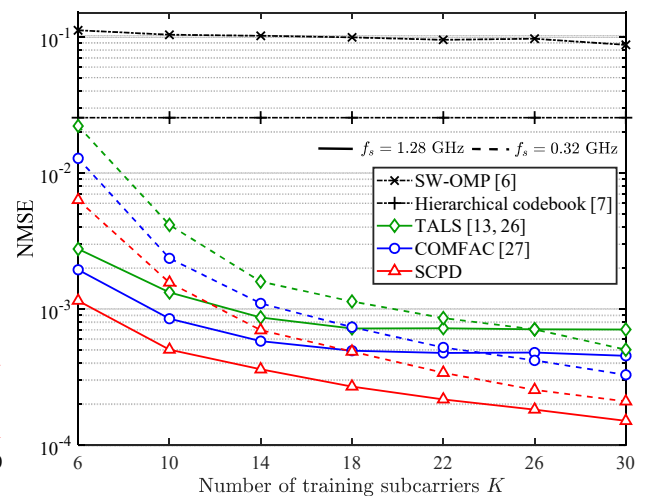


Fig. 5. NMSE performance of channel matrices vs. number of training subcarriers,  $Q = P = 6$ , SNR = 20 dB.

Fig. 6 and Fig. 7 plot the NMSE curves versus the number of precoding subframes and

measurement RF chains, respectively, where  $K = 6$  and  $\text{SNR} = 20$  dB. It demonstrates that increasing training measurements, i.e.,  $Q$  and  $P$ , contribute to the improvement of estimation performance, which corroborates to the analysis of the uniqueness condition.

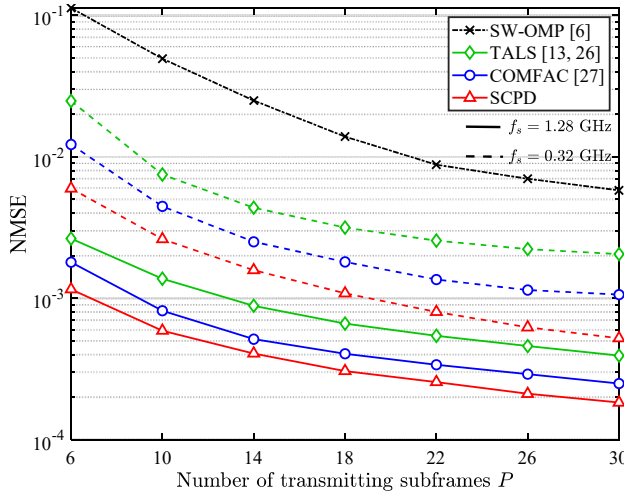


Fig. 6. NMSE performance of channel matrices vs. number of transmitting subframes,  $Q = K = 6$ ,  $\text{SNR} = 20$  dB.

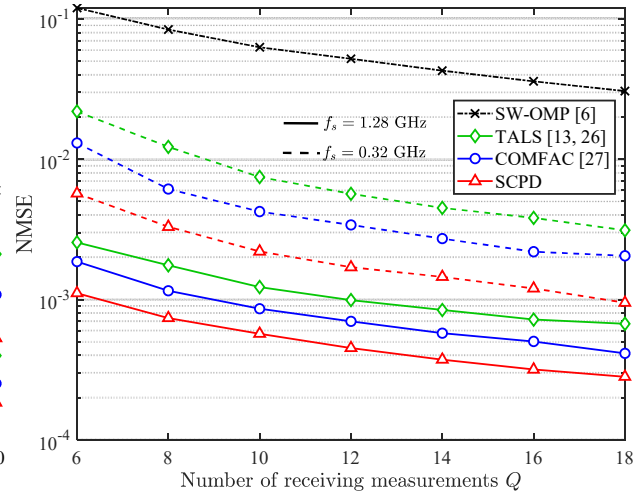


Fig. 7. NMSE performance of channel matrices vs. number of receiving measurements,  $P = K = 6$ ,  $\text{SNR} = 20$  dB.

2) *Performance With Dual-Wideband Effects:* We now turn to the case with significant dual-wideband effects, where the NMSE metric of channel matrices for the DWE-SCPD scheme is  $\min_k \sum_{t=1}^T \|\widehat{\mathbf{H}}_{t,k} - \mathbf{H}_{t,k}\|_F^2 / \sum_{t=1}^T \|\mathbf{H}_{t,k}\|_F^2$ .

Fig. 8 depicts the NMSE curves of DWE-SCPD schemes versus the system SNR with  $Q = P = T = K = 6$ ,  $f_c = 4$  GHz, which is compared with the TALS, COMFAC and SCPD. It shows that as the SNR tends to larger levels, the performance of TALS, COMFAC and SCPD hits a bottleneck, which comes from the interference of dual-wideband effects. With identical SNR level, DWE-SCPD-1 outperforms DWE-SCPD-2. This gap comes from the different beamforming method employed by the BS, where  $\mathcal{M}_2$  discards a portion of the useful signal data and becomes more noise-sensitive than  $\mathcal{M}_0$ . DWE-SCPD-0 suffers from noise interference, especially in the low-SNR region. This can be attributed to the finely designed structure of  $\mathcal{M}_1$  in (47) that is extremely sensitive to disturbance, i.e., noise with even small power may cause serious deviation of (48), which results in deteriorating estimation performance of the multipath parameters, as well as, channel matrices. Furthermore, in the low-SNR region, the SCPD schemes achieve better performance than DWE-SCPD; this is due to the fact that the

factor matrix  $\mathbf{B}^{(3)}$  in (27c) contributes more structure information than  $\mathbf{C}_k^{(3)}$  in (45c) to the CP decomposition procedure.

Fig. 9 plots the NMSE curves of DWE-SCPD versus the number of training time slots  $T$  with  $f_c = 4$  GHz and SNR = 20 dB. It illustrates that as  $T$  increases, the DWE-SCPD schemes obtain greater performance, where DWE-SCPD-1 outperforms DWE-SCPD-2, corroborating the noise-sensitive property of specially designed beamformers. Moreover, the NMSEs of DWE-SCPD schemes versus the number of measurements  $\{Q, P\}$  similarly follow the tendency of channel estimation performance curves in Fig. 6 and Fig. 7. These results are omitted due to space limitation.

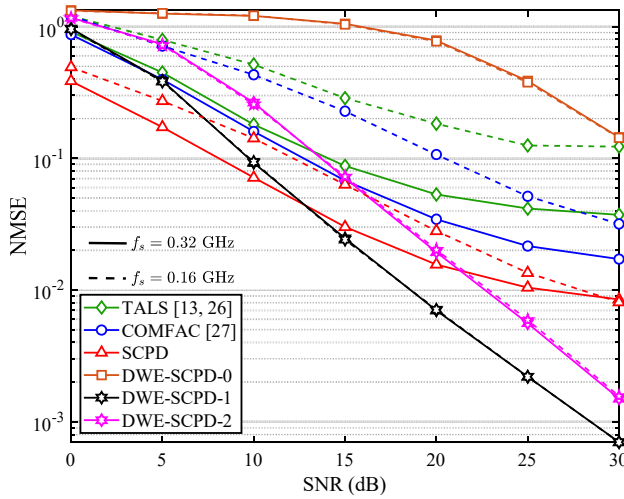


Fig. 8. NMSE performance of channel matrices vs. system SNR,  $Q = P = T = K = 6$ ,  $f_c = 4$  GHz.

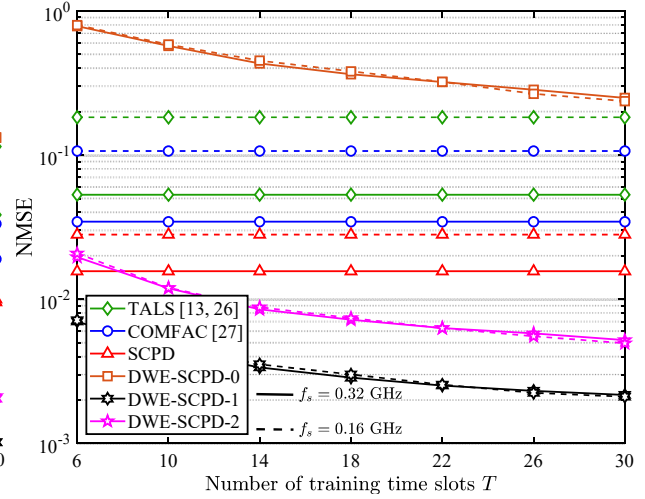


Fig. 9. NMSE performance of channel matrices vs. number of training time slots,  $Q = P = K = 6$ ,  $f_c = 4$  GHz, SNR = 20 dB.

Fig. 10 depicts the NMSE curves versus the carrier frequency  $f_c$  with  $Q = P = T = K = 6$  and SNR = 30 dB. It shows that with the same carrier frequency  $f_c$ , as the bandwidth  $f_s$  increases, SCPD, TALS and COMFAC yield improved performance, which aligns with the observation in Fig. 4. Meanwhile, with the identical  $f_s$ , as  $f_c$  decreases, their NMSE performance gradually deteriorates. This phenomenon follows from the fact that decreasing  $f_c$  aggravates the dual-wideband effects  $\{e^{-j2\pi f_k \theta_l / f_c}, e^{-j2\pi f_k \phi_l / f_c}\}$  in (16). We also observe that with extremely low levels of  $f_c$ , the negative impact from dual-wideband effects on the performance of SCPD etc. even suppresses the positive impact from the phase rarefaction effect. In contrast, the DWE-SCPD

schemes work robustly across different carrier frequencies.

Fig. 11 plots the NMSE curves versus the transmission bandwidth  $f_s$  with SNR = 30 dB. It shows that as  $f_s$  rises from small values to medium values, the SCPD, TALS and COMFAC schemes achieve improved performance, owing to the phase rarefaction effect observed in Fig. 4. As  $f_s$  increases to extremely large values, their performance gradually deteriorates due to the aggravation of dual-wideband effects. More specifically, it shows that with lower carrier frequency, the inflection points of the NMSE curve tend to smaller values of transmission bandwidth. In contrast, the performance of DWE-SCPD schemes is not sensitive to the varying bandwidth.

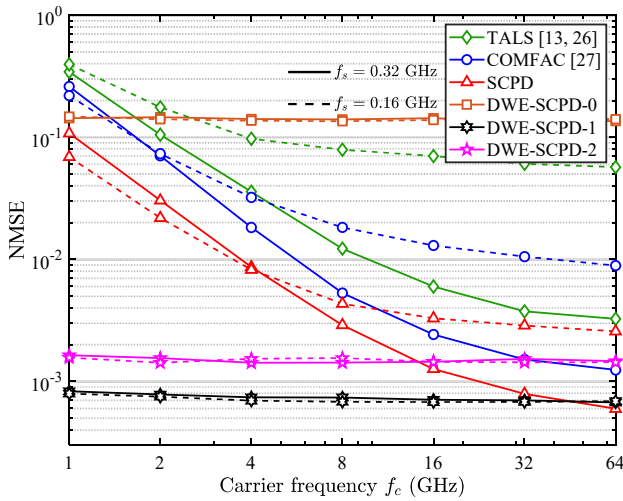


Fig. 10. NMSE performance of channel matrices vs. carrier frequency,  $Q = P = T = K = 6$ , SNR = 30 dB.

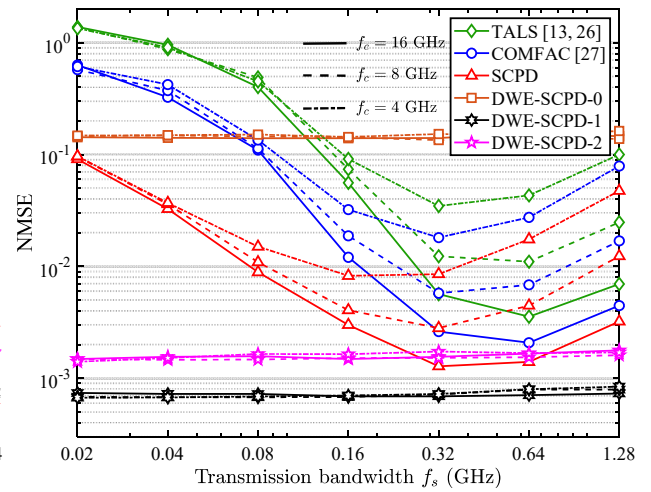


Fig. 11. NMSE performance of channel matrices vs. transmission bandwidth,  $Q = P = T = K = 6$ , SNR = 30 dB.

## VII. CONCLUSIONS

In this paper, we have researched the critical dual-wideband effects in mmWave massive MIMO scenarios, and derive a new spatial-frequency channel model. We developed a tensor decomposition-based scheme to estimate the channel parameters for the MIMO-OFDM systems. We formulated the training signal as a third-order tensor fitting the CP model with factor matrices containing channel parameters. We leveraged both the special Vandermonde constraint and spatial smoothing method to develop a structured CP decomposition-based channel estimation scheme. Two dedicated algorithm versions with particular tensor processing procedures were proposed

for the cases with negligible and severe dual-wideband effects, respectively. The uniqueness of structured CP decomposition is guaranteed with relaxed system parameter constraints. Numerical results showed that the proposed channel estimation strategy outperforms the traditional scheme in terms of estimation accuracy and computational complexity. The dual-wideband effects-oriented scheme achieves robust performance against the carrier frequency and bandwidth. On the other hand, our proposed scheme against dual-wideband effects has some drawbacks. For instance, the specially designed beamformers are proved to worsen the performance. Meanwhile, they do not benefit much from the phase rarefaction effect due to its frequency-independent property. These issues are the directions that we are going to pursue in our future work.

## REFERENCES

- [1] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Structured tensor decomposition-based channel estimation for wideband millimeter wave MIMO," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Dec. 2019.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *P. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [4] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sept. 2014.
- [5] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [6] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.
- [7] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [8] J. Mo, P. Schniter, N. N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2014, pp. 957–961.
- [9] H. Ghach, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.
- [10] A. Liao, Z. Gao, Y. Wu, H. Wang, and M. S. Alouini, "2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding," *IEEE Access*, vol. 5, pp. 24747–24757, Nov. 2017.
- [11] J. Yang, C. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmWave systems via cospase image reconstruction technique," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4767–4782, Oct. 2018.
- [12] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.
- [13] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

- [14] C. Qian, X. Fu, and N. D. Sidiropoulos, "Algebraic channel estimation algorithms for FDD massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 961–973, Sept. 2019.
- [15] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor-based channel estimation for dual-polarized massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6390–6403, Dec. 2018.
- [16] L. N. Ribeiro, A. L. F. de Almeida, N. J. Myers, and R. W. Heath, Jr., "Tensor-based estimation of mmWave MIMO channels with carrier frequency offset," in *Proc. Inter. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 4155–4159.
- [17] D. C. Araújo, A. L. F. de Almeida, J. P. C. L. da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, Apr. 2019.
- [18] J. P. Gonzalez-Coma, W. Utschick, and L. Castedo, "Hybrid LISA for wideband multiuser millimeter-wave communication systems under beam squint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1277–1288, Feb. 2019.
- [19] B. Wang, *et al.*, "Spatial-wideband effect in massive MIMO with application in mmWave systems," *IEEE Commun. Lett.*, vol. 56, no. 12, pp. 134–141, Dec. 2018.
- [20] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial-wideband effect in massive MIMO systems," in *Proc. Asia-Pacific Conf. Commun.*, Dec. 2017, pp. 1–6.
- [21] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
- [22] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "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.
- [23] S. Park, A. Ali, N. González-Prelcic, and R. W. Heath, Jr., "Spatial channel covariance estimation for hybrid architectures based on tensor decompositions," *IEEE Trans. Wireless Commun. (Early Access)*, doi: 10.1109/TWC.2019.2951135.
- [24] M. Sørensen and L. De Lathauwer, "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.
- [25] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-on-arrival of coherent signals," *IEEE Trans. Acoust. Syst. Signal Process.*, vol. 33, no. 8, pp. 806–811, Aug. 1985.
- [26] C. For and R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
- [27] R. Bro, N. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. Int. Workshop Independ. Compon. Anal. Blind Signal Separation*, Jan. 1999, pp. 289–294.
- [28] X. Liu and N. D. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2366–2368, Sept. 2002.
- [29] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4543–4556, Dec. 2006.
- [30] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.
- [31] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.
- [32] K. Liu, J. P. C. L. da Costa, H. C. So, L. Huang, and J. Ye, "Detection of number of components in CANDECOMP/PARAFAC models via minimum description length," *Digital Signal Process.*, vol. 51, pp. 110–123, Jan. 2016.
- [33] T. Yokota, N. Lee, and A. Cichocki, "Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria," *IEEE Trans. Signal Process.*, vol. 65, no. 5, pp. 1196–1206, Mar. 2017.



- [34] R. Boyer and P. Comon, "Rectified ALS algorithm for multidimensional harmonic retrieval," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
- [35] Y. Gao, Y. Wan, S. Tang, B. Xu, and Q. Wan, "A CPD-based AOA estimation algorithm with Vandermonde-constrained preprocessing approach," in *Proc. Inter. Conf. Infor., Commun. Signal Process. (ICICS)*, Dec. 2015, pp. 1–5.
- [36] J. Liu, X. Liu, and X. Ma, "Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5179–5194, Nov. 2007.
- [37] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, Nov. 2010.
- [38] A. Paulraj, R. Roy, and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [39] G. H. Golub, and C. F. Van Loan, *Matrix Computations*, 4th ed., Baltimore, MD: Johns Hopkins Univ. Press, 2013.
- [40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Upper Saddle River, NJ, USA: Prentice-Hall, 1993.