Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems

Zhou Zhou, Jun Fang, *Member, IEEE*, Linxiao Yang, Hongbin Li, *Senior Member, IEEE*, Zhi Chen, and Rick S. Blum, *Fellow, IEEE*

***Abstract*— We consider the problem of downlink channel estimation for millimeter wave (mmWave) MIMO-OFDM systems, where both the base station (BS) and the mobile station (MS) employ large antenna arrays for directional precoding/beamforming. Hybrid analog and digital beamforming structures are employed in order to offer a compromise between hardware complexity and system performance. Different from most existing studies that are concerned with narrowband channels, we consider estimation of wideband mmWave channels with frequency selectivity, which is more appropriate for mmWave MIMO-OFDM systems. By exploiting the sparse scattering nature of mmWave channels, we propose a CANDECOMP/PARAFAC (CP) decomposition-based method for channel parameter estimation (including angles of arrival/departure, time delays, and fading coefficients). In our proposed method, the received signal at the MS is expressed as a third-order tensor. We show that the tensor has the form of a low-rank CP, and the channel parameters can be estimated from the associated factor matrices. Our analysis reveals that the uniqueness of the CP decomposition can be guaranteed even when the size of the tensor is small. Hence the proposed method has the potential to achieve substantial training overhead reduction. We also develop Cramér-Rao bound (CRB) results for channel parameters and compare our proposed method with a compressed sensing-based method. Simulation results show that the proposed method attains mean square errors that are very close to their associated CRBs and present a clear advantage over the compressed sensing-based method.**

***Index Terms*— MmWave MIMO-OFDM systems, channel estimation, CANDECOMP/PARAFAC (CP) decomposition, Cramér-Rao bound (CRB).**

Manuscript received November 1, 2016; revised February 18, 2017; accepted March 6, 2017. Date of publication April 28, 2017; date of current version June 19, 2017. This work was supported in part by the National Science Foundation of China under Grant 61522104, Grant 61428103, and Grant 91438118, in part by the Key Project of Fundamental and Frontier Research of Chongqing under Grant cstc2015jcyjBX0085, in part by the National Science Foundation under Grant ECCS-1408182, Grant ECCS-1609393, and Grant ECCS-1405579, and in part by the Air Force Office of Scientific Research under Grant FA9550-16-1-0243. *(Corresponding author: Jun Fang.)*

Z. Zhou, J. Fang, L. Yang, and Z. Chen are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: junfang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hongbin.li@stevens.edu).

R. S. Blum is with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: rblum@lehigh.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2017.2699338

I. Introduction

MILLIMETER-WAVE (mmWave) communication is a promising technology for future cellular networks [1]–[4]. The large bandwidth available in mmWave bands can offer gigabit-per-second communication data rates. However, high signal attenuation at such high frequency presents a major challenge for system design [5]. To compensate for the significant path loss, large antenna arrays should be used at both the base station (BS) and the mobile station (MS) to provide sufficient beamforming gain for mmWave communications [6], [7]. This requires accurate channel estimation which is essential for the proper operation of directional precoding/beamforming in mmWave systems. Note that the high directivity makes mmWave communications vulnerable to blockage events, which can be frequent in indoor environments. To address this issue, multiple relays can be employed and a path selection technique was proposed in [8] to select the optimal path that maximizes the throughput.

Channel estimation in mmWave systems is challenging due to hybrid precoding structures and the large number of antennas. A primary challenge is that hybrid precoding structures [9]–[13] employed in mmWave systems prevent the digital baseband from directly accessing the entire channel dimension. This is also referred to as the channel subspace sampling limitation [14], [15], which makes it difficult to acquire useful channel state information (CSI) during a practical channel coherence time. To address this issue, fast beam scanning and searching techniques have been extensively studied, e.g. [14], [16], [17]. The objective of beam scanning is to search for the best beamformer-combiner pair by letting the transmitter and receiver scan the adaptive sounding beams or coded beams chosen from pre-determined sounding beam codebooks. Nevertheless, as the number of antennas increases, the size of the codebook should be enlarged accordingly, which in turn results in an increase in the sounding/training overhead.

Unlike beam scanning techniques whose objective is to find the best beam pair, another approach is to directly estimate the mmWave channel or its associated parameters, e.g. [15], [18]–[21]. In particular, by exploiting the sparse scattering nature of the mmWave channels, mmWave channel estimation can be formulated as a sparse signal recovery problem, and it has been shown [18], [19] that substantial reduction in training overhead can be achieved via compressed

0733-8716 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

sensing methods. In [19], an adaptive compressed sensing method was developed for mmWave channel estimation based on a hierarchical multi-resolution beamforming codebook. Compared to the standard compressed sensing method, the adaptive method is more efficient as the training precoding is adaptively adjusted according to the outputs of earlier stages. Nevertheless, this improved efficiency comes at the expense of requiring feedback from the MS to the BS. Other compressed sensing-based mmWave channel estimation methods include [22]–[28]. Most of the above existing methods are concerned with estimation of narrowband channels. MmWave systems, however, are very likely to operate on wideband channels with frequency selectivity [29]. In [30], the authors considered the problem of multi-user uplink channel estimation in mmWave MIMO-OFDM systems and proposed a distributed compressed sensing-based scheme by exploiting the angular domain structured sparsity of mmWave wideband frequency-selective fading channels. Precoding design, with limited feedback for frequency selective wideband mmWave channels, was studied in [29].

In this paper, we study the problem of downlink channel estimation for mmWave MIMO-OFDM systems, where wideband frequency-selective fading channels are considered. We propose a CANDECOMP/PARAFAC (CP) decomposition-based method for downlink channel estimation. The proposed method is based on the following three key observations. First, by adopting a simple setup at the transmitter, the received signal can be organized into a third-order tensor which admits a CP decomposition. Second, due to the sparse scattering nature of mmWave channels, the tensor has an intrinsic low CP rank that guarantees the uniqueness of the CP decomposition. Third, the channel parameters, including angles of arrival/departure, time delays, and fading coefficients, can be easily extracted based on the decomposed factor matrices. We conduct a rigorous analysis on the uniqueness of the CP decomposition. Analyses show that the uniqueness of the CP decomposition can be guaranteed even when the size of the tensor is small. This result implies that our proposed method can achieve a substantial training overhead reduction. The Cramér-Rao bound (CRB) results for channel parameters are also developed, which provides a benchmark for the performance of our proposed method, and also describes the best asymptotically achievable performance. Our experiments show that the mean square errors attained by the proposed method are close to their corresponding CRBs.

Our proposed CP decomposition-based method enjoys the following advantages as compared with the compressed sensing-based method. Firstly, unlike compressed sensing techniques which require to discretize the continuous parameter space into a finite set of grid points, our proposed method is essentially a gridless approach and therefore is free of the grid discretization errors. Secondly, the proposed method captures the intrinsic multi-dimensional structure of the multiway data, which helps achieve a performance improvement. Thirdly, the use of tensors for data representation and processing leads to a very low computational complexity, whereas most compressed sensing methods are usually plagued by high computational complexity. Our simulation results show that our proposed method has a computational complexity as low as the simplest compressed sensing method, i.e. the orthogonal matching pursuit (OMP) method [31], [32], while achieving a much higher estimation accuracy than the OMP. Lastly, the conditions for the uniqueness of the CP decomposition are easy to analyze, and can be employed to determine the exact amount of training overhead required for unique decomposition. In contrast, it is usually difficult to analyze and check the exact recovery condition for generic dictionaries for compressed sensing techniques.

We would like to emphasize the difference between the current work and our previous work [21]. Although both works employ the CP decomposition for mmWave channel estimation, their formulations are quite different. Specifically, our previous work [21] considered the estimation of multi-user uplink narrowband mmWave channels, which requires a delicate layered transmission scheme such that the received signal can be expressed as a tensor that admits a CP decomposition. In contrast, our current work considers the estimation of downlink wideband frequency-selective mmWave channels. Subcarriers in the OFDM systems provide an additional mode that naturally leads to a tensor representation of the data. In the current paper, we also provide CRB results for channel parameters, which is not available in our previous work.

The rest of the paper is organized as follows. In Section II, we provide notations and basics on the CP decomposition. The system model and the channel estimation problem are discussed in Section III. In Section IV, we propose a CP decomposition-based method for mmWave channel estimation. The uniqueness of the CP decomposition is also analyzed. Section V develops CRB results for the estimation of channel parameters. A compressed sensing-based channel estimation method is discussed in Section VI. Computational complexity of the proposed method and the compressed sensing-based method is analyzed in Section VII. Simulation results are provided in Section VIII, followed by concluding remarks in Section IX.

II. Preliminaries

To make the paper self-contained, we provide a brief review on tensors and the CP decomposition. More details regarding the notations and basics on tensors can be found in [33]–[36]. Simply speaking, a tensor is a generalization of a matrix to higher-order dimensions, also known as ways or modes. Vectors and matrices can be viewed as special cases of tensors with one and two modes, respectively. Throughout this paper, we use symbols $\otimes$, $\circ$, and $\odot$ to denote the Kronecker, outer, and Khatri-Rao product, respectively.

Let $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ denote an Nth-order tensor with its $(i_1, \ldots, i_N)$th entry denoted by $\mathcal{X}_{i_1 \cdots i_N}$. Here the order N of a tensor is the number of dimensions. Fibers are a higher-order analogue of matrix rows and columns. The mode-n fibers of $\boldsymbol{\mathcal{X}}$ are I_n-dimensional vectors obtained by fixing every index but i_n. Slices are two-dimensional sections of a tensor, defined by fixing all but two indices. Unfolding or matricization is an operation that turns a tensor into a matrix. The mode-n unfolding of a tensor $\boldsymbol{\mathcal{X}}$, denoted as $\boldsymbol{X}_{(n)}$, arranges the mode-n fibers to be the columns of the resulting matrix.

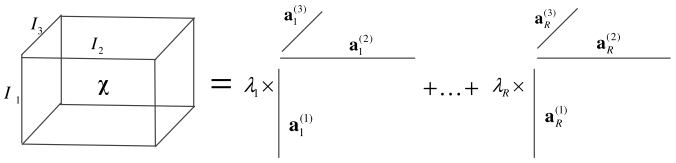


Fig. 1. Schematic of CP decomposition.

The CP decomposition decomposes a tensor into a sum of rank-one component tensors (see Fig. 1), i.e.

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \lambda_r \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \tag{1}$$

where $\boldsymbol{a}_r^{(n)} \in \mathbb{C}^{I_n}$, the minimum achievable R is referred to as the rank of the tensor, and $\boldsymbol{A}^{(n)} \triangleq [\boldsymbol{a}_1^{(n)} \ \ldots \ \boldsymbol{a}_R^{(n)}] \in \mathbb{C}^{I_n \times R}$ denotes the factor matrix along the n-th mode. Elementwise, we have

$$\mathcal{X}_{i_1 i_2 \cdots i_N} = \sum_{r=1}^{R} \lambda_r a_{i_1 r}^{(1)} a_{i_2 r}^{(2)} \cdots a_{i_N r}^{(N)} \tag{2}$$

The mode-n unfolding of $\boldsymbol{\mathcal{X}}$ can be expressed as

$$\boldsymbol{X}_{(n)} = \boldsymbol{A}^{(n)} \boldsymbol{\Lambda} \left(\boldsymbol{A}^{(N)} \odot \cdots \boldsymbol{A}^{(n+1)} \odot \boldsymbol{A}^{(n-1)} \odot \cdots \boldsymbol{A}^{(1)} \right)^T \tag{3}$$

where $\boldsymbol{\Lambda} \triangleq \mathrm{diag}(\lambda_1, \ldots, \lambda_R)$.

III. System Model

Consider a mmWave massive MIMO-OFDM system consisting of a base station (BS) and multiple mobile stations (MSs). To facilitate the hardware implementation, hybrid analog and digital beamforming structures are employed by both the BS and the MS (see Fig. 2). We assume that the BS is equipped with N_{BS} antennas and M_{BS} RF chains, and the MS is equipped with N_{MS} antennas and M_{MS} RF chains. The number of RF chains is less than the number of antennas, i.e. $M_{\mathrm{BS}} < N_{\mathrm{BS}}$ and $M_{\mathrm{MS}} < N_{\mathrm{MS}}$. In particular, we assume $M_{\mathrm{MS}} = 1$, i.e. each MS has only one RF chain. The total number of OFDM tones (subcarriers) is assumed to be $\bar{K}$, among which K subcarriers are selected for training. For simplicity, here we assume subcarriers $\{1, 2, \ldots, K\}$ are assigned for training. Nevertheless, our formulation and method can be easily extended to other subset choices. In the downlink scenario, we only need to consider a single user system because the channel estimation is conducted by each user individually [37]. To simplify our problem, we ignore the inter-cell interference resulting from frequency reuse from neighboring cells. To mitigate inter-cell interference, many useful techniques such as the coordinated transmission scheme with large-scale CSI at the transmitter [38] and the cognitive transmission scheme [39] were recently developed.

We adopt a downlink training scheme similar to [18], [19]. For each subcarrier, the BS employs T different beamforming vectors at T successive time frames. Each time frame is divided into M sub-frames, and at each sub-frame, the MS uses an individual combining vector to detect the transmitted signal. The beamforming vector associated with the kth subcarrier at the tth time frame can be expressed as

$$\boldsymbol{x}_k(t) = \boldsymbol{F}_{\mathrm{RF}}(t) \boldsymbol{F}_k(t) \boldsymbol{s}_k(t) \quad \forall k = 1, \ldots, K \tag{4}$$

where $\boldsymbol{s}_k(t) \in \mathbb{C}^r$ denotes the pilot symbol vector, $\boldsymbol{F}_k(t) \in \mathbb{C}^{M_{\mathrm{BS}} \times r}$ denotes the digital precoding matrix for the kth subcarrier, and $\boldsymbol{F}_{\mathrm{RF}}(t) \in \mathbb{C}^{N_{\mathrm{BS}} \times M_{\mathrm{BS}}}$ is a common RF precoder for all subcarriers. The procedure to generate the beamforming vector (4) is elaborated as follows. The pilot symbol vector $\boldsymbol{s}_k(t)$ at each subcarrier is first precoded using a digital precoding matrix $\boldsymbol{F}_k(t)$. The symbol blocks are transformed to the time-domain using $\bar{K}$-point inverse discrete Fourier transform (IDFT). A cyclic prefix is then added to the symbol blocks, finally a common RF precoder $\boldsymbol{F}_{\mathrm{RF}}(t)$ is applied to all subcarriers.

At each time frame, the MS successively employs M RF combining vectors $\{\boldsymbol{q}_m\}$ to detect the transmitted signal. Note that these combining vectors are common to all subcarriers. At each sub-frame, the received signal is first combined in the RF domain. Then, the cyclic prefix is removed and symbols are converted back to the frequency domain by performing a discrete Fourier transform (DFT). After processing, the received signal associated with the kth subcarrier at the mth sub-frame can be expressed as [29]

$$y_{k,m}(t) = \boldsymbol{q}_m^T \boldsymbol{H}_k \boldsymbol{x}_k(t) + w_{k,m}(t) \tag{5}$$

where $\boldsymbol{q}_m \in \mathbb{C}^{N_{\mathrm{MS}}}$ denotes the combining vector used at the mth sub-frame, $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{MS}} \times N_{\mathrm{BS}}}$ is the channel matrix associated with the kth subcarrier, and $w_{k,m}(t)$ denotes the additive Gaussian noise. Collecting the M received signals $\{y_{k,m}(t)\}_{m=1}^{M}$ at each time frame, we have

$$\begin{aligned} \boldsymbol{y}_k(t) &= \boldsymbol{Q}^T \boldsymbol{H}_k \boldsymbol{x}_k(t) + \boldsymbol{w}_k(t) \\ &= \boldsymbol{Q}^T \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}}(t) \boldsymbol{F}_k(t) \boldsymbol{s}_k(t) + \boldsymbol{w}_k(t) \end{aligned} \tag{6}$$

where

$$\begin{aligned} \boldsymbol{y}_k(t) &\triangleq [y_{k,1}(t) \ \ldots \ y_{k,M}(t)]^T \\ \boldsymbol{w}_k(t) &\triangleq [w_{k,1}(t) \ \ldots \ w_{k,M}(t)]^T \\ \boldsymbol{Q} &\triangleq [\boldsymbol{q}_1 \ \ldots \ \boldsymbol{q}_M] \end{aligned} \tag{7}$$

Measurement campaigns in dense-urban NLOS environments reveal that mmWave channels typically exhibit limited scattering characteristics [40]. Also, considering the wideband nature of mmWave channels, we adopt a geometric wideband mmWave channel model with L scatterers between the MS and the BS. Each scatterer is characterized by a time delay τ_l, angles of arrival and departure (AoA/AoD), $\theta_l, \phi_l \in [0, 2\pi]$. With these parameters, the channel matrix in the delay domain can be written as [29], [30]

$$\boldsymbol{H}(\tau) = \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_{\mathrm{MS}}(\theta_l) \boldsymbol{a}_{\mathrm{BS}}^T(\phi_l) \delta(\tau - \tau_l) \tag{8}$$

where α_l is the complex path gain associated with the lth path, $\boldsymbol{a}_{\mathrm{MS}}(\theta_l)$ and $\boldsymbol{a}_{\mathrm{BS}}(\phi_l)$ are the antenna array response vectors of the MS and BS, respectively, and $\delta(\cdot)$ represents the delta function. Throughout this paper, we assume

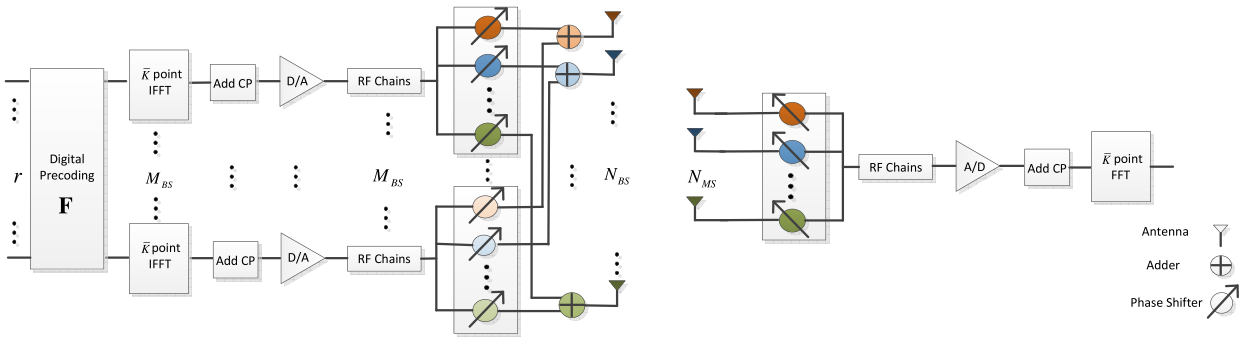


Fig. 2. A block diagram of the MIMO-OFDM transceiver that employs hybrid analog/digital precoding.

A1 Different scatterers have different angles of arrival, angles of departure as well as time delays, i.e. $\theta_i \neq \theta_j$, $\phi_i \neq \phi_j$, $\tau_i \neq \tau_j$ for $i \neq j$.

Since scatterers are randomly distributed in space, this assumption is usually valid in practice.

Given the delay-domain channel model, the frequency-domain channel matrix $\boldsymbol{H}_k$ associated with the kth subcarrier can be obtained as

$$\boldsymbol{H}_k = \sum_{l=1}^{L} \alpha_l \exp(-j2\pi\tau_l f_s k/\bar{K}) \boldsymbol{a}_{\mathrm{MS}}(\theta_l)\boldsymbol{a}_{\mathrm{BS}}^T(\phi_l) \tag{9}$$

where f_s denotes the sampling rate. Our objective is to estimate the channel matrices $\{\boldsymbol{H}_k\}_{k=1}^{\bar{K}}$ from the received signals $\boldsymbol{y}_k(t), \forall k = 1, \ldots, K, \forall t = 1, \ldots, T$. In particular, we wish to provide a reliable channel estimate by using as few measurements as possible because the number of measurements is linearly proportional to the number of time frames and the number of sub-frames, both of which are expected to be minimized. To facilitate our algorithm development, we assume that the digital precoding matrices and the pilot symbols remain the same for different subcarriers, i.e. $\boldsymbol{F}_k(t) = \boldsymbol{F}(t)$, $\boldsymbol{s}_k(t) = \boldsymbol{s}(t), \forall k = 1, \ldots, K$. As will be shown later, this simplification enables us to develop an efficient tensor factorization-based method to extract the channel state information from very few number of measurements.

IV. Proposed CP Decomposition-Based Method

Suppose $\boldsymbol{F}_k(t) = \boldsymbol{F}(t)$ and $\boldsymbol{s}_k(t) = \boldsymbol{s}(t), \forall k = 1, \ldots, K$. Let $\boldsymbol{S} \triangleq [\boldsymbol{s}(1) \ \ldots \ \boldsymbol{s}(T)]$. The received signal at the kth subcarrier can be written as

$$\boldsymbol{Y}_k = \boldsymbol{Q}^T \boldsymbol{H}_k \boldsymbol{P} + \boldsymbol{W}_k \quad k = 1, \ldots, K \tag{10}$$

where

$$\begin{aligned} \boldsymbol{Y}_k &\triangleq [\boldsymbol{y}_k(1) \ \ldots \ \boldsymbol{y}_k(T)] \\ \boldsymbol{W}_k &\triangleq [\boldsymbol{w}_k(1) \ \ldots \ \boldsymbol{w}_k(T)] \\ \boldsymbol{P} &\triangleq [\boldsymbol{p}(1) \ \ldots \ \boldsymbol{p}(T)] \end{aligned} \tag{11}$$

in which $\boldsymbol{p}(t) \triangleq \boldsymbol{F}_{\mathrm{RF}}(t)\boldsymbol{F}(t)\boldsymbol{s}(t)$.

Since signals from multiple subcarriers are available at the MS, the received signal can be expressed by a third-order tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M \times T \times \bar{K}}$ whose three modes respectively stand for the sub-frame, the time frame, and the subcarrier, and its (m, t, k)th entry is given by $y_{k,m}(t)$. Substituting (9) into (10), we obtain

$$\begin{aligned} \boldsymbol{Y}_k &= \sum_{l=1}^{L} \tilde{\alpha}_{l,k} \boldsymbol{Q}^T \boldsymbol{a}_{\mathrm{MS}}(\theta_l)\boldsymbol{a}_{\mathrm{BS}}^T(\phi_l)\boldsymbol{P} + \boldsymbol{W}_k \\ &= \sum_{l=1}^{L} \tilde{\alpha}_{l,k} \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l)\tilde{\boldsymbol{a}}_{\mathrm{BS}}^T(\phi_l) + \boldsymbol{W}_k \end{aligned} \tag{12}$$

where $\tilde{\alpha}_{l,k} \triangleq \alpha_l \exp(-j2\pi\tau_l f_s k/\bar{K})$, $\tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \triangleq \boldsymbol{Q}^T \boldsymbol{a}_{\mathrm{MS}}(\theta_l)$, and $\tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_l) \triangleq \boldsymbol{P}^T \boldsymbol{a}_{\mathrm{BS}}(\phi_l)$. We see that each slice of the tensor $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{Y}_k$, is a weighted sum of a common set of rank-one outer products. The tensor $\boldsymbol{\mathcal{Y}}$ thus admits the following CANDECOMP/PARAFAC (CP) decomposition which decomposes a tensor into a sum of rank-one component tensors, i.e.

$$\boldsymbol{\mathcal{Y}} = \sum_{l=1}^{L} \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \circ \tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_l) \circ (\alpha_l \boldsymbol{g}(\tau_l)) + \boldsymbol{\mathcal{W}} \tag{13}$$

in which

$$\boldsymbol{g}(\tau_l) \triangleq [\exp(-j2\pi\tau_l f_s(1/\bar{K})) \ \ldots \ \exp(-j2\pi\tau_l f_s(K/\bar{K}))]^T \tag{14}$$

Due to the sparse scattering nature of the mmWave channel, the number of paths, L, is usually small relative to the dimensions of the tensor. Hence the tensor $\boldsymbol{\mathcal{Y}}$ has an intrinsic low-rank structure. As will be discussed later, this low-rank structure ensures that the CP decomposition of $\boldsymbol{\mathcal{Y}}$ is unique up to scaling and permutation ambiguities. Therefore an estimate of the parameters $\{\alpha_l, \phi_l, \theta_l, \tau_l\}$ can be obtained by performing a CP decomposition of the received signal $\boldsymbol{\mathcal{Y}}$. Define

$$\boldsymbol{A} \triangleq [\tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_1) \ \ldots \ \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_L)] \tag{15}$$

$$\boldsymbol{B} \triangleq [\tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_1) \ \ldots \ \tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_L)] \tag{16}$$

$$\boldsymbol{C} \triangleq [\alpha_1 \boldsymbol{g}(\tau_1) \ \ldots \ \alpha_L \boldsymbol{g}(\tau_L)] \tag{17}$$

These three matrices $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ are factor matrices associated with a noiseless version of $\boldsymbol{\mathcal{Y}}$.

A. CP Decomposition

If the number of paths, L, is known or estimated *a priori*, the CP decomposition of $\boldsymbol{\mathcal{Y}}$ can be accomplished by solving

$$\min_{\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}} \quad \|\boldsymbol{\mathcal{Y}} - \sum_{l=1}^{L} \hat{\boldsymbol{a}}_l \circ \hat{\boldsymbol{b}}_l \circ \hat{\boldsymbol{c}}_l\|_F^2 \tag{18}$$

where we let $\hat{\boldsymbol{A}} = [\hat{\boldsymbol{a}}_1 \ \ldots \ \hat{\boldsymbol{a}}_L]$, $\hat{\boldsymbol{B}} = [\hat{\boldsymbol{b}}_1 \ \ldots \ \hat{\boldsymbol{b}}_L]$, and $\hat{\boldsymbol{C}} = [\hat{\boldsymbol{c}}_1 \ \ldots \ \hat{\boldsymbol{c}}_L]$. The above optimization can be efficiently solved by an alternating least squares (ALS) procedure which alternatively minimizes the data fitting error with respect to one of the factor matrices, with the other two factor matrices fixed

$$\hat{\boldsymbol{A}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{A}}} \left\| \boldsymbol{Y}_{(1)}^T - (\hat{\boldsymbol{C}}^{(t)} \odot \hat{\boldsymbol{B}}^{(t)})\hat{\boldsymbol{A}}^T \right\|_F^2 \tag{19}$$

$$\hat{\boldsymbol{B}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{B}}} \left\| \boldsymbol{Y}_{(2)}^T - (\hat{\boldsymbol{C}}^{(t)} \odot \hat{\boldsymbol{A}}^{(t+1)})\hat{\boldsymbol{B}}^T \right\|_F^2 \tag{20}$$

$$\hat{\boldsymbol{C}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{C}}} \left\| \boldsymbol{Y}_{(3)}^T - (\hat{\boldsymbol{B}}^{(t+1)} \odot \hat{\boldsymbol{A}}^{(t+1)})\hat{\boldsymbol{C}}^T \right\|_F^2 \tag{21}$$

Note that (19)–(21) are least squares problems whose solutions can be easily obtained.

If the knowledge of the number of paths L is unavailable, more sophisticated CP decomposition techniques (e.g. [41]–[43]) can be employed to jointly estimate the model order and the factor matrices. The basic idea of these CP decomposition techniques is to use sparsity-promoting priors or functions to find a low-rank representation of the observed tensor. Specifically, let $\hat{L} \gg L$ denote an overestimated CP rank, the following optimization can be employed for CP decomposition [41]

$$\min_{\hat{\boldsymbol{A}},\hat{\boldsymbol{B}},\hat{\boldsymbol{C}}} \ \|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}\|_F^2 + \mu \left(\mathrm{tr}(\hat{\boldsymbol{A}}\hat{\boldsymbol{A}}^H) + \mathrm{tr}(\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}^H) + \mathrm{tr}(\hat{\boldsymbol{C}}\hat{\boldsymbol{C}}^H) \right)$$

$$\text{s.t.} \ \ \boldsymbol{\mathcal{X}} = \sum_{l=1}^{\hat{L}} \hat{\boldsymbol{a}}_l \circ \hat{\boldsymbol{b}}_l \circ \hat{\boldsymbol{c}}_l \tag{22}$$

where μ is a regularization parameter to control the tradeoff between low-rankness and the data fitting error, $\hat{\boldsymbol{A}} = [\hat{\boldsymbol{a}}_1 \ \ldots \ \hat{\boldsymbol{a}}_{\hat{L}}]$, $\hat{\boldsymbol{B}} = [\hat{\boldsymbol{b}}_1 \ \ldots \ \hat{\boldsymbol{b}}_{\hat{L}}]$, and $\hat{\boldsymbol{C}} = [\hat{\boldsymbol{c}}_1 \ \ldots \ \hat{\boldsymbol{c}}_{\hat{L}}]$. The above optimization (22) can still be solved by an ALS procedure as follows

$$\hat{\boldsymbol{A}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{A}}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(1)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{C}}^{(t)} \odot \hat{\boldsymbol{B}}^{(t)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \hat{\boldsymbol{A}}^T \right\|_F^2$$

$$\hat{\boldsymbol{B}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{B}}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(2)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{C}}^{(t)} \odot \hat{\boldsymbol{A}}^{(t+1)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \hat{\boldsymbol{B}}^T \right\|_F^2$$

$$\hat{\boldsymbol{C}}^{(t+1)} = \arg\min_{\hat{\boldsymbol{C}}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(3)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{B}}^{(t+1)} \odot \hat{\boldsymbol{A}}^{(t+1)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \hat{\boldsymbol{C}}^T \right\|_F^2$$

The true CP rank of the tensor, L, can be estimated by removing those negligible rank-one tensor components after convergence.

B. Channel Estimation

We discuss how to estimate the mmWave channels based on the estimated factor matrices $\{\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}\}$. As shown in the next subsection, the CP decomposition is unique up to scaling and permutation ambiguities under a mild condition. More precisely, the estimated factor matrices and the true factor matrices are related as

$$\hat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Lambda}_1\boldsymbol{\Pi} + \boldsymbol{E}_1 \tag{23}$$

$$\hat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Lambda}_2\boldsymbol{\Pi} + \boldsymbol{E}_2 \tag{24}$$

$$\hat{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\Lambda}_3\boldsymbol{\Pi} + \boldsymbol{E}_3 \tag{25}$$

where $\{\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3\}$ are unknown nonsingular diagonal matrices which satisfy $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \boldsymbol{I}$; $\boldsymbol{\Pi}$ is an unknown permutation matrix; and $\boldsymbol{E}_1$, $\boldsymbol{E}_2$, and $\boldsymbol{E}_3$ denote the estimation errors associated with the three estimated factor matrices, respectively. The permutation matrix $\boldsymbol{\Pi}$ can be ignored because it is common to all three factor matrices. Note that each column of $\boldsymbol{A}$ is characterized by the associated angle of arrival θ_l. Hence the angle of arrival θ_l can be estimated via a simple correlation-based scheme

$$\hat{\theta}_l = \arg\max_{\theta_l} \ \frac{|\hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l)|}{\|\hat{\boldsymbol{a}}_l\|_2 \|\tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l)\|_2} \tag{26}$$

where $\hat{\boldsymbol{a}}_l$ denotes the lth column of $\hat{\boldsymbol{A}}$. It can be shown in Appendix A that this simple correlation-based scheme is a maximum likelihood (ML) estimator, provided that entries in the estimation error matrix, $\boldsymbol{E}_1$, follow an i.i.d. circularly symmetric Gaussian distribution. The angle of departure ϕ_l can be obtained similarly as

$$\hat{\phi}_l = \arg\max_{\phi_l} \ \frac{|\hat{\boldsymbol{b}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_l)|}{\|\hat{\boldsymbol{b}}_l\|_2 \|\tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_l)\|_2} \tag{27}$$

where $\hat{\boldsymbol{b}}_l$ denotes the lth column of $\hat{\boldsymbol{B}}$. We now discuss how to estimate the time delay τ_l from the estimated factor matrix $\hat{\boldsymbol{C}}$. Note that the lth column of $\boldsymbol{C}$ is given by $\alpha_l \boldsymbol{g}(\tau_l)$. Therefore the time delay τ_l can be estimated via

$$\hat{\tau}_l = \arg\min_{\tau_l} \ \frac{|\hat{\boldsymbol{c}}_l^H \boldsymbol{g}(\tau_l)|}{\|\hat{\boldsymbol{c}}_l\|_2 \|\boldsymbol{g}(\tau_l)\|_2} \tag{28}$$

where $\hat{\boldsymbol{c}}_l$ denotes the lth column of $\hat{\boldsymbol{C}}$. The maximization problems (26)–(28) involve one-dimensional search which can be efficiently performed by first employing a coarse grid and then gradually refining the search in the vicinity of possible grid points. Substituting the estimated $\{\theta_l\}$ and $\{\phi_l\}$ back into (23) and (24), an estimate of the nonsingular diagonal matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ can be obtained. An estimate of $\boldsymbol{\Lambda}_3$ can then be calculated from the equality $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \boldsymbol{I}$. Finally, the fading coefficients $\{\alpha_l\}$ can be estimated from (25). The channel matrices $\{\boldsymbol{H}_k\}$ can now be recovered from the estimated parameters $\{\hat{\theta}_l, \hat{\phi}_l, \hat{\tau}_l, \hat{\alpha}_l\}$.

C. Uniqueness

We discuss the uniqueness of the CP decomposition. It is well known that the essential uniqueness of CP decomposition can be guaranteed by Kruskal's condition [44]. Let $k_{\boldsymbol{X}}$ denote the k-rank of a matrix $\boldsymbol{X}$, which is defined as the largest value of $k_{\boldsymbol{X}}$ such that every subset of $k_{\boldsymbol{X}}$ columns of the matrix $\boldsymbol{X}$ is linearly independent. We have the following theorem.

Theorem 1: Let $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ be a CP solution which decomposes a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M\times N\times K}$ into R rank-one

arrays, where $\boldsymbol{X} \in \mathbb{C}^{M\times R}$, $\boldsymbol{Y} \in \mathbb{C}^{N\times R}$, and $\boldsymbol{Z} \in \mathbb{C}^{K\times R}$. Suppose the following Kruskal's condition

$$k_{\boldsymbol{X}} + k_{\boldsymbol{Y}} + k_{\boldsymbol{Z}} \geq 2R + 2 \tag{29}$$

holds and there is an alternative CP solution $(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}, \bar{\boldsymbol{Z}})$ which also decomposes $\boldsymbol{\mathcal{X}}$ into R rank-one arrays. Then we have $\bar{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{\Pi}\boldsymbol{\Lambda}_a$, $\bar{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{\Pi}\boldsymbol{\Lambda}_b$, and $\bar{\boldsymbol{Z}} = \boldsymbol{Z}\boldsymbol{\Pi}\boldsymbol{\Lambda}_c$, where $\boldsymbol{\Pi}$ is a unique permutation matrix and $\boldsymbol{\Lambda}_a$, $\boldsymbol{\Lambda}_b$, and $\boldsymbol{\Lambda}_c$ are unique diagonal matrices such that $\boldsymbol{\Lambda}_a\boldsymbol{\Lambda}_b\boldsymbol{\Lambda}_c = \boldsymbol{I}$.

Proof: A rigorous proof can be found in [45]. ■

Note that Kruskal's condition cannot hold when $R = 1$. However, in that case the uniqueness has been proven by Harshman [46]. Kruskal's sufficient condition is also necessary for $R = 2$ and $R = 3$, but not for $R > 3$ [45].

From the above theorem, we know that if

$$k_{\boldsymbol{A}} + k_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2L + 2 \tag{30}$$

then the CP decomposition of $\boldsymbol{\mathcal{Y}}$ is essentially unique.

We first examine the k-rank of $\boldsymbol{A}$. Note that

$$\boldsymbol{A} = \boldsymbol{Q}^T[\boldsymbol{a}_{\mathrm{MS}}(\theta_1) \ \ \ldots \ \ \boldsymbol{a}_{\mathrm{MS}}(\theta_L)] \triangleq \boldsymbol{Q}^T\boldsymbol{A}_{\mathrm{MS}} \tag{31}$$

where $\boldsymbol{A}_{\mathrm{MS}} \in \mathbb{C}^{N_{\mathrm{MS}}\times L}$ is a Vandermonte matrix when a uniform linear array is employed. Suppose assumption A1 holds valid. For a randomly generated $\boldsymbol{Q}$ whose entries are chosen uniformly from a unit circle, we can show that the k-rank of $\boldsymbol{A}$ is equal to (details can be found in Appendix B)

$$k_{\boldsymbol{A}} = \min(M, L) \tag{32}$$

with probability one. Consider the k-rank of $\boldsymbol{B}$. We have

$$\boldsymbol{B} = \boldsymbol{P}^T[\boldsymbol{a}_{\mathrm{BS}}(\phi_1) \ \ \ldots \ \ \boldsymbol{a}_{\mathrm{BS}}(\phi_L)] \triangleq \boldsymbol{P}^T\boldsymbol{A}_{\mathrm{BS}} \tag{33}$$

Similarly, for a randomly generated $\boldsymbol{P}$ whose entries are uniformly chosen from a unit circle, we can deduce that the k-rank of $\boldsymbol{B}$ is equal to

$$k_{\boldsymbol{B}} = \min(T, L) \tag{34}$$

with probability one. Now let us examine the k-rank of $\boldsymbol{C}$. Recall that $\boldsymbol{C}$ can be expressed as

$$\boldsymbol{C} = [\boldsymbol{g}(\tau_1) \ \ \ldots \ \ \boldsymbol{g}(\tau_L)]\boldsymbol{D}_\alpha \tag{35}$$

where $\boldsymbol{D}_\alpha \triangleq \mathrm{diag}(\alpha_1, \ldots, \alpha_L)$, and $\boldsymbol{g}(\tau_l)$ is defined in (14). We see that $\boldsymbol{C}$ is a columnwise-scaled Vandermonte matrix. Therefore the k-rank of $\boldsymbol{C}$ is

$$k_{\boldsymbol{C}} = \min(K, L) \tag{36}$$

Since L is usually small, it is reasonable to assume that the number of subcarriers used for training is greater than L, i.e. $K \geq L$. Hence we have $k_{\boldsymbol{C}} = L$. To meet Kruskal's condition (30), we only need $k_{\boldsymbol{A}} + k_{\boldsymbol{B}} \geq L + 2$. Recalling (32)–(34), we can either choose $\{T = L, M = 2\}$ or $\{M = L, T = 2\}$ to satisfy Kruskal's condition. In summary, for randomly generated beamforming matrix $\boldsymbol{P}$ and combining matrix $\boldsymbol{Q}$ whose entries are chosen uniformly from a unit circle, our proposed method only needs $T = L$ (or $T = 2$) time frames and $M = 2$ (or $M = L$) sub-frames to enable reliable estimation of channel parameters, thus achieving a substantial training overhead reduction. In practice, due to the observation noise and estimation errors, we may need a slightly larger T and M to yield an accurate channel estimate.

Note that although we assume entries of $\boldsymbol{P}$ are chosen from a unit circle, the hybrid precoder $\boldsymbol{P}$ can be otherwise devised, as long as Kruskal's condition is satisfied. Moreover, besides random coding, coded beams [17] which steer the antenna array towards multiple beam directions simultaneously can also be used to serve as the beamforming and combining vectors $\{\boldsymbol{p}_t\}$ and $\{\boldsymbol{q}_m\}$. The k-ranks of $\boldsymbol{A}$ and $\boldsymbol{B}$ may still obey (32) and (34) if $\boldsymbol{P}$ and $\boldsymbol{Q}$ are carefully designed. This is an important issue and will be explored in our future work.

V. CRB

In this section, we develop Cramér-Rao bound (CRB) results for the channel parameter (i.e. $\{\hat{\theta}_l, \hat{\phi}_l, \hat{\tau}_l, \hat{\alpha}_l\}$) estimation problem considered in (13). Details of the derivation can be found in Appendix C. Throughout our analysis, the observation noise in (13) is assumed to be complex circularly symmetric i.i.d. Gaussian noise. As is well known, the CRB is a lower bound on the variance of any unbiased estimator [47]. It provides a benchmark for evaluating the performance of our proposed method. In addition, the CRB results illustrate the behavior of the resulting bounds, which helps understand the effect of different system parameters, including the beamforming and combining matrices, on the estimation performance.

Note that our proposed method involves two steps: the first step employs an ALS algorithm to perform the CP decomposition, and based on the decomposed factor matrices, the second step uses a simple correlation-based method to estimate the channel parameters. For zero-mean i.i.d. Gaussian noise, the ALS yields maximum likelihood estimates [48], provided that the global minimum is reached. Also, it can be proved that the correlation-based method used in the second step is a maximum likelihood estimator if the estimation errors associated with the factor matrices are i.i.d. Gaussian random variables. Therefore our proposed method can be deemed as a quasi-maximum likelihood estimator for the channel parameters. Under mild regularity conditions, the maximum likelihood estimator is asymptotically (in terms of the sample size) unbiased and asymptotically achieves the CRB. It therefore makes sense to compare our proposed CP-decomposition-based method with the CRB results.

VI. Compressed Sensing-Based Channel Estimation

By exploiting the sparse scattering nature, the downlink channel estimation problem considered in this paper can also be formulated as a sparse signal recovery problem. In the following, we briefly discuss the compressed sensing-based channel estimation method.

Taking the mode-3 unfolding of $\boldsymbol{\mathcal{Y}}$ (c.f. (13)), we have

$$\begin{aligned}\boldsymbol{Y}_{(3)} &= \boldsymbol{C}(\boldsymbol{B} \odot \boldsymbol{A})^T + \boldsymbol{W}_{(3)} \\ &= \boldsymbol{G}\boldsymbol{D}_\alpha\boldsymbol{\Sigma}^T(\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)^T + \boldsymbol{W}_{(3)}\end{aligned} \tag{37}$$

where

$$\boldsymbol{G} \triangleq [\boldsymbol{g}(\tau_1) \ \ \ldots \ \ \boldsymbol{g}(\tau_L)], \quad \boldsymbol{D}_\alpha \triangleq \mathrm{diag}(\alpha_1, \ldots, \alpha_L),$$
$$\boldsymbol{\Sigma} \triangleq [\boldsymbol{a}_{\mathrm{BS}}(\phi_1) \otimes \boldsymbol{a}_{\mathrm{MS}}(\theta_1) \ \ \cdots \ \ \boldsymbol{a}_{\mathrm{BS}}(\phi_L) \otimes \boldsymbol{a}_{\mathrm{MS}}(\theta_L)]$$

Taking the transpose of $\boldsymbol{Y}_{(3)}$, we have

$$\boldsymbol{Y}_{(3)}^T = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\boldsymbol{\Sigma}\boldsymbol{D}_\alpha \boldsymbol{G}^T + \boldsymbol{W}_{(3)}^T \tag{38}$$

We see that both $\boldsymbol{\Sigma}$ and $\boldsymbol{G}$ are characterized by unknown parameters which need to be estimated. To convert the estimation problem into a sparse signal recovery problem, we discretize the AoA-AoD space into an $N_1 \times N_2$ grid, in which each grid point is given by $\{\bar{\theta}_i, \bar{\phi}_j\}$ for $i = 1, ..., N_1$ and $j = 1, ..., N_2$, where $N_1 \gg L$, and $N_2 \gg L$. The true angles of arrival/departure are assumed to lie on the grid. Also, we discretize the time-delay domain into a finite set of grid points $\{\bar{\tau}_l\}_{l=1}^{N_3}$ ($N_3 \gg L$), and assume that the true time-delays $\{\tau_l\}$ lie on the discretized grid. Thus (38) can be re-expressed as

$$\boldsymbol{Y}_{(3)}^T = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\bar{\boldsymbol{\Sigma}}\bar{\boldsymbol{D}}_\alpha \bar{\boldsymbol{G}}^T + \boldsymbol{W}_{(3)}^T \tag{39}$$

where $\bar{\boldsymbol{\Sigma}} \in \mathbb{C}^{N_{\mathrm{MS}}N_{\mathrm{BS}} \times N_1 N_2}$ is an overcomplete dictionary consisting of $N_1 N_2$ columns, with its $(i + (j-1)N_1)$th column given by $\boldsymbol{a}_{\mathrm{BS}}(\bar{\phi}_j) \otimes \boldsymbol{a}_{\mathrm{MS}}(\bar{\theta}_i)$, and $\bar{\boldsymbol{G}} \in \mathbb{C}^{K \times N_3}$ is an overcomplete dictionary, with its nth column given by $\boldsymbol{g}(\bar{\tau}_n)$. $\bar{\boldsymbol{D}}_\alpha$ is a sparse matrix obtained by augmenting $\boldsymbol{D}_\alpha$ with zero rows and columns. Let $\boldsymbol{y} \triangleq \mathrm{vec}(\boldsymbol{Y}_{(3)}^T)$, (39) can be formulated as a conventional sparse signal recovery problem

$$\boldsymbol{y} = \bar{\boldsymbol{G}} \otimes ((\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\bar{\boldsymbol{\Sigma}})\boldsymbol{d} + \boldsymbol{w} \tag{40}$$

where $\boldsymbol{d} \triangleq \mathrm{vec}(\bar{\boldsymbol{D}}_\alpha)$ is an unknown sparse vector. Many efficient algorithms such as the orthogonal matching pursuit (OMP) [31] or the fast iterative shrinkage-thresholding algorithm (FISTA) [49] can be employed to solve the above sparse signal recovery problem.

Although all channel parameters can be jointly estimated via (40), this formulation involves a very large dictionary which in turn results in a high computational complexity. In fact, a more efficient compressed sensing scheme can be developed by dealing with the AoAs, AoDs and time delays separately. For example, to estimate the time delay parameters, we rewrite (37) as

$$\boldsymbol{Y}_{(3)} = \bar{\boldsymbol{G}}\boldsymbol{D}_\tau + \boldsymbol{W}_{(3)} \tag{41}$$

where $\boldsymbol{D}_\tau \triangleq \tilde{\boldsymbol{D}}_\alpha \boldsymbol{\Sigma}^T (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)^T$, $\tilde{\boldsymbol{D}}_\alpha$ is a row-sparse matrix, i.e. it has only a few nonzero rows, and the time delays can be estimated from the indices of its nonzero rows. Note that $\tilde{\boldsymbol{D}}_\alpha$ and $\boldsymbol{D}_\tau$ share the same row sparsity pattern. Therefore the problem reduces to estimating $\boldsymbol{D}_\tau$, which is a multiple measurement vector (MMV) compressed sensing problem that can be solve by the simultaneous-OMP (S-OMP) method [32], an extension of the OMP to the MMV framework. Similarly, we can estimate the AoAs and the AoDs from $\boldsymbol{Y}_{(2)}$ and $\boldsymbol{Y}_{(1)}$, respectively.

VII. Computational Complexity Analysis

We analyze the computational complexity of the proposed CP decomposition-based method and the compressed sensing method discussed in the previous section. The major computational task of our proposed method involves solving the three least squares problems (19)–(21) at each iteration. Considering the calculation of $\hat{\boldsymbol{A}}$, we have

$$\hat{\boldsymbol{A}}^{(t+1)} = \boldsymbol{Y}_{(1)}\boldsymbol{V}^*(\boldsymbol{V}^T\boldsymbol{V}^*)^{-1} \tag{42}$$

where $\boldsymbol{V} \triangleq \hat{\boldsymbol{C}}^{(t)} \odot \hat{\boldsymbol{B}}^{(t)} \in \mathbb{C}^{TK \times L}$ is a tall matrix since we usually have $TK > L$. Noting that $\boldsymbol{Y}_{(1)} \in M \times TK$, the number of flops required to compute $\hat{\boldsymbol{A}}^{(t+1)}$ is of order $\mathcal{O}(MTKL + MKL^2 + L^3)$. When L is small, the dominant term has a computational complexity of order $\mathcal{O}(MTK)$, which scales linearly with the size of the observed tensor $\boldsymbol{\mathcal{Y}}$. It can also be shown that solving the other two least squares problems requires flops of order $\mathcal{O}(MTK)$ as well.

Now consider the computational complexity of the compressed sensing method. The joint compressed sensing method involves finding a sparse solution to the linear equation (40). It can be easily verified that to solve (40), the computational complexity of the OMP is of order $\mathcal{O}(MTK + N_1N_2N_3)$. For the FISTA [49], the main computational task at each iteration is to evaluate the proximal operator whose computational complexity is of the order $\mathcal{O}(n^2)$, where n denotes the number of columns of the overcomplete dictionary. For the problem (40), we have $n = N_1N_2N_3$. Thus the required number of flops at each iteration is of order $\mathcal{O}(N_1^2N_2^2N_3^2)$, which scales quadratically with $N_1N_2N_3$. In order to achieve a substantial overhead reduction, the parameters $\{M, T, K\}$ are usually chosen such that the number of measurements is far less than the dimension of the sparse signal, i.e., $MTK \ll N_1N_2N_3$. Therefore this joint compressed sensing method has a higher computational complexity than our proposed method.

For the separate compressed sensing scheme, it involves solving (41) to obtain an estimate of time delays. Since $\boldsymbol{Y}_{(3)} \in \mathbb{C}^{K \times MT}$ and $\bar{\boldsymbol{G}} \in \mathbb{C}^{K \times N_3}$, the number of flops required to solve (41) using S-OMP is of order $\mathcal{O}(MTKN_3)$. Similarly, we can arrive at the numbers of flops required to estimate AoAs and AoDs are of orders $\mathcal{O}(MTKN_1)$ and $\mathcal{O}(MTKN_2)$, respectively. Thus the total computational complexity for S-OMP is of order $\mathcal{O}(MTK(N_1 + N_2 + N_3))$. Therefore the separate compressed sensing scheme which uses S-OMP has a computational complexity similar to our proposed method.

VIII. Simulation Results

We present simulation results to illustrate the performance of our proposed CP decomposition-based method (referred to as CP). We consider a scenario where the BS employs a uniform linear array with $N_{\mathrm{BS}} = 64$ antennas and the MS employs a uniform linear array with $N_{\mathrm{MS}} = 32$ antennas. The distance between neighboring antenna elements is assumed to be half the wavelength of the signal. In our simulations, the mmWave channel is generated according to the wideband geometric channel model, in which the AoAs and AoDs are randomly distributed in $[0, 2\pi]$, the number of paths is set equal to $L = 4$, the delay spread τ_l for each path is uniformly distributed between 0 and 100 nanoseconds, and the complex gain α_l is a random variable following a circularly-symmetric Gaussian distribution $\alpha_{u,l} \sim \mathcal{CN}(0, 1/\rho)$. Here ρ is given by $\rho = (4\pi D f_c/c)^2$, where c represents the speed of light, D denotes the distance between the MS and the BS, and f_c is the carrier frequency. We set $D = 30$m, $f_c = 28$GHz. The total number of subcarriers is set to $\bar{K} = 128$, out of which K subcarriers are selected for training. The sampling rate is set to $f_s = 0.32$GHz. Also, in our experiments, the

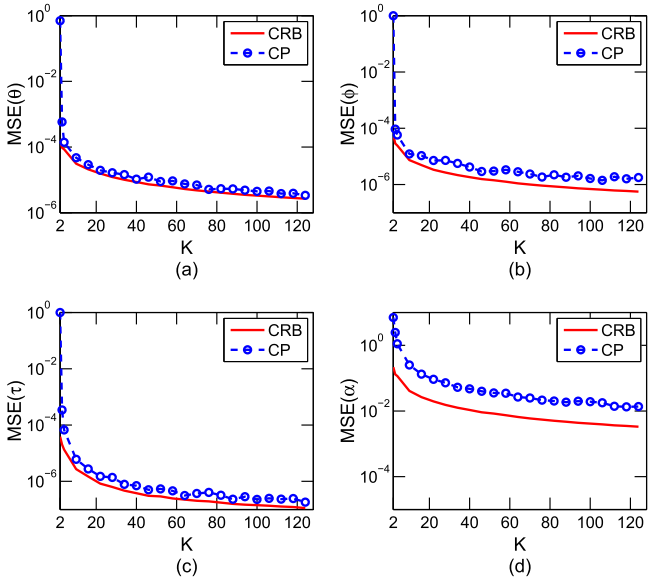


Fig. 3. MSEs and CRBs associated with different sets of parameters vs. the number of subcarriers, K.

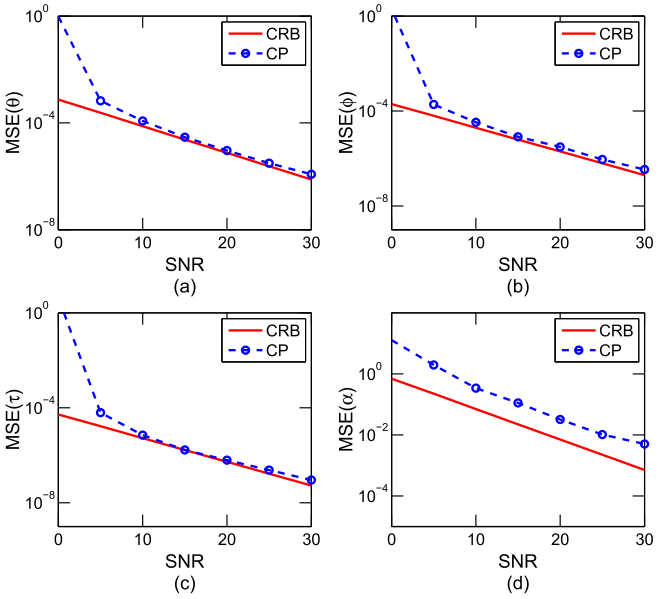


Fig. 4. MSEs and CRBs associated with different sets of parameters vs. SNR.

beamforming matrix $\boldsymbol{P}$ and the combining matrix $\boldsymbol{Q}$ are randomly generated with their entries uniformly chosen from a unit circle. We assume that the number of paths, L, is known *a priori* in our simulations because the OMP algorithm used for comparison requires the knowledge of L to decide when to terminate the greedy process. As discussed in Section IV, our proposed method can also deal with the case where L is unknown. The signal-to-noise ratio (SNR) is defined as the ratio of the signal component to the noise component, i.e.

$$\text{SNR} \triangleq \frac{\|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{W}}\|_F^2}{\|\boldsymbol{\mathcal{W}}\|_F^2} \tag{43}$$

where $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{W}}$ represent the received signal and the additive noise in (13), respectively. In our simulations, we assume a reasonably large SNR (0dB-30dB) at the receiver. Due to the significant path loss, the SNR could drop below 0dB, particularly if the distance between the MS and the BS, D, is large. To cope with this issue, the beamforming matrix $\boldsymbol{P}$ and the combining matrix $\boldsymbol{Q}$ can be devised to form directional beams to compensate for the path loss.

We first examine the estimation accuracy of the channel parameters $\{\theta_l, \phi_l, \tau_l, \alpha_l\}$. Mean square errors (MSEs) are calculated separately for each set of parameters, i.e.

$$\text{MSE}(\theta) = \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2 \quad \text{MSE}(\phi) = \|\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}\|_2^2$$
$$\text{MSE}(\tau) = \|\boldsymbol{\tau} - \hat{\boldsymbol{\tau}}\|_2^2 \quad \text{MSE}(\alpha) = \|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\|_2^2$$

where $\boldsymbol{\theta} \triangleq [\theta_1 \ \ldots \ \theta_L]^T$, $\boldsymbol{\phi} \triangleq [\phi_1 \ \ldots \ \phi_L]^T$, $\boldsymbol{\tau} \triangleq [\tau_1 \ \ldots \ \tau_L]^T$, and $\boldsymbol{\alpha} \triangleq [\alpha_1 \ \ldots \ \alpha_L]^T$. Fig. 3 plots the MSEs of our proposed method as a function of the number of subcarriers used for training, K, where we set $M = 6$, $T = 6$, and SNR $= 10$dB. The CRB results for different sets of parameters are also included for comparison. We see that our proposed method yields accurate estimates of the channel parameters even for small values of M, T, and K. This result indicates that our proposed method is able to achieve a substantial training overhead reduction. We also notice that the MSEs attained by our proposed method are very close to their corresponding CRBs, particularly for the AoA, AoD, and the time delay parameters. This result corroborates the optimality of the proposed method. As indicated earlier, the optimality of the proposed method comes from the fact that the ASL and the correlation-based scheme used in our proposed method are all maximum likelihood estimators under mild conditions. Specifically, it has been shown in [48] that the ALS yields maximum likelihood estimates and achieves its associated CRB in the presence of zero-mean i.i.d. Gaussian noise. We also proved that the simple correlation-based scheme employed in the second stage of our proposed method is a maximum likelihood estimator, provided that the estimator errors of the factor matrices are i.i.d. Gaussian random variables. Although the estimator errors may not strictly follow an i.i.d. Gaussian distribution, the correlation-based scheme is still an effective estimator that achieves near-optimality. Lastly, we observe that our proposed method fails when the number of subcarriers $K \leq 2$. This is because, for the case where $T = 6 > L$ and $M = 6 > L$, Kruskal's condition is satisfied only if $K \geq 2$. Thus our result roughly coincides with our previous analysis regarding the uniqueness of the CP decomposition. Fig. 4 depicts the MSEs and CRBs vs. SNR, where we set $T = 6$, $M = 6$, and $K = 6$. From Fig. 4, we see that the CRBs decrease exponentially with increasing SNR, and the estimation accuracy achieved by our proposed method has similar tendency as the CRBs. The MSEs attained by our proposed method, again, are close to their corresponding CRBs, except in the low SNR regime.

We now examine the channel estimation performance of our proposed method and its comparison with the compressed sensing method discussed in Section VI. The joint compressed sensing scheme is considered and an orthogonal matching pursuit (OMP) algorithm is employed to solve the sparse signal recovery problem (40). The separate estimation scheme (41) yields performance worse than (40), and thus is not included. Note that the dimension of the signal to be recovered in (40) is equal to $N_1 N_2 N_3$, where N_1, N_2, and N_3 denote the number

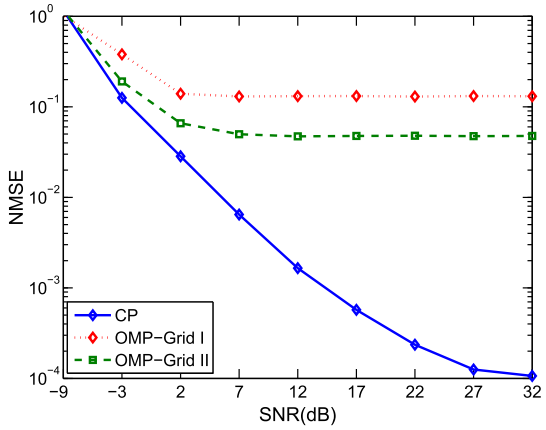


Fig. 5. NMSEs of respective algorithms vs. SNR, $M = 6$, $T = 6$, $K = 6$.

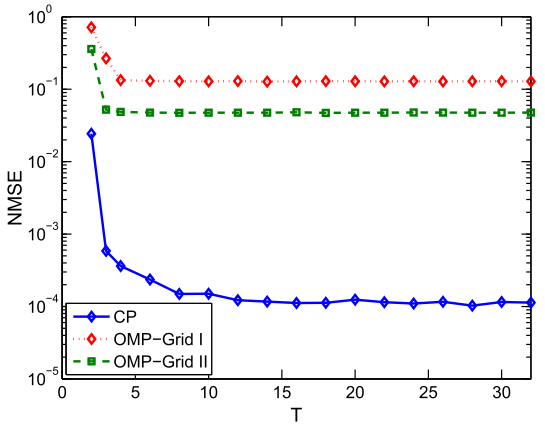


Fig. 7. NMSEs of respective algorithms vs. the number of frames T, $M = 6$, $K = 6$.

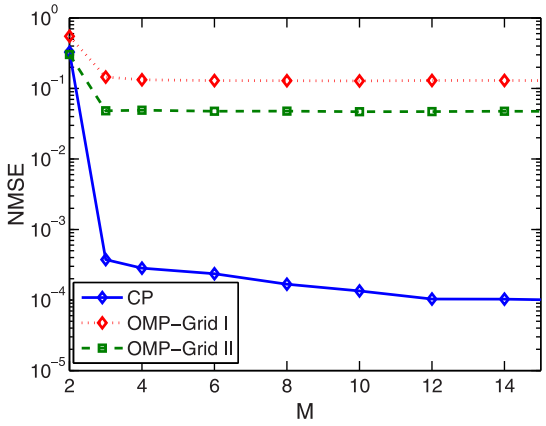


Fig. 6. NMSEs of respective algorithms vs. the number of sub-frames M, $T = 6$, $K = 6$.

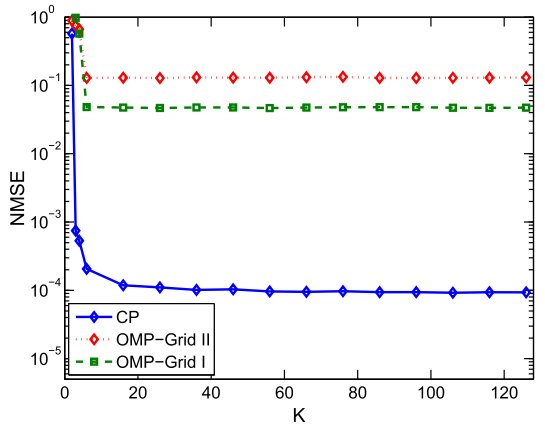


Fig. 8. NMSEs of respective algorithms vs. the number of subcarriers for training K, $M = 6$, $T = 6$.

of grid points used to discretize the AoA, AoD, and time delay domain, respectively. For a typical choice of $N_1 = 32$, $N_2 = 64$ and $N_3 = 32$, the dimension of the signal is of order $O(10^4)$. In this case, more sophisticated sparse recovery algorithms such as the fast iterative shrinkage-thresholding algorithm (FISTA) have a prohibitive computational complexity and thus are not included. Also, for the OMP, we employ two different grids to discretize the continuous parameter space: the first grid (referred to as Grid-I) discretizes the AoA-AoD-time delay space into $64 \times 128 \times 256$ grid points, and the second grid (referred to as Grid-II) discretizes the AoA-AoD-time delay space into $128 \times 256 \times 512$ grid points.

In Fig. 5, we show the NMSE results for our proposed method and the OMP algorithm as a function of SNR, where we set $M = 6$, $T = 6$, and $K = 6$. Here the NMSE is calculated as

$$\text{NMSE} = \frac{\sum_{k=1}^{K} \|\boldsymbol{H}_k - \hat{\boldsymbol{H}}_k\|_F^2}{\sum_{k=1}^{K} \|\boldsymbol{H}_k\|_F^2} \tag{44}$$

where $\boldsymbol{H}_k$ denotes the frequency-domain channel matrix associated with the kth subcarrier, and $\hat{\boldsymbol{H}}_k$ is its estimate. We see that our proposed method achieves a substantial performance improvement over the compressed sensing algorithm. The performance gain is primarily due to the following two reasons. First, unlike compressed sensing techniques, our proposed CP decomposition-based method is essentially a gridless approach which is free from grid discretization errors. Second, the CP decomposition-based method captures the intrinsic multi-dimensional structure of the multiway data, which helps lead to a performance improvement. From Fig. 6 to Fig. 8, we plot the NMSEs of respective methods vs. M, T, and K, respectively, where the SNR is set to 20dB. These results, again, demonstrate the superiority of the proposed method over the compressed sensing method. We also observe that these results corroborate our theoretical analysis concerning the uniqueness of the CP decomposition. For example, in Fig. 6, since we have $T = 6 > L$ and $K = 6 > L$, we only need $M \geq 2$ to satisfy Kruskal's condition. We see that our proposed method achieves an accurate channel estimate only when $M > 2$, which roughly coincides with our analysis.

Table I shows the average run times of our proposed method and the OMP method. To provide a glimpse of other more sophisticated compressed sensing method's computational complexity, the average run times of the FISTA are also included, from which we can see that sophisticated compressed sensing methods have a prohibitive computational complexity, and thus are not suitable for our channel estimation problem. We also see that our proposed method has a computational complexity as low as the OMP method. It takes similar run

TABLE I
AVERAGE RUN TIMES OF RESPECTIVE ALGORITHMS: $T = 6$, $M = 6$, $K = 6$, AND SNR = 20dB

ALG	Grid	NMSE	Average Run Time(s)
OMP	$64 \times 128 \times 256$	$2.3e-1$	0.2
	$160 \times 320 \times 640$	$3.5e-2$	0.8
FISTA	$32 \times 64 \times 128$	$2.3e-1$	$9e2$
	$64 \times 128 \times 256$	$9e-2$	$6e3$
CP	-	$1.2e-4$	0.2

times as the OMP method which employs the coarser grid of the two choices, meanwhile achieving a much higher estimation accuracy than the OMP method that uses the finer grid.

IX. CONCLUSIONS

We proposed a CP decomposition-based method for downlink channel estimation in mm-Wave MIMO-OFDM systems, where wideband mmWave channels with frequency selectivity were considered. The proposed method exploited the intrinsic multi-dimensional structure of the multiway data received at the BS. Specifically, the received signal at the BS was expressed as a third-order tensor. We showed that the tensor has a form of a low-rank CP decomposition, and the channel parameters can be easily extracted from the decomposed factor matrices. The uniqueness of the CP decomposition was investigated, which revealed that the uniqueness of the CP decomposition can be guaranteed even with a small number of measurements. Thus the proposed method is able to achieve a substantial training overhead reduction. CRB results for channel parameters were also developed. We compared our proposed method with a compressed sensing-based channel estimation method. Simulation results showed that our proposed method presents a clear performance advantage over the compressed sensing method in terms of both estimation accuracy and computational complexity.

APPENDIX A

In (23), for the lth column, we have

$$\hat{\boldsymbol{a}}_l = \lambda_l \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) + \boldsymbol{e}_l \tag{45}$$

where θ_l and λ_l are unknown parameters. We assume $\boldsymbol{e}_l$ satisfies circularly symmetric complex Gaussian distribution with zero mean and covariance matrix $\epsilon^2 \boldsymbol{I}$. Thus the log-likelihood function is given by

$$\begin{aligned} L(\theta_l, \lambda_l) &= -M \ln(\pi \epsilon^2) - \frac{1}{\epsilon^2} \left\| \hat{\boldsymbol{a}}_l - \lambda_l \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|_F^2 \\ &\propto - \left\| \hat{\boldsymbol{a}}_l - \lambda_l \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|_2^2 \end{aligned}$$

Given a fixed θ_l, the optimal λ_l can be obtained by taking the partial derivative of the log-likelihood function with respect to λ_l and setting the partial derivative equal to zero, i.e.

$$\frac{\partial L(\theta_l, \lambda_l)}{\partial \lambda_l^*} = (\hat{\boldsymbol{a}}_l - \lambda_l \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l))^T \tilde{\boldsymbol{a}}_{\mathrm{MS}}^*(\theta_l) = 0$$

which leads to

$$\lambda_l^\star = \frac{\hat{\boldsymbol{a}}_l{}^T \tilde{\boldsymbol{a}}_{\mathrm{MS}}^*(\theta_l)}{\left\| \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2}$$

Note that (45) can be rewritten as

$$\left\| \hat{\boldsymbol{a}}_l \right\|^2 = \lambda_l \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) + \hat{\boldsymbol{a}}_l^H \boldsymbol{e}_l$$

Then the log-likelihood function becomes

$$L(\theta_l, \lambda_l) \propto - \left\| \left\| \hat{\boldsymbol{a}}_l \right\|^2 - \lambda_l \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2 \tag{46}$$

Substituting $\lambda_l^\star$ into the above log-likelihood function, we arrive at

$$L(\theta_l, \lambda_l^\star) \propto - \left\| \left\| \hat{\boldsymbol{a}}_l \right\|^2 - \frac{\left| \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right|^2}{\left\| \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2} \right\|^2 \tag{47}$$

$$\propto - \left\| 1 - \frac{\left| \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right|^2}{\left\| \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2 \left\| \hat{\boldsymbol{a}}_l \right\|^2} \right\|^2 \tag{48}$$

Due to the Cauchy-Schwarz inequality, we have

$$0 \le \frac{\left| \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right|^2}{\left\| \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2 \left\| \hat{\boldsymbol{a}}_l \right\|^2} \le 1$$

Therefore maximizing the log-likelihood with respect to θ_l is equivalent to

$$\theta_l^\star = \arg\max_{\theta_l} L(\theta_l, \lambda_l^\star) = \arg\max_{\theta_l} \frac{\left| \hat{\boldsymbol{a}}_l^H \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right|^2}{\left\| \hat{\boldsymbol{a}}_l \right\|^2 \left\| \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \right\|^2} \tag{49}$$

The proof is completed here.

APPENDIX B

For a uniform linear array, the steering vector $\boldsymbol{a}_{\mathrm{MS}}(\theta_i)$ can be written as

$$\boldsymbol{a}_{\mathrm{MS}}(\theta_i) \triangleq [1 \;\; e^{j(2\pi/\lambda)d\sin(\theta_i)} \;\; \cdots \;\; e^{j(N_{\mathrm{MS}}-1)(2\pi/\lambda)d\sin(\theta_i)}]^T$$

where λ is the signal wavelength, and d denotes the distance between neighboring antenna elements. We assume each entry of $\boldsymbol{Q} \in \mathbb{C}^{N_{\mathrm{MS}} \times M}$ is chosen uniformly from a unit circle scaled by a constant $1/N_{\mathrm{MS}}$, i.e. $q_{m,n} = (1/N_{\mathrm{MS}}) e^{j\vartheta_{m,n}}$, where $\vartheta_{m,n} \in [-\pi, \pi]$ follows a uniform distribution. Let $a_{m,i} \triangleq \boldsymbol{q}_m^T \boldsymbol{a}_{\mathrm{MS}}(\theta_i)$ denote the (m,i)th entry of $\boldsymbol{A}$, in which $\boldsymbol{q}_m$ denotes the mth column of $\boldsymbol{Q}$. It can be readily verified that $\mathbb{E}[a_{m,i}] = 0, \forall m, i$ and

$$\mathbb{E}[a_{m,i} a_{n,j}^*] = \begin{cases} 0 & m \ne n \\ \dfrac{1}{N_{\mathrm{MS}}^2} \boldsymbol{a}_{\mathrm{MS}}^H(\theta_j) \boldsymbol{a}_{\mathrm{MS}}(\theta_i) & m = n \end{cases} \tag{50}$$

When the number of antennas at the MS is sufficiently large, the steering vectors $\{\boldsymbol{a}_{\mathrm{MS}}(\theta_i)\}$ become mutually quasi-orthogonal, i.e. $(1/N_{\mathrm{MS}}) \boldsymbol{a}_{\mathrm{MS}}^H(\theta_j) \boldsymbol{a}_{\mathrm{MS}}(\theta_j) \to \delta(\theta_i - \theta_j)$, which implies that the entries of $\boldsymbol{A}$ are uncorrelated with each other. On the other hand, according to the central limit theorem, we know that each entry $a_{m,i}$ approximately follows

a Gaussian distribution. Therefore entries of $\boldsymbol{A}$ can be considered as i.i.d. Gaussian variables with zero mean and variance $1/N_{\mathrm{MS}}$. Thus we can reach that the k-rank of $\boldsymbol{A}$ is equivalent to the number of columns or the number of rows, whichever is smaller, with probability one.

APPENDIX C
DERIVATION OF CRAMÉR RAO LOWER BOUND

Consider the $M \times T \times K$ observation tensor $\boldsymbol{\mathcal{Y}}$ in (13)

$$\boldsymbol{\mathcal{Y}} = \sum_{l=1}^{L} \alpha_l \tilde{\boldsymbol{a}}_{\mathrm{MS}}(\theta_l) \circ \tilde{\boldsymbol{a}}_{\mathrm{BS}}(\phi_l) \circ \boldsymbol{g}(\tau_l) + \boldsymbol{\mathcal{W}} \tag{51}$$

where $\{\alpha_l, \theta_l, \phi_l, \tau_l\}$ are the unknown channel parameters to be estimated. We assume that entries of $\boldsymbol{\mathcal{W}}$ are i.i.d zero mean, circular symmetric Gaussian random variables with variance σ^2. For ease of exposition, let $\boldsymbol{\theta} \triangleq [\theta_1 \ \cdots \ \theta_L]^T$, $\boldsymbol{\phi} \triangleq [\phi_1 \ \cdots \ \phi_L]^T$, $\boldsymbol{\tau} \triangleq [\tau_1 \ \cdots \ \tau_L]^T$, $\boldsymbol{\alpha} \triangleq [\alpha_1 \ \cdots \ \alpha_L]^T$, and $\boldsymbol{p} \triangleq [\boldsymbol{\theta}^T \ \boldsymbol{\phi}^T \ \boldsymbol{\tau}^T \ \boldsymbol{\alpha}^T]$. Thus, the log-likelihood function of $\boldsymbol{p}$ can be expressed as

$$L(\boldsymbol{p}) = f(\boldsymbol{\mathcal{Y}}; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \tag{52}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, defined in (15), (16) and (17) respectively, are functions of the parameter vector $\boldsymbol{p}$, and $f(\boldsymbol{\mathcal{Y}}; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is given by

$$\begin{aligned} &f(\boldsymbol{\mathcal{Y}}; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \\ &\quad = -MTK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T \right\|_F^2 \\ &\quad = -MTK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \boldsymbol{Y}_{(2)}^T - (\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^T \right\|_F^2 \\ &\quad = -MTK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \boldsymbol{Y}_{(3)}^T - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^T \right\|_F^2 \end{aligned}$$

The complex Fisher information matrix (FIM) for $\boldsymbol{p}$ is given by [47], [48]

$$\boldsymbol{\Omega}(\boldsymbol{p}) = \mathbb{E}\left\{ \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{p}} \right)^H \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{p}} \right) \right\}. \tag{53}$$

In the next, to calculate $\boldsymbol{\Omega}(\boldsymbol{p})$, we first compute the partial derivative of $L(\boldsymbol{p})$ with respect to $\boldsymbol{p}$ and then calculate the expectation with respect to $p(\boldsymbol{\mathcal{Y}}; \boldsymbol{p})$.

A. Partial Derivative of $L(\boldsymbol{p})$ W.R.T $\boldsymbol{p}$

The partial derivative of $L(\boldsymbol{p})$ with respect to θ_l can be computed as

$$\frac{\partial L(\boldsymbol{p})}{\partial \theta_l} = \mathrm{tr}\left\{ \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{A}} \right)^T \frac{\partial \boldsymbol{A}}{\partial \theta_l} + \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{A}^*} \right)^T \frac{\partial \boldsymbol{A}^*}{\partial \theta_l} \right\} \tag{54}$$

where

$$\begin{aligned} \frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{A}} &= \frac{1}{\sigma^2}(\boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T)^H (\boldsymbol{C} \odot \boldsymbol{B}) \\ \frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{A}^*} &= \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{A}} \right)^* \\ \frac{\partial \boldsymbol{A}}{\partial \theta_l} &= \begin{bmatrix} \boldsymbol{0} & \cdots & \tilde{\boldsymbol{a}}_l & \cdots & \boldsymbol{0} \end{bmatrix} \end{aligned} \tag{55}$$

For a uniform linear array with the element spacing equal to half of the signal wavelength, we have $\tilde{\boldsymbol{a}}_l \triangleq j\boldsymbol{Q}^T \boldsymbol{D}_a \boldsymbol{a}_{\mathrm{MS}}(\theta_l)$, and

$$\boldsymbol{D}_a \triangleq \pi \cos(\theta_l) \mathrm{diag}(0, 1, \cdots, N_{\mathrm{MS}} - 1) \tag{56}$$

Therefore, we have

$$\begin{aligned} \frac{\partial L(\boldsymbol{p})}{\partial \theta_l} &= \boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{C} \odot \boldsymbol{B})^T (\boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T)^* \tilde{\boldsymbol{a}}_l \\ &\quad + \boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{C} \odot \boldsymbol{B})^H (\boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T) \tilde{\boldsymbol{a}}_l^* \\ &= 2\mathrm{Re}\{\boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{C} \odot \boldsymbol{B})^T (\boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T)^* \tilde{\boldsymbol{a}}_l\} \\ &= 2\mathrm{Re}\{\boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{C} \odot \boldsymbol{B})^T (\boldsymbol{Y}_{(1)}^T - (\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^T)^* \tilde{\boldsymbol{A}} \boldsymbol{e}_l\} \end{aligned} \tag{57}$$

where $\mathrm{Re}\{\cdot\}$ is an operator which takes the real part of a complex number, $\boldsymbol{e}_l \in \mathbb{C}^{L \times 1}$ is the canonical vector whose non-zero entry is indexed as l, and

$$\tilde{\boldsymbol{A}} \triangleq \begin{bmatrix} \tilde{\boldsymbol{a}}_1 & \tilde{\boldsymbol{a}}_2 & \cdots & \tilde{\boldsymbol{a}}_L \end{bmatrix} \tag{58}$$

Similarly, we can obtain the partial derivatives with respect to other parameters as follows

$$\begin{aligned} \frac{\partial L(\boldsymbol{p})}{\partial \phi_l} &= 2\mathrm{Re}\{\boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{C} \odot \boldsymbol{A})^T (\boldsymbol{Y}_{(2)}^T - (\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^T)^* \tilde{\boldsymbol{B}} \boldsymbol{e}_l\} \\ \frac{\partial L(\boldsymbol{p})}{\partial \tau_l} &= 2\mathrm{Re}\{\boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{B} \odot \boldsymbol{A})^T (\boldsymbol{Y}_{(3)}^T - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^T)^* \tilde{\boldsymbol{C}} \boldsymbol{e}_l\} \\ \frac{\partial L(\boldsymbol{p})}{\partial \alpha_l} &= \boldsymbol{e}_l^T \frac{1}{\sigma^2}(\boldsymbol{B} \odot \boldsymbol{A})^T (\boldsymbol{Y}_{(3)}^T - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^T)^* \boldsymbol{G} \boldsymbol{e}_l \end{aligned}$$

where

$$\tilde{\boldsymbol{B}} \triangleq \begin{bmatrix} \tilde{\boldsymbol{b}}_1 & \tilde{\boldsymbol{b}}_2 & \cdots & \tilde{\boldsymbol{b}}_L \end{bmatrix} \tag{59}$$

$$\tilde{\boldsymbol{C}} \triangleq \begin{bmatrix} \tilde{\boldsymbol{c}}_1 & \tilde{\boldsymbol{c}}_2 & \cdots & \tilde{\boldsymbol{c}}_L \end{bmatrix} \tag{60}$$

$$\boldsymbol{G} \triangleq \begin{bmatrix} \boldsymbol{g}_1 & \boldsymbol{g}_2 & \cdots & \boldsymbol{g}_L \end{bmatrix} \tag{61}$$

in which $\tilde{\boldsymbol{b}}_l \triangleq j\boldsymbol{P}^T \boldsymbol{D}_b \boldsymbol{a}_{\mathrm{BS}}(\phi_l)$, $\tilde{\boldsymbol{c}}_l \triangleq j\boldsymbol{D}_c \boldsymbol{c}_l$, and

$$\boldsymbol{D}_b \triangleq \pi \cos(\phi_l) \mathrm{diag}(0, 1, \cdots, N_{\mathrm{BS}} - 1) \tag{62}$$

$$\boldsymbol{D}_c \triangleq -2\pi \, \mathrm{diag}(0, f_s/\bar{K}, \cdots, (K-1)f_s/\bar{K}) \tag{63}$$

B. Calculation of Fisher Information Matrix

We first calculate the entries in the principal minors of $\boldsymbol{\Omega}(\boldsymbol{p})$. For instance, the (l_1, l_2)th entry of

$$\mathbb{E}\left\{ \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial L(\boldsymbol{p})}{\partial \boldsymbol{\theta}} \right) \right\}.$$

is given by

$$\begin{aligned} &\mathbb{E}\left\{ \left(\frac{\partial L(\boldsymbol{p})}{\partial \theta_{l_1}} \right)^* \left(\frac{\partial L(\boldsymbol{p})}{\partial \theta_{l_2}} \right) \right\} \\ &= 4\mathbb{E}\left[\mathrm{Re}\{\boldsymbol{e}_{l_1}^T \boldsymbol{N}^a \boldsymbol{e}_{l_1}\} \mathrm{Re}\{\boldsymbol{e}_{l_2}^T \boldsymbol{N}^a \boldsymbol{e}_{l_2}\} \right] \\ &= \mathbb{E}\left[\left(\boldsymbol{N}^a(l_1, l_1) + \boldsymbol{N}^a(l_1, l_1)^* \right) \left(\boldsymbol{N}^a(l_2, l_2) + \boldsymbol{N}^a(l_2, l_2)^* \right) \right] \end{aligned}$$

where $\boldsymbol{N}^a(l_1,l_1)$ stands for the (l_1,l_1)th entry of $\boldsymbol{N}^a \in \mathbb{C}^{L\times L}$ and

$$\begin{aligned}\boldsymbol{N}^a &\triangleq \frac{1}{\sigma^2}(\boldsymbol{C}\odot\boldsymbol{B})^T(\boldsymbol{Y}_{(1)}^T-(\boldsymbol{C}\odot\boldsymbol{B})\boldsymbol{A}^T)^*\tilde{\boldsymbol{A}}\\ &= \frac{1}{\sigma^2}(\boldsymbol{C}\odot\boldsymbol{B})^T(\boldsymbol{W}_{(1)}^T)^*\tilde{\boldsymbol{A}}.\end{aligned}$$

Letting $\boldsymbol{n}^a \triangleq \mathrm{vec}(\boldsymbol{N}^a)$, we have

$$\boldsymbol{n}^a = \left(\tilde{\boldsymbol{A}}^T\otimes(\boldsymbol{C}\odot\boldsymbol{B})^T\right)\mathrm{vec}(\boldsymbol{W}_{(1)}^H). \tag{64}$$

where $\boldsymbol{W}_{(1)}$ is the mode-1 unfolding of $\boldsymbol{\mathcal{W}}$, thus $\mathrm{vec}(\boldsymbol{W}_{(1)}^H)$ is a zero mean circularly symmetric complex Gaussian vector whose covariance matrix is given by $\sigma^2\boldsymbol{I}$. Since $\boldsymbol{n}^a$ is the linear transformation of $\mathrm{vec}(\boldsymbol{W}_{(1)}^H)$, $\boldsymbol{n}^a$ also follows a circularly symmetric complex Gaussian distribution. Its covariance matrix $\boldsymbol{C}_{\boldsymbol{n}^a} \in \mathbb{C}^{L^2\times L^2}$ and second-order moments $\boldsymbol{M}_{\boldsymbol{n}^a} \in \mathbb{C}^{L^2\times L^2}$ are respectively given by

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{n}^a} &= \mathbb{E}\left[(\boldsymbol{n}^a)(\boldsymbol{n}^a)^H\right]\\ &= \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{A}}^T\otimes(\boldsymbol{C}\odot\boldsymbol{B})^T\right)\left(\tilde{\boldsymbol{A}}^*\otimes(\boldsymbol{C}\odot\boldsymbol{B})^*\right)\\ &= \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{A}}^T\tilde{\boldsymbol{A}}^*\right)\otimes\left((\boldsymbol{C}\odot\boldsymbol{B})^T(\boldsymbol{C}\odot\boldsymbol{B})^*\right)\end{aligned} \tag{65}$$

and

$$\boldsymbol{M}_{\boldsymbol{n}^a} = \mathbb{E}\left[(\boldsymbol{n}^a)(\boldsymbol{n}^a)^T\right] = \boldsymbol{0} \tag{66}$$

Therefore, we have

$$\mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\theta_{l_1}}\right)^*\left(\frac{\partial L(\boldsymbol{p})}{\partial\theta_{l_2}}\right)\right\} = 2\mathrm{Re}\{\boldsymbol{C}_{\boldsymbol{n}^a}(m,n)\} \tag{67}$$

where $m \triangleq L(l_1-1)+l_1$ and $n \triangleq L(l_2-1)+l_2$. Similarly, we can arrive at

$$\begin{aligned}\mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\phi_{l_1}}\right)^*\left(\frac{\partial L(\boldsymbol{p})}{\partial\phi_{l_2}}\right)\right\} &= 2\mathrm{Re}\{\boldsymbol{C}_{\boldsymbol{n}^b}(m,n)\}\\ \mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\tau_{l_1}}\right)^*\left(\frac{\partial L(\boldsymbol{p})}{\partial\tau_{l_2}}\right)\right\} &= 2\mathrm{Re}\{\boldsymbol{C}_{\boldsymbol{n}^c}(m,n)\}\\ \mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\alpha_{l_1}}\right)^*\left(\frac{\partial L(\boldsymbol{p})}{\partial\alpha_{l_2}}\right)\right\} &= \boldsymbol{C}_{\tilde{\boldsymbol{n}}^c}(m,n)^*,\end{aligned}$$

in which

$$\boldsymbol{C}_{\boldsymbol{n}^b} \triangleq \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{B}}^T\tilde{\boldsymbol{B}}^*\right)\otimes\left((\boldsymbol{C}\odot\boldsymbol{A})^T(\boldsymbol{C}\odot\boldsymbol{A})^*\right) \tag{68}$$

$$\boldsymbol{C}_{\boldsymbol{n}^c} \triangleq \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{C}}^*\right)\otimes\left((\boldsymbol{B}\odot\boldsymbol{A})^T(\boldsymbol{B}\odot\boldsymbol{A})^*\right) \tag{69}$$

$$\boldsymbol{C}_{\tilde{\boldsymbol{n}}^c} \triangleq \frac{1}{\sigma^2}\left(\boldsymbol{G}^T\boldsymbol{G}^*\right)\otimes\left((\boldsymbol{B}\odot\boldsymbol{A})^T(\boldsymbol{B}\odot\boldsymbol{A})^*\right) \tag{70}$$

For the elements in the off-principal minors of $\boldsymbol{\Omega}(\boldsymbol{p})$, such as the (l_1,l_2)th entry of

$$\mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\boldsymbol{\theta}}\right)^H\left(\frac{\partial L(\boldsymbol{p})}{\partial\boldsymbol{\phi}}\right)\right\}.$$

is given by

$$\begin{aligned}&\mathbb{E}\left\{\left(\frac{\partial L(\boldsymbol{p})}{\partial\theta_{l_1}}\right)^*\left(\frac{\partial L(\boldsymbol{p})}{\partial\phi_{l_2}}\right)\right\}\\ &= 4\mathbb{E}\left[\mathrm{Re}\left\{\mathbf{e}_{l_1}^T\boldsymbol{N}^a\mathbf{e}_{l_1}\right\}\mathrm{Re}\{\mathbf{e}_{l_2}^T\boldsymbol{N}^b\boldsymbol{e}_{l_2}\}\right]\\ &= \mathbb{E}\left[(\boldsymbol{N}^a(l_1,l_1)+\boldsymbol{N}^a(l_1,l_1)^*)(\boldsymbol{N}^b(l_2,l_2)+\boldsymbol{N}^b(l_2,l_2)^*)\right]\\ &= 2\mathrm{Re}\left\{\boldsymbol{C}_{\boldsymbol{n}^a,\boldsymbol{n}^b}(m,n)\right\}\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{n}^a,\boldsymbol{n}^b} &\triangleq \mathbb{E}\left[(\boldsymbol{n}^a)(\boldsymbol{n}^b)^H\right]\\ &= \frac{1}{\sigma^4}(\tilde{\boldsymbol{A}}\otimes(\boldsymbol{C}\odot\boldsymbol{B}))^T\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_2}(\tilde{\boldsymbol{B}}^*\otimes(\boldsymbol{C}\odot\boldsymbol{A})^*)\end{aligned} \tag{71}$$

in which

$$\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_2} \triangleq \mathbb{E}\{\mathrm{vec}(\boldsymbol{W}_{(1)}^H)\mathrm{vec}(\boldsymbol{W}_{(2)}^T)^T\} \tag{72}$$

Similarly, we can obtain

$$\begin{aligned}\mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{p})}{\partial\theta_{l_1}}\right)^*\left(\frac{\partial f(\boldsymbol{p})}{\partial\tau_{l_2}}\right)\right\} &= 2\mathrm{Re}\{\boldsymbol{C}_{\boldsymbol{n}^a,\boldsymbol{n}^c}(m,n)\}\\ \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{p})}{\partial\theta_{l_1}}\right)^*\left(\frac{\partial f(\boldsymbol{p})}{\partial\alpha_{l_2}}\right)\right\} &= \boldsymbol{C}_{\boldsymbol{n}^a,\tilde{\boldsymbol{n}}^c}(m,n)^*\\ \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{p})}{\partial\phi_{l_1}}\right)^*\left(\frac{\partial f(\boldsymbol{p})}{\partial\tau_{l_2}}\right)\right\} &= 2\mathrm{Re}\{\boldsymbol{C}_{\boldsymbol{n}^b,\boldsymbol{n}^c}(m,n)\}\\ \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{p})}{\partial\phi_{l_1}}\right)^*\left(\frac{\partial f(\boldsymbol{p})}{\partial\alpha_{l_2}}\right)\right\} &= \boldsymbol{C}_{\boldsymbol{n}^b,\tilde{\boldsymbol{n}}^c}(m,n)^*\\ \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{p})}{\partial\tau_{l_1}}\right)^*\left(\frac{\partial f(\boldsymbol{p})}{\partial\alpha_{l_2}}\right)\right\} &= \boldsymbol{C}_{\boldsymbol{n}^c,\tilde{\boldsymbol{n}}^c}(m,n)^*\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{n}^a,\boldsymbol{n}^c} &\triangleq \frac{1}{\sigma^4}(\tilde{\boldsymbol{A}}\otimes(\boldsymbol{C}\odot\boldsymbol{B}))^T\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_3}(\tilde{\boldsymbol{C}}^*\otimes(\boldsymbol{B}\odot\boldsymbol{A})^*)\\ \boldsymbol{C}_{\boldsymbol{n}^a,\tilde{\boldsymbol{n}}^c} &\triangleq \frac{1}{\sigma^4}(\tilde{\boldsymbol{A}}\otimes(\boldsymbol{C}\odot\boldsymbol{B}))^T\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_3}(\boldsymbol{G}^*\otimes(\boldsymbol{B}\odot\boldsymbol{A})^*)\\ \boldsymbol{C}_{\boldsymbol{n}^b,\boldsymbol{n}^c} &\triangleq \frac{1}{\sigma^4}(\tilde{\boldsymbol{B}}\otimes(\boldsymbol{C}\odot\boldsymbol{A}))^T\boldsymbol{C}_{\boldsymbol{w}_2,\boldsymbol{w}_3}(\tilde{\boldsymbol{C}}^*\otimes(\boldsymbol{B}\odot\boldsymbol{A})^*)\\ \boldsymbol{C}_{\boldsymbol{n}^b,\tilde{\boldsymbol{n}}^c} &\triangleq \frac{1}{\sigma^4}(\tilde{\boldsymbol{B}}\otimes(\boldsymbol{C}\odot\boldsymbol{A}))^T\boldsymbol{C}_{\boldsymbol{w}_2,\boldsymbol{w}_3}(\boldsymbol{G}^*\otimes(\boldsymbol{B}\odot\boldsymbol{A})^*)\\ \boldsymbol{C}_{\boldsymbol{n}^c,\tilde{\boldsymbol{n}}^c} &\triangleq \frac{1}{\sigma^2}(\tilde{\boldsymbol{C}}\otimes(\boldsymbol{B}\odot\boldsymbol{A}))^T(\boldsymbol{G}^*\otimes(\boldsymbol{B}\odot\boldsymbol{A})^*)\end{aligned}$$

in which

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_3} &\triangleq \mathbb{E}\{\mathrm{vec}(\boldsymbol{W}_{(1)}^H)\mathrm{vec}(\boldsymbol{W}_{(3)}^T)^T\}\\ \boldsymbol{C}_{\boldsymbol{w}_2,\boldsymbol{w}_3} &\triangleq \mathbb{E}\{\mathrm{vec}(\boldsymbol{W}_{(2)}^H)\mathrm{vec}(\boldsymbol{W}_{(3)}^T)^T\}.\end{aligned}$$

The computation of $\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_2}$ is elaborated as follows. Note that the (m,t,k)th entry in $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{M\times T\times K}$ corresponds to the $(m,t+(k-1)T)$th entry of $\boldsymbol{W}_{(1)}$ and also corresponds to the $(t,m+(k-1)M)$th entry of $\boldsymbol{W}_{(2)}$. Furthermore, the $(m,t+(k-1)T)$th entry of $\boldsymbol{W}_{(1)}$ corresponds to the $(t+(k-1)T+(m-1)TK)$th entry of $\mathrm{vec}(\boldsymbol{W}_{(1)}^H)$ and the $(t,m+(k-1)M)$th entry of $\boldsymbol{W}_{(2)}$ corresponds to the

$(m + (k-1)M + (t-1)MK)$th entry of $\text{vec}(\boldsymbol{W}_{(2)}^T)$. Since entries in $\boldsymbol{\mathcal{W}}$ are i.i.d. random variables, i.e.,

$$\mathbb{E}\{w_{m_1,t_1,k_1} w^*_{m_2,t_2,k_2}\} = \begin{cases} \sigma^2; & m_1 = m_2, t_1 = t_2, k_1 = k_2 \\ 0; & \text{otherwise} \end{cases}$$

where $w_{m,t,k}$ represents the (m,t,k)th entry of $\boldsymbol{\mathcal{W}}$. Therefore, in $\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_2} \in \mathbb{C}^{TKM \times MKT}$, the number of nonzero entries is MTK and the corresponding indexes, $\{(n_1,n_2)|\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_2}(n_1,n_2) \neq 0\}$, is equal to

$$\{(n_1,n_2)|n_1 = t + (k-1)T + (m-1)TK, \\ n_2 = m + (k-1)M + (t-1)MK, \forall m,t,k\}$$

Similarly, the index of the nonzero elements in $\boldsymbol{C}_{\boldsymbol{w}_1,\boldsymbol{w}_3} \in \mathbb{C}^{TKM \times MTK}$ and $\boldsymbol{C}_{\boldsymbol{w}_2,\boldsymbol{w}_3} \in \mathbb{C}^{MKT \times MTK}$ are respectively belongs to

$$\{(n_1,n_2)|n_1 = t + (k-1)T + (m-1)TK, \\ n_2 = m + (t-1)M + (k-1)MT, \forall m,t,k\}$$

and

$$\{(n_1,n_2)|n_1 = m + (k-1)M + (t-1)MK, \\ n_2 = m + (t-1)M + (k-1)MT, \forall m,t,k\}$$

C. Cramér-Rao Bound

After obtaining the fisher information matrix, the CRB for the parameters $\boldsymbol{p}$ can be calculated as [47]

$$\text{CRB}(\boldsymbol{p}) = \boldsymbol{\Omega}^{-1}(\boldsymbol{p}). \tag{73}$$

REFERENCES

[1] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[4] G. Yang, M. Xiao, J. Gross, H. Al-Zubaidy, and Y. Huang. (Jul. 2016). "Delay and backlog analysis for 60 GHz wireless networks." [Online]. Available: https://arxiv.org/abs/1608.00120

[5] 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, Sep. 2014.

[6] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[7] Q. Xue, X. Fang, M. Xiao, and L. Yan, "Multi-user millimeter wave communications with nonorthogonal beams," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2016.2617083.

[8] G. Yang, J. Du, and M. Xiao, "Maximum throughput path selection with random blockage for indoor 60 GHz relay networks," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3511–3524, Oct. 2015.

[9] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[10] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[11] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[12] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.

[13] J. Zhang, L. Dai, X. Zhang, E. Björnson, and Z. Wang, "Achievable rate of Rician large-scale MIMO channels with transceiver hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8800–8806, Oct. 2016.

[14] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[15] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. 16th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun./Jul. 2015, pp. 146–150.

[16] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[17] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Globel Commun. Conf. (Globecom)*, Houston, TX, USA, Dec. 2011, pp. 1–6.

[18] A. Alkhateeb, G. Leus, and R. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. 40th IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[19] A. Alkhateeb, O. El 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.

[20] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 273–277.

[21] 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.

[22] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2012, pp. 234–239.

[23] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 690–697.

[24] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-Wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[25] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[26] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[27] X. Gao, L. Dai, and A. M. Sayeed. (Jul 2016). "Low RF-complexity technologies for 5G millimeter-wave MIMO systems with large antenna arrays." [Online]. Available: https://arxiv.org/abs/1607.04559

[28] L. Dai, X. Gao, X. Su, S. Han, C.-L. I, and Z. Wang, "Low-complexity soft-output signal detection based on Gauss–Seidel method for uplink multiuser large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4839–4845, Oct. 2015.

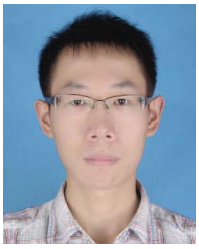
[29] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[30] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[31] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[32] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.

[33] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[34] T. G. Kolda, *Multilinear Operators for Higher-Order Decompositions*, United States Department of Energy, Washington, DC, USA, 2006.
[35] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
[36] L. D. Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM. J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, Sep. 2008.
[37] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.
[38] W. Feng, Y. Wang, N. Ge, J. Lu, and J. Zhang, "Virtual MIMO in multi-cell distributed antenna systems: Coordinated transmissions with large-scale CSIT," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2067–2081, Oct. 2013.
[39] L. Zhang, M. Xiao, G. Wu, S. Li, and Y.-C. Liang, "Energy-efficient cognitive transmission with imperfect spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1320–1335, May 2016.
[40] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[41] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Rank regularization and Bayesian inference for tensor completion and extrapolation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5689–5703, Nov. 2013.
[42] P. Rai, Y. Wang, S. Guo, G. Chen, D. Dunson, and L. Carin, "Scalable Bayesian low-rank decomposition of incomplete multiway tensors," in *Proc. 31st Int. Conf. Mach. Learn. (ICML-14)*, Beijing, China, 2014, pp. 1800–1808.
[43] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1751–1763, Sep. 2015.
[44] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.
[45] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the candecomp/parafac decomposition," *Linear Algebra Appl.*, vol. 420, nos. 2–3, pp. 540–552, Jan. 2007.
[46] R. A. Harshman, "Determination and proof of minimum uniqueness conditions for PARAFAC1," UCLA Working Papers in Phonetics, vol. 22, pp. 111–117, (University Microfilms, Ann Arbor, No. 10,085), 1972.
[47] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[48] X. Liu and N. D. Sidiropoulos, "Cramer–Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
[49] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.



Zhou Zhou received the B.Sc. degree from the University of Electronic Science and Technology of China in 2011, where he is currently pursuing the Ph.D. degree. His current research interests include compressed sensing, sparse theory, and tensor signal processing.

Jun Fang (M'08) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

In 2006, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic Science and Technology of China. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference.

Dr. Fang received the IEEE Jack Neubauer Memorial Award in 2013 for the Best Systems Paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is an Associate Technical Editor of the *IEEE Communications Magazine* and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Linxiao Yang received the B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2013. He is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China. His current research interests include compressed sensing, sparse theory, tensor analysis, and machine learning.

Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. He was a summer Visiting Faculty Member with the Air Force Research Laboratory in 2003, 2004, and 2009. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li was a member of the IEEE SPS Sensor Array and Multichannel Technical Committee (TC) from 2006 to 2012. He has been a member of the IEEE SPS Signal Processing Theory and Methods TC since 2011. He is a member of Tau Beta Pi and Phi Kappa Phi. He received the IEEE Jack Neubauer Memorial Award in 2013 for the Best Systems Paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003, the Jess H. Davis Memorial Award for excellence in research in 2001 from the Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been involved in various conference organization activities, including serving as a General Co-Chair of the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, in 2012. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2009, the IEEE SIGNAL PROCESSING LETTERS from 2005 to 2006, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2003 to 2006. He has been an Associate Editor of *Signal Processing* (Elsevier) since 2013 and the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2014, and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *EURASIP Journal on Applied Signal Processing*.

Zhi Chen received the B.Eng., M.Eng., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC) in 1997, 2000, and 2006, respectively. In 2006, he joined the National Key Laboratory on Communications, UESTC, where he is currently a Professor. He was a Visiting Scholar with the University of California at Riverside, Riverside, USA, from 2010 to 2011. His current research interests include relay and cooperative communications, multiuser beamforming in cellular networks, interference coordination and cancellation, and terahertz communication. He has served as a Reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Rick S. Blum (F'05) received the B.S. degree from Pennsylvania State University in 1984 and the M.S. and Ph.D. degrees from the University of Pennsylvania in 1987 and 1991, respectively, all in electrical engineering. From 1984 to 1991, he was a member of Technical Staff, General Electric Aerospace, Valley Forge, PA, USA, and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA, USA, where he is currently a Professor and holds the Robert W. Wieseman Chaired Research Professorship in electrical engineering. His research interests include signal processing for smart grid, communications, sensor networking, radar, and sensor processing. He was a member of the Signal Processing for Communications Technical Committee (TC) of the IEEE Signal Processing Society. He is a member of the SAM TC of the IEEE Signal Processing Society and the Communications Theory TC of the IEEE Communication Society. He was on the Awards Committee of the IEEE Communication Society. He was an Associate Editor of the *IEEE Transactions on Signal Processing* and the IEEE COMMUNICATIONS LETTERS. He has edited special issues for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is on the Editorial Board of the *Journal of Advances in Information Fusion* of the International Society of Information Fusion.

Dr. Blum holds several patents. He was an IEEE Signal Processing Society Distinguished Lecturer. He is a member of Eta Kappa Nu and Sigma Xi. He received the ONR Young Investigator Award and the NSF Research Initiation Award. He received the IEEE Third Millennium Medal. His IEEE Fellow Citation for scientific contributions to detection, data fusion and signal processing with multiple sensors acknowledges contributions to the field of sensor networking.