Channel Estimation for Millimeter-Wave Multiuser MIMO Systems via PARAFAC Decomposition

Zhou Zhou, Jun Fang, Linxiao Yang, Hongbin Li, Zhi Chen, and Shaoqian Li, *Fellow, IEEE*

***Abstract*— We consider the problem of uplink channel estimation for millimeter wave (mmWave) systems, where the base station (BS) and mobile stations (MSs) are equipped with large antenna arrays to provide sufficient beamforming gain for outdoor wireless communications. Hybrid analog and digital beamforming structures are employed by both the BS and the MS due to hardware constraints. We propose a layered pilot transmission scheme and a CANDECOMP/PARAFAC (CP) decomposition-based method for joint estimation of the channels from multiple users (i.e., MSs) to the BS. The proposed method exploits the intrinsic low-rank structure of the multiway data collected from multiple modes, where the low-rank structure is a result of the sparse scattering nature of the mmWave channel. The uniqueness of the CP decomposition is studied, and the sufficient conditions for essential uniqueness are obtained. The conditions shed light on the design of the beamforming matrix, the combining matrix, and the pilot sequences, and meanwhile provide general guidelines for choosing system parameters. Our analysis reveals that our proposed method can achieve a substantial training overhead reduction by leveraging the low-rank structure of the received signal. Simulation results show that the proposed method presents a clear advantage over a compressed sensing-based method in terms of both estimation accuracy and computational complexity.**

***Index Terms*— Mm-Wave systems, channel estimation, CANDECOMP/PARAFAC (CP) decomposition, compressed sensing.**

I. Introduction

MILLIMETER-WAVE (mmWave) communication is a promising technology for future 5G cellular networks [1], [2]. It has the potential to offer gigabit-per-second data rates by exploiting the large bandwidth available at mmWave frequencies. However, communication at such high frequencies also suffers from high attenuation and signal absorption [3]. To compensate for the significant path loss, very large antenna arrays can be used at the base station (BS) and the mobile station (BS) to exploit beam steering to increase the link gain [4]. Due to the small wavelength at the mmWave frequencies, the antenna size is very small and a large number of array elements can be packed into a small area. Directional precoding/beamforming with large antenna arrays is essential for providing sufficient beamforming gain for mmWave communications [5]–[7]. On the other hand, the design of the precoding matrix requires complete channel state information. Reliable mmWave channel estimation, however, is challenging due to the large number of antennas and the low signal-to-noise ratio (SNR) before beamforming. The problem becomes exacerbated when considering multi-user MIMO systems. Multi-user MIMO operation was advocated in [8] which considers a single-cell time-division duplex (TDD) scenario. The time-slot over which the channel can be assumed constant is divided between uplink pilot transmission and downlink data transmission. The BS, through channel reciprocity, obtains an estimate of the downlink channel, and then generates a linear precoder for transmitting data to multiple terminals simultaneously. The time required for pilots, in this case, increases linearly with the number of terminals served. Moreover, due to higher Doppler spread, the coherence time for the mmWave bands could be much shorter than that for cellular frequencies below 6GHz. Therefore training overhead reduction is crucial to support multi-user MIMO operation in mmWave systems.

The sparse scattering nature of the mm-Wave channel can be utilized to reduce the training overhead for channel estimation [9]–[15]. Specifically, it was shown (e.g. [9], [10]) that compressed sensing-based methods achieve a significant training overhead reduction via leveraging the poor scattering nature of mmWave channels. In [16], a hierarchical multi-resolution beamforming codebook was first employed for training. Based on this hierarchical codebook, an adaptive compressed sensing method was proposed in [10] for channel estimation. The main idea of adaptive compressed sensing-based channel estimation method is to divide the training process into a number of stages, with the training precoding used at each stage determined by the output of earlier stages. Compared to the standard compressed sensing method, the adaptive method is more efficient and yields better performance at low signal-to-noise ratio (SNR). Nevertheless, this performance improvement requires a feedback channel from the MS to the BS, which may not be available before the communication between the BS and the MS is established.

Manuscript received March 6, 2016; revised August 22, 2016 and July 1, 2016; accepted August 25, 2016. Date of publication August 30, 2016; date of current version November 9, 2016. This work was supported in part by the National Science Foundation of China under Grant 61428103, Grant 61522104, and Grant 91438118, in part by the National Science Foundation under Grant ECCS-1408182 and Grant ECCS-1609393, and in part by the Air Force Office of Scientific Research under Grant FA9550-16-1-0243. The associate editor coordinating the review of this paper and approving it for publication was A. Zajic.

Z. Zhou, J. Fang, L. Yang, Z. Chen, and S. Li 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).

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/TWC.2016.2604259

In [13]–[15], a spatial angle refinement technique was proposed to improve the channel estimation accuracy. Channel estimation and precoding design for mmWave communications were also considered in [11], where aperture shaping was used to ensure a sparse virtual-domain MIMO channel representation.

In this paper, we consider the problem of multi-user uplink mmWave channel estimation. Such a problem arises in multi-user massive MIMO systems [17], [18] where the BS simultaneously serves a number of independent users sharing the same time-frequency bandwidth, and thus requires to acquire the channel state information of multiple users via uplink pilots (channel reciprocity is assumed). To jointly estimate channels from multiple users to the BS, we propose a layered pilot transmission scheme in which the training phase consists of a number of frames and each frame is divided into a number of sub-frames. In each sub-frame, users employ a common beamforming vector to simultaneously transmit their respective pilot symbols. With this layered transmission scheme, the received signal at the BS can be organized to construct a third-order tensor. We show that the third-order tensor can be expressed as a CANDECOMP/PARAFAC (CP) decomposition which decomposes the tensor into a linear combination of a number of rank-one tensors. CP decomposition, also referred to as tensor rank decomposition, may be regarded as a generalization of the matrix singular value decomposition to tensors [19]. As will be shown in our paper, the decomposition allows the channels to be estimated from the decomposed factor matrices. We analyze the uniqueness of the CP decomposition. Our analysis shows that our proposed method, by exploiting the tensor low-rank structure resulting from the sparse scattering nature of the channel, can achieve a significant training overhead reduction.

Compressed-sensing based techniques can be employed to solve the joint uplink channel estimation problem, and one such method is briefly discussed in this paper. As compared with compressed sensing-based techniques, our proposed CP decomposition-based method has the following advantages. Firstly, due to the use of tensors for data representation and processing, our proposed method is computationally more efficient than compressed sensing techniques. In particular, CP decomposition allows the multi-user channel estimation to be decomposed into a number of single-user channel estimation problems. Thus the dimension of each problem, along with the computational complexity, is significantly reduced. In contrast, compressed sensing techniques formulate the multi-user channel estimation problem as a single sparse signal recovery problem which involves very high dimensional matrix-vector multiplication operations. Secondly, it is noted that CP decomposition enjoys a nice uniqueness theoretical guarantee under simple and mild conditions. These conditions are easy to analyze, and can be employed to determine the exact amount of training overhead required for unique decomposition. In addition, these conditions provide useful guidelines for beamforming/combining matrices design and system parameter selection. While for compressed sensing techniques, it is usually difficult to analyze and check the exact recovery condition for generic dictionaries.

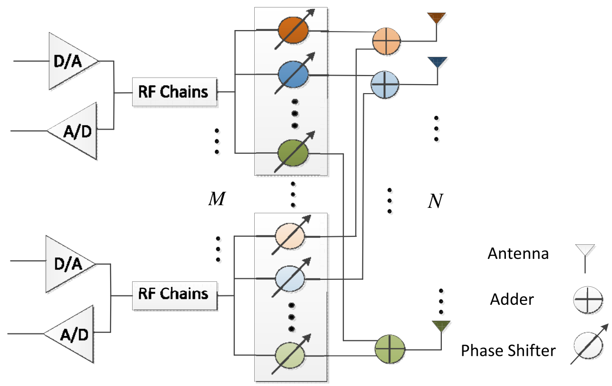


Fig. 1. The hybrid precoding structure for the base station and the mobile station.

We note that mutlilinear tensor algebra, as a powerful tool, has been widely used in a variety of applications in signal processing and wireless communications, such as multiuser detection in direct-sequence code-division multiple access (DS-CDMA) [20], blind spatial signature estimation [21], two-way relaying MIMO communications [22], etc. In particular, the uniqueness of CP decomposition has proven useful in solving many array processing problems from the multiple invariance sensor array processing [23] to the detection and localization of multiple targets in MIMO radar [24]. Another important application is the multidimensional harmonic retrieval, where significant improvements of parameter estimation accuracy can be achieved by using multilinear algebra [25]. Recent years have seen a resurgence of interest in tensor [19], motivated by a number of applications involving real-world multiway data.

The rest of the paper is organized as follows. In Section II, we introduce the system model and a layered pilot transmission scheme. Section III provides notations and basics on tensors. In Section IV, a tensor decomposition-based method is developed for jointly estimating the channels from multiple users to the BS. The uniqueness of the CP decomposition is studied and sufficient conditions for the uniqueness of the CP decomposition are derived in Section V. 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. System Model and Problem Formulation

Consider a mmWave system consisting of a base station (BS) and U mobile stations (MSs). We assume that hybrid analog and digital beamforming structures (Fig. 1) are employed by both the BS and the MS. The BS is equipped with N_{BS} antennas and M_{BS} RF chains, and each MS is equipped with N_{MS} antennas and M_{MS} RF chains. Since the RF chain is expensive and power consuming, the number of RF chains is usually less than the number of antennas, i.e. $M_{\text{BS}} < N_{\text{BS}}$ and $M_{\text{MS}} < N_{\text{MS}}$. We also assume $M_{\text{MS}} = 1$, i.e. each user only transmits one data stream.

In this paper, we consider the problem of estimating the uplink mmWave channels from users to the BS. MmWave channels are expected to have very limited scattering. Measurement campaigns in dense-urban NLOS environments reveals that mmWave channels typically exhibit only 3-4 scattering clusters, with relatively little delay/angle spreading within each cluster [26]. Following [10], we assume a geometric channel model with L_u scatterers between the uth user and the BS. Under this model, the channel $\boldsymbol{H}_u \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{MS}}}$ from the uth user to the BS can be expressed as

$$\boldsymbol{H}_u = \sum_{l=1}^{L_u} \alpha_{u,l} \boldsymbol{a}_{\text{BS}}(\theta_{u,l}) \boldsymbol{a}_{\text{MS}}^T(\phi_{u,l}) \tag{1}$$

where $\alpha_{u,l}$ is the complex path gain associated with the lth path of the uth user, $\theta_{u,l} \in [0, 2\pi]$ and $\phi_{u,l} \in [0, 2\pi]$ are the associated azimuth angle of arrival (AoA) and azimuth angle of departure (AoD), respectively, $\boldsymbol{a}_{\text{BS}}(\theta_{u,l}) \in \mathbb{C}^{N_{\text{BS}}}$ and $\boldsymbol{a}_{\text{MS}}(\phi_{u,l}) \in \mathbb{C}^{N_{\text{MS}}}$ denote the antenna array response vectors associated with the BS and the MS, respectively. In this paper, for simplicity, a uniform linear array is assumed, though its extension to arbitrary antenna arrays is possible. The steering vectors at the BS and the MS can thus be written as follows respectively

$$\boldsymbol{a}_{\text{BS}}(\theta_{u,l}) \triangleq \frac{1}{\sqrt{N_{\text{BS}}}}[1 \;\; e^{j(2\pi/\lambda)d\sin(\theta_{u,l})} \;\; \dots \;\; e^{j(N_{\text{BS}}-1)(2\pi/\lambda)d\sin(\theta_{u,l})}]^T$$

$$\boldsymbol{a}_{\text{MS}}(\phi_{u,l}) \triangleq \frac{1}{\sqrt{N_{\text{MS}}}}[1 \;\; e^{j(2\pi/\lambda)d\sin(\phi_{u,l})} \;\; \dots \;\; e^{j(N_{\text{MS}}-1)(2\pi/\lambda)d\sin(\phi_{u,l})}]^T$$

where λ is the signal wavelength, and d denotes the distance between neighboring antenna elements.

Note that the problem of single-user mmWave channel estimation has been studied in [9] and [10]. Specifically, to estimate the downlink channel, the BS employs P different beamforming vectors at P successive time frames, and at each time frame, the MS uses Q combining vectors to detect the signal transmitted over each beamforming vector. By exploiting the sparse scattering nature of mmWave channels, the problem of estimating the mmWave channel can be formulated as a sparse signal recovery problem and the training overhead can be considerably reduced. The above method can also be used to solve our uplink channel estimation problem if channels from users to the BS are estimated separately. Nevertheless, we will show that a joint estimation (of multiusers' channels) scheme may lead to an additional training overhead reduction.

We first propose a layered pilot transmission scheme which is elaborated as follows. The training phase consists of T consecutive frames, and each frame is divided into T' sub-frames. In each sub-frame $t' = 1, \dots, T'$, users employ a common beamforming vector $\boldsymbol{p}_{t'} \in \mathbb{C}^{N_{\text{MS}}}$ to simultaneously transmit their respective pilot symbols $s_{u,t}$, where $s_{u,t}$ denotes the pilot symbol used by the uth user at the tth frame. Thus, at each frame, each user employs a total of T' beamforming vectors to transmit its pilot symbol. These T' beamforming vectors can be devised to achieve beam scanning in all directions. Note that each user can use an individual beamforming vector, instead of a common beamforming vector, at each sub-frame, but such a scheme does not necessarily lead to better performance. At the BS, the transmitted signal can be received simultaneously via M_{BS} RF chains associated with different receiving vectors $\{\boldsymbol{q}_m\}_{m=1}^{M_{\text{BS}}}$, where $\boldsymbol{q}_m \in \mathbb{C}^{N_{\text{BS}}}$. Therefore the signal received by the mth RF chain at the t'th sub-frame of the tth frame can be expressed as

$$y_{m,t',t} = \boldsymbol{q}_m^T \sum_{u=1}^{U} \boldsymbol{H}_u \boldsymbol{p}_{t'} s_{u,t} + w_{m,t',t} \tag{2}$$

where $w_{m,t',t}$ denotes the additive white Gaussian noise associated with the mth RF chain at the t'th sub-frame of the tth frame. We see that the proposed layered pilot transmission scheme requires synchronization among users, and assumes the same number of antennas at each user. It should be noted that in TDD multi-user MIMO systems, pilot synchronization among users is usually assumed (see, e.g. [27]). Synchronization among users may be relaxed if we use time multiplexed pilots and estimate each user's channel separately. This, however, comes at the expense of reduced spectral efficiency. In practice, synchronization may be achieved via time synchronization signals conveyed to all users through a broadcast channel.

Our objective is to estimate the channels $\{\boldsymbol{H}_u\}$ from the received signal $\{y_{m,t',t}\}$. We wish to achieve a reliable channel estimation by using as few measurements as possible. Particularly the number of pilot symbols T is assumed to be less than U, i.e. $T < U$, otherwise orthogonal pilots (such as time multiplexed pilots) can be employed and the joint channel estimation problem can be decomposed into a number of single-user channel estimation problems. In the following, we show that the received data can be represented as a tensor and such a representation allows a more efficient algorithm to extract the channel state information with minimum number of measurements. Before proceeding, we first provide a brief review of tensor and the CANDECOMP/PARAFAC (CP) decomposition.

III. Preliminaries

We first provide a brief review on tensor and the CP decomposition. 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 $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ denote an Nth-order tensor with its $(i_1, \dots, i_N)$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 higher-order analogue of matrix rows and columns. The mode-n fibers of $\mathcal{X}$ 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 $\mathcal{X}$, denoted as $\boldsymbol{X}_{(n)}$, arranges the mode-n fibers to be the columns of the resulting matrix. More precisely, the $(i_1, \dots, i_N)$th entry of $\mathcal{X}$ corresponds to

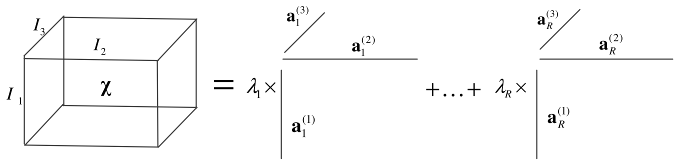


Fig. 2. Schematic of CP decomposition.

the (i_n, j)th entry of $\boldsymbol{X}_{(n)}$, where

$$j = 1 + \sum_{k=1, k\neq n}^{N} (i_k - 1) J_k \tag{3}$$

in which $J_k = \prod_{m=1, m\neq n}^{k-1} I_m$. The n-mode product of $\mathcal{X}$ with a matrix $\boldsymbol{A} \in \mathbb{C}^{J\times I_n}$ is denoted by $\mathcal{X} \times_n \boldsymbol{A}$ and is of size $I_1 \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$, with each mode-$n$ fiber multiplied by the matrix $\boldsymbol{A}$, i.e.

$$\mathcal{Y} = \mathcal{X} \times_n \boldsymbol{A} \Leftrightarrow \boldsymbol{Y}_{(n)} = \boldsymbol{A}\boldsymbol{X}_{(n)} \tag{4}$$

The CP decomposition decomposes a tensor into a sum of rank-one component tensors (see Fig. 2), i.e.

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \tag{5}$$

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

$$\mathcal{X}_{i_1 i_2 \cdots i_N} = \sum_{r=1}^{R} \lambda_r a_{i_1 r}^{(1)} a_{i_2 r}^{(2)} \cdots a_{i_N r}^{(N)} \tag{6}$$

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

$$\boldsymbol{X}_{(n)} = \boldsymbol{A}^{(n)} \boldsymbol{\Lambda} \left(\boldsymbol{A}^{(N)} \odot \cdots \boldsymbol{A}^{(n+1)} \odot \boldsymbol{A}^{(n-1)} \odot \cdots \boldsymbol{A}^{(1)} \right)^T \tag{7}$$

where $\boldsymbol{\Lambda} \triangleq \text{diag}(\lambda_1, \ldots, \lambda_R)$. The inner product of two tensors with the same size is defined as

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N} y^*_{i_1 i_2 \ldots i_N}$$

where * denotes the complex conjugate. The Frobenius norm of a tensor $\mathcal{X}$ is the square root of the inner product with itself, i.e.

$$\|\mathcal{X}\|_F = \langle \mathcal{X}, \mathcal{X}^* \rangle^{\frac{1}{2}}$$

To help gain a better understanding, we use a toy example to show the CP decomposition and the mode-n unfolding of $\mathcal{X}$. Let

$$\mathcal{X} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \circ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \circ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8}$$

Clearly, the three factor matrices are given respectively as $\boldsymbol{A}^{(1)} \triangleq [1 \ 1; 0 \ 1]$, $\boldsymbol{A}^{(2)} \triangleq [0 \ 1; 1 \ 0]$, $\boldsymbol{A}^{(3)} \triangleq [1 \ 0; 1 \ 1]$. The mode-3 unfolding of $\mathcal{X}$ can be written as

$$\boldsymbol{X}_{(3)} = \boldsymbol{A}^{(3)} \boldsymbol{\Lambda} (\boldsymbol{A}^{(2)} \odot \boldsymbol{A}^{(1)})^T = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 1 & 2 & 0 \end{bmatrix} \tag{9}$$

where $\boldsymbol{\Lambda} \triangleq [2 \ 0; 0 \ 1]$.

IV. Proposed CP Decomposition-Based Channel Estimation Method

Tensors provide a natural representation of data with multiple modes. Note that in our data model, the received signal $y_{m,t',t}$ has three modes which respectively stand for the RF chain, the sub-frame and the frame. Therefore the received data $\{y_{m,t',t}\}$ can be naturally represented by a three-mode tensor $\mathcal{Y} \in \mathbb{C}^{M_{\text{BS}}\times T'\times T}$, with its (m, t', t)th entry given by $y_{m,t',t}$. Combining (1) and (2), $y_{m,t',t}$ can be rewritten as

$$\begin{aligned} y_{m,t',t} &= \sum_{u=1}^{U} \sum_{j=1}^{L_u} \alpha_{u,j} \boldsymbol{q}_m^T \boldsymbol{a}_{\text{BS}}(\theta_{u,j}) \boldsymbol{a}_{\text{MS}}^T(\phi_{u,j}) \boldsymbol{p}_{t'} s_{u,t} + w_{m,t',t} \\ &= \sum_{l=1}^{L} \alpha_l \boldsymbol{q}_m^T \boldsymbol{a}_{\text{BS}}(\theta_l) \boldsymbol{a}_{\text{MS}}^T(\phi_l) \boldsymbol{p}_{t'} \bar{s}_{l,t} + w_{m,t',t} \end{aligned} \tag{10}$$

where with a slight abuse of notation, we let $\alpha_l = \alpha_{u,j}$, $\theta_l = \theta_{u,j}$, and $\phi_l = \phi_{u,j}$, in which $l = \sum_{i=1}^{u-1} L_i + j$; $L \triangleq \sum_{u=1}^{U} L_u$ denotes the total number of paths associated with all users, and $\bar{s}_{l,t} = s_{u,t}$ if the lth path comes from the uth user, i.e.

$$\bar{s}_{l,t} = s_{u,t} \quad \forall l \in \left[\sum_{i=1}^{u-1} L_i + 1, \sum_{i=1}^{u} L_i \right] \tag{11}$$

Define

$$\boldsymbol{Q} \triangleq [\boldsymbol{q}_1 \ \ldots \ \boldsymbol{q}_{M_{\text{BS}}}]$$
$$\boldsymbol{P} \triangleq [\boldsymbol{p}_1 \ \ldots \ \boldsymbol{p}_{T'}]$$

Since both $\boldsymbol{Q}$ and $\boldsymbol{P}$ are implemented using analog phase shifters, their entries are of constant modulus. Let $\boldsymbol{Y}_t \in \mathbb{C}^{M_{\text{BS}}\times T'}$ denote a matrix obtained by fixing the index t of the tensor $\mathcal{Y}$, we have

$$\begin{aligned} \boldsymbol{Y}_t &= \sum_{l=1}^{L} \alpha_l \bar{s}_{l,t} \boldsymbol{Q}^T \boldsymbol{a}_{\text{BS}}(\theta_l) \boldsymbol{a}_{\text{MS}}^T(\phi_l) \boldsymbol{P} + \boldsymbol{W}_t \\ &= \sum_{l=1}^{L} \bar{s}_{l,t} \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_l) \tilde{\boldsymbol{a}}_{\text{MS}}^T(\phi_l) + \boldsymbol{W}_t \end{aligned} \tag{12}$$

where

$$\tilde{\boldsymbol{a}}_{\text{BS}}(\theta_l) \triangleq \alpha_l \boldsymbol{Q}^T \boldsymbol{a}_{\text{BS}}(\theta_l)$$
$$\tilde{\boldsymbol{a}}_{\text{MS}}(\phi_l) \triangleq \boldsymbol{P}^T \boldsymbol{a}_{\text{MS}}(\phi_l)$$

We see that each slice of $\mathcal{Y}$ (i.e. $\boldsymbol{Y}_t$) is a weighted sum of a common set of rank-one outer products. The tensor $\mathcal{Y}$ thus admits the following CP decomposition which decomposes a tensor into a sum of rank-one component tensors, i.e.

$$\mathcal{Y} = \sum_{l=1}^{L} \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_l) \circ \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_l) \circ \bar{\boldsymbol{s}}_l + \mathcal{W} \tag{13}$$

where $\bar{\boldsymbol{s}}_l \triangleq [\bar{s}_{l,1} \ \cdots \ \bar{s}_{l,T}]^T$. The above equation suggests that an estimate of the mmWave channels $\{\boldsymbol{H}_u\}$ may be obtained by performing a CP decomposition of the tensor $\mathcal{Y}$. Due to the sparse scattering nature of the mmWave channel, L is usually small relative to the dimensions of the tensor, which implies that the tensor $\mathcal{Y}$ has an intrinsic low-rank structure. As analyzed in Section V, this low-rank property is crucial to guarantee the uniqueness of the CP decomposition, and in turn, the channel estimation. Define

$$\boldsymbol{A}_Q \triangleq [\tilde{\boldsymbol{a}}_{\text{BS}}(\theta_1) \ \cdots \ \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_L)] \tag{14}$$

$$\boldsymbol{A}_P \triangleq [\tilde{\boldsymbol{a}}_{\text{MS}}(\phi_1) \ \cdots \ \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_L)] \tag{15}$$

$$\boldsymbol{S}_L \triangleq [\bar{\boldsymbol{s}}_1 \ \cdots \ \bar{\boldsymbol{s}}_L] \tag{16}$$

Clearly, $\{\boldsymbol{A}_Q, \boldsymbol{A}_P, \boldsymbol{S}_L\}$ are factor matrices associated with a noiseless version of $\mathcal{Y}$. Also, let

$$\boldsymbol{S} \triangleq [\boldsymbol{s}_1 \ \cdots \ \boldsymbol{s}_U] \tag{17}$$

where

$$\boldsymbol{s}_u \triangleq [s_{u,1} \ \cdots \ s_{u,T}]^T \tag{18}$$

then we have $\boldsymbol{S}_L = \boldsymbol{S}\boldsymbol{O}$, where

$$\boldsymbol{O} \triangleq \begin{bmatrix} \boldsymbol{1}_{L_1}^T & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{L_2}^T & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{1}_{L_U}^T \end{bmatrix} \tag{19}$$

where $\boldsymbol{1}_l$ denotes an l-dimensional column vector with all entries equal to one.

A. CP Decomposition

We first consider the case where the number of total paths, L, is known *a priori*.[1] In this case, the CP decomposition can be accomplished by solving the following optimization problem

$$\min_{\boldsymbol{A}_Q, \boldsymbol{A}_P, \boldsymbol{S}_L} \quad \|\mathcal{Y} - \sum_{l=1}^{L} \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_l) \circ \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_l) \circ \bar{\boldsymbol{s}}_l\|_F^2 \tag{20}$$

The above optimization can be efficiently solved by an alternating least squares (ALS) procedure which iteratively minimizes the data fitting error with respect to the three factor matrices:

$$\boldsymbol{A}_Q^{(t+1)} = \arg\min_{\boldsymbol{A}_Q} \left\| \boldsymbol{Y}_{(1)}^T - (\boldsymbol{S}_L^{(t)} \odot \boldsymbol{A}_P^{(t)}) \boldsymbol{A}_Q^T \right\|_F^2 \tag{21}$$

$$\boldsymbol{A}_P^{(t+1)} = \arg\min_{\boldsymbol{A}_P} \left\| \boldsymbol{Y}_{(2)}^T - (\boldsymbol{S}_L^{(t)} \odot \boldsymbol{A}_Q^{(t+1)}) \boldsymbol{A}_P^T \right\|_F^2 \tag{22}$$

$$\boldsymbol{S}_L^{(t+1)} = \arg\min_{\boldsymbol{S}_L} \left\| \boldsymbol{Y}_{(3)}^T - (\boldsymbol{A}_P^{(t+1)} \odot \boldsymbol{A}_Q^{(t+1)}) \boldsymbol{S}_L^T \right\|_F^2 \tag{23}$$

For the general case where the total number of paths L is unknown *a priori*, more sophisticated CP decomposition techniques can be used to jointly estimate the model order and the factor matrices. Since the tensor $\mathcal{Y}$ has a low-rank structure, the CP decomposition can be cast as a rank minimization problem as

$$\begin{aligned} \min_{\mathcal{X}} \quad & \text{rank}(\mathcal{X}) \\ \text{s.t.} \quad & \|\mathcal{Y} - \mathcal{X}\|_F^2 \leq \varepsilon \end{aligned} \tag{24}$$

[1]This could be the case if there is only a direct line-of-sight path between each user and the BS, in which case we have $L = U$.

where ε is an error tolerance parameter related to noise statistics. Note that the CP rank is the minimum number of rank-one tensor components required to represent the tensor. Thus the search for a low rank $\mathcal{X}$ can be converted to the optimization of its associated factor matrices. Let

$$\mathcal{X} = \sum_{k=1}^{K} \boldsymbol{a}_k \circ \boldsymbol{b}_k \circ \boldsymbol{c}_k \tag{25}$$

where $K \gg L$ denotes an upper bound of the total number of paths, and

$$\boldsymbol{A} \triangleq [\boldsymbol{a}_1 \ \ldots \ \boldsymbol{a}_K]$$
$$\boldsymbol{B} \triangleq [\boldsymbol{b}_1 \ \ldots \ \boldsymbol{b}_K]$$
$$\boldsymbol{C} \triangleq [\boldsymbol{c}_1 \ \ldots \ \boldsymbol{c}_K]$$

The optimization (24) can be re-expressed as

$$\begin{aligned} \min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \quad & \|\boldsymbol{z}\|_0 \\ \text{s.t.} \quad & \|\mathcal{Y} - \mathcal{X}\|_F^2 \leq \varepsilon \\ & \mathcal{X} = \sum_{k=1}^{K} \boldsymbol{a}_k \circ \boldsymbol{b}_k \circ \boldsymbol{c}_k \end{aligned} \tag{26}$$

where $\boldsymbol{z}$ is a K-dimensional vector with its kth entry given by

$$z_k \triangleq \|\boldsymbol{a}_k \circ \boldsymbol{b}_k \circ \boldsymbol{c}_k\|_F \tag{27}$$

We see that $\|\boldsymbol{z}\|_0$ equals to the number of nonzero rank-one tensor components. Therefore minimizing the ℓ_0-norm of $\boldsymbol{z}$ is equivalent to minimizing the rank of the tensor $\mathcal{X}$.

The optimization (26) is an NP-hard problem. Nevertheless, alternative sparsity-promoting functions can be used to replace ℓ_0-norm to find a sparse solution of $\boldsymbol{z}$ more efficiently. In this paper, we use $\|\boldsymbol{z}\|_{2/3}^{2/3}$ as the relaxation of $\|\boldsymbol{z}\|_0$, where

$$\|\boldsymbol{z}\|_{2/3}^{2/3} \triangleq \left(\sum_{k=1}^{K} |z_k|^{2/3} \right)^{2/3} \tag{28}$$

It was shown [28] that the resulting optimization is equivalent to the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \quad & \|\mathcal{Y} - \mathcal{X}\|_F^2 + \mu \left(\text{tr}(\boldsymbol{A}\boldsymbol{A}^H) + \text{tr}(\boldsymbol{B}\boldsymbol{B}^H) + \text{tr}(\boldsymbol{C}\boldsymbol{C}^H) \right) \\ \text{s.t.} \quad & \mathcal{X} = \sum_{k=1}^{K} \boldsymbol{a}_k \circ \boldsymbol{b}_k \circ \boldsymbol{c}_k \end{aligned} \tag{29}$$

where μ is a regularization parameter whose choice will be discussed later in this paper. Again, the above optimization can be efficiently solved by an alternating least squares (ALS) procedure which iteratively minimizes (29) with respect to the three factor matrices:

$$\boldsymbol{A}^{(t+1)} = \arg\min_{\boldsymbol{A}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(1)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{C}^{(t)} \odot \boldsymbol{B}^{(t)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \boldsymbol{A}^T \right\|_F^2 \tag{30}$$

$$\boldsymbol{B}^{(t+1)} = \arg\min_{\boldsymbol{B}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(2)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{C}^{(t)} \odot \boldsymbol{A}^{(t+1)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \boldsymbol{B}^T \right\|_F^2 \tag{31}$$

$$\boldsymbol{C}^{(t+1)} = \arg\min_{\boldsymbol{C}} \left\| \begin{bmatrix} \boldsymbol{Y}_{(3)}^T \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{B}^{(t+1)} \odot \boldsymbol{A}^{(t+1)} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \boldsymbol{C}^T \right\|_F^2 \tag{32}$$

We can repeat the above iterations until the difference between estimated factor matrices of successive iterations is negligible, i.e. smaller than a pre-specified tolerance value. The rank of the tensor can be estimated by removing those negligible rank-one tensor components. Note that during the decomposition, we do not need to impose a specific structure on the estimates of the factor matrices since the CP decomposition is unique provided some mild conditions are satisfied.

B. Channel Estimation

We now discuss how to estimate the mmWave channel based on the estimated factor matrices $\{\hat{\boldsymbol{A}}_Q, \hat{\boldsymbol{A}}_P, \hat{\boldsymbol{S}}_L\}$. As to be shown in (65)–(72), under a mild condition, the estimated factor matrices and the true factor matrices are related as follows

$$\hat{\boldsymbol{A}}_Q = \boldsymbol{A}_Q \boldsymbol{\Lambda}_1 \boldsymbol{\Pi} + \boldsymbol{E}_1 \tag{33}$$

$$\hat{\boldsymbol{A}}_P = \boldsymbol{A}_P \boldsymbol{\Lambda}_2 \boldsymbol{\Pi} + \boldsymbol{E}_2 \tag{34}$$

$$\hat{\boldsymbol{S}}_L = \boldsymbol{S}_L \boldsymbol{\Lambda}_3 \boldsymbol{\Pi} + \boldsymbol{E}_3 \tag{35}$$

where $\boldsymbol{\Lambda}_3$ is a nonsingular diagonal matrix, $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are nonsingular block diagonal matrices compatible with the block structure of $\boldsymbol{A}_Q$ and $\boldsymbol{A}_P$ defined in (36)–(37), and we have $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_3\boldsymbol{\Lambda}_2^T = \boldsymbol{I}$; $\boldsymbol{\Pi}$ is a permutation matrix, $\boldsymbol{E}_1$, $\boldsymbol{E}_2$, and $\boldsymbol{E}_3$ denote the estimation errors associated with the three estimated factor matrices, respectively. Note that both $\boldsymbol{A}_Q$ and $\boldsymbol{A}_P$ can be partitioned into U blocks with each block consisting of column vectors associated with each user, i.e.

$$\boldsymbol{A}_Q = [\boldsymbol{A}_{Q,1} \;\; \boldsymbol{A}_{Q,2} \;\; \dots \;\; \boldsymbol{A}_{Q,U}] \tag{36}$$

$$\boldsymbol{A}_P = [\boldsymbol{A}_{P,1} \;\; \boldsymbol{A}_{P,2} \;\; \dots \;\; \boldsymbol{A}_{P,U}] \tag{37}$$

in which

$$\boldsymbol{A}_{Q,u} \triangleq [\tilde{\boldsymbol{a}}_{\text{BS}}(\theta_{u,1}) \;\; \dots \;\; \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_{u,L_u})] \tag{38}$$

$$\boldsymbol{A}_{P,u} \triangleq [\tilde{\boldsymbol{a}}_{\text{MS}}(\phi_{u,1}) \;\; \dots \;\; \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_{u,L_u})] \tag{39}$$

The block-diagonal structure of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ is compatible with the block structure of $\boldsymbol{A}_Q$ and $\boldsymbol{A}_P$. Thus we have

$$\boldsymbol{\Lambda}_1 = \text{diag}(\boldsymbol{\Lambda}_1^{(1)}, \dots, \boldsymbol{\Lambda}_1^{(U)}) \tag{40}$$

$$\boldsymbol{\Lambda}_2 = \text{diag}(\boldsymbol{\Lambda}_2^{(1)}, \dots, \boldsymbol{\Lambda}_2^{(U)}) \tag{41}$$

and

$$\boldsymbol{A}_Q\boldsymbol{\Lambda}_1 = [\boldsymbol{A}_{Q,1}\boldsymbol{\Lambda}_1^{(1)} \;\; \dots \;\; \boldsymbol{A}_{Q,U}\boldsymbol{\Lambda}_1^{(U)}] \tag{42}$$

$$\boldsymbol{A}_P\boldsymbol{\Lambda}_2 = [\boldsymbol{A}_{P,1}\boldsymbol{\Lambda}_2^{(1)} \;\; \dots \;\; \boldsymbol{A}_{P,U}\boldsymbol{\Lambda}_2^{(U)}] \tag{43}$$

The diagonal matrix $\boldsymbol{\Lambda}_3$ can also be partitioned according to the structure of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$:

$$\boldsymbol{\Lambda}_3 = \text{diag}(\boldsymbol{\Lambda}_3^{(1)}, \dots, \boldsymbol{\Lambda}_3^{(U)}) \tag{44}$$

From $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_3\boldsymbol{\Lambda}_2^T = \boldsymbol{I}$, we can readily arrive at

$$\boldsymbol{\Lambda}_1^{(u)}\boldsymbol{\Lambda}_3^{(u)}(\boldsymbol{\Lambda}_2^{(u)})^T = \boldsymbol{I} \quad \forall u = 1, \dots, U \tag{45}$$

To estimate the channel, we first estimate the number of paths associated with each user, the diagonal matrix $\boldsymbol{\Lambda}_3$ and the permutation matrix $\boldsymbol{\Pi}$ from (35). Suppose there are no estimation errors, each column of $\hat{\boldsymbol{S}}_L$ is a scaled version of a training sequence associated with an unknown user. Since the training sequences of all users are known *a priori*, a simple correlation-based matching method can be used to determine the unknown scaling factor and the permutation ambiguity for each column of $\hat{\boldsymbol{S}}_L$, based on which the number of paths associated with each user, the diagonal matrix $\boldsymbol{\Lambda}_3$ and the permutation matrix $\boldsymbol{\Pi}$ can be readily obtained.

Suppose the diagonal matrix $\boldsymbol{\Lambda}_3$ and the permutation matrix $\boldsymbol{\Pi}$ are perfectly recovered. The permutation ambiguity for the estimated factor matrices $\hat{\boldsymbol{A}}_Q$ and $\hat{\boldsymbol{A}}_P$ can be removed using the estimated permutation matrix. Thus we have

$$\hat{\boldsymbol{A}}_Q = \boldsymbol{A}_Q \boldsymbol{\Lambda}_1 + \boldsymbol{E}_1 \tag{46}$$

$$\hat{\boldsymbol{A}}_P = \boldsymbol{A}_P \boldsymbol{\Lambda}_2 + \boldsymbol{E}_2 \tag{47}$$

Given $\hat{\boldsymbol{A}}_Q$, $\hat{\boldsymbol{A}}_P$ and $\boldsymbol{\Lambda}_3$, the uth user's channel matrix $\boldsymbol{H}_u$ can be estimated from $\hat{\boldsymbol{A}}_{Q,u}\boldsymbol{\Lambda}_3^{(u)}\hat{\boldsymbol{A}}_{P,u}$ since we have

$$\begin{aligned}
\hat{\boldsymbol{A}}_{Q,u}\boldsymbol{\Lambda}_3^{(u)}\hat{\boldsymbol{A}}_{P,u} &= \boldsymbol{A}_{Q,u}\boldsymbol{\Lambda}_1^{(u)}\boldsymbol{\Lambda}_3^{(u)}(\boldsymbol{\Lambda}_2^{(u)})^T\boldsymbol{A}_{P,u}^T + \boldsymbol{E} \\
&= \boldsymbol{A}_{Q,u}\boldsymbol{A}_{P,u}^T + \boldsymbol{E} \\
&= \sum_{l=1}^{L_u} \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_{u,l})\tilde{\boldsymbol{a}}_{\text{MS}}(\phi_{u,l})^T + \boldsymbol{E} \\
&= \boldsymbol{Q}^T \sum_{l=1}^{L_u} \alpha_{u,l}\boldsymbol{a}_{\text{BS}}(\theta_{u,l})\boldsymbol{a}_{\text{MS}}(\phi_{u,l})^T \boldsymbol{P} + \boldsymbol{E} \\
&= \boldsymbol{Q}^T\boldsymbol{H}_u\boldsymbol{P} + \boldsymbol{E}
\end{aligned} \tag{48}$$

where $\boldsymbol{E}$ denotes the estimation error caused by $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$. We see that the joint multiuser channel estimation has been decoupled into U single-user channel estimation problems via the CP factorization. In the following section, we will show that the uniqueness of the decomposition can be guaranteed even when $T \ll U$. This enables a significant training overhead reduction since traditional estimation methods rely on the use of orthogonal pilot sequences (which requires $T = U$) to decouple the multiuser channel estimation problem into a set of single-user channel estimation problems. Let $\boldsymbol{z}_u \triangleq \text{vec}(\hat{\boldsymbol{A}}_{Q,u}\boldsymbol{\Lambda}_3^{(u)}\hat{\boldsymbol{A}}_{P,u})$. We have

$$\boldsymbol{z}_u = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\tilde{\boldsymbol{H}}_u\boldsymbol{\alpha}_u + \boldsymbol{e} \tag{49}$$

where $\boldsymbol{\alpha}_u \triangleq [\alpha_{u,1} \;\; \dots \;\; \alpha_{u,L_u}]$, and

$$\tilde{\boldsymbol{H}}_u \triangleq [\boldsymbol{a}_{\text{MS}}(\phi_{u,1}) \otimes \boldsymbol{a}_{\text{BS}}(\theta_{u,1}) \;\; \dots \;\; \boldsymbol{a}_{\text{MS}}(\phi_{u,L_u}) \otimes \boldsymbol{a}_{\text{BS}}(\theta_{u,L_u})] \tag{50}$$

The estimation of $\tilde{\boldsymbol{H}}_u$ can be cast as a compressed sensing problem by discretizing the continuous parameter space into an $N_1 \times N_2$ two dimensional grid with each grid point given by $\{\bar{\theta}_i, \bar{\phi}_j\}$ for $i = 1, \dots, N_1$ and $j = 1, \dots, N_2$ and assuming that $\{\phi_{u,l}, \theta_{u,l}\}_{l=1}^{L_u}$ lie on the grid. Thus (49) can be re-expressed as

$$\boldsymbol{z}_u = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\bar{\boldsymbol{\Sigma}}\bar{\boldsymbol{\alpha}}_u + \boldsymbol{e} \tag{51}$$

where $\bar{\boldsymbol{\Sigma}}$ is an overcomplete dictionary consisting of $N_1 \times N_2$ columns, with its $((i-1)N_1 + j)$th column given by $\boldsymbol{a}_{\text{MS}}(\bar{\phi}_i) \otimes \boldsymbol{a}_{\text{BS}}(\bar{\theta}_j)$, $\bar{\boldsymbol{\alpha}}_u \in \mathbb{C}^{N_1N_2\times 1}$ is a sparse vector obtained by augmenting $\boldsymbol{\alpha}_u$ with zero elements.

V. Uniqueness

In this section, we discuss under what conditions the uniqueness of the CP decomposition and, in turn, the channel estimation can be guaranteed.

A. Uniqueness for the Single-Path Geometric Model

We first consider the special case where there is a direct line-of-sight path between the BS and each user, in which case we have $L = U$ and $\boldsymbol{S}_L = \boldsymbol{S}$ (recalling $\boldsymbol{S}_L = \boldsymbol{S}\boldsymbol{O}$). It is well known that essential uniqueness of the CP decomposition can be guaranteed by the Kruskal's condition [29]. Let $k_{\boldsymbol{A}}$ denote the k-rank of a matrix $\boldsymbol{A}$, which is defined as the largest value of $k_{\boldsymbol{A}}$ such that every subset of $k_{\boldsymbol{A}}$ columns of the matrix $\boldsymbol{A}$ is linearly independent. Kruskal showed that a CP decomposition $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ of a third-order tensor is essentially unique if [29]

$$k_{\boldsymbol{A}} + k_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2R + 2 \tag{52}$$

where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are factor matrices, and R denotes the CP rank. More formally, we have the following theorem.

Theorem 1: Let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ be a CP solution which decomposes a three-mode tensor $\mathcal{X} \in \mathbb{C}^{M\times N\times K}$ into R rank-one arrays, where $\boldsymbol{A} \in \mathbb{C}^{M\times R}$, $\boldsymbol{B} \in \mathbb{C}^{N\times R}$, and $\boldsymbol{C} \in \mathbb{C}^{K\times R}$. Suppose Kruskal's condition (52) holds and there is an alternative CP solution $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$ which also decomposes $\mathcal{X}$ into R rank-one arrays. Then we have $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_a$, $\bar{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Pi}\boldsymbol{\Lambda}_b$, and $\bar{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_c$, 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 [30]. In Appendix A, we show that Theorem 1 is in fact a special case of Theorem 2. Thus the proof for Theorem 2 can be easily adapted to prove Theorem 1. ■

From Theorem 1, we know that if the following condition holds

$$k_{\boldsymbol{A}_Q} + k_{\boldsymbol{A}_P} + k_{\boldsymbol{S}} \geq 2U + 2 \tag{53}$$

then the CP decomposition of $\mathcal{Y}$ is unique and in the noiseless case, we can ensure that the factor matrices can be estimated up to a permutation and scaling ambiguity, i.e. $\hat{\boldsymbol{A}}_Q = \boldsymbol{A}_Q\boldsymbol{\Pi}\boldsymbol{\Lambda}_1$, $\hat{\boldsymbol{A}}_P = \boldsymbol{A}_P\boldsymbol{\Pi}\boldsymbol{\Lambda}_2$, and $\hat{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3$, with $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \boldsymbol{I}$.

We now discuss how to design the beamforming matrix $\boldsymbol{P} \in \mathbb{C}^{N_{\text{MS}}\times T'}$, the combining matrix $\boldsymbol{Q} \in \mathbb{C}^{N_{\text{BS}}\times M_{\text{BS}}}$, and the pilot symbol matrix $\boldsymbol{S} \in \mathbb{C}^{T\times U}$ such that the Kruskal's condition (53) can be met. In our following discussion, we assume $M_{\text{BS}} \geq U$ and $T' \geq U$ in order to find the minimum value of T such that the Kruskal's condition can be met. Note that $\boldsymbol{A}_Q = \boldsymbol{Q}^T\boldsymbol{A}_{\text{BS}}$, where $\boldsymbol{A}_{\text{BS}}$ is a Vandermonte matrix whose k-rank is equivalent to the number of columns, U, when the angles of arrival $\{\theta_u\}$ are distinct. The k-rank of $\boldsymbol{A}_Q$, therefore, is no greater than U, i.e. $k_{\boldsymbol{A}_Q} \leq U$. The problem now becomes whether we can design a combining matrix $\boldsymbol{Q}$ such that $k_{\boldsymbol{A}_Q}$ achieves its upper bound U. We will show that the answer is affirmative for a randomly generated $\boldsymbol{Q}$ with i.i.d. entries. Specifically, we assume each entry of $\boldsymbol{Q}$ is chosen uniformly from a unit circle scaled by a constant $1/N_{\text{BS}}$, i.e. $q_{m,n} = (1/N_{\text{BS}})e^{j\vartheta_{m,n}}$, where $\vartheta_{m,n} \in [-\pi, \pi]$ follows a uniform distribution. Let $a_{m,i} \triangleq \boldsymbol{q}_m^T\boldsymbol{a}_{\text{BS}}(\theta_i)$ denote the (m, i)th entry of $\boldsymbol{A}_Q$. It can be readily verified that $\mathbb{E}[a_{m,i}] = 0, \forall m, i$ and

$$\mathbb{E}[a_{m,i}a_{n,j}^*] = \begin{cases} 0 & m \neq n \\ \frac{1}{N_{\text{BS}}^2}\boldsymbol{a}_{\text{BS}}^H(\theta_i)\boldsymbol{a}_{\text{BS}}(\theta_j) & m = n \end{cases} \tag{54}$$

When the number of antennas at the BS is sufficiently large, the steering vectors $\{\boldsymbol{a}_{\text{BS}}(\theta_i)\}$ become mutually quasi-orthogonal, i.e. $\boldsymbol{a}_{\text{BS}}^H(\theta_i)\boldsymbol{a}_{\text{BS}}(\theta_j) \to \delta(\theta_i - \theta_j)$, which implies that the entries of $\boldsymbol{A}_Q$ 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 $\boldsymbol{A}_Q$ can be considered as i.i.d. Gaussian variables, and $\boldsymbol{A}_Q$ is full column rank with probability one. Thus we can reach that the k-rank of $\boldsymbol{A}_Q$ is equivalent to U with probability one.

Following a similar derivation, we can arrive at the following conclusion: if each entry of the beamforming matrix $\boldsymbol{P}$ is chosen uniformly from a unit circle scaled by a constant $1/N_{\text{MS}}$, then the k-rank of $\boldsymbol{A}_P$ is equivalent to U with probability one. Thus we can guarantee that the Kruskal's condition (53) is met with probability one as long as $k_{\boldsymbol{S}} \geq 2$, i.e. any two columns of $\boldsymbol{S}$ are linearly independent. For the single path geometric model, $\boldsymbol{S}$ consists of U columns, with the uth column constructed by pilot symbols of the uth user. Therefore the condition $k_{\boldsymbol{S}} \geq 2$ can be ensured provided that $T \geq 2$, and pilot symbol vectors of users are mutually independent. Specifically, we can design the pilot symbols by minimizing the mutual coherence of $\boldsymbol{S}$, i.e.

$$\min_{\boldsymbol{S}} \quad \mu(\boldsymbol{S}) \tag{55}$$

where

$$\mu(\boldsymbol{S}) \triangleq \max_{i\neq j}\left|\frac{\langle \boldsymbol{s}_i, \boldsymbol{s}_j\rangle}{\|\boldsymbol{s}_i\|\|\boldsymbol{s}_j\|}\right|$$

The solution of above problem can be found in [31] and [32]. For the case $k_{\boldsymbol{A}_Q} = U$ and $k_{\boldsymbol{A}_P} = U$, the Kruskal's condition can be met by choosing the length of the pilot sequence equal to two, i.e. $T = 2$, irrespective of the value of U. This allows a considerable training overhead reduction, particularly when U is large. Note that besides random coding, the beamforming and combining matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ can also be devised to form a certain number of transmit/receive beams. The k-rank of the resulting matrices $\boldsymbol{A}_P$ and $\boldsymbol{A}_Q$ may also achieve the upper bound U.

B. Uniqueness for the General Geometric Model

For the general geometric model where there are more than one path between each user and the BS, the Kruskal's condition becomes

$$k_{\boldsymbol{A}_Q} + k_{\boldsymbol{A}_P} + k_{\boldsymbol{S}_L} \geq 2L + 2 \tag{56}$$

Since the k-rank of $\boldsymbol{A}_Q \in \mathbb{C}^{M_{\text{BS}}\times L}$ and $\boldsymbol{A}_P \in \mathbb{C}^{T'\times L}$ is at most equal to L, we need $k_{\boldsymbol{S}_L} \geq 2$ to satisfy the above Kruskal's condition. However, for the general geometric model, the k-rank of $\boldsymbol{S}_L$ is always equal to one because multiple column vectors associated with a common user are linearly dependent. Thus the Kruskal's condition can never be satisfied in this case. Nevertheless, this does not mean that the uniqueness of the CP decomposition does not hold for the general geometric model. In fact, considering the special form of the decomposition (13), the uniqueness can be guaranteed under a less restrictive condition.

We first write (13) as follows

$$\begin{aligned}\mathcal{Y} &= \sum_{u=1}^{U}\sum_{l=1}^{L_u} \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_{u,l}) \circ \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_{u,l}) \circ \boldsymbol{s}_u + \mathcal{W} \\ &= \sum_{u=1}^{U} (\boldsymbol{A}_{Q_u}\boldsymbol{A}_{P_u}^T) \circ \boldsymbol{s}_u + \mathcal{W} \end{aligned} \tag{57}$$

where $\boldsymbol{A}_{Q_u}$ and $\boldsymbol{A}_{P_u}$ are defined in (38) and (39), respectively, and $\boldsymbol{s}_u$ is defined in (18). We see that the tensor $\mathcal{Y}$ can be expressed as a sum of matrix-vector outer products, more specifically, a sum of rank-$(L_u, L_u, 1)$ terms since $\boldsymbol{A}_{Q_u}$ and $\boldsymbol{A}_{P_u}$ are both rank-L_u. For this block term decomposition, we have the following generalized version of the Kruskal's condition.

Before proceeding, we define $\boldsymbol{A} \triangleq [\boldsymbol{A}_1 \ \ldots \ \boldsymbol{A}_R]$, $\boldsymbol{B} \triangleq [\boldsymbol{B}_1 \ \ldots \ \boldsymbol{B}_R]$, and $\boldsymbol{C} \triangleq [\boldsymbol{c}_1 \ \ldots \ \boldsymbol{c}_R]$, and generalize the k-rank concept to the above partitioned matrices. Specifically, the k'-rank of a partitioned matrix $\boldsymbol{A}$, denoted by $k'_{\boldsymbol{A}}$, is the maximal number r such that any set of r submatrices of $\boldsymbol{A}$ yields a set of linearly independent columns.

We have the following theorem.

Theorem 2: Let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ represent a decomposition of $\mathcal{X} \in \mathbb{C}^{M\times N\times K}$ in rank-$(L_r, L_r, 1)$ terms, i.e.

$$\mathcal{X} = \sum_{r=1}^{R} (\boldsymbol{A}_r\boldsymbol{B}_r^T) \circ \boldsymbol{c}_r$$

We assume $M \geq \max_r L_r$, $N \geq \max_r L_r$, $\text{rank}(\boldsymbol{A}_r) = L_r$, and $\text{rank}(\boldsymbol{B}_r) = L_r$. Suppose the following conditions

$$MN \geq \sum_{r=1}^{R} L_r^2 \tag{58}$$

$$k'_{\boldsymbol{A}} + k'_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2R + 2 \tag{59}$$

hold and we have an alternative decomposition of $\mathcal{X}$, represented by $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$, with $k'_{\bar{\boldsymbol{A}}}$ and $k'_{\bar{\boldsymbol{B}}}$ maximal under the given dimensionality constraints. Then $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ and $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$ are essentially equal, i.e. $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_a$, $\bar{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Pi}\boldsymbol{\Lambda}_b$ and $\bar{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\Pi}_c\boldsymbol{\Lambda}_c$, in which $\boldsymbol{\Pi}$ is a block permutation matrix whose block structure is consistent with that of $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{\Pi}_c$ is permutation matrix whose permutation pattern is the same as that of $\boldsymbol{\Pi}$, $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ are nonsingular block-diagonal matrices, compatible with the block structure of $\boldsymbol{A}$ and $\boldsymbol{B}$, and $\boldsymbol{\Lambda}_c$ is a nonsingular diagonal matrix. Also, let $\boldsymbol{\Lambda}_{a,r}$ and $\boldsymbol{\Lambda}_{b,r}$ denote the rth diagonal block of $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$, respectively, and λ_r denote the rth diagonal element of $\boldsymbol{\Lambda}_c$. We have $\lambda_r\boldsymbol{\Lambda}_{a,r}\boldsymbol{\Lambda}_{b,r}^T = \boldsymbol{I}, \forall r$.

Proof: A rigorous proof can be found in [33]. We also provide a sketch of the proof in Appendix B. ■

From Theorem 2, we know that if the following conditions hold

$$M_{\text{BS}}T' \geq \sum_{u=1}^{U} L_u^2 \tag{60}$$

$$k'_{\boldsymbol{A}_Q} + k'_{\boldsymbol{A}_P} + k_{\boldsymbol{S}} \geq 2U + 2 \tag{61}$$

then the essential uniqueness of the CP decomposition of $\mathcal{Y}$ in (13) can be guaranteed. Following an analysis similar to our previous subsection, we can arrive at the k'-ranks of $\boldsymbol{A}_Q$ and $\boldsymbol{A}_P$ are equivalent to U with probability one. Therefore we only need $k_{\boldsymbol{S}} \geq 2$ in order to satisfy the above generalized Kruskal's condition (61). This condition can be easily satisfied by assigning pairwise independent pilot symbol vectors to users (provided $T \geq 2$).

Since the proposed algorithm yields a canonical form of CP decomposition represented as a sum of rank-one tensor components, we need further explore the relationship between the true factor matrices and the estimated factor matrices. We write

$$\begin{aligned}\mathcal{X} &= \sum_{r=1}^{R} (\boldsymbol{A}_r\boldsymbol{B}_r^T) \circ \boldsymbol{c}_r \\ &= \sum_{r=1}^{R}\sum_{j=1}^{L_R} \boldsymbol{A}_r[:, j] \circ \boldsymbol{B}_r[:, j] \circ \boldsymbol{c}_r \\ &= \sum_{l=1}^{L} \boldsymbol{a}_l \circ \boldsymbol{b}_l \circ \boldsymbol{f}_l \end{aligned} \tag{62}$$

where $L \triangleq \sum_{r=1}^{R} L_r$, $\boldsymbol{X}[:, j]$ denotes the jth column of $\boldsymbol{X}$, $\boldsymbol{a}_l$ and $\boldsymbol{b}_l$ denote the lth column of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and

$$\boldsymbol{f}_l = \boldsymbol{c}_r \quad \forall l \in \left[\sum_{i=1}^{r-1} L_i + 1, \sum_{i=1}^{r} L_i\right] \tag{63}$$

Define $\boldsymbol{F} \triangleq [\boldsymbol{f}_1 \ \ldots \ \boldsymbol{f}_L]$. Clearly, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{F}$ are true factor matrices of $\mathcal{X}$. The CP decomposition of $\mathcal{X}$ can also be expressed as

$$\begin{aligned}\mathcal{X} &= \sum_{r=1}^{R} (\bar{\boldsymbol{A}}_r\bar{\boldsymbol{B}}_r^T) \circ \bar{\boldsymbol{c}}_r \\ &= \sum_{r=1}^{R}\sum_{j=1}^{L_r} (\bar{\boldsymbol{A}}_r[:, j]\bar{\boldsymbol{B}}_r[:, j]^T) \circ (\lambda_r\boldsymbol{c}_r) \\ &= \sum_{r=1}^{R}\sum_{j=1}^{L_r} (\beta_{r,j}\bar{\boldsymbol{A}}_r[:, j]\bar{\boldsymbol{B}}_r[:, j]^T) \circ (\beta_{r,j}^{-1}\lambda_r\boldsymbol{c}_r) \\ &= \sum_{l=1}^{L} \tilde{\boldsymbol{a}}_l \circ \tilde{\boldsymbol{b}}_l \circ \tilde{\boldsymbol{f}}_l \end{aligned} \tag{64}$$

where

$$\begin{aligned} \tilde{\boldsymbol{a}}_l &\triangleq \beta_{r,j}\bar{\boldsymbol{A}}_r[:, j] \quad l = \textstyle\sum_{i=1}^{r-1} L_r + j \\ \tilde{\boldsymbol{b}}_l &\triangleq \quad \bar{\boldsymbol{B}}_r[:, j] \quad l = \textstyle\sum_{i=1}^{r-1} L_r + j \\ \tilde{\boldsymbol{f}}_l &\triangleq \quad \beta_{r,j}^{-1}\lambda_r\boldsymbol{c}_r \quad l = \textstyle\sum_{i=1}^{r-1} L_r + j \end{aligned}$$

Define

$$\begin{aligned} \tilde{\boldsymbol{A}} &\triangleq [\tilde{\boldsymbol{a}}_1 \ \tilde{\boldsymbol{a}}_2 \ \ldots \ \tilde{\boldsymbol{a}}_L] \\ \tilde{\boldsymbol{B}} &\triangleq [\tilde{\boldsymbol{b}}_1 \ \tilde{\boldsymbol{b}}_2 \ \ldots \ \tilde{\boldsymbol{b}}_L] \\ \tilde{\boldsymbol{F}} &\triangleq [\tilde{\boldsymbol{f}}_1 \ \tilde{\boldsymbol{f}}_2 \ \ldots \ \tilde{\boldsymbol{f}}_L] \end{aligned}$$

Clearly, $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{F}})$ is an alternative solution which decomposes $\mathcal{X}$ into L rank-one tensor components. It is easy to verify that the true factor matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{F})$ and the estimated factor matrices $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{F}})$ are related as follows:

$$\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Lambda}_1\boldsymbol{\Pi} \tag{65}$$

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Lambda}_2\boldsymbol{\Pi} \tag{66}$$

$$\tilde{\boldsymbol{F}} = \boldsymbol{F}\boldsymbol{\Lambda}_3\boldsymbol{\Pi} \tag{67}$$

where $\boldsymbol{\Pi}$ is a permutation matrix, and

$$\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_a \boldsymbol{D}_\beta \tag{68}$$
$$\boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}_b \tag{69}$$
$$\boldsymbol{\Lambda}_3 = \boldsymbol{D}_\beta^{-1} \boldsymbol{D}_\lambda \tag{70}$$

in which $\boldsymbol{D}_\beta$ is a diagonal matrix with its lth ($l = \sum_{i=1}^{r-1} L_r + j$) diagonal element equal to $\beta_{r,j}$, and

$$\boldsymbol{D}_\lambda \triangleq \text{diag}(\lambda_1 \boldsymbol{I}_{L_1}, \ldots, \lambda_R \boldsymbol{I}_{L_R}) \tag{71}$$

where $\boldsymbol{I}_n$ is an $n \times n$ identity matrix. It is easy to verify that

$$\boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_3 \boldsymbol{\Lambda}_2^T = \boldsymbol{\Lambda}_a \boldsymbol{D}_\lambda \boldsymbol{\Lambda}_b^T = \boldsymbol{I} \tag{72}$$

since we have $\lambda_r \boldsymbol{\Lambda}_{a,r} \boldsymbol{\Lambda}_{b,r}^T = \boldsymbol{I}, \forall r$.

VI. A Direct Compressed Sensing-Based Channel Estimation Method

The multiuser channel estimation problem considered in this paper can also be formulated as a sparse signal recovery problem by exploiting the poor scattering nature of the mmWave channel, without resorting to the CP decomposition. Such a direct compressed sensing-based method is discussed in the following. Let $\boldsymbol{Y}_{(3)}$ denote the mode-3 unfolding of the tensor $\mathcal{Y}$ defined in (13). We have

$$\begin{aligned}\boldsymbol{Y}_{(3)} &= \boldsymbol{S}_L(\boldsymbol{A}_P \odot \boldsymbol{A}_Q)^T + \boldsymbol{W}_{(3)} \\ &= \boldsymbol{S}_L[\tilde{\boldsymbol{a}}_{\text{MS}}(\phi_1) \otimes \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_1) \ \ \ldots \ \ \tilde{\boldsymbol{a}}_{\text{MS}}(\phi_L) \otimes \tilde{\boldsymbol{a}}_{\text{BS}}(\theta_L)]^T \\ &\quad + \boldsymbol{W}_{(3)} \\ &\overset{(a)}{=} \boldsymbol{S}_L \boldsymbol{D} \boldsymbol{\Sigma}^T (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)^T + \boldsymbol{W}_{(3)}\end{aligned} \tag{73}$$

where (a) comes from the mixed-product property: $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$, and

$$\begin{aligned}\boldsymbol{\Sigma} &\triangleq [\boldsymbol{a}_{\text{MS}}(\phi_1) \otimes \boldsymbol{a}_{\text{BS}}(\theta_1) \ \ \ldots \ \ \boldsymbol{a}_{\text{MS}}(\phi_L) \otimes \boldsymbol{a}_{\text{BS}}(\theta_L)] \\ \boldsymbol{D} &\triangleq \text{diag}(\alpha_1, \ldots, \alpha_L)\end{aligned} \tag{74}$$

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

$$\boldsymbol{Y}_{(3)}^T = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\boldsymbol{\Sigma} \boldsymbol{D} \boldsymbol{O}^T \boldsymbol{S}^T + \boldsymbol{W}_{(3)} \tag{75}$$

The dictionary $\boldsymbol{\Sigma}$ is characterized by a number of unknown parameters $\{\theta_l, \phi_l\}$ which need to be estimated. To formulate the channel estimation as a sparse signal recovery problem, we discretize the continuous parameter space into an $N_1 \times N_2$ two dimensional grid with each grid point given by $\{\bar{\theta}_i, \bar{\phi}_j\}$ for $i = 1, \ldots, N_1$ and $j = 1, \ldots, N_2$. Assume that the true parameters $\{\theta_l, \phi_l\}$ lie on the two-dimensional grid. Hence (75) can be re-expressed as

$$\boldsymbol{Y}_{(3)}^T = (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\bar{\boldsymbol{\Sigma}} \bar{\boldsymbol{D}} \boldsymbol{S}^T + \boldsymbol{W}_{(3)} \tag{76}$$

where $\bar{\boldsymbol{\Sigma}}$ is an overcomplete dictionary consisting of $N_1 \times N_2$ columns, with its $((i-1)N_1 + j)$th column given by $\boldsymbol{a}_{\text{MS}}(\bar{\phi}_i) \otimes \boldsymbol{a}_{\text{BS}}(\bar{\theta}_j)$, $\bar{\boldsymbol{D}} \in \mathbb{C}^{N_1 N_2 \times U}$ is a sparse matrix obtained by augmenting $\boldsymbol{D}\boldsymbol{O}^T$ with zero rows. Let $\boldsymbol{y} \triangleq \text{vec}(\boldsymbol{Y}_{(3)}^T)$ and define $\boldsymbol{\Phi} \triangleq (\boldsymbol{P}^T \otimes \boldsymbol{Q}^T)\bar{\boldsymbol{\Sigma}}$. We have

$$\boldsymbol{y} = (\boldsymbol{S} \otimes \boldsymbol{\Phi})\boldsymbol{d} + \boldsymbol{w} \tag{77}$$

where $\boldsymbol{d} \triangleq \text{vec}(\bar{\boldsymbol{D}})$ is an unknown sparse vector, and $\boldsymbol{w} \triangleq \text{vec}(\boldsymbol{W}_{(3)})$ denotes the additive noise. We see that the channel estimation problem has now been formulated as a conventional sparse signal recovery problem. The problem can be further recast as an ℓ_1-regularized optimization problem

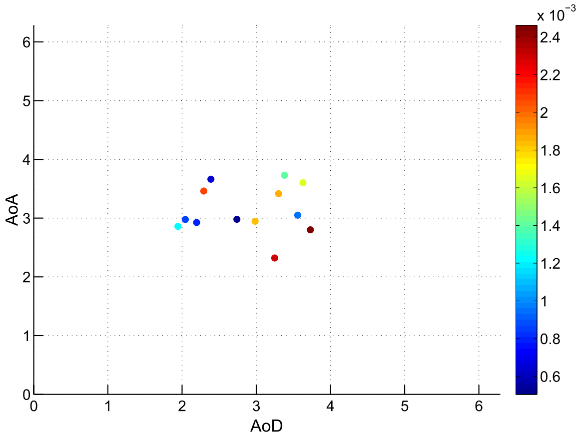
$$\min_{\boldsymbol{d}} \quad \|\boldsymbol{y} - (\boldsymbol{S} \otimes \boldsymbol{\Phi})\boldsymbol{d}\|_2^2 + \eta\|\boldsymbol{d}\|_1 \tag{78}$$

and many efficient algorithms such as fast iterative shrinkage-thresholding algorithm (FISTA) [34] can be employed to solve the above ℓ_1-regularized optimization problem. In practice, the true parameters may not be aligned on the presumed grid. This error, also referred to as the grid mismatch, leads to deteriorated performance. Finer grids can certainly be used to reduce grid mismatch and improve the reconstruction accuracy. Nevertheless, recovery algorithms may become numerically instable and computationally prohibitive when very fine discretized grids are employed.

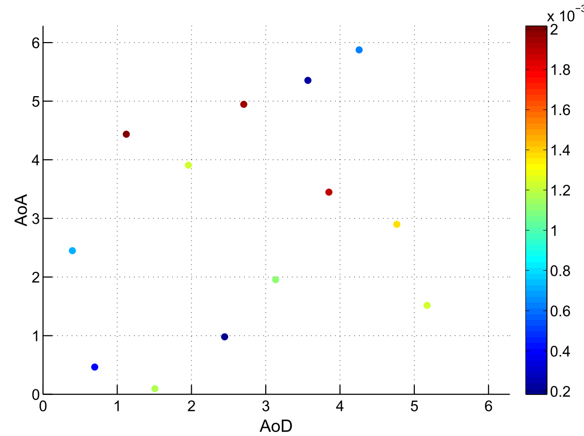
VII. Computational Complexity Analysis

We discuss the computational complexity of the proposed CP decomposition-based method and its comparison with the direct compressed sensing-based method. The computational task of our proposed method involves solving the least squares problems (30)–(32) at each iteration and solving the compressed sensing problem (51) after the factor matrices are estimated. Let $\boldsymbol{A} = \boldsymbol{A}_Q$, $\boldsymbol{B} = \boldsymbol{A}_P$, $\boldsymbol{C} = \boldsymbol{S}_L$ in (30)–(32). Considering the update of $\boldsymbol{A}_Q$, we have $\boldsymbol{A}_Q^T = (\boldsymbol{V}^H\boldsymbol{V} + \mu\boldsymbol{I})^{-1}\boldsymbol{V}^H\boldsymbol{Y}_{(1)}^T$, where $\boldsymbol{V} \triangleq (\boldsymbol{S}^{(t)} \odot \boldsymbol{A}_P^{(t)}) \in \mathbb{C}^{TT' \times K}$ is a tall matrix as we usually have $TT' > K$. Noting that $\boldsymbol{Y}_{(1)}^T \in \mathbb{C}^{TT' \times M_{\text{BS}}}$, it can be easily verified that the number of flops required to calculate $\boldsymbol{A}_Q^T$ is of order $\mathcal{O}(KT'TM_{\text{BS}} + K^2T'T + K^3)$. K is usually of the same order of magnitude as the value of L. When L is small, the order of the dominant term will be $\mathcal{O}(T'TM_{\text{BS}})$ which scales linearly with the size of observed tensor $\mathcal{Y}$. We can also easily show that solving the least squares problems (31) and (32) requires flops of order $\mathcal{O}(T'TM_{\text{BS}})$ as well. To solve (51), a fast iterative shrinkage-thresholding algorithm (FISTA) [34] can be used. The main computational task associated with the FISTA algorithm at each iteration is to evaluate a so-called 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 (51), we have $n = N_1N_2$ and the computational complexity is thus of order $\mathcal{O}(N_1^2N_2^2)$. Therefore the overall computational complexity is $\mathcal{O}(N_1^2N_2^2 + T'TM_{\text{BS}})$.

For the direct compressed sensing-based method discussed in Section VI, the main computational task associated with the FISTA algorithm at each iteration is to evaluate the proximal operator whose computational complexity, as indicated earlier, is of the order $\mathcal{O}(n^2)$, where n denotes the number of columns of the overcomplete dictionary. For the compressed sensing problem considered in (78), we have $n = N_1N_2U$. Thus the required number of flops at each iteration of the FISTA is of order $\mathcal{O}(N_1^2N_2^2U^2)$, which scales quadratically with N_1N_2U. Note that the overcomplete dictionary $\boldsymbol{S} \otimes \boldsymbol{\Phi}$ in (78) is of dimension $TT'M_{\text{BS}} \times N_1N_2U$. In order to achieve a substantial overhead reduction, the parameters $\{M_{\text{BS}}, T, T'\}$ are usually chosen such that the number of measurements is far less



(a) The set of AoAs and AoDs associated with the first channel



(b) The set of AoAs and AoDs associated with the second channel

Fig. 3. Two sets of AoAs/AoDs realizations.

than the dimension of the sparse signal, i.e. $TT'M_{\text{BS}} \ll UN_1N_2$. Therefore the compressed sensing-based method has a higher computational complexity than the proposed CP decomposition-based method.

VIII. Simulation Results

We now present simulation results to illustrate the performance of our proposed CP decomposition-based method (referred to as CP), and its comparison with the direct compressed sensing-based method (referred to as CS) discussed in Section VI. For the CP method, μ in (29) is chosen to be 3×10^{-3} throughout our experiments. In fact, empirical results suggest that stable recovery performance can be achieved when μ is set in the range $[10^{-3}, 10^{-2}]$. For the direct compressed sensing-based method, the parameter η used in (78) is tuned carefully to ensure the best performance is achieved. We consider a system model consisting of a BS and U MSs, with the BS employing a uniform linear array of $N_{\text{BS}} = 64$ antennas and each MS employing a uniform linear array of $N_{\text{MS}} = 32$ antennas. We set $U = 8$ and the distance between neighboring antennas is set to be half the wavelength of the signal. The mmWave channel is assumed to follow a geometric channel model with the AoAs and AoDs distributed in $[0, 2\pi]$. The complex gain $\alpha_{u,l}$ is assumed to be a random variable following a circularly-symmetric Gaussian distribution $\alpha_{u,l} \sim \mathcal{CN}(0, N_{\text{BS}}N_{\text{MS}}/\rho)$, where ρ is given by $\rho = (4\pi Df_c/c)^2$. Here c represents the speed of light, D denotes the distance between the MS and the BS, and f_c is the carrier frequency. We assume $D = 30$m and $f_c = 28$GHz. In our simulations, we investigate the performance of the proposed method under two randomly generated mmWave channels. For the first mmWave channel, the AoAs and AoDs associated with the U users are closely-spaced (see Fig. 3 (a)), while the AoAs and AoDs associated with the U users are sufficiently separated for the other mmWave channel (see Fig. 3 (b)). The total number of paths is set to $L = 13$ and the number of scatterers between each MS and the BS, L_u, is set equal to one or two. The beamforming matrix $\boldsymbol{P}$ and the combining matrix $\boldsymbol{Q}$ are generated according to the way described in Section V. The pilot symbol matrix $\boldsymbol{S}$ is chosen from the codebook of Grassmannian beamforming [35] for $T = 2$, while for $T = 3$, $T = 4$ and $T = 6$, $\boldsymbol{S}$ can be calculated by the algorithm proposed in [36]. When $T = 8$, $\boldsymbol{S}$ is simply chosen as a DFT matrix.

The estimation performance is evaluated by the normalized mean squared error (NMSE) which is calculated as

$$\text{NMSE} = \frac{\sum_{u=1}^{U} \|\boldsymbol{H}_u - \hat{\boldsymbol{H}}_u\|_F^2}{\sum_{u=1}^{U} \|\boldsymbol{H}_u\|_F^2} \tag{79}$$

where $\hat{\boldsymbol{H}}_u$ denotes the estimated channel. The signal-to-noise ratio (SNR) is defined as the ratio of the signal component to the noise component, i.e.

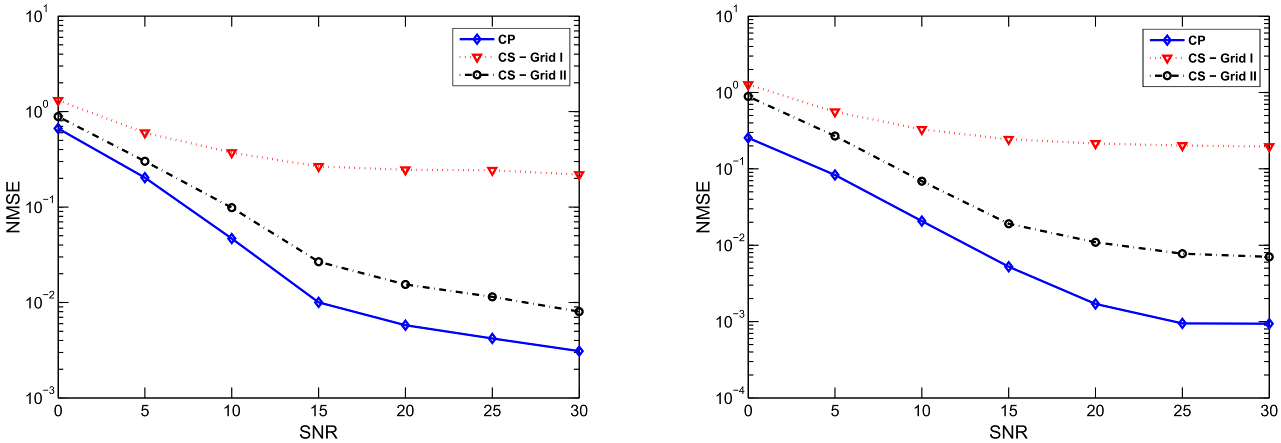
$$\text{SNR} \triangleq \frac{\|\mathcal{Y} - \mathcal{W}\|_F^2}{\|\mathcal{W}\|_F^2} \tag{80}$$

where $\mathcal{Y}$ and $\mathcal{W}$ represent the received signal and the additive noise in (13), respectively. Throughout our simulations, we usually set SNR = 15dB and SNR = 30dB. The choice of SNR = 15dB can be justified by a typical scenario as follows. Suppose the carrier frequency is 28GHz, the bandwidth of the signal is 50MHz. The radius of the small cell is assumed to be 30 meters. According to [37], the power of the received signal, in dBm, is given by

$$P_{RX} = P_T - PL_d + G_T + G_R \tag{81}$$

where P_T is the transmitted power in dBm, and a typical value of P_T in practical systems is 14dBm, PL_d is the path loss in decibels for the transmitter-receiver separation of distance d, G_T and G_R denote the transmit and receive antenna gain in decibels, respectively. We assume $G_R = G_T = 0$dB. By Friis' law, the pathloss can be calculated as $PL_d = 20\log 10(4\pi d/\lambda) = 90$dB, where λ denotes the signal wavelength. Hence the power of the received signal is $P_{RX} = -76$dBm. The noise power at the receiver can be calculated as

$$P_{\text{noise}} = 10\log(kT_{\text{syst}}) + 10\log(B) + NF_{RX} \tag{82}$$



(a) Channel I

(b) Channel II

Fig. 4. NMSE versus SNR, $M_{\text{BS}} = 16$, $T' = 16$, and $T = 4$.

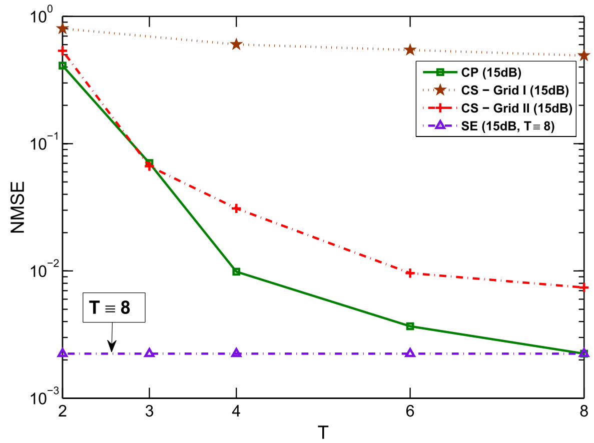
where $10\log(kT_{\text{syst}})$ is equal to -174dBm/Hz for a system temperature of 17°C, B is the bandwidth of the signal in hertz, and NF_{RX} is the noise figure of the receiver in decibels whose typical value is 6dB. Thus the noise power at the receiver front end is given by $P_{\text{noise}} = -174\text{dBm/Hz} + 10\log(50\text{MHz}) + 6\text{dB} = -91\text{dBm}$. The SNR of the received signal is equal to $P_{RX} - P_{\text{noise}} = 15\text{dB}$ for this typical scenario. We also consider a high SNR of 30dB in our experiments because numerical results in the high SNR regime could be used to corroborate our theoretical results regarding the uniqueness of the CP decomposition.

We first examine the channel estimation performance under different SNRs. Fig. 4 plots the estimation accuracy as a function of SNR. Note that the compressed sensing method requires to discretize the parameter space into a finite set of grid points, and the true parameters may not lie on the discretized grid. To illustrate the tradeoff between the estimation accuracy and the computational complexity for the CS method, we employ two different grids to discretize the continuous parameter space: the first grid (referred to as Grid-I) discretizes the AoA-AoD space into 64×32 grid points, and the second grid (referred to as Grid-II) discretizes the AoA-AoD space into 128×64 grid points. For our proposed CP method, after the factor matrices are estimated, a compressed sensing method is also used to estimate each user's channel. Nevertheless, since the problem has been decoupled into a set of single-user channel estimation problems via CP factorization, the size of the overcomplete dictionary involved in compressed sensing is now much smaller. Hence a finer grid can be employed. In our simulations, we use a grid of 256×128 for our proposed method. Table I shows that even using such a fine grid, our proposed method still consumes much less average run times as compared with the CS method which uses a grid of 128×64. From Fig. 4, we see that our proposed method presents a clear performance advantage over the CS method that employs the finer grid of the two choices. The performance gain is possibly due to the reason that the CP decomposition, which serves as a critical step of our method, enjoys a nice uniqueness theoretical guarantee and is more robust against noise. We also observe that the CS method achieves a performance improvement by employing a finer grid. Nevertheless, the required average runtime increases drastically when a finer grid is used (see Table I).

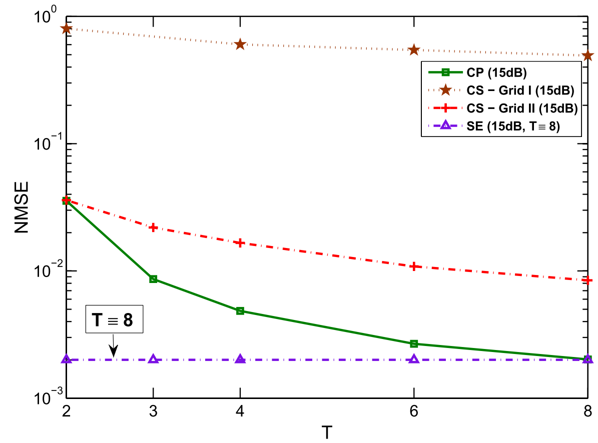
TABLE I
AVERAGE RUN TIMES OF RESPECTIVE ALGORITHMS, $T' = 16$, $M_{\text{BS}} = 16$, $T = 4$, AND SNR = 30dB

ALG	Grid	NMSE		Average Run Time(s)	
		Channel I	Channel II	Channel I	Channel II
CS	64×32	$2.5e-1$	$2.3e-1$	16.5	11
	128×64	$6.7e-3$	$6.4e-3$	270	220
CP	-	$2.7e-3$	$1.5e-3$	23	19

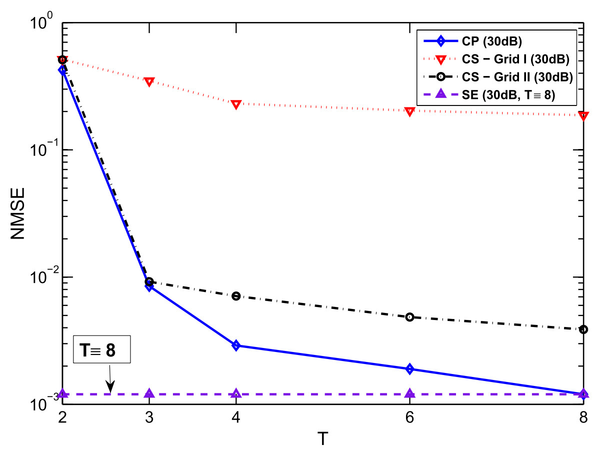
Next, we examine how the estimation performance depends on the parameters T, M_{BS} and T'. Fig. 5 shows the NMSEs of respective algorithms as a function of T, where the other two parameters T' and M_{BS} are fixed to be $T' = 16$ and $M_{\text{BS}} = 16$, and the signal-to-noise ratio (SNR) is set to 15dB and 30dB in our simulations. A separate estimation strategy (referred to as SE) was also included in Fig. 5 for comparison. The separate estimation strategy uses time multiplexed pilots to decompose the joint channel estimation into a number of single user's channel estimation sub-problems. For each sub-problem, a compressed sensing technique can be employed to estimate each user's channel. Clearly, for this separate estimation scheme, the number of required pilot symbols (i.e. frames) is equivalent to the number of users, i.e. $T = U$. The separate estimation strategy can serve as a reference method for evaluating the proposed joint estimation methods (including our proposed CP decomposition-based method and the compressed sensing-based technique developed in Section VI). Since $T' > L$ and $M_{\text{BS}} > L$, the generalized Kruskal's condition (61) can be satisfied when $k_S \geq 2$, that is, $T \geq 2$. From Fig. 5, we see that when $T > 2$, our proposed method is able to provide an accurate channel estimate under a high SNR. This result roughly coincides with our previous analysis regarding the uniqueness of the CP decomposition. Also, it can be seen that our proposed method achieves a substantial training overhead reduction at the expense of very mild performance loss: it requires only $T = 4$ pilot symbols to



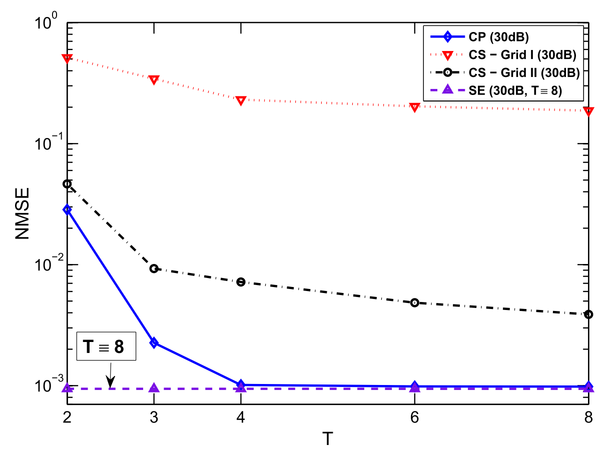
(a) Channel I



(b) Channel II

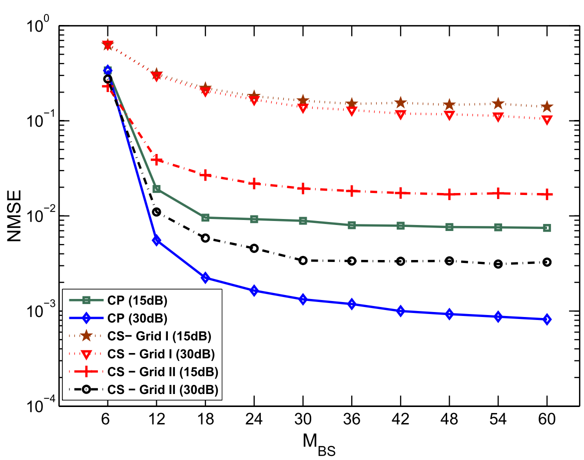


(c) Channel I

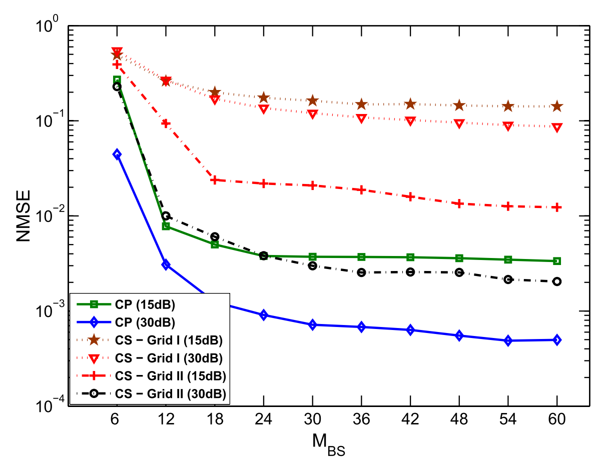


(d) Channel II

Fig. 5. NMSE versus T, $M_{\text{BS}} = 16$, $T' = 16$.



(a) Channel I

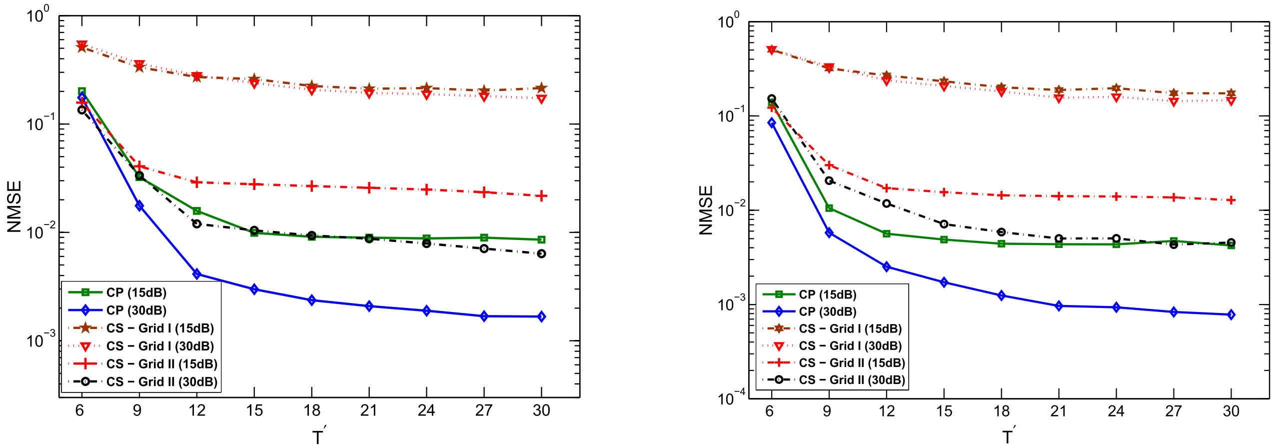


(b) Channel II

Fig. 6. NMSE versus M_{BS}, $T' = 16$, $T = 4$.

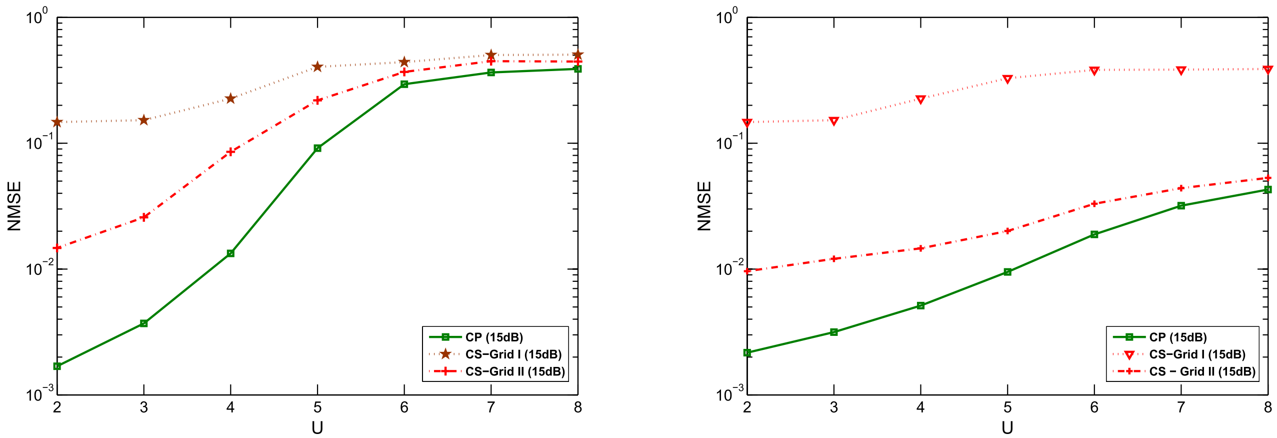
achieve performance similar to the separate estimation strategy that uses $T = U = 8$ pilot symbols. In addition, we observe that better estimation performance can be achieved for the latter mmWave channel. This is expected since the mutual coherence of the factor matrices $\boldsymbol{A}_Q$, $\boldsymbol{A}_P$ becomes lower as the AoAs/AoDs are more sufficiently separated. As a result, the CP factorization can be accomplished with a higher accuracy.

Fig. 6 depicts the NMSEs of respective algorithms as a function of M_{BS}, where we set $T' = 16$, $T = 4$, and the SNR is set to be 15dB and 30dB, respectively. To satisfy (61), it is



(a) Channel I (b) Channel II

Fig. 7. NMSE versus T', $M_{\text{BS}} = 16$, $T = 4$.



(a) Channel I (b) Channel II

Fig. 8. NMSE versus the number of users U, $M_{\text{BS}} = 16$, $T' = 16$, $T = 2$, SNR=15dB.

easy to know that M_{BS} should be greater than or equal to 11. From Fig. 6, we see that our simulation results again roughly corroborate our analysis: the transition in the NMSE curve for our proposed method takes place around $M_{\text{BS}} = 11$. Also, our proposed method outperforms the compressed sensing method by a considerable margin. In Fig. 7, we plot the estimation accuracy of respective algorithms as a function of T', where we set $M_{\text{BS}} = 16$ and $T = 4$. Similar conclusions can be made from this figure. The NMSEs of respective algorithms as a function of the number of users are depicted in Fig. 8, where we set $T = 2$, $T' = 16$, $M_{\text{BS}} = 16$, and SNR = 15dB. We see that with the same amount of training overhead, our proposed method is able to serve up to $U = 5$ users given a target NMSE of 0.01, whereas the compressed sensing method can only support $U = 2$ users in order to attain the same target estimation performance.

Table I shows the average run times of our proposed method and the compressed sensing method. We see that the computational complexity of the compressed sensing method grows dramatically as the dimension of the grid increases. Our proposed method is more computationally efficient than the compressed sensing method. It takes similar run times as the direct compressed sensing method which employs the coarser grid of the two choices, meanwhile achieving a better estimation accuracy than the compressed sensing method that uses the finer grid.

IX. Conclusions

We proposed a layered pilot transmission scheme and a CANDECOMP/PARAFAC (CP) decomposition-based method for uplink multiuser channel estimation in mm-Wave MIMO systems. The joint uplink multiuser channel estimation was formulated as a tensor decomposition problem. The uniqueness of the CP decomposition was investigated for both the single-path geometric model and the general geometric model. The conditions for the uniqueness of the CP decomposition shed light on the design of the beamforming matrix and the combining matrix, and meanwhile provide general guidelines for choosing the system parameters. The proposed method is able to achieve an additional training overhead reduction as compared with a conventional scheme which separately estimates multiple users' channels. Simulation results show that our proposed method presents a clear performance advantage over the compressed sensing method, and meanwhile achieving a substantial computational complexity reduction.

Appendix A
Proof of Theorem 1

We show that Theorem 1 is a special case of Theorem 2, i.e. Theorem 2 reduces to Theorem 1 when $L_r = 1, \forall r$. Note that when $L_r = 1$, the generalized k-rank is simplified as the k-rank. As a result, the generalized Kruskal's condition (59) in Theorem 2 reduces to the Kruskal's condition (52) in Theorem 1. Now we show that the condition (58) required by Theorem 2 is automatically satisfied when $L_r = 1, \forall r$. For the case $L_r = 1, \forall r$, the condition (58) becomes $MN \geq R$. Due to the fact that $k_{\boldsymbol{A}} \leq M$, $k_{\boldsymbol{B}} \leq N$ and $k_{\boldsymbol{C}} \leq R$, the Kruskal's condition (52) yields $M + N \geq R + 2$. In addition, from the fact $(M-1)(N-1) \geq 0$, we have $MN \geq M+N-1 \geq R+1$. Thus if the Kruskal's condition holds, then the condition (58) is automatically satisfied for the case $L_r = 1, \forall r$. It can also be easily verified that the conclusion in Theorem 2 is exactly identical to that in Theorem 1 when $L_r = 1, \forall r$.

Appendix B
Proof of Theorem 2

Theorem 2 corresponds to [33, Th. 4.7] . A rigorous proof can be found in [33]. Here we only provide a sketch of the proof. In summary, the proof of Theorem 2 proceeds in three steps. In the first step, the following result is proved.

Lemma 1: Let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ represent a decomposition of $\mathcal{X}$ in rank-$(L_r, L_r, 1)$ terms. Suppose the condition

$$k'_{\boldsymbol{A}} + k'_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2R + 2 \tag{83}$$

holds and that we have an alternative decomposition of $\mathcal{X}$, represented by $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$, with $k'_{\bar{\boldsymbol{A}}}$ and $k'_{\bar{\boldsymbol{B}}}$ maximal under the given dimensionality constraints. Then there holds $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}_a\boldsymbol{\Lambda}_a$, in which $\boldsymbol{\Pi}_a$ is a block permutation matrix and $\boldsymbol{\Lambda}_a$ a square nonsingular block-diagonal matrix, compatible with the block structure of $\boldsymbol{A}$. There also holds $\bar{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Pi}_b\boldsymbol{\Lambda}_b$, in which $\boldsymbol{\Pi}_b$ is a block permutation matrix and $\boldsymbol{\Lambda}_b$ a square nonsingular block-diagonal matrix, compatible with the block structure of $\boldsymbol{B}$.

Proof: This lemma corresponds to [33, Lemma 4.3]. Its proof can be found in [33]. ■

Based on Lemma 1, we show that under a generalized Kruskal's condition, the submatrices of $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ in an alternative decomposition of $\mathcal{X}$ are ordered in the same way. The results are summarized in the following lemma.

Lemma 2: Let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ represent a decomposition of $\mathcal{X}$ in rank-$(L_r, L_r, 1)$ terms. Suppose that we have an alternative decomposition of $\mathcal{X}$, represented by $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$, with $k'_{\bar{\boldsymbol{A}}}$ and $k'_{\bar{\boldsymbol{B}}}$ maximal under the given dimensionality constraints. If the condition

$$k'_{\boldsymbol{A}} + k'_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2R + 2 \tag{84}$$

holds, then $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_a$ and $\bar{\boldsymbol{B}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_b$, in which $\boldsymbol{\Pi}$ is a block permutation matrix and $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ nonsingular block-diagonal matrices, compatible with the block structure of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Proof: This lemma corresponds to [33, Lemma 4.6]. Its proof can be found in [33]. ■

We also introduce the following lemma before proceeding to prove Theorem 2.

Lemma 3: Consider partitioned matrices $\boldsymbol{A} = [\boldsymbol{A}_1 \ \cdots \ \boldsymbol{A}_R]$ with $\boldsymbol{A}_r$ of size $I \times L_r$, $1 \leq r \leq R$, and $\boldsymbol{B} = [\boldsymbol{B}_1 \ \cdots \ \boldsymbol{B}_R]$ with $\boldsymbol{B}_r$ of size $J \times M_r$, $1 \leq r \leq R$. Generically we have that $\text{rank}(\boldsymbol{A} \odot_b \boldsymbol{B}) = \min(IJ, \sum_{r=1}^{R} L_r M_r)$, where $\boldsymbol{A} \odot_b \boldsymbol{B} = [\boldsymbol{A}_1 \otimes \boldsymbol{B}_1 \ \ \boldsymbol{A}_2 \otimes \boldsymbol{B}_2 \ \cdots \ \boldsymbol{A}_R \otimes \boldsymbol{B}_R]$.

Proof: This lemma corresponds to [38, Lemma 3.3]. Its proof can be found in [38]. ■

Now we are ready to prove Theorem 2. From Lemma 2 we have that $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_a$ and $\bar{\boldsymbol{B}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_b$. Put the submatrices of $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ in the same order as the submatrices of $\boldsymbol{A}$ and $\boldsymbol{A}$. After reordering, we have $\bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Lambda}_a$, with $\boldsymbol{\Lambda}_a = \text{blockdiag}(\boldsymbol{\Lambda}_{a,1}, \ldots, \boldsymbol{\Lambda}_{a,R})$, and $\bar{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Lambda}_b$, with $\boldsymbol{\Lambda}_b = \text{blockdiag}(\boldsymbol{\Lambda}_{b,1}, \ldots, \boldsymbol{\Lambda}_{b,R})$. Then we have that

$$\begin{aligned} \boldsymbol{X}_{(3)}^T &= (\boldsymbol{B} \odot \boldsymbol{A})\text{blockdiag}(\boldsymbol{1}_{L_1}, \ldots, \boldsymbol{1}_{L_R})\boldsymbol{C}^T && (85) \\ &= (\boldsymbol{B} \odot_b \boldsymbol{A})\text{blockdiag}(\text{vec}(\boldsymbol{I}_{L_1}), \ldots, \text{vec}(\boldsymbol{I}_{L_R}))\boldsymbol{C}^T \\ &= (\bar{\boldsymbol{B}} \odot_b \bar{\boldsymbol{A}})\text{blockdiag}(\text{vec}(\boldsymbol{I}_{L_1}), \ldots, \text{vec}(\boldsymbol{I}_{L_R}))\bar{\boldsymbol{C}}^T \\ &= (\boldsymbol{B} \odot_b \boldsymbol{A})\text{blockdiag}(\text{vec}(\boldsymbol{\Lambda}_{a,1}\boldsymbol{\Lambda}_{b,1}^T), \ldots, \\ &\qquad \text{vec}(\boldsymbol{\Lambda}_{a,R}\boldsymbol{\Lambda}_{b,R}^T))\bar{\boldsymbol{C}}^T && (86) \end{aligned}$$

where $\boldsymbol{1}_{L_r}$ denotes a vector of size $L_r \times 1$ whose entries equal to 1 and $\boldsymbol{I}_{L_r}$ denotes an identity matrix of size $L_r \times L_r$.

From Lemma 3 we have that, under condition (58), $\boldsymbol{B} \odot_b \boldsymbol{A}$ is generically full column rank. Equation (86) then implies that there exist nonzero scalars α_r such that $\boldsymbol{\Lambda}_{a,1}\boldsymbol{\Lambda}_{b,1}^T = \alpha_r \boldsymbol{I}_{L_r}$ (i.e. $\boldsymbol{\Lambda}_{a,1} = \alpha_r \boldsymbol{\Lambda}_{b,1}^{-T}$) and $\boldsymbol{c}_r = \alpha_r \bar{\boldsymbol{c}}_r$. In other words, $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ and $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{C}})$ are equal up to trivial indeterminacies.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "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.

[7] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "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, QLD, Australia, Apr. 2015, pp. 2909–2913.

[10] 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.

[11] 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.

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

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

[14] 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.*, Oct. 2012, pp. 690–697.

[15] 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.

[16] 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.

[17] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

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

[19] 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.

[20] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.

[21] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1697–1710, May 2005.

[22] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5720–5735, Nov. 2010.

[23] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.

[24] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, Nov. 2010.

[25] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order svd-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

[26] 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.

[27] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[28] 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.

[29] 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.

[30] 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.

[31] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experim. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[32] T. Strohmer and R. W. Heath, Jr., "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.

[33] L. De 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.

[34] 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, Mar. 2009.

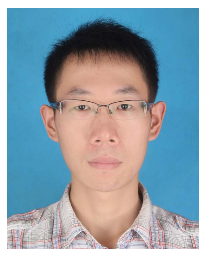
[35] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[36] A. Medra and T. N. Davidson, "Flexible codebook design for limited feedback systems via sequential smooth optimization on the Grassmannian manifold," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1305–1318, Mar. 2014.

[37] 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.

[38] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, Sep. 2008.

[39] 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.*, to be published.



Zhou Zhou received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2011. He is currently pursuing the Ph.D. degree with the National Key Laboratory on Communications. His current research interests include massive MIMO, millimeter-wave communications, sparse theory, and tensor analysis.

Jun Fang (M'08) received the B.S. and M.S. degrees from the 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 current research interests include sparse theory and compressed sensing, and Bayesian inference for data analysis.

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

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. 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. He was a Visiting Faculty Member with the Air Force Research Laboratory in 2003, 2004, and 2009. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li has been a member of the IEEE SPS Signal Processing Theory and Methods (2011-) Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC from 2006 to 2012. He has been an Associate Editor of the *Signal Processing* (Elsevier) since 2013, the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006 to 2009 and since 2014), the IEEE SIGNAL PROCESSING LETTERS from 2005 to 2006, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2003 to 2006, and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and the *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities. He served as a General Co-Chair of the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, NJ, USA, in 2012. 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 and the Jess H. Davis Memorial Award for excellence in research from the Stevens Institute of Technology in 2001, and the Sigma Xi Graduate Research Award from the University of Florida in 1999.

Zhi Chen received the B.Eng., M. Eng., and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC) in 1997, 2000, 2006, respectively, all in electrical engineering. 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 from 2010 to 2011. His current research interests include relay and cooperative communications, multi-user beamforming in cellular networks, interference coordination and cancellation, and THz 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.

Shaoqian Li (F'16) received the B.S.E. degree in communication technology from the Northwest Institute of Telecommunication, Xidian University in 1982, and the M.S.E. degree in communication system from the University of Electronic Science and Technology of China (UESTC) in 1984. He is currently a Professor, a Ph.D. Supervisor, the Director of National Key Laboratory of Communication, UESTC, and a member of National High Technology Research and Development Program (863 Program) Communications Group. His research includes wireless communication theory, anti-interference technology for wireless communications, spread-spectrum and frequency hopping technology, mobile, and personal communications.