

Article

PARAFAC-Based Multiuser Channel Parameter Estimation for MmWave Massive MIMO Systems over Frequency Selective Fading Channels

Rui Chang¹, Chaowei Yuan¹ and Jianhe Du^{2,*} 

¹ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; chiong@bupt.edu.cn (R.C.); yuancw2000@bupt.edu.cn (C.Y.)

² School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

* Correspondence: dujianhe1@gmail.com

Abstract: Channel estimation is crucial in millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems, especially with a few training sequences. To solve the problem of uplink channel estimation in mmWave massive MIMO systems, a PARAFAC-based algorithm is proposed for joint estimation of multiuser channels. The orthogonal frequency divisional multiplexing (OFDM) technique is exploited to combat the frequency selective fading channels. In this paper, the received signal at the base station (BS) is formulated as a third-order parallel factor (PARAFAC) tensor, and then a low-complexity algorithm is designed for fast estimation of the factor matrices related to channel parameters, thus leading to joint estimation of multiuser channel parameters via one-dimensional search. Moreover, the Cramér–Rao Bound (CRB) results for multiuser channel parameters are derived for evaluation. Theoretical analysis and numerical results reveal that the algorithm performs well with a few training sequences. Compared with existing algorithms, the proposed algorithm has clear advantages both in estimation accuracy and computational complexity.

Keywords: mmWave; massive MIMO; multiuser; sparsity; PARAFAC



Citation: Chang, R.; Yuan, C.; Du, J. PARAFAC-Based Multiuser Channel Parameter Estimation for MmWave Massive MIMO Systems over Frequency Selective Fading Channels. *Electronics* **2021**, *10*, 2983. <https://doi.org/10.3390/electronics10232983>

Academic Editor: Dinh-Thuan Do

Received: 1 November 2021

Accepted: 27 November 2021

Published: 30 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a key technology for 5G wireless communications, millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) has received significant attention in recent years [1,2]. mmWave techniques provide the ability of fast speed data delivery and wide bandwidth [3], making them valued in information ages. The large number of spatial degrees of freedom created by the massive MIMO contribute to high user throughput and high spectral efficiency for extended mobile broadband (eMBB). Furthermore, massive MIMO can be used to efficiently support ultra-reliable low-latency communication (URLLC) and massive machine-type communication (mMTC) in the various types of Internet of Things (IoT) connectivities [4–6]. In mmWave massive MIMO systems, large antenna arrays implemented at the base station (BS) and mobile station (MS) are utilized to supply adequate beamforming gains, which offers a compensation for high signal attenuation at mmWave frequencies [7–9]. In such a context, the number of radio frequency (RF) chains is much smaller than that of antennas due to lower hardware cost and power consumption. Considering the RF limitations, the hybrid analog/digital precoding scheme [10–14] is widely adopted to provide larger precoding gains. However, the overall channel state information is hard to observe because of massive antennas and hybrid precoding structures. In practice, mmWave channels exhibit the characteristic of broad-band frequency-selective fading (FSF) owing to the large bandwidth and distinct multipath delays [15]. While the whole bandwidth may be subject to FSF, each sub-band undergoes flat fading. When orthogonal frequency division multiplexing (OFDM) is integrated into massive

MIMO architectures, mmWave channels are viewed as the aggregation of a number of parallel flat fading channels, which urges us to achieve large multiplexing and diversity gains [16–18] with a large number of spatial degrees of freedom. Nevertheless, the overall system performance still hugely depends on the channel estimation, which is a tricky task for the channel subspace sampling limitation [19,20].

In the aspect of channel estimation in massive MIMO systems, several relevant schemes [21–25] have been developed in various scenarios. In general, massive MIMO channels are efficiently parameterized with a series of associated parameters including fading coefficients, angles of arrival or departure (AoAs/AoDs) and delays. Owing to the intrinsic sparsity of mmWave channels, some approaches [10], Refs. [26–29] based on compressing sensing (CS) techniques convert the issue of channel estimation into the recovery of a line sparse signal. For example, a conventional CS-based algorithm is proposed for channel estimation with much training overhead reduction in a multiuser mmWave system [28]. Further in the work [10], with a novel hierarchical multi-resolution codebook, an adaptive CS-based algorithm is proposed to estimate the channel parameters both in single-path and multipath mmWave environments. According to the CS-based methods, the continuous parameter space is discretized into a finite of grid points [30–32], and the estimated angles are assumed to lie on the prespecified grid points. In fact, actual angles may be not in accordance with this assumption, which seriously degrades the channel estimation performance. In addition, some tensor decomposition-based methods stand out in the fields of mmWave channel estimation. The work [33] achieves the joint channel parameter estimation via parallel factor (PARAFAC) analysis in massive MIMO systems. As the improvement of [33], Ref. [34] proposes a more accurate PARAFAC-based estimator in the presence of pilot contamination. For the joint multiuser channel estimation, [35] proposes a layered pilot transmission scheme and decomposes the problem of multiuser channel estimation into a number of problems of single-user channel estimation. In the aforementioned scenarios [33–35], simulations demonstrate that the tensor-based algorithms have better estimation performance compared with these CS-based methods. However, most of the existing works concentrate on the narrowband systems while realistic mmWave channels exhibit the broad-band FSF. The work [36] proposes a simultaneous orthogonal matching pursuit (SOMP) scheme and regards the broadband mmWave channel estimation as a multiple measurement vector (MMV) problem with a common support (i.e., the channel support at different frequencies is assumed to be the same). Unlike the proposed scheme in [36], the work [37] estimates the support of the angle-domain channel at some frequencies independently by the orthogonal matching pursuit (OMP) algorithm and combines them into the common support at all frequencies. Moreover, a CANDECOMP/PARAFAC (CP) decomposition-based method for downlink channel estimation is proposed in a mmWave MIMO-OFDM system, where the received signal at individual user is formulated as a low-rank PARAFAC tensor. So far, to the best of our knowledge, there are few tensor decomposition-based methods for joint multiuser mmWave channel estimation, especially over the FSF channels.

In our paper, we consider the broad-band FSF channel and address the problem of multiuser uplink channel estimation in a massive MIMO system, where the hybrid precoding strategy is adopted. Due to the sparsity of mmWave channels, a PARAFAC decomposition-based scheme is proposed for multiuser mmWave channel estimation. In the successive time slots, all the MSs simultaneously transmit respective pilots to a common BS, and the signals received at the BS can be fitted in a low-rank PARAFAC structure. Table 1 presents some reviewed works along their main distinctive characteristics to better position our present work. Unlike the system [38] where individual users are considered for downlink transmission, our proposed algorithm concentrates on the joint estimation of multiuser uplink channels, which leads to wider identifiability and more relaxed requirements for unique PARAFAC decomposition. The main contributions of our work are summarized as follows.

(1) For the sparsity of mmWave FSF channels, the signals transmitted from multiple users can be modeled as a low-rank PARAFAC structure at the BS, which promotes the joint channel estimation in contrast with [38]. It's worth noting that the massive MIMO channel is characterized with path gains, AoAs/AoDs and delays. This tensor modeling not only takes full advantage of diversities in space, time and frequency dimensions, but also has the benefits of multidimensional structures.

(2) In this context, a PARAFAC decomposition-based algorithm is proposed for accurate estimates of multiuser mmWave channels parameters including path gains, AoAs/AoDs and delays. The proposed algorithm consists of two stages. In the first stage, the accelerated trilinear alternating least squares (ATALS) algorithm is put forward for efficient PARAFAC decomposition. In the second stage, we make further efforts to extract unknown channel parameters of multiple users from factor matrices. The uniqueness of PARAFAC decomposition is also studied, which gives the guidance on the design of precoding and combining matrices. What's more, our proposed scheme has wider identifiability in terms of unique PARAFAC decomposition.

(3) To provide a benchmark for system evaluation, the Cramér -Rao Bound (CRB) results for channel parameters of multiple users are derived. Simulations reveal that the mean square errors (MSEs) of channel parameters obtained by our algorithm are nearly close to their corresponding CRB results, which indicates our proposed PARAFAC-based algorithm yields accurate channel estimation. While our proposed scheme is put forward at the assumption of uniform linear array (ULA) antennas, it is also suitable for uniform planar array (UPA) antennas.

(4) The computational complexity of the proposed scheme is analyzed. For comparison, we also analyze the computational complexities of some CS-based channel estimation schemes including the SOMP [36], OMP [37] and adaptive CS [10] algorithms. Numerical results demonstrate that our proposed algorithm outperforms its counterparts in estimation accuracy and computational complexity, particularly with a substantial training overhead reduction. Specially, our proposed scheme still performs well, even though the number of users is large.

Table 1. Some reviewed channel estimation works.

| Works | System Scenario | Algorithm |
|----------|---|--|
| [36] | multiuser mmWave massive MIMO-OFDM system | SOMP |
| [37] | mmWave massive MIMO-OFDM system | OMP |
| [38] | mmWave massive MIMO-OFDM system | adaptive CS |
| [10] | mmWave massive MIMO system | CP decomposition-based method |
| our work | multiuser mmWave massive MIMO-OFDM system | PARAFAC-based channel parameter estimation |

The rest of this paper is organized as follows. Section 2 first presents the fundamentals of PARAFAC model and formulates the PARAFAC model of multiuser mmWave massive MIMO-OFDM architecture. The uniqueness issue of constructed PARAFAC model is also discussed in Section 2. In Section 3, a two-staged PARAFAC-based algorithm is proposed for multiuser mmWave channel estimation. The CRB derivation are given in Section 4. Section 5 presents some numerical results for the system performance analysis. Finally, some conclusions are drawn in Section 6.

Notation : Scalars, vectors, matrices and tensors are denoted by lowercase, boldface lowercase, boldface uppercase and calligraphic uppercase letters, e.g., a , \mathbf{a} , \mathbf{A} and \mathcal{A} , respectively. \mathbf{A}^T , \mathbf{A}^H , \mathbf{A}^* and \mathbf{A}^\dagger denote the transpose, conjugate transpose, inverse and pseudo-inverse of matrix \mathbf{A} , respectively. The operators \odot , \otimes and \circ denote the Khatri-Rao, Kronecker and outer products, respectively. \mathbf{I}_K denotes the $K \times K$ identity matrix and $\|\cdot\|_F$ denotes the Frobenius norm. The operator $\text{diag}(\mathbf{a})$ constructs a diagonal matrix with the vector \mathbf{a} and the operator $\text{vec}(\cdot)$ changes a matrix into a column vector by stacking the columns. For convenience, all the abbreviations used in this paper are given in Abbreviations .

2. System Model

As shown in Figure 1, we consider a multiuser mmWave massive MIMO-OFDM system composed of BS and U MSs. The BS is equipped with N antennas and R RF chains, and the u -th MS is equipped with M_u antennas and $R_u = 1$ RF chain. Since hybrid precoding structure is adopted in this scenario, the number of RF chains is less than that of antennas, i.e., $R \leq N$, $R_u \leq M_u$. Specially, the number of antennas for each MS is assumed to be equal, i.e., $M_u = M$, $\forall u = 1, \dots, U$. The number of subcarriers for transmission is assumed to be K .

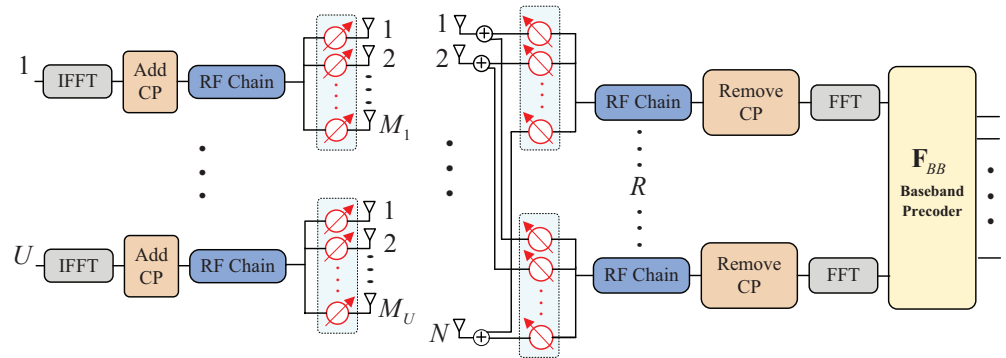


Figure 1. Architecture of multiuser mmWave massive MIMO-OFDM system.

In the uplink, all the MSs simultaneously transmit respective pilot signals to the common BS via a set of K subcarriers. As mmWave channels exhibit the inherent feature of sparsity, a geometric channel model with finite scatterings is utilized to account for mmWave channels. In terms of the uplink transmission via the k -th subcarrier, the channel of L_u scatterings from the u -th MS to BS can be expressed as $\mathbf{H}_{u,k} \in \mathbb{C}^{N \times M}$ in the frequency domain

$$\mathbf{H}_{u,k} = \sum_{l_u=1}^{L_u} \mathbf{H}_{u,l_u} \exp(-j2\pi f_s \tau_{u,l_u} k/K), \quad (1)$$

with the l_u -th path

$$\mathbf{H}_{u,l_u} = \beta_{u,l_u} \mathbf{a}_{BS}(\varphi_{u,l_u}) \mathbf{a}_{MS}^T(\psi_{u,l_u}) \in \mathbb{C}^{N \times M}, \quad (2)$$

for $l_u = 1, \dots, L_u$, $u = 1, \dots, U$, where f_s is the sampling rate, L_u denotes the number of propagations and τ_{u,l_u} denotes the time delay of the l_u -th path. In this setting of L_u scatterings, a single propagation path \mathbf{H}_{u,l_u} is parameterized by a set of channel parameters $(\beta_{u,l_u}, \varphi_{u,l_u}, \psi_{u,l_u})$, where β_{u,l_u} , φ_{u,l_u} and ψ_{u,l_u} denote the complex gain, AoA and AoD of the l_u -th path, respectively. The vectors $\mathbf{a}_{BS}(\varphi_{u,l_u})$ and $\mathbf{a}_{MS}(\psi_{u,l_u})$ are the receive and transmit steering vectors, respectively. For the ULA, the steering vectors are given by

$$\mathbf{a}_{BS}(\varphi_{u,l_u}) = \frac{1}{\sqrt{N}} \left[1, e^{j(2\pi/\lambda)d \sin(\varphi_{u,l_u})}, \dots, e^{j(N-1)(2\pi/\lambda)d \sin(\varphi_{u,l_u})} \right]^T \in \mathbb{C}^{N \times 1}, \quad (3)$$

$$\mathbf{a}_{MS}(\psi_{u,l_u}) = \frac{1}{\sqrt{M}} \left[1, e^{j(2\pi/\lambda)d \sin(\psi_{u,l_u})}, \dots, e^{j(M-1)(2\pi/\lambda)d \sin(\psi_{u,l_u})} \right]^T \in \mathbb{C}^{M \times 1}, \quad (4)$$

where λ is signal wavelength and d is the antenna spacing. Assume that $\mathbf{A}_{BS,u} = [\mathbf{a}_{BS}(\varphi_{u,1}), \dots, \mathbf{a}_{BS}(\varphi_{u,L_u})] \in \mathbb{C}^{N \times L_u}$, $\mathbf{A}_{MS,u} = [\mathbf{a}_{MS}(\psi_{u,1}), \dots, \mathbf{a}_{MS}(\psi_{u,L_u})] \in \mathbb{C}^{M \times L_u}$, $\boldsymbol{\beta}_{u,k} = [\beta_{u,1,k}, \dots, \beta_{u,L_u,k}] \in \mathbb{C}^{1 \times L_u}$ with the term $\beta_{u,l_u,k} = \beta_{u,l_u} \exp(-j2\pi f_s \tau_{u,l_u} k/K)$, we derive

$$\mathbf{H}_{u,k} = \mathbf{A}_{BS,u} \text{diag}(\boldsymbol{\beta}_{u,k}) \mathbf{A}_{MS,u}^T \quad (5)$$

2.1. PARAFAC Model

In this subsection, we give a brief overview of PARAFAC model and discuss the unique issue of PARAFAC decomposition. The PARAFAC decomposition amounts to

decomposing the N -way tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \dots \times I_N}$ into a sum of components, each component being a rank-one tensor. For a three-way tensor $\mathcal{X} \in \mathbb{C}^{I \times J \times K}$ of R components, PARAFAC decomposition can be expressed as

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \Leftrightarrow [\mathcal{X}]_{i,j,k} = \sum_{r=1}^R [\mathbf{A}]_{i,r} [\mathbf{B}]_{j,r} [\mathbf{C}]_{k,r}, \quad (6)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_R]$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R]$. For brevity, we have the expression $\mathcal{X} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$. Alternatively, it can be written as the mode- n product:

$$\mathcal{X} = \mathcal{I}_{3,R} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}, \quad (7)$$

where $\mathcal{I}_{3,R}$ is the three-way identity tensor of dimensions $R \times R \times R$ which represents a diagonal hypercubic tensor with its nonzero elements $\delta_{i,j,k} = 1$ when and only when $i = j = k$. Furthermore, slicing \mathcal{X} along each dimension comes to three distinct slabs $\mathcal{X}(i, :, :) = \mathbf{B} \mathbf{D}_i(\mathbf{A}) \mathbf{C}^T$, $\mathcal{X}(:, j, :) = \mathbf{A} \mathbf{D}_j(\mathbf{B}) \mathbf{C}^T$ and $\mathcal{X}(:, :, k) = \mathbf{A} \mathbf{D}_k(\mathbf{C}) \mathbf{B}^T$. Arranging the slices in some manner leads to the three corresponding matrix unfoldings: $\mathbf{X}^{(1)} = (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \in \mathbb{C}^{KJ \times I}$, $\mathbf{X}^{(2)} = (\mathbf{C} \odot \mathbf{A}) \mathbf{B}^T \in \mathbb{C}^{IK \times J}$ and $\mathbf{X}^{(3)} = (\mathbf{B} \odot \mathbf{A}) \mathbf{C}^T \in \mathbb{C}^{IJ \times K}$.

Extended to a N -way tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \dots \times I_N}$ of rank R , PARAFAC analysis has the form:

$$\mathcal{X} = \mathcal{I}_{N,R} \times_{n=1}^N \mathbf{A}^{(n)}, \quad (8)$$

with its element $[\mathcal{X}]_{i_1, \dots, i_N} = \sum_{r=1}^R \prod_{n=1}^N [\mathbf{a}_r^{(n)}]_{i_n}$, where $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$, $i_n = 1, \dots, I_n$, $n = 1, \dots, N$, $r = 1, \dots, R$. Furthermore, the columns of matrix $\mathbf{A}^{(n)}$ are assumed to be normalized, i.e., with a unit norm, for $1 \leq n \leq N$, we derive that

$$[\mathcal{X}]_{i_1, \dots, i_N} = \sum_{r=1}^R g_r \prod_{n=1}^N [\mathbf{A}^{(n)}]_{i_n, r}, \quad (9)$$

where the identity tensor $\mathcal{I}_{N,R}$ is replaced by the diagonal tensor $\mathcal{G} \in \mathbb{C}^{R \times \dots \times R}$ whose nonzero elements are equal to scaling factors g_r lying on its diagonal.

Then we investigate the uniqueness of PARAFAC- N model in (9). Since the condition for essential uniqueness of PARAFAC model involves the notion of k -rank of a matrix, we first recall the definition of k -rank. The k -rank of a matrix $\mathbf{A} \in \mathbb{C}^{I \times R}$ is denoted by $k_{\mathbf{A}}$, where $k_{\mathbf{A}} = r$ on condition that any r columns of \mathbf{A} are linearly independent, where k -rank stands for Kruskal-rank. It can be shown by using the identifiability theorem for the PARAFAC- N model if

$$\sum_{r=1}^R k_{\mathbf{A}^{(n)}} \geq 2R + (N - 1), \quad (10)$$

then $\mathcal{X} \in \mathbb{C}^{I_1 \times \dots \times I_N}$ is essentially unique and its loading matrices $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$, $n = 1, \dots, N$ are unique to column permutation and scaling. Hence, there exists an alternative set of matrices $\hat{\mathbf{A}}^{(n)} \in \mathbb{C}^{I_n \times R}$ with the relation $\hat{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)} \mathbf{\Pi} \mathbf{\Delta}^{(n)}$, for $n = 1, \dots, N$, where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{\Delta}^{(n)}$ is a nonsingular diagonal matrix such as $\prod_{n=1}^N \mathbf{\Delta}^{(n)} = \mathbf{I}_R$.

2.2. Constructed PARAFAC Model

According to the upstream communication protocol, multiple MSs encode their respective pilots to the BS through the FSF channels via K subcarriers in P successive time slots. At the destination, the BS employs the combining vectors to detect the transmitted signals from the MSs. Define $s_{u,k} \in \mathbb{C}^{R_u \times 1}$ as the pilot for the u -th MS with k -th the subcarrier,

and $\mathbf{e}_{u,k}^{(p)} \in \mathbb{C}^{M \times R_u}$ as the precoding vector in the p -th time slot, the signal vector received at the BS in the p -th time slot can be expressed as

$$\mathbf{x}_k^{(p)} = \sum_{u=1}^U \mathbf{H}_{u,k} \mathbf{e}_{u,k}^{(p)} s_{u,k} + \mathbf{v}_k^{(p)}, \quad (11)$$

where $\mathbf{v}_k^{(p)}$ is additive white Gaussian noise. Considering the hybrid precoding structure, the BS employs the combining vectors $\{\mathbf{f}_{r,k}^{(p)}\}_{r=1}^R$ to detect the received signal $\mathbf{x}_k^{(p)}$. For the r -th RF chain at the BS, we derive

$$y_{r,k}^{(p)} = \mathbf{f}_{r,k}^{(p)T} \sum_{u=1}^U \mathbf{H}_{u,k} \mathbf{e}_{u,k}^{(p)} s_{u,k} + v_{r,k}^{(p)}, \quad (12)$$

where $y_{r,k}^{(p)}$ is the received signal observed from the r -th RF via the k -th subcarrier in the p -th time slot, and $v_{r,k}^{(p)}$ is additive White Gaussian noise. Arrange the R combining vectors at the BS to form the composite combining matrix $\mathbf{F}_k^{(p)} \in \mathbb{C}^{R \times N}$, i.e., $\mathbf{F}_k^{(p)} = [\mathbf{f}_{1,k}^{(p)}, \dots, \mathbf{f}_{R,k}^{(p)}]^T$, which is composed of baseband combiner $\mathbf{F}_{BB,k}^{(p)} \in \mathbb{C}^{R \times R}$ and RF combiner $\mathbf{F}_{RF}^{(p)} \in \mathbb{C}^{N \times R}$, i.e., $\mathbf{F}_k^{(p)} = \mathbf{F}_{RF}^{(p)} \mathbf{F}_{BB,k}^{(p)} \in \mathbb{C}^{N \times R}$. Similarly, the combined precoder $\mathbf{e}_{u,k}^{(p)}$ at the u -th MS is composed of baseband precoder $\mathbf{e}_{MM,u,k}^{(p)} \in \mathbb{C}^{R_u \times R_u}$ and RF precoder $\mathbf{e}_{RF,u}^{(p)} \in \mathbb{C}^{M \times R_u}$, i.e., $\mathbf{e}_{u,k}^{(p)} = \mathbf{e}_{RF,u}^{(p)} \mathbf{e}_{MM,u,k}^{(p)} \in \mathbb{C}^{M \times R_u}$. It should be mentioned that the elements of RF precoding/combining matrices $\mathbf{e}_{RF,u}^{(p)}$ and $\mathbf{F}_{RF}^{(p)}$ meet the constant modulus property. In addition, all K subcarriers share the same RF precoding/combining in consideration of flat analog precoders which are fixed over the band.

Our goal is to achieve reliable estimation of channel parameters $(\beta_{u,l_u}, \varphi_{u,l_u}, \psi_{u,l_u}, \tau_{u,l_u})$ from $y_{r,k}^{(p)}$ with as few pilots as possible. Since the elements of matrices $\mathbf{e}_{u,k}^{(p)}$ and $\mathbf{F}_k^{(p)}$ are randomly chosen from a unit circle, we assume that $\mathbf{e}_{u,k}^{(p)} = \mathbf{e}^{(p)}$ and $\mathbf{F}_k^{(p)} = \mathbf{F}$. Substituting (5) into (12) yields

$$\begin{aligned} y_{r,k}^{(p)} &= \mathbf{f}_r^T \sum_{u=1}^U \mathbf{A}_{BS,u} \text{diag}(\beta_{u,k}) \mathbf{A}_{MS,u}^T \mathbf{e}^{(p)} s_{u,k} + v_{r,k}^{(p)} \\ &= \mathbf{f}_r^T \mathbf{A}_{BS} \text{diag}(\beta_k) \mathbf{A}_{MS}^T \mathbf{e}^{(p)} + v_{r,k}^{(p)}, \end{aligned} \quad (13)$$

where $\Gamma = \sum_{u=1}^U L_u$, $\mathbf{A}_{BS} = [\mathbf{A}_{BS,1}, \dots, \mathbf{A}_{BS,U}] \in \mathbb{C}^{N \times \Gamma}$, $\mathbf{A}_{MS} = [\mathbf{A}_{MS,1}, \dots, \mathbf{A}_{MS,U}] \in \mathbb{C}^{M \times \Gamma}$, $\beta_k = [\beta_{1,k}, \beta_{2,k}, \dots, \beta_{U,k}] \in \mathbb{C}^{1 \times \Gamma}$. Define $\mathbf{Z}_k = \mathbf{A}_{BS} \text{diag}(\beta_k) \mathbf{A}_{MS}^T$, \mathbf{Z}_k corresponds to the frontal slice of third order PARAFAC model $\mathcal{Z} \in \mathbb{C}^{N \times M \times K}$. We derive that

$$\mathcal{Z} = \mathcal{I}_{3,R \times 1} \mathbf{A}_{BS} \times_2 \mathbf{A}_{MS} \times_3 \mathbf{B}, \quad (14)$$

where $\mathbf{B} = [\beta_1^T, \dots, \beta_K^T]^T \in \mathbb{C}^{K \times \Gamma}$. Clearly, the aggregate channel tensor $\mathcal{Z} \in \mathbb{C}^{N \times M \times K}$ includes all the channels of U MSs via K subcarriers with its three modes standing for the number of antennas at the BS, antennas at the MS and subcarriers for transmission.

Collecting the received signal $y_{r,k}^{(p)}$ from R RF chains contributes to $\mathbf{y}_k^{(p)} = [y_{1,k}^{(p)}, \dots, y_{R,k}^{(p)}]^T \in \mathbb{C}^{R \times 1}$, and arranging the vectors $\{\mathbf{y}_k^{(1)}, \dots, \mathbf{y}_k^{(P)}\}$ in columns leads to the received signal at the k -th subcarrier. By concatenating the received data via K subcarriers, a three-way signal tensor $\mathcal{Y} \in \mathbb{C}^{R \times P \times K}$ can be expressed as

$$\mathcal{Y} = \mathcal{I}_{3,R \times 1} \mathbf{F}_{BS} \times_2 \mathbf{E}_{MS} \times_3 \mathbf{B} + \mathcal{V}, \quad (15)$$

where $\mathbf{F}_{BS} = \mathbf{F}\mathbf{A}_{BS} \in \mathbb{C}^{R \times \Gamma}$, $\mathbf{B} \in \mathbb{C}^{K \times \Gamma}$, $\mathbf{E}_{MS} = \mathbf{E}^T \mathbf{A}_{MS} \in \mathbb{C}^{P \times \Gamma}$ and $\mathbf{E} = [\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(P)}] \in \mathbb{C}^{M \times P}$. Notice that $\mathcal{Y} \in \mathbb{C}^{R \times P \times K}$ and $\mathcal{V} \in \mathbb{C}^{R \times P \times K}$ are the signal and noise tensors characterized with three dimensions standing for numbers of RF chains, time slots and subcarriers in space, time and frequency domains, respectively. Moreover, the three forms of unfolding $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ and $\mathbf{Y}^{(3)}$ are given by

$$\begin{aligned}\mathbf{Y}^{(1)} &= (\mathbf{B} \odot \mathbf{E}_{MS}) \mathbf{F}_{BS}^T + \mathbf{V}^{(1)} \in \mathbb{C}^{PK \times R}, \\ \mathbf{Y}^{(2)} &= (\mathbf{B} \odot \mathbf{F}_{BS}) \mathbf{E}_{MS}^T + \mathbf{V}^{(2)} \in \mathbb{C}^{RK \times P}, \\ \mathbf{Y}^{(3)} &= (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \mathbf{B}^T + \mathbf{V}^{(3)} \in \mathbb{C}^{RP \times K},\end{aligned}\quad (16)$$

where the slices $\mathbf{V}^{(1)} \in \mathbb{C}^{PK \times R}$, $\mathbf{V}^{(2)} \in \mathbb{C}^{RK \times P}$ and $\mathbf{V}^{(3)} \in \mathbb{C}^{RP \times K}$ represent the three forms of unfolding of tensor $\mathcal{V} \in \mathbb{C}^{R \times P \times K}$.

Since the number of resolvable paths K_{path} from MS to BS approximately satisfies the Poisson distribution $K_{\text{path}} \sim \max\{\text{Poisson}(\lambda), 1\}$ with $\lambda = 1.8$ when mmWave works at 28 GHz [39], each MS can be assumed to have 1 or 2 scatterings in the mmWave communications. Thus, the value of Γ is usually relatively smaller to the value of each dimension $\{R, P, K\}$, which enables the signal tensor $\mathcal{Y} \in \mathbb{C}^{R \times P \times K}$ to exhibit a low-rank PARAFAC structure.

2.3. Uniqueness Issue

According to the above derivations, it has been proven that the received signal tensor $\mathcal{Y} \in \mathbb{C}^{R \times P \times K}$ has a low-rank PARAFAC structure. In this subsection, we study the uniqueness of PARAFAC analysis and provide some sufficient conditions for the unique decomposition of $\mathcal{Y} \in \mathbb{C}^{R \times P \times K}$.

Inequality (10) establishes the sufficient condition for the uniqueness of PARAFAC- N model. In the case of our mmWave massive MIMO architecture, if the sum of k -ranks satisfies

$$k_{\mathbf{F}_{BS}} + k_{\mathbf{E}_{MS}} + k_{\mathbf{B}} \geq 2\Gamma + 2, \quad (17)$$

there exists the unique triple $(\hat{\mathbf{F}}_{BS}, \hat{\mathbf{E}}_{MS}, \hat{\mathbf{B}})$ up to permutation and scaling ambiguities, i.e., $\hat{\mathbf{F}}_{BS} = \mathbf{F}_{BS} \mathbf{\Pi} \mathbf{\Delta}^{(1)}$, $\hat{\mathbf{E}}_{MS} = \mathbf{E}_{MS} \mathbf{\Pi} \mathbf{\Delta}^{(2)}$, $\hat{\mathbf{B}} = \mathbf{B} \mathbf{\Pi} \mathbf{\Delta}^{(3)}$, where $\mathbf{\Delta}^{(1)} \mathbf{\Delta}^{(2)} \mathbf{\Delta}^{(3)} = \mathbf{I}$ and $\hat{\mathbf{F}}_{BS}$, $\hat{\mathbf{E}}_{MS}$ and $\hat{\mathbf{B}}$ are unique estimates of \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} . In the generic case, the sufficient condition becomes

$$\min(R, \Gamma) + \min(P, \Gamma) + \min(K, \Gamma) \geq 2\Gamma + 2. \quad (18)$$

This indicates that matrices \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} must be of full k -rank for the reason that the k -rank is always less than or equal to the rank, i.e., $k_{\mathbf{A}} \leq \text{rank}(\mathbf{A}) \leq \min(I, J)$, for any matrix $\mathbf{A} \in \mathbb{C}^{I \times J}$.

Then we make a profound study of the k -ranks of $(\mathbf{F}_{BS}, \mathbf{E}_{MS}, \mathbf{B})$. Considering the ULA implemented at the BS and MSs, matrices \mathbf{A}_{BS} and \mathbf{A}_{MS} have the Vandermonde structure, and thus we have $k_{\mathbf{A}_{BS}} = \min(N, \Gamma)$, $k_{\mathbf{A}_{MS}} = \min(M, \Gamma)$. For the requirement of full k -rank of \mathbf{F}_{BS} and \mathbf{E}_{MS} , an ingenious design of combining matrix \mathbf{F} and precoding matrix \mathbf{E} becomes an indispensable factor. In practice, both \mathbf{F}_{RF} and $\mathbf{e}_{RF}^{(p)}$ are accomplished by analog phase shifters, which only change the phase of signals. Therefore, the constant modulus constraint is imposed on the elements of \mathbf{F}_{RF} and $\mathbf{e}_{RF}^{(p)}$. Since the total transmission power constraint is enforced by normalizing $\mathbf{e}_{MM}^{(p)}$ such that $\|\mathbf{e}_{RF}^{(p)} \mathbf{e}_{MM}^{(p)}\|_F^2 = M$, we assume that the entries of combining matrix \mathbf{F} and precoding matrix \mathbf{E} are chosen uniformly from a unit circle by a constant, i.e., $[\mathbf{F}]_{i_1, j_1} = \frac{1}{\sqrt{N}} e^{jv_{i_1, j_1}}$, $1 \leq i_1 \leq R$, $1 \leq j_1 \leq N$, $[\mathbf{E}]_{i_2, j_2} = \frac{1}{\sqrt{M}} e^{jv_{i_2, j_2}}$, $1 \leq i_2 \leq M$, $1 \leq j_2 \leq P$, where v_{i_1, j_1} and v_{i_2, j_2} follow the independent and identically distribution (i.i.d) uniform distribution $\mathcal{U} \sim [0, 2\pi)$. Thus, the entries of \mathbf{F}_{BS} and \mathbf{E}_{MS} are viewed as i.i.d Gaussian variables, and we have $k_{\mathbf{F}_{BS}} = \min(R, \Gamma)$, $k_{\mathbf{E}_{MS}} = \min(P, \Gamma)$.

In addition, due to the shift-invariant vector $\mathbf{d}(\tau_{u,l_u})$ of all U MSs, \mathbf{B} is full k -rank, i.e., $k_{\mathbf{B}} = \min(K, \Gamma)$, where $\mathbf{d}(\tau_{u,l_u}) = [\exp(-j2\pi f_s \tau_{u,l_u}/K), \dots, \exp(-j2\pi f_s \tau_{u,l_u} K/K)]^T$ can be seen as a frequency-domain steering vector pointing towards the delay τ_{u,l_u} .

In practice, the number of propagation paths Γ is largely limited for mmWave circumstances of poor scatterings. Moreover, the scatterings appear in groups with similar delays, AoAs and AoDs because of massive antennas, which limits effective numbers of active paths. Generally, the number of subcarriers K is larger than Γ , which turns the uniqueness condition into

$$\begin{aligned} & \min(R, \Gamma) + \min(P, \Gamma) + \min(K, \Gamma) \\ & = \min(R, \Gamma) + \min(P, \Gamma) + \Gamma \geq 2\Gamma + 2 \rightarrow \min(R, \Gamma) + \min(P, \Gamma) \geq \Gamma + 2. \end{aligned} \quad (19)$$

From this condition, we can conclude that the BS has the potential to achieve the accurate channel estimation in as few as two successive time slots, with the help of Γ combining vectors. Hence, our proposed scheme can realize the accurate multiuser channel estimation with much reduced training overhead. However, a slightly larger pair of values for $\{R, P\}$ is chosen for better performance in our simulations.

3. Proposed PARAFAC Decomposition-Based Channel Estimation Scheme

In the presence of inescapable white noise, we are seeking for some appropriate solutions to the channel estimation issue. To make the most of low-rank PARAFAC structure of tensor \mathcal{Y} , an efficient and low-complexity multiuser channel estimation via PARAFAC decomposition is put forward. The algorithm is divided into two stages. In the first stage, three factor matrices $(\hat{\mathbf{F}}_{BS}, \hat{\mathbf{E}}_{MS}, \hat{\mathbf{B}})$ including channel parameters are uniquely decomposed from \mathcal{Y} by the proposed ATALS algorithm. In the second stage, channel parameters $(\hat{\beta}_{u,l_u}, \hat{\phi}_{u,l_u}, \hat{\psi}_{u,l_u}, \hat{\tau}_{u,l_u})$ are extracted from these factor matrices through the one-dimensional search method.

3.1. The ATALS Algorithm

As the workhorse technique for low-rank decomposition of three and higher dimensions, the alternating least squares (ALS) algorithm is widely applied in the field of PARAFAC analysis. In order to estimate the channel associated matrices \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} in an efficient way, the ATALS algorithm is proposed as a speed-up version of the classical ALS algorithm. In our simulations, all the channel state information is assumed to be unknown from the MSs or the BS, and the numbers of propagation paths at each MS, the matrices \mathbf{E} and \mathbf{F} are provided in advance.

In terms of the ALS algorithm, the criteria is that when estimating each matrix in the LS algorithm, the other matrices should maintain the previous estimation until the following cost function converges to the minimum value. At the i -th iteration, the cost function $\delta^{(i)}$ is expressed as:

$$\delta^{(i)} = \left\| \mathbf{Y}^{(1)} - (\hat{\mathbf{B}}^{(i)} \odot \hat{\mathbf{E}}_{MS}^{(i)}) \mathbf{F}_{BS}^{(i)T} \right\|_F^2, \quad (20)$$

where $\hat{\mathbf{F}}_{BS}^{(i)}$, $\hat{\mathbf{E}}_{MS}^{(i)}$ and $\hat{\mathbf{B}}^{(i)}$ are the estimated values of \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} acquired from the i -th iteration, respectively. At the i -th iteration, the factor matrices are updated as follows:

$$\mathbf{F}_{BS}^{(i)} = \left[(\hat{\mathbf{B}}^{(i-1)} \odot \hat{\mathbf{E}}_{MS}^{(i-1)})^\dagger \mathbf{Y}^{(1)} \right]^T, \mathbf{E}_{MS}^{(i)} = \left[(\hat{\mathbf{B}}^{(i-1)} \odot \hat{\mathbf{F}}_{BS}^{(i)})^\dagger \mathbf{Y}^{(2)} \right]^T, \mathbf{B}^{(i)} = \left[(\hat{\mathbf{E}}_{MS}^{(i)} \odot \hat{\mathbf{F}}_{BS}^{(i)})^\dagger \mathbf{Y}^{(3)} \right]^T \quad (21)$$

where $\hat{\mathbf{F}}_{BS}^{(i-1)}$, $\hat{\mathbf{E}}_{MS}^{(i-1)}$ and $\hat{\mathbf{B}}^{(i-1)}$ are the estimated values of \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} at the $(i-1)$ -th iteration, respectively. However, in some cases, simulations reveal that the convergence is slow and the estimates of \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} show an increasing tendency towards a certain direction.

To speed up the convergence of the ALS algorithm, a line search process of calculating the linear interpolations for $(\mathbf{F}_{BS}, \mathbf{E}_{MS}, \mathbf{B})$ is carried out before the ALS algorithm. The interpolated values for the i -th iteration are expressed as

$$\hat{\mathbf{F}}_{BS}^{(new)} = \hat{\mathbf{F}}_{BS}^{(i-2)} + \rho(\hat{\mathbf{F}}_{BS}^{(i-1)} - \hat{\mathbf{F}}_{BS}^{(i-2)}), \quad (22)$$

$$\hat{\mathbf{E}}_{MS}^{(new)} = \hat{\mathbf{E}}_{MS}^{(i-2)} + \rho(\hat{\mathbf{E}}_{MS}^{(i-1)} - \hat{\mathbf{E}}_{MS}^{(i-2)}), \quad (23)$$

$$\hat{\mathbf{B}}^{(new)} = \hat{\mathbf{B}}^{(i-2)} + \rho(\hat{\mathbf{B}}^{(i-1)} - \hat{\mathbf{B}}^{(i-2)}), \quad (24)$$

where ρ is the relaxation factor and the differences $\mathbf{W}_{\mathbf{F}_{BS}}^{(i)} = \hat{\mathbf{F}}_{BS}^{(i-1)} - \hat{\mathbf{F}}_{BS}^{(i-2)}$, $\mathbf{W}_{\mathbf{E}_{MS}}^{(i)} = \hat{\mathbf{E}}_{MS}^{(i-1)} - \hat{\mathbf{E}}_{MS}^{(i-2)}$ and $\mathbf{W}_{\mathbf{B}}^{(i)} = \hat{\mathbf{B}}^{(i-1)} - \hat{\mathbf{B}}^{(i-2)}$ denote the search directions at the i -th iteration. To speed up the convergence, we choose an appropriate step size ρ ($\rho > 1$) [40], which is viewed as $\rho = 1$ in the classic ALS algorithm. Assume that $\rho = i^{1/\alpha}$ and $\alpha = 3$, the cost function comes to

$$\begin{aligned} \delta_{\rho}^{(i)} &= \left\| \mathbf{Y}^{(1)} - (\hat{\mathbf{B}}^{(new)} \odot \hat{\mathbf{E}}_{MS}^{(new)}) \mathbf{F}_{BS}^{(new)T} \right\|_F^2 \\ &= \left\| \mathbf{Y}^{(1)} - ((\hat{\mathbf{B}}^{(i-2)} + \rho \mathbf{W}_{\mathbf{B}}^{(i)}) \odot (\hat{\mathbf{E}}_{MS}^{(i-2)} + \rho \mathbf{W}_{\mathbf{E}_{MS}}^{(i)})) (\mathbf{F}_{BS}^{(i-2)} + \rho \mathbf{W}_{\mathbf{F}_{BS}}^{(i)})^T \right\|_F^2. \end{aligned} \quad (25)$$

The relaxation factor $\rho = i^{1/\alpha}$ is accepted on the condition that $\delta_{\rho}^{(i)} \leq \delta^{(i)}$, and then $\hat{\mathbf{E}}_{MS}^{(new)}$, $\hat{\mathbf{E}}_{MS}^{(new)}$ and $\hat{\mathbf{B}}^{(new)}$ are chosen for the i -th iteration in the following ALS algorithm. Otherwise, $\hat{\mathbf{F}}_{BS}^{(i-1)}$, $\hat{\mathbf{E}}_{MS}^{(i-1)}$ and $\hat{\mathbf{B}}^{(i-1)}$ are chosen for further estimates at the i -th iteration. Nevertheless, if this acceleration experiences three consecutive failures, a smaller step size should be chosen as $\rho = i^{1/(\alpha+1)}$.

3.2. Channel Parameter Extraction

On the basis of estimates $(\hat{\mathbf{F}}_{BS}, \hat{\mathbf{E}}_{MS}, \hat{\mathbf{B}})$ from the previous stage, we aim to extract the channel parameters from these estimated factor matrices in this subsection. With the permutation and scaling ambiguities, we have

$$\hat{\mathbf{F}}_{BS} = \mathbf{F}_{BS} \mathbf{\Pi} \mathbf{\Delta}^{(1)}, \hat{\mathbf{E}}_{MS} = \mathbf{E}_{MS} \mathbf{\Pi} \mathbf{\Delta}^{(2)}, \hat{\mathbf{B}} = \mathbf{B} \mathbf{\Pi} \mathbf{\Delta}^{(3)}, \quad (26)$$

where $\mathbf{\Pi}$ is the permutation ambiguity. $\mathbf{\Delta}^{(1)}$, $\mathbf{\Delta}^{(2)}$ and $\mathbf{\Delta}^{(3)}$ are the diagonal scaling matrices with $\mathbf{\Delta}^{(1)} \mathbf{\Delta}^{(2)} \mathbf{\Delta}^{(3)} = \mathbf{I}$. To achieve the estimation of the channel parameters of all U MSs, we resort to a one-dimensional search. For the l_u -th path of the u -th MS, we obtain the estimated channel parameters $(\hat{\phi}_{u,l_u}, \hat{\psi}_{u,l_u}, \hat{\tau}_{u,l_u})$:

$$\hat{\phi}_{u,l_u} = \arg \max_{\phi_{u,l_u}} \frac{\left| [\hat{\mathbf{F}}_{BS} \mathbf{e}_{u,l_u}]^H [\mathbf{a}_{BS}(\phi_{u,l_u})] \right|}{\left\| \hat{\mathbf{F}}_{BS} \mathbf{e}_{u,l_u} \right\|_2 \left\| \mathbf{a}_{BS}(\phi_{u,l_u}) \right\|_2}, \quad (27)$$

$$\hat{\psi}_{u,l_u} = \arg \max_{\psi_{u,l_u}} \frac{\left| [\hat{\mathbf{E}}_{MS} \mathbf{e}_{u,l_u}]^H [\mathbf{a}_{MS}(\psi_{u,l_u})] \right|}{\left\| \hat{\mathbf{E}}_{MS} \mathbf{e}_{u,l_u} \right\|_2 \left\| \mathbf{a}_{MS}(\psi_{u,l_u}) \right\|_2}, \quad (28)$$

$$\hat{\tau}_{u,l_u} = \arg \max_{\tau_{u,l_u}} \frac{\left| [\hat{\mathbf{B}} \mathbf{e}_{u,l_u}]^H [\mathbf{d}(\tau_{u,l_u})] \right|}{\left\| \hat{\mathbf{B}} \mathbf{e}_{u,l_u} \right\|_2 \left\| \mathbf{d}(\tau_{u,l_u}) \right\|_2}, \quad (29)$$

where \mathbf{e}_{u,l_u} is the l -th unit coordinate vector with $l = \sum_{i=1}^{u-1} L_i + l_u$. Relying on the estimates $\hat{\phi}_{u,l_u}$ and $\hat{\psi}_{u,l_u}$ of U MSs, we can obtain the scaling matrices $\mathbf{\Delta}^{(1)}$ and $\mathbf{\Delta}^{(2)}$. According to the equality $\mathbf{\Delta}^{(1)} \mathbf{\Delta}^{(2)} \mathbf{\Delta}^{(3)} = \mathbf{I}$, matrix $\mathbf{\Delta}^{(3)}$ can be computed. With the known parameters τ_{u,l_u} and $\mathbf{\Delta}^{(3)}$, the complex gain β_{u,l_u} finally can be obtained. So far, we are capable of estimating

the channel parameters $(\hat{\beta}_{u,l_u}, \hat{\varphi}_{u,l_u}, \hat{\psi}_{u,l_u}, \hat{\tau}_{u,l_u})$ of U MSs, and then the channel tensor \mathcal{Z} can be recovered. The detailed steps of our proposed algorithm are given in Algorithm 1:

Algorithm 1 PARAFAC decomposition-based channel estimation algorithm

Input: Received tensor \mathcal{Y} , precoding matrix \mathbf{E} , combining matrix \mathbf{F} , one-dimensional search number J , the number of paths L_u of U MSs and error threshold ε .

Output: Estimates $\hat{\beta}_{u,l_u}, \hat{\varphi}_{u,l_u}, \hat{\psi}_{u,l_u}$ and $\hat{\tau}_{u,l_u}$.

First stage (the ATALS algorithm):

Initialization: Randomly initialize $\hat{\mathbf{F}}_{BS}^{(0)}, \hat{\mathbf{F}}_{BS}^{(1)}, \hat{\mathbf{B}}^{(0)}, \hat{\mathbf{B}}^{(1)}, \hat{\mathbf{E}}_{MS}^{(0)}, \hat{\mathbf{E}}_{MS}^{(1)}$. Set $i = 2, \alpha = 3$ and $\delta^{(0)} = \infty$.

Step 1.1. Compute $\delta_\rho^{(i)}$ from (25) with $\rho = i^{1/\alpha}$, and compare it with $\delta^{(i)}$ from (20).

If $\delta_\rho^{(i)} \leq \delta^{(i)}$, construct $\hat{\mathbf{F}}_{BS}^{(new)}, \hat{\mathbf{E}}_{MS}^{(new)}, \hat{\mathbf{B}}^{(new)}$ with $\rho = i^{1/\alpha}$;

else, construct $\hat{\mathbf{F}}_{BS}^{(new)}, \hat{\mathbf{E}}_{MS}^{(new)}, \hat{\mathbf{B}}^{(new)}$ with $\rho = 1$.

Step 1.2. Update $\hat{\mathbf{F}}_{BS}^{(i)}$ using $\hat{\mathbf{B}}^{(new)}, \hat{\mathbf{E}}_{MS}^{(new)}, \mathbf{Y}^{(1)}$ by $\hat{\mathbf{F}}_{BS}^{(i)} = \left[(\hat{\mathbf{B}}^{(new)} \odot \hat{\mathbf{E}}_{MS}^{(new)})^\dagger \mathbf{Y}^{(1)} \right]^T$.

Step 1.3. Update $\hat{\mathbf{E}}_{MS}^{(i)}$ using $\hat{\mathbf{B}}^{(new)}, \hat{\mathbf{F}}_{BS}^{(i)}, \mathbf{Y}^{(2)}$ by $\hat{\mathbf{E}}_{MS}^{(i)} = \left[(\hat{\mathbf{B}}^{(new)} \odot \hat{\mathbf{F}}_{BS}^{(i)})^\dagger \mathbf{Y}^{(2)} \right]^T$.

Step 1.4. Update $\hat{\mathbf{B}}^{(i)}$ using $\hat{\mathbf{E}}_{MS}^{(i)}, \hat{\mathbf{F}}_{BS}^{(i)}, \mathbf{Y}^{(3)}$ by $\hat{\mathbf{B}}^{(i)} = \left[(\hat{\mathbf{E}}_{MS}^{(i)} \odot \hat{\mathbf{F}}_{BS}^{(i)})^\dagger \mathbf{Y}^{(3)} \right]^T$.

Step 1.5. Compute the new error $\delta^{(i)}$. If $|\delta^{(i)} - \delta^{(i-1)}| / \delta^{(i)} \leq \varepsilon$, then end.

Otherwise, set $i = i + 1$ and go to **Step 1.1**.

Second stage (Channel parameter extraction via one-dimension search):

for $u = 1, \dots, U$, and for $l_u = 1, \dots, L_u$

Step 2.1. Estimate the AoA $\hat{\varphi}_{u,l_u}$ by (27) with $\varphi_j \in [0, 2\pi], 1 \leq j \leq J$.

Step 2.2. Estimate the AoD $\hat{\psi}_{u,l_u}$ by (28) with $\psi_j \in [0, 2\pi], 1 \leq j \leq J$.

Step 2.3. Estimate the delay $\hat{\tau}_{u,l_u}$ by (29) with $\tau_j \in [0, \tau_{\max}], 1 \leq j \leq J$.

Step 2.4. Compute the diagonal matrices $\Delta^{(1)}, \Delta^{(2)}$ and $\Delta^{(3)}$.

Step 2.5. Estimate the path gain $\hat{\beta}_{u,l_u}$ according to $\hat{\mathbf{B}} = \mathbf{B} \mathbf{\Pi} \mathbf{\Delta}^{(3)}$.

4. Cramér–Rao Bound Derivation

In this section, the derivation of CRB results of channel parameters $(\beta_{u,l_u}, \varphi_{u,l_u}, \psi_{u,l_u}, \tau_{u,l_u})$ is given. As known to all, the CRB is a lower bound of the variance of arbitrary unbiased estimator [41], which in turn makes it a reliable benchmark. In this system, the derived CRB results provide a lower bound on the channel prediction error. Therein, the covariance of any biased estimator of the canonical parameter vector $\boldsymbol{\eta}$ satisfies

$$\text{Cov}(\boldsymbol{\eta}) \geq \boldsymbol{\Omega}^{-1}(\boldsymbol{\eta}), \quad (30)$$

where $\boldsymbol{\Omega}(\boldsymbol{\eta})$ is the Fisher information matrix.

In view of three equivalent forms of \mathcal{Y} in (16), the entries of noise matrices $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}$ and $\mathbf{V}^{(3)}$ are i.i.d zero mean, circularly symmetric complex Gaussian noise, of variance σ^2 . As a result, the log-likelihood function of signal tensor \mathcal{Y} is expressed as

$$\begin{aligned} \ln L(\mathcal{Y}) &= -RPK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{r=1}^R \left\| \mathbf{y}_r^{(1)} - (\mathbf{B} \odot \mathbf{E}_{MS}) \mathbf{f}_r^{(1)} \right\|_2^2 \\ &= -RPK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{p=1}^P \left\| \mathbf{y}_p^{(2)} - (\mathbf{B} \odot \mathbf{F}_{BS}) \mathbf{e}_p^{(2)} \right\|_2^2 \\ &= -RPK \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{k=1}^K \left\| \mathbf{y}_k^{(3)} - (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \mathbf{b}_k^{(3)} \right\|_2^2, \end{aligned} \quad (31)$$

where $\mathbf{y}_r^{(1)}$, $\mathbf{y}_p^{(2)}$ and $\mathbf{y}_k^{(3)}$ represent the r -th, p -th, k -th column of $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ and $\mathbf{Y}^{(3)}$, respectively; $\mathbf{f}_r^{(1)}$, $\mathbf{e}_p^{(2)}$ and $\mathbf{b}_k^{(3)}$ represent the r -th, p -th, k -th column of \mathbf{F}_{BS}^T , \mathbf{E}_{MS}^T and \mathbf{B}^T , respectively. The matrices \mathbf{F}_{BS} , \mathbf{E}_{MS} and \mathbf{B} contain unknown parameter sets $\{\boldsymbol{\varphi}_u\}_{u=1}^U$, $\{\boldsymbol{\psi}_u\}_{u=1}^U$ and $\{\boldsymbol{\beta}_u, \boldsymbol{\tau}_u\}_{u=1}^U$, respectively. Define

$$\boldsymbol{\varphi}_u = [\varphi_{u,1}, \dots, \varphi_{u,L_u}], \boldsymbol{\psi}_u = [\psi_{u,1}, \dots, \psi_{u,L_u}], \boldsymbol{\beta}_u = [\beta_{u,1}, \dots, \beta_{u,L_u}], \boldsymbol{\tau}_u = [\tau_{u,1}, \dots, \tau_{u,L_u}],$$

and collect the vectors $(\boldsymbol{\beta}_u, \boldsymbol{\varphi}_u, \boldsymbol{\psi}_u, \boldsymbol{\tau}_u)$ for all U MSs, we derive the parameter vector $\boldsymbol{\eta}$

$$\boldsymbol{\eta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_U, \boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_U, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_U, \boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_U] \in \mathbb{C}^{1 \times 4\Gamma},$$

which consists of all the unknown channel parameters. Furthermore, the log-likelihood function is written as $\ln L(\boldsymbol{\eta}) = \ln L(\mathcal{Y})$. Thereby, the complex Fisher information matrix is given by

$$\begin{aligned} \boldsymbol{\Omega}(\boldsymbol{\eta}) &= E \left\{ \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)^H \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right) \right\} \\ &= \begin{bmatrix} \Omega_{\beta\beta} & \Omega_{\beta\varphi} & \Omega_{\beta\psi} & \Omega_{\beta\tau} \\ \Omega_{\varphi\beta} & \Omega_{\varphi\varphi} & \Omega_{\varphi\psi} & \Omega_{\varphi\tau} \\ \Omega_{\psi\beta} & \Omega_{\psi\varphi} & \Omega_{\psi\psi} & \Omega_{\psi\tau} \\ \Omega_{\tau\beta} & \Omega_{\tau\varphi} & \Omega_{\tau\psi} & \Omega_{\tau\tau} \end{bmatrix} \end{aligned} \quad (32)$$

Any matrix block of $\boldsymbol{\Omega}(\boldsymbol{\eta}) \in \mathbb{C}^{4\Gamma \times 4\Gamma}$ can be expressed as $\Omega_{xy} \in \mathbb{C}^{\Gamma \times \Gamma}$, where x and y represent any symbol of $(\beta, \varphi, \psi, \tau)$. The (l_1, l_2) -th element of Ω_{xy} is expressed as

$$[\Omega_{xy}]_{l_1, l_2} = E \left\{ \frac{\partial \ln L(\boldsymbol{\eta})}{\partial x_{u_1, l_{u_1}}}^* \frac{\partial \ln L(\boldsymbol{\eta})}{\partial y_{u_2, l_{u_2}}} \right\}. \quad (33)$$

where $l_1 = \sum_{i=1}^{u_1-1} L_i + l_{u_1}$, $l_{u_1} = 1, \dots, L_{u_1}$, for $u_1 = 1, \dots, U$; $l_2 = \sum_{i=1}^{u_2-1} L_i + l_{u_2}$, $l_{u_2} = 1, \dots, L_{u_2}$, for $u_2 = 1, \dots, U$.

Further precise derivation can be found in the Appendix A.

5. System Performance Analysis

5.1. Analysis of Computational Complexity

In this subsection, we analyze the computational complexity of our proposed PARAFAC decomposition-based algorithm. Since the traditional CS-based methods are common and effective solutions to the problem of mmWave channel estimation, we compare our method with the SOMP, OMP and adaptive CS-based algorithms.

As shown in Table 2, the overall computational complexity of our proposed algorithm consists of two parts. In the first stage, the major computational task is the ATALS algorithm, which is an iterative algorithm in Step 1.2.–Step 1.4. at each iteration. Take the Step 1.2. for example, to compute the matrix $\hat{\mathbf{F}}_{BS}^{(i)}$ at the i -th iteration, this step has the complexity in order of $\mathcal{O}(RPKT + RPKT^2 + \Gamma^3)$, which is the same for Step 1.3. and Step 1.4.. In the second stage, the complexity mainly lies in the one-dimensional searching (27)–(29) for channel parameters of each path. The calculation of delay τ_{u, l_u} has the complexity in order of $\mathcal{O}(JK)$ for each path, similarly for the calculation of AoAs and AoDs. The total number of paths Γ is so small that the main complexity order for the one-dimensional search can be ignored. Finally, the overall computational complexity is represented as $\mathcal{O}(RPKT + RPKT^2 + \Gamma^3)$.

As can be seen from Table 2, the complexity of the SOMP, OMP and adaptive CS-based algorithms can finally be written as $\mathcal{O}(RPKT^2) + \mathcal{O}(RPMUNKT)$, $\mathcal{O}(RPKT^3) + \mathcal{O}(RPMUNKT)$ and $\mathcal{O}(SRPKT^2) + \mathcal{O}(SRPMUNKT)$, respectively. S is the number of adaptive stages for each path estimation referred in [10]. Considering the smaller value of

Γ , theoretical analysis shows that our proposed algorithm has a lower computational complexity than the SOMP, OMP and adaptive CS-based algorithms. In our simulations, the average run time of the four algorithms is also recorded for comprehensive comparison.

Table 2. Computational complexity of respective algorithms.

| Algorithm | | Complexity |
|-------------|--------------|--|
| SOMP | | $\mathcal{O}(RPK\Gamma^2) + \mathcal{O}(RPMUNK\Gamma)$ |
| OMP | | $\mathcal{O}(RPK\Gamma^3) + \mathcal{O}(RPMUNK\Gamma)$ |
| Adaptive CS | | $\mathcal{O}(SRPK\Gamma^2) + \mathcal{O}(SRPMUNK\Gamma)$ |
| Proposed | First stage | $\mathcal{O}(RPK\Gamma + RPK\Gamma^2 + \Gamma^3)$ |
| | Second stage | $\mathcal{O}(JK)$ |
| | Total | $\mathcal{O}(RPK\Gamma + RPK\Gamma^2 + \Gamma^3)$ |

5.2. Simulation Results

In this subsection, numerical results prove the superiority of our proposed algorithm over other CS-based algorithms through computer emulation. For convenience, system parameter settings in our simulations are represented in Table 3. In this context, the BS and the MSs are equipped with the ULA with $N = 64$ and $M = 32$, respectively. The total number of transmission subcarriers is assumed to be $K = 128$. The carrier frequency and the sampling rate are set as $f_c = 28$ GHz and $f_s = 0.25$ GHz, respectively. The separation between antenna elements d is half of signal wavelength λ . According to the wideband geometric channel model, the mmWave channel parameters are set as follows: $\beta_{u,l_u} \sim \mathcal{CN}(0, 1)$, $\varphi_{u,l_u} \sim [0, 2\pi]$, $\psi_{u,l_u} \sim [0, 2\pi]$, $\tau_{u,l_u} \sim [0, \tau_{\max}]$ with the maximum delay $\tau_{\max} = 100$ ns. In addition, we assume a reasonably large signal-to-noise ratio (SNR) range of 0 dB to 30 dB at the MSs with $U = 6$, $\Gamma = 8$, and the number of scatterings between the BS and each MS is 1 or 2. For representing the estimation accuracy of channel parameters, the MSEs are introduced:

$$\begin{aligned} \text{MSE}(\beta) &= \|\beta - \hat{\beta}\|_2^2, \text{MSE}(\varphi) = \|\varphi - \hat{\varphi}\|_2^2, \\ \text{MSE}(\psi) &= \|\psi - \hat{\psi}\|_2^2, \text{MSE}(\tau) = \|\tau - \hat{\tau}\|_2^2. \end{aligned} \quad (34)$$

where $\beta = [\beta_1, \dots, \beta_U] \in \mathbb{C}^{1 \times \Gamma}$, $\varphi = [\varphi_1, \dots, \varphi_U] \in \mathbb{C}^{1 \times \Gamma}$, $\psi = [\psi_1, \dots, \psi_U] \in \mathbb{C}^{1 \times \Gamma}$, $\tau = [\tau_1, \dots, \tau_U] \in \mathbb{C}^{1 \times \Gamma}$. The vectors $\hat{\beta}$, $\hat{\varphi}$, $\hat{\psi}$ and $\hat{\tau}$ are the estimates of true channel parameter vectors β , φ , ψ and τ , respectively. Furthermore, the estimation accuracy of multiuser mmWave channels can also be evaluated by the normalized mean square error (NMSE):

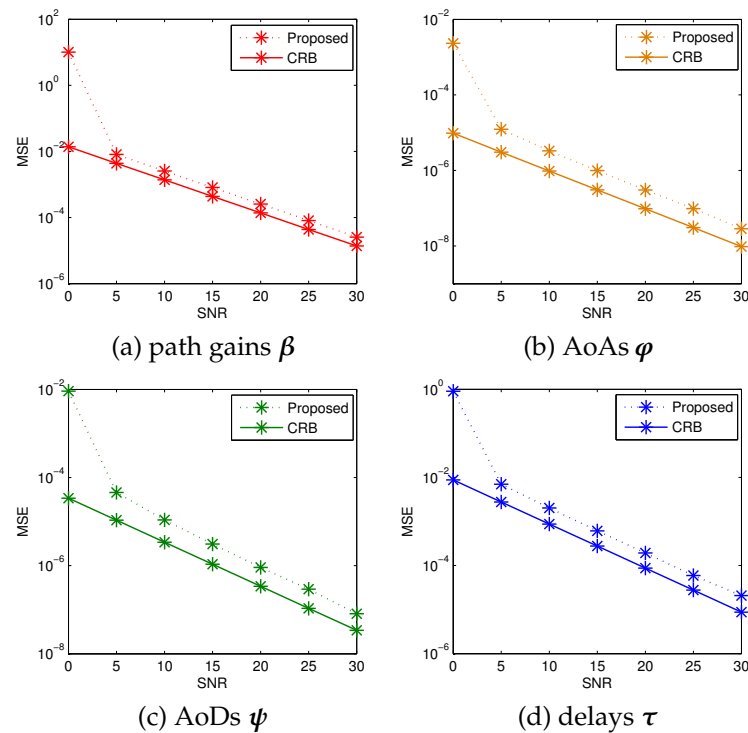
$$\text{NMSE} = \sum_{k=1}^K \frac{\|\mathbf{Z}_k - \hat{\mathbf{Z}}_k\|_F^2}{\|\mathbf{Z}_k\|_F^2}, \quad (35)$$

where \mathbf{Z}_k represents the multiuser channel matrix via the k -th subcarrier and $\hat{\mathbf{Z}}_k$ is its estimated matrix.

In the first example, we examine the performance of our proposed algorithm at different SNR changes with $P = 8$. The Figure 2 depicts the MSEs and CRB results of associated channel parameters versus SNR. From Figure 2, we observe that the MSEs of channel parameters exhibit a significant decline on the increase of SNR. As the benchmark for evaluation, the corresponding CRB results have the similar downward trend and perform approximately as a linear function of SNR. As expected, the MSEs performance of channel parameters are very close to their corresponding CRBs, especially in a higher SNR region. The simulation results indicate that our proposed algorithm enjoys an accurate estimation of multiuser channel parameters.

Table 3. Simulation configuration of system parameters.

| System Parameter | Configuration |
|--|----------------------|
| BS antennas N / RF chains R | 64 / 10 |
| MS antennas M / RF chains R_u | 32 / 1 |
| Transmission subcarriers K | 128 |
| Total number of MSs U / paths Γ | 6 / 8 |
| Carrier frequency f_c | 28GHz |
| Sampling rate f_s | 0.25GHz |
| AoA / AoD $\varphi_{u,l_u} / \psi_{u,l_u}$ | $[0, 2\pi]$ |
| Path gain β_{u,l_u} | $\mathcal{CN}(0, 1)$ |
| Delay τ_{u,l_u} | $[0, \tau_{\max}]$ |

**Figure 2.** MSEs and CRBs of channel parameters versus SNR.

In the second example, we investigate the MSEs performance of our proposed algorithm as a function of the number of subcarriers K , where we set $\text{SNR} = 30$ dB and $P = 8$. The CRB results of different sets of channel parameters are also represented for comparison. It can be seen from Figure 3 that the proposed algorithm yields accurate estimates even for a small value of K . As the value of K increases, the CRB results of a certain set of channel parameters show a downward trend, and the MSEs performance are close to their corresponding CRBs. We also observe that when the number of subcarriers $K \leq 2$, our proposed algorithm fails. In fact, the phenomenon is consistent with our previous analysis on unique PARAFAC decomposition. For the case $P = 8 \geq \Gamma$ and $R = 10 \geq \Gamma$, the condition (18) is satisfied only if $K \geq 2$.

In the third example, we study the NMSE performance of our proposed algorithm when varying the number of subcarriers K , where we assume $\text{SNR} = 30$ dB. The Figure 4 displays the NMSE performance comparison with the SOMP and OMP algorithms for different time slots $P = 8$ and $P = 16$. Apparently, with the increasing of K , all the three algorithms have a more and more accurate channel estimation performance. Among these algorithms, the proposed PARAFAC-based estimator presents a clear advantage over the SOMP and OMP algorithms, and it has a much lower error floor. In particular, compare the NMSE performance with different P for a certain algorithm, and we can see that it performs better when $P = 16$, especially for the proposed PARAFAC-based and OMP algorithm.

Clearly, the increasing number of time slots requires more pilots, which leads to a better estimation performance. However, our goal is to reach a reliable channel estimation with as few measurements as possible. Besides, the NMSE curve obtained by our proposed method accords with previous analysis on unique PARAFAC decomposition. For the case $P = 8 \geq \Gamma$ and $R = 10 \geq \Gamma$, when $K \geq 2$, the uniqueness condition (18) holds. This indicates that our proposed algorithm performs well even at the minimum number of subcarriers $K = 2$.

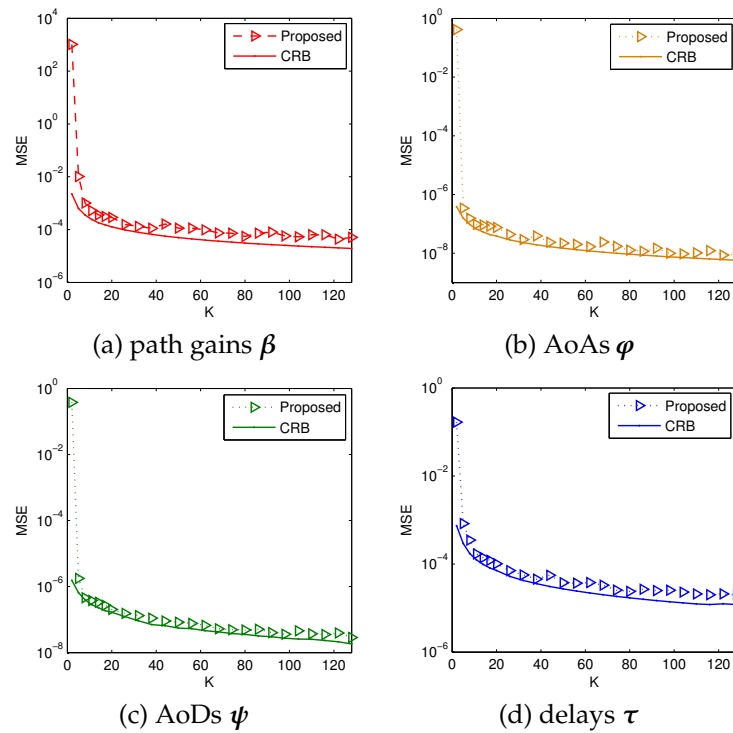


Figure 3. MSEs and CRBs of channel parameters versus K .

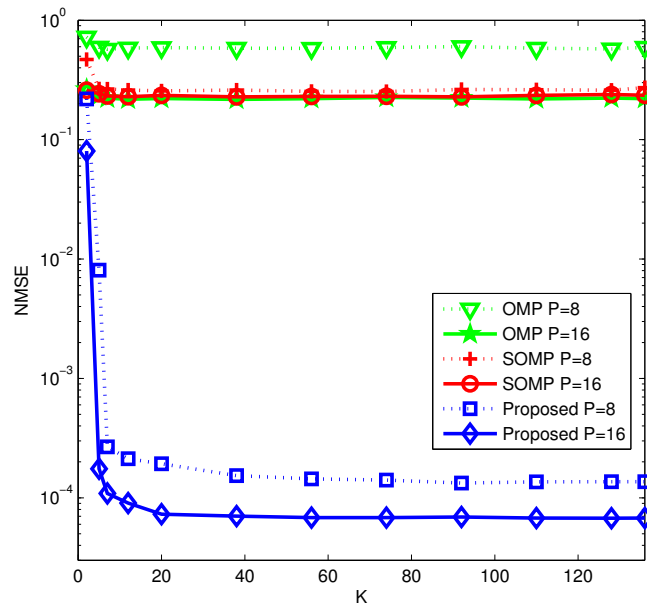


Figure 4. NMSE comparison among different algorithms versus the number of subcarriers K .

In the fourth example, we investigate the system performance with different numbers of MSs. The impact of different numbers of MSs on the NMSE performance is shown in Figure 5, where the numbers of MSs are assumed to be $U = 2$, $U = 4$, $U = 8$, $U = 16$ and $U = 32$, with single path propagation for per MS. To guarantee the uniqueness of PARAFAC analysis, we set $P = 22$ and $R = 12$. As the number of MSs increases, the NMSE performance decreases. There is a reason that channel parameters to be estimated grow with the rise of U , thus leading to a decline in estimation accuracy. Moreover, the estimation performance of factor matrices in the first stage suffers from degradation for the increasing unknown matrix elements. However, our proposed scheme still has a good performance even though the number of MSs is large to $U = 32$. While our proposed algorithm is put forward for multiuser architecture, it is also available in a single-user environment of multipaths.

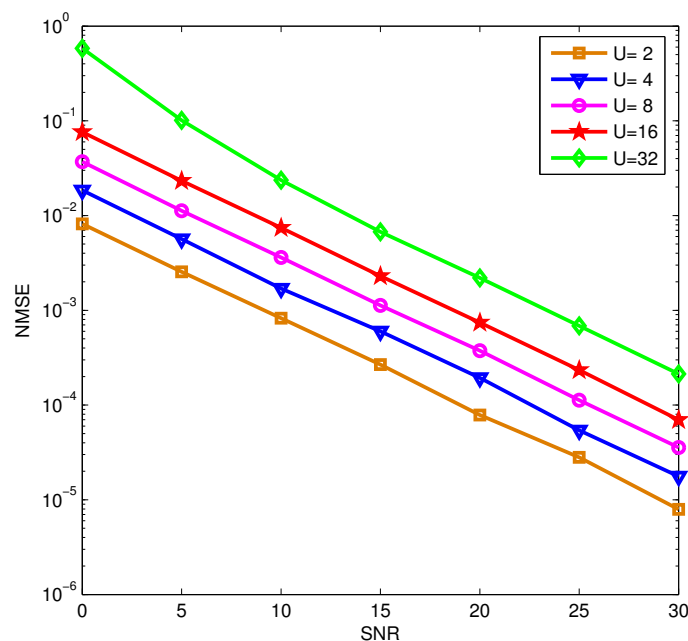


Figure 5. NMSE comparison of the proposed algorithm versus the number of MSs.

In the final example, we compare our proposed algorithm with the SOMP, OMP and adaptive CS algorithms with different numbers of RF chains $R = 8$ and $R = 16$. As shown in Figure 6, it describes the NMSE comparison among the four algorithms versus SNR. As the SNR increases, all the algorithms have more and more accurate channel estimation. Our proposed algorithm is superior to the other three CS-based algorithms in terms of estimation accuracy with the same R . Especially in a higher SNR region, the three CS-based algorithms have a poor performance with a higher error floor while the PARAFAC-based method still yields more accurate channel estimates. Such a performance gap comes from that these CS-based algorithms assume the estimated angles are discretely distributed on the pre-defined grids. In fact, actual angles are continuous and may not necessarily lie on the grids, which results in an inevitable grid mismatch, and thus degrading the performance. Generally speaking, the finer the grid is, the smaller the grid mismatch error is. However, it also brings an unstable performance with higher computational complexity. By contrast, our proposed method is not only free of grid mismatch but also fully utilizes the low-rank structure of mmWave channels, which leads to a better and stable performance. From Figure 6, we can also gain some insights into how different numbers of RF chains effect the system performance. As R increases from 8 to 16, the NMSE performance degrades. While more RF chains brings about a better system performance, it suffers from higher hardware cost and power consumption. Therefore, a smaller number of RF chains should be selected for a compromise between hardware cost and system performance.

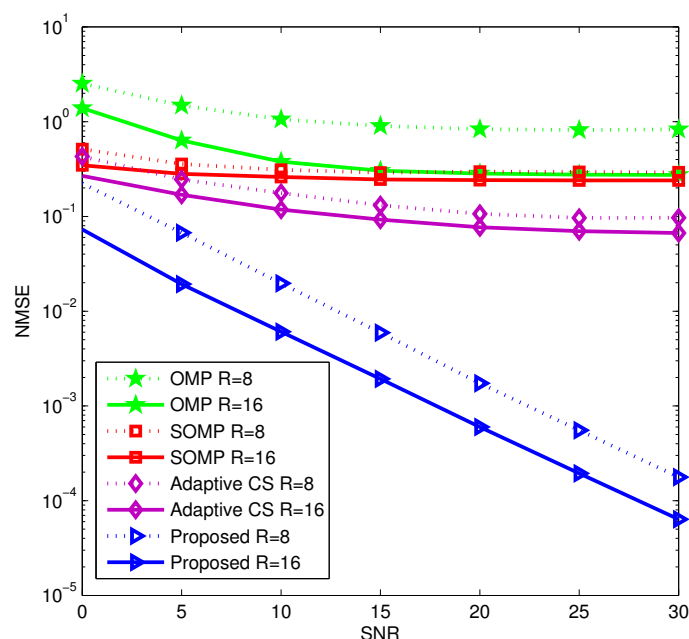


Figure 6. NMSE comparison among different algorithms versus SNR.

For this example, the average run time of the SOMP, OMP, adaptive CS and proposed PARAFAC-based algorithms at different SNR changes is shown in Table 4. Numerical results demonstrate that the average run time of our proposed algorithm is shorter than that of other three algorithms. In contrast with the CS-based algorithms, our PARAFAC-based algorithm takes advantage of its inherent structure of multidimensional data and achieves a stable and low-complexity channel estimation.

Table 4. Average run time of respective algorithms.

| SNR(dB) | 0 | 5 | 10 | 15 | 20 | 25 | 30 |
|----------------------|----------|----------|----------|----------|----------|----------|----------|
| SOMP (R = 8) | 1.1239 s | 1.1892 s | 1.1923 s | 1.2035 s | 1.2315 s | 1.2174 s | 1.2248 s |
| OMP (R = 8) | 2.2667 s | 2.2075 s | 2.2738 s | 2.2493 s | 2.2512 s | 2.2555 s | 2.2542 s |
| Adaptive CS (R = 8) | 1.0299 s | 0.9888 s | 1.1948 s | 1.1877 s | 1.2464 s | 1.2162 s | 1.1911 s |
| Proposed (R = 8) | 0.7117 s | 0.6612 s | 0.595 s | 0.7766 s | 0.7377 s | 0.6304 s | 0.6987 s |
| SOMP (R = 16) | 1.635 s | 1.5626 s | 1.6055 s | 1.6113 s | 1.5785 s | 1.6153 s | 1.6129 s |
| OMP (R = 16) | 5.3028 s | 5.3577 s | 5.2848 s | 5.2799 s | 5.328 s | 5.4576 s | 5.3256 s |
| Adaptive CS (R = 16) | 1.4869 s | 1.679 s | 1.6707 s | 1.678 s | 1.6739 s | 1.7011 s | 1.6774 s |
| Proposed (R = 16) | 0.7749 s | 0.8524 s | 0.8869 s | 0.8478 s | 0.8686 s | 0.8329 s | 0.8868 s |

6. Conclusions

In this paper, a PARAFAC decomposition-based algorithm has been provided for multiuser uplink channel estimation over the broad-band FSF mmWave channels. Owing to the sparsity of mmWave channels, the received signal at the BS is formulated as a low-rank PARAFAC tensor, and thus multiuser channel parameter estimation can be achieved via PARAFAC decomposition. Therein, the uniqueness issue is studied. Moreover, the CRB results are derived for evaluation in the situation of multiple users. Simulations reveal that the algorithm performs well with a few training sequences. Compared with other CS-based algorithms, our proposed algorithm outperforms its counterparts both in estimation accuracy and computational complexity. For future research, we will pay more attention to the time-varying mmWave channels, and try to seek more precise tensor models, such as the nested PARAFAC model and the nested Tucker model. Novel tensor fitting algorithms such as the Bayesian inference-based algorithm [42] will be developed in future work for better performance.

Author Contributions: Conceptualization R.C. and J.D.; methodology, R.C. and J.D.; software, R.C.; validation, R.C.; writing—original draft, R.C.; writing—review and editing, J.D. and C.Y.; supervision, J.D. and C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1800800, the National Key Research and Development Program of China under Grant 2018YFB1800805 and the National Natural Science Foundation of China under Grants 61601414.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editor for their careful reviews and constructive suggestions to help us improve the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---------|---|
| ALS | alternating least squares |
| AoAs | angles of arrival |
| AoDs | angles of departure |
| ATALS | accelerated trilinear alternating least squares |
| BS | base station |
| CP | CANDECOMP/PARAFAC |
| CRB | Cramér–Rao Bound |
| CS | compressing sensing |
| eMBB | extended mobile broadband |
| FSF | frequency-selective fading |
| IoT | internet of things |
| MIMO | multiple-input multiple-output |
| mMTC | massive machine-type communication |
| MMV | multiple measurement vector |
| mmWave | millimeter wave |
| MS | mobile station |
| MSEs | mean square errors |
| OFDM | orthogonal frequency divisional multiplexing |
| OMP | orthogonal matching pursuit |
| PARAFAC | parallel factor |
| RF | radio frequency |
| SOMP | simultaneous orthogonal matching pursuit |
| ULA | uniform linear array |
| UPA | uniform planar array |
| URLLC | ultra-reliable low-latency communication |

Appendix A. Derivation of Cramér–Rao Bound

Taking the partial derivatives of $\ln L(\boldsymbol{\eta})$ with respect to unknown parameters $(\beta_{u,l_u}, \varphi_{u,l_u}, \psi_{u,l_u}, \tau_{u,l_u})$, we derive that

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \beta_{u,l_u}} = \frac{1}{\sigma^2} \sum_{k=1}^K \left[\left(\mathbf{y}_k^{(3)} - (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \mathbf{b}_k^{(3)} \right)^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{d}_k \mathbf{e}_{u,l_u} \right], \quad (\text{A1})$$

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \varphi_{u,l_u}} = \frac{2}{\sigma^2} \sum_{r=1}^R \text{Re} \left[\left(\mathbf{y}_r^{(1)} - (\mathbf{B} \odot \mathbf{E}_{MS}) \mathbf{f}_r^{(1)} \right)^H (\mathbf{B} \odot \mathbf{E}_{MS}) \bar{f}_r \mathbf{e}_{u,l_u} \right], \quad (\text{A2})$$

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \psi_{u,l_u}} = \frac{2}{\sigma^2} \sum_{p=1}^P \text{Re} \left[\left(\mathbf{y}_p^{(2)} - (\mathbf{B} \odot \mathbf{F}_{BS}) \mathbf{e}_p^{(2)} \right)^H (\mathbf{B} \odot \mathbf{F}_{BS}) \bar{e}_p \mathbf{e}_{u,l_u} \right], \quad (\text{A3})$$

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \tau_{u,l_u}} = \frac{2}{\sigma^2} \sum_{k=1}^K \text{Re} \left[\left(\mathbf{y}_k^{(3)} - (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \mathbf{b}_k^{(3)} \right)^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{b}_k \mathbf{e}_{u,l_u} \right], \quad (\text{A4})$$

where $\bar{d}_k = [\mathbf{D}]_{k,l}$, $\bar{f}_r = j\pi \cos \varphi_{u,l_u} \mathbf{f}_r^T \mathbf{D}_1 \mathbf{a}_{BS}(\varphi_{u,l_u})$, $\bar{e}_p = j\pi \cos \psi_{u,l_u} \mathbf{e}_p^T \mathbf{D}_2 \mathbf{a}_{MS}(\psi_{u,l_u})$, $\bar{b}_k = -(j2\pi f_s k/K) [\mathbf{B}]_{k,l}$ with the definitions $l = \sum_{i=1}^{u-1} L_i + l_u$, $\mathbf{D}_1 = \text{diag}(0, 1, \dots, N-1)$, $\mathbf{D}_2 = \text{diag}(0, 1, \dots, M-1)$. The vector $\mathbf{e}_{u,l_u} \in \mathbb{C}^{\Gamma \times 1}$ is the l -th unit coordinate vector. In addition, we derive that

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \beta_{u,l_u}^*} = \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \beta_{u,l_u}} \right)^*, \quad \frac{\partial \ln L(\boldsymbol{\eta})}{\partial \varphi_{u,l_u}^*} = \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \varphi_{u,l_u}} \right)^*,$$

$$\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \psi_{u,l_u}^*} = \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \psi_{u,l_u}} \right)^*, \quad \frac{\partial \ln L(\boldsymbol{\eta})}{\partial \tau_{u,l_u}^*} = \left(\frac{\partial \ln L(\boldsymbol{\eta})}{\partial \tau_{u,l_u}} \right)^*.$$

Next, we give some derivations for the aforementioned matrix blocks of $\boldsymbol{\Omega}(\boldsymbol{\eta})$. Define the noise vectors $\mathbf{v}_r^{(1)} = \mathbf{y}_r^{(1)} - (\mathbf{B} \odot \mathbf{E}_{MS}) \mathbf{f}_r^{(1)}$, $\mathbf{v}_p^{(2)} = \mathbf{y}_p^{(2)} - (\mathbf{B} \odot \mathbf{F}_{BS}) \mathbf{e}_p^{(2)}$, $\mathbf{v}_k^{(3)} = \mathbf{y}_k^{(3)} - (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \mathbf{b}_k^{(3)}$, and we focus on the matrix blocks on the diagonal of $\boldsymbol{\Omega}(\boldsymbol{\eta})$:

$$[\boldsymbol{\Omega}_{\beta\beta}]_{l_1,l_2} = \frac{1}{\sigma^2} \sum_{k_1=1}^K \sum_{k_2=1}^K \mathbf{e}_{u_1,l_{u_1}}^H \bar{d}_{k_1}^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS})^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{d}_{k_2} \mathbf{e}_{u_2,l_{u_2}} \delta_{k_1,k_2}, \quad (\text{A5})$$

$$[\boldsymbol{\Omega}_{\varphi\varphi}]_{l_1,l_2} = \frac{2}{\sigma^2} \sum_{r_1=1}^R \sum_{r_2=1}^R \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{f}_{r_1}^H (\mathbf{B} \odot \mathbf{E}_{MS})^H (\mathbf{B} \odot \mathbf{E}_{MS}) \bar{f}_{r_2} \mathbf{e}_{u_2,l_{u_2}} \delta_{r_1,r_2} \right], \quad (\text{A6})$$

$$[\boldsymbol{\Omega}_{\psi\psi}]_{l_1,l_2} = \frac{2}{\sigma^2} \sum_{p_1=1}^P \sum_{p_2=1}^P \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{e}_{p_1}^H (\mathbf{B} \odot \mathbf{F}_{BS})^H (\mathbf{B} \odot \mathbf{F}_{BS}) \bar{e}_{p_2} \mathbf{e}_{u_2,l_{u_2}} \delta_{p_1,p_2} \right], \quad (\text{A7})$$

$$[\boldsymbol{\Omega}_{\tau\tau}]_{l_1,l_2} = \frac{2}{\sigma^2} \sum_{k_1=1}^K \sum_{k_2=1}^K \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{b}_{k_1}^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS})^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{b}_{k_2} \mathbf{e}_{u_2,l_{u_2}} \delta_{k_1,k_2} \right]. \quad (\text{A8})$$

Furthermore, we obtain the other matrix blocks of $\boldsymbol{\Omega}(\boldsymbol{\eta})$:

$$[\boldsymbol{\Omega}_{\beta\varphi}]_{l_1,l_2} = \frac{1}{\sigma^4} \sum_{k=1}^K \sum_{r=1}^R \mathbf{e}_{u_1,l_{u_1}}^H \bar{d}_k^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS})^H \mathbf{E} \left[\mathbf{v}_k^{(3)} \mathbf{v}_r^{(1)H} \right] (\mathbf{B} \odot \mathbf{E}_{MS}) \bar{f}_r \mathbf{e}_{u_2,l_{u_2}}, \quad (\text{A9})$$

$$[\boldsymbol{\Omega}_{\beta\psi}]_{l_1,l_2} = \frac{1}{\sigma^4} \sum_{k=1}^K \sum_{p=1}^P \mathbf{e}_{u_1,l_{u_1}}^H \bar{d}_k^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS})^H \mathbf{E} \left[\mathbf{v}_k^{(3)} \mathbf{v}_p^{(2)H} \right] (\mathbf{B} \odot \mathbf{F}_{BS}) \bar{e}_p \mathbf{e}_{u_2,l_{u_2}}, \quad (\text{A10})$$

$$[\boldsymbol{\Omega}_{\beta\tau}]_{l_1,l_2} = \frac{1}{\sigma^4} \sum_{k_1=1}^K \sum_{k_2=1}^K \mathbf{e}_{u_1,l_{u_1}}^H \bar{d}_{k_1}^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS})^H (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{b}_{k_2} \mathbf{e}_{u_2,l_{u_2}} \delta_{k_1,k_2}, \quad (\text{A11})$$

$$[\boldsymbol{\Omega}_{\varphi\psi}]_{l_1,l_2} = \frac{2}{\sigma^4} \sum_{r=1}^R \sum_{p=1}^P \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{f}_r^H (\mathbf{B} \odot \mathbf{E}_{MS})^H \mathbf{E} \left[\mathbf{v}_r^{(1)} \mathbf{v}_p^{(2)H} \right] (\mathbf{B} \odot \mathbf{F}_{BS}) \bar{e}_p \mathbf{e}_{u_2,l_{u_2}} \right], \quad (\text{A12})$$

$$[\boldsymbol{\Omega}_{\varphi\tau}]_{l_1,l_2} = \frac{2}{\sigma^4} \sum_{r=1}^R \sum_{k=1}^K \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{f}_r^H (\mathbf{B} \odot \mathbf{E}_{MS})^H \mathbf{E} \left[\mathbf{v}_r^{(1)} \mathbf{v}_k^{(3)H} \right] (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{b}_k \mathbf{e}_{u_2,l_{u_2}} \right], \quad (\text{A13})$$

$$[\boldsymbol{\Omega}_{\psi\tau}]_{l_1,l_2} = \frac{2}{\sigma^4} \sum_{p=1}^P \sum_{k=1}^K \text{Re} \left[\mathbf{e}_{u_1,l_{u_1}}^H \bar{e}_p^H (\mathbf{B} \odot \mathbf{F}_{BS})^H \mathbf{E} \left[\mathbf{v}_p^{(2)} \mathbf{v}_k^{(3)H} \right] (\mathbf{E}_{MS} \odot \mathbf{F}_{BS}) \bar{b}_k \mathbf{e}_{u_2,l_{u_2}} \right], \quad (\text{A14})$$

and we have $\boldsymbol{\Omega}_{\varphi\beta} = \boldsymbol{\Omega}_{\beta\varphi}^H$, $\boldsymbol{\Omega}_{\psi\beta} = \boldsymbol{\Omega}_{\beta\psi}^H$, $\boldsymbol{\Omega}_{\tau\beta} = \boldsymbol{\Omega}_{\beta\tau}^H$, $\boldsymbol{\Omega}_{\varphi\psi} = \boldsymbol{\Omega}_{\psi\varphi}^H$, $\boldsymbol{\Omega}_{\tau\varphi} = \boldsymbol{\Omega}_{\varphi\tau}^H$, $\boldsymbol{\Omega}_{\tau\psi} = \boldsymbol{\Omega}_{\psi\tau}^H$.

Note that the noise tensor \mathcal{V} is independent in three dimensions, i.e., $\mathbf{E} \left\{ [\mathcal{V}]_{r_1,p_1,k_1} [\mathcal{V}]_{r_2,p_2,k_2}^* \right\} = \sigma^2 \delta_{r_1,r_2} \delta_{p_1,p_2} \delta_{k_1,k_2}$, where $\delta_{i,j}$ denotes the Kronecker delta, and

$\delta_{i,j} = 1$ when $i = j$, $\delta_{i,j} = 0$ when $i \neq j$. In consideration of the (i, j, k) -th entry of \mathcal{V} , $[\mathcal{V}]_{r,p,k}$ corresponds to the following indexes of the noise vectors $\mathbf{v}_r^{(1)}$, $\mathbf{v}_p^{(2)}$ and $\mathbf{v}_k^{(3)}$:

$$[\mathcal{V}]_{r,p,k} = [\mathbf{v}_r^{(1)}]_{(p-1)K+k} = [\mathbf{v}_p^{(2)}]_{(k-1)R+r} = [\mathbf{v}_k^{(3)}]_{(r-1)P+p}.$$

Hence, we can compute the product terms of noise vectors in (A9)–(A14). For example, $E[\mathbf{v}_r^{(1)}\mathbf{v}_p^{(2)H}]$ is given by

$$E[\mathbf{v}_r^{(1)}\mathbf{v}_p^{(2)H}] = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \leftarrow (p-1)K+1 \\ \leftarrow (p-1)K+2 \\ \leftarrow (p-1)K+K \end{matrix}$$

$\begin{matrix} \uparrow & & \uparrow & & \uparrow \\ r & & R+r & & (K-1)R+r \end{matrix}$

Similarly, the computation of $E[\mathbf{v}_r^{(1)}\mathbf{v}_k^{(3)H}]$ and $E[\mathbf{v}_p^{(2)}\mathbf{v}_k^{(3)H}]$ can be obtained.

In the end, the CRB is written as the inverse of $\Omega(\boldsymbol{\eta})$:

$$\begin{bmatrix} \mathbf{CRB}_{\beta\beta} & \mathbf{CRB}_{\beta\varphi} & \mathbf{CRB}_{\beta\psi} & \mathbf{CRB}_{\beta\tau} \\ \mathbf{CRB}_{\varphi\beta} & \mathbf{CRB}_{\varphi\varphi} & \mathbf{CRB}_{\varphi\psi} & \mathbf{CRB}_{\varphi\tau} \\ \mathbf{CRB}_{\psi\beta} & \mathbf{CRB}_{\psi\varphi} & \mathbf{CRB}_{\psi\psi} & \mathbf{CRB}_{\psi\tau} \\ \mathbf{CRB}_{\tau\beta} & \mathbf{CRB}_{\tau\varphi} & \mathbf{CRB}_{\tau\psi} & \mathbf{CRB}_{\tau\tau} \end{bmatrix} = \Omega^{-1}(\boldsymbol{\eta}). \quad (\text{A15})$$

The CRB results of the four sets of channel parameters $\{\beta_u\}_{u=1}^U$, $\{\varphi_u\}_{u=1}^U$, $\{\psi_u\}_{u=1}^U$ and $\{\tau_u\}_{u=1}^U$ can be obtained from $\mathbf{CRB}_{\beta\beta}$, $\mathbf{CRB}_{\varphi\varphi}$, $\mathbf{CRB}_{\psi\psi}$ and $\mathbf{CRB}_{\tau\tau}$, respectively.

References

1. Rappaport, T.S.; Sun, S.; Mayzus, R.; Zhao, H.; Azar, Y.; Wang, K.; Wong, G.N.; Schulz, J.K.; Samimi, M.; Gutierrez, F. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access* **2013**, *1*, 335–349. [\[CrossRef\]](#)
2. Swindlehurst, A.L.; Ayanoglu, E.; Heydari, P.; Capolino, F. Millimeter-wave massive MIMO: The next wireless revolution? *IEEE Commun. Mag.* **2014**, *52*, 56–62. [\[CrossRef\]](#)
3. Pi, Z.; Choi, J.; Heath, R. Millimeter-wave gigabit broadband evolution toward 5G: Fixed access and backhaul. *IEEE Commun. Mag.* **2016**, *54*, 138–144. [\[CrossRef\]](#)
4. Ciuonzo, D.; Rossi, P.S.; Dey, S. Massive MIMO Channel-Aware Decision Fusion. *IEEE Trans. Signal Process.* **2015**, *63*, 604–619. [\[CrossRef\]](#)
5. Bana, A.S.; Carvalho, E.D.; Soret, B.; Abrão, T.; Marinello, J.C.; Larsson, E.G.; Popovski, P. Massive MIMO for Internet of Things (IoT) connectivity. *Phys. Commun.* **2019**, *37*, 100859–100875. [\[CrossRef\]](#)
6. Dey, I.; Ciuonzo, D.; Rossi, P.S. Wideband Collaborative Spectrum Sensing Using Massive MIMO Decision Fusion. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5246–5260. [\[CrossRef\]](#)
7. Tsang, Y.M.; Poon, A.S.Y.; Addepalli, S. Coding the Beams: Improving Beamforming Training in mmWave Communication System. In Proceedings of the 2011 IEEE Global Telecommunications Conference—GLOBECOM 2011, Houston, TX, USA, 5–9 December 2011; pp. 1–6.
8. Kutty, S.; Sen, D. Beamforming for Millimeter Wave Communications: An Inclusive Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 949–973. [\[CrossRef\]](#)
9. Han, S.; Chih-Lin, I.; Xu, Z.; Rowell, C. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun. Mag.* **2015**, *53*, 186–194. [\[CrossRef\]](#)
10. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846. [\[CrossRef\]](#)

11. Gao, Z.; Dai, L.; Mi, D.; Wang, Z.; Imran, M.A.; Shakir, M.Z. MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network. *IEEE Wirel. Commun.* **2015**, *22*, 13–21. [\[CrossRef\]](#)
12. Gao, X.; Dai, L.; Han, S.; Chih-Lin, I.; Heath, R.W. Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems with Large Antenna Arrays. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 998–1009. [\[CrossRef\]](#)
13. Kassam, J.; Miri, M.; Magueta, R.; Castanheira, D.; Gameiro, A. Two-Step Multiuser Equalization for Hybrid mmWave Massive MIMO GFDM Systems. *Electronics* **2020**, *9*, 1220. [\[CrossRef\]](#)
14. Magueta, R.; Castanheira, D.; Pedrosa, P.; Silva, A.; Dinis, R.; Gameiro, A. Iterative Analog–Digital Multi-User Equalizer for Wideband Millimeter Wave Massive MIMO Systems. *Sensors* **2020**, *20*, 575. [\[CrossRef\]](#)
15. Alkhateeb, A.; Heath, R.W. Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems. *IEEE Trans. Commun.* **2016**, *64*, 1801–1818. [\[CrossRef\]](#)
16. Ngo, H.Q.; Larsson, E.G.; Marzetta, T.L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Trans. Commun.* **2013**, *61*, 1436–1449.
17. Zhang, J.; Dai, L.; Sun, S.; Wang, Z. On the Spectral Efficiency of Massive MIMO Systems with Low-Resolution ADCs. *IEEE Commun. Lett.* **2016**, *20*, 842–845. [\[CrossRef\]](#)
18. Jin, S.; Liang, X.; Wong, K.K.; Gao, X.; Zhu, Q. Ergodic Rate Analysis for Multipair Massive MIMO Two-Way Relay Networks. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 1480–1491. [\[CrossRef\]](#)
19. Hur, S.; Kim, T.; Love, D.J.; Krogmeier, J.V.; Thomas, T.A.; Ghosh, A. Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks. *IEEE Trans. Commun.* **2013**, *61*, 4391–4403. [\[CrossRef\]](#)
20. Kim, T.; Love, D.J. Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels. In Proceedings of the 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Stockholm, Sweden, 27 June–1 July 2015; pp. 146–150.
21. Heath, R.W.; González-Prelcic, N.; Rangan, S.; Roh, W.; Sayeed, A.M. An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 436–453. [\[CrossRef\]](#)
22. Schniter, P.; Sayeed, A. Channel estimation and precoder design for millimeter-wave communications: The sparse way. In Proceedings of the 2014 48th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2–5 November 2014; pp. 273–277.
23. Gao, X.; Dai, L.; Zhou, S.; Sayeed, A.M.; Hanzo, L. BeamSpace Channel Estimation for Wideband Millimeter-Wave MIMO with Lens Antenna Array. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
24. Cheng, L.; Yue, G.; Yu, D.; Liang, Y.; Li, S. Millimeter Wave Time-Varying Channel Estimation via Exploiting Block-Sparse and Low-Rank Structures. *IEEE Access* **2019**, *7*, 123355–123366. [\[CrossRef\]](#)
25. Fazal-E-Asim; Antreich, F.; Cavalcante, C.C.; de Almeida, A.L.F.; Nossek, J.A. Two-Dimensional Channel Parameter Estimation for Millimeter-Wave Systems Using Butler Matrices. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 2670–2684. [\[CrossRef\]](#)
26. Bajwa, W.U.; Haupt, J.; Sayeed, A.M.; Nowak, R. Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels. *Proc. IEEE* **2010**, *98*, 1058–1076. [\[CrossRef\]](#)
27. Nguyen, S.L.H.; Ghayeb, A. Compressive sensing-based channel estimation for massive multiuser MIMO systems. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 7–10 April 2013; pp. 2890–2895.
28. Alkhateeb, A.; Leus, G.; Heath, R.W. Compressed sensing based multi-user millimeter wave systems: How many measurements are needed? In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 2909–2913.
29. Xie, H.; Gao, F.; Zhang, S.; Jin, S. A Unified Transmission Strategy for TDD/FDD Massive MIMO Systems with Spatial Basis Expansion Model. *IEEE Trans. Veh. Technol.* **2017**, *66*, 3170–3184. [\[CrossRef\]](#)
30. Wang, S.; Li, Y.; Wang, J. Multiuser Detection in Massive Spatial Modulation MIMO with Low-Resolution ADCs. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 2156–2168. [\[CrossRef\]](#)
31. Zhang, J.; Podkurkov, I.; Haardt, M.; Nadeev, A. Channel Estimation and Training Design for Hybrid Analog-Digital Multi-Carrier Single-User Massive MIMO Systems. In Proceedings of the WSA 2016; 20th International ITG Workshop on Smart Antennas, Munich, Germany, 9–11 March 2016; pp. 1–8.
32. You, L.; Gao, X.; Swindlehurst, A.L.; Zhong, W. Channel Acquisition for Massive MIMO-OFDM with Adjustable Phase Shift Pilots. *IEEE Trans. Signal Process.* **2016**, *64*, 1461–1476. [\[CrossRef\]](#)
33. Wei, X.; Peng, W.; Ng, D.; Robert, S.; Jiang, T. Joint Estimation of Channel Parameters in Massive MIMO Systems via PARAFAC Analysis. In Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC), Maui, HI, USA, 5–8 March 2018; pp. 496–502.
34. Wei, X.; Peng, W.; Chen, D.; Ng, D.; Jiang, T. Joint Channel Parameter Estimation in Multi-Cell Massive MIMO System. *IEEE Trans. Commun.* **2019**, *67*, 3251–3264. [\[CrossRef\]](#)
35. Zhou, Z.; Fang, J.; Yang, L.; Li, H.; Chen, Z.; Li, S. Channel Estimation for Millimeter-Wave Multiuser MIMO Systems via PARAFAC Decomposition. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7501–7516. [\[CrossRef\]](#)
36. Gao, Z.; Hu, C.; Dai, L.; Wang, Z. Channel Estimation for Millimeter-Wave Massive MIMO with Hybrid Precoding Over Frequency-Selective Fading Channels. *IEEE Commun. Lett.* **2016**, *20*, 1259–1262. [\[CrossRef\]](#)

-
37. Venugopal, K.; Alkhateeb, A.; González Precic, N.; Heath, R.W. Channel Estimation for Hybrid Architecture-Based Wideband Millimeter Wave Systems. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1996–2009. [[CrossRef](#)]
 38. Zhou, Z.; Fang, J.; Yang, L.; Li, H.; Chen, Z.; Blum, R.S. Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1524–1538. [[CrossRef](#)]
 39. Akdeniz, M.R.; Liu, Y.; Samimi, M.K.; Sun, S.; Rangan, S.; Rappaport, T.S.; Erkip, E. Millimeter Wave Channel Modeling and Cellular Capacity Evaluation. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1164–1179. [[CrossRef](#)]
 40. Rajih, M.; Comon, P. Enhanced Line Search: A novel method to accelerate Parafac. In Proceedings of the 2005 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005; pp. 1–4.
 41. Liu, X.; Sidiropoulos, N. Cramer–Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Trans. Signal Process.* **2001**, *49*, 2074–2086.
 42. Zhou, Y.; Cheung, Y.M. Bayesian Low-Tubal-Rank Robust Tensor Factorization with Multi-Rank Determination. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 62–76. [[CrossRef](#)]