Joint Channel Parameter Estimation in Multi-Cell Massive MIMO System

Xiao Wei, Wei Peng, Da Chen, Derrick Wing Kwan Ng, and Tao Jiang

***Abstract*— In this paper, we consider the uplink channel parameter estimation problem in the presence of pilot contamination for massive multiple-input-multiple-output (MIMO) systems. We propose a *parallel factor* (PARAFAC)-based estimation scheme, which exploits the low-rank property of massive MIMO channels caused by the finite scattering in a physical environment. Specifically, we first parameterize the channel in terms of three parameters, i.e., fading coefficients, directions of arrival (DOAs), and delays; thereby, the channel is characterized via three equivalent PARAFAC models. Then, the proposed PARAFAC-based scheme is developed, which jointly estimates these three channel parameters using an alternating least squares (ALS) algorithm. Therein, we certify the *identifiability* of the three channel parameters of the PARAFAC models to mitigate the pilot contamination and state the convergence of the ALS algorithm, which guarantees that the three channel parameters can be uniquely determined with the proposed scheme. Moreover, to further reduce the computational complexity, two advanced schemes are proposed by antenna selection and reducing the estimation frequency of DOAs and delays, respectively. Simulation results show that the proposed schemes can achieve both low computational complexities and close to optimal Cramer–Rao Bound performance.**

***Index Terms*— Massive MIMO, channel parameter estimation, pilot contamination, computational complexity, CRB, PARAFAC analysis.**

I. Introduction

RADIO spectrum is a scarce resource in wireless communication, and thus high spectral efficiency is a fundamental goal of communication system design. In addition, next generation communication systems are expected to support multiple users and to guarantee stringent quality of service (QoS) requirements. As a result, a spectrally-efficient communication system design is important to overcome the spectrum scarcity. In particular, massive multiple-input multiple-output (MIMO) has received significant attention in recent years [1], since it offers a large number of spatial degrees of freedom which facilitate large multiplexing and diversity gains [2]–[4].

In a massive MIMO system, a base station (BS) is generally equipped with hundreds or even thousands of antennas to simultaneously serve multiple users. To fully embrace the benefits of massive MIMO, the BS needs to acquire a large number of channel parameters associated to each user, including fading coefficients, directions of arrival (DOAs), and delays, before performing precoding, resource allocation, coherent detection, and etc [5]. Moreover, pilot interference from adjacent cells, which gives rise to so-called pilot contamination, severely limits the channel parameter estimation performance. Pilot contamination is a fundamental and intractable problem for massive MIMO systems as it is the only remaining impairment with unlimited number of antennas [6]. Hence, channel parameter estimation in the presence of pilot contamination is an important but challenging problem in massive MIMO system design.

The delay or/and DOA estimation problem are widely studied for massive MIMO systems taking into account the correlation of massive MIMO channels across different times and antennas [7], [8], respectively. For example, Chen and Yang [9] utilized channel sparsity caused by channel correlation in the time domain to acquire the delay of each path in massive MIMO systems. The sparse nature of massive MIMO channels is also exploited for delay estimation by minimizing the least-squares error function in [10]. On the other hand, DOA estimation is investigated by exploiting another feature of massive MIMO channels caused by channel correlation across different antennas, i.e., the columns of the covariance matrix of the received useful signal are orthogonal to those of the noise matrix [11]. Utilizing this feature, subspace algorithms including the Multiple Signal Classification (MUSIC) and the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [12] have been widely adopted for DOA estimation in massive MIMO systems. By spectral-peak searching, the MUSIC algorithm obtains a high accuracy for DOA estimation in [13], however, it takes a lot of computation. Hence, ESPRIT is widely adopted to reduce the complexity. For instance, [14] proposed an ESPRIT-like approach for DOA estimation, and achieves a good performance. On this basis, [15] investigated the estimation of DOAs and delays

Manuscript received July 30, 2018; revised November 11, 2018; accepted January 6, 2019. Date of publication January 18, 2019; date of current version May 15, 2019. This work was supported in part by the NSFC with Grants 61771216, 61771214, 61601191, 61631015, 61501193, and 61325004, Fundamental Research Funds for the Central Universities with Grant 2015ZDTD012, Major Program of Natural Science Foundation of Hubei in China with Grant 2016CFA009, and the Australian Research Council's Discovery Early Career Researcher Award funding scheme with Grant DE170100137. This paper was presented in part at the IEEE ICNC 2018. The associate editor coordinating the review of this paper and approving it for publication was X. Wang. *(Corresponding author: Tao Jiang.)*

X. Wei, W. Peng, D. Chen, and T. Jiang are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: weixiao1991@hust.edu.cn; pengwei@hust.edu.cn; chenda@hust.edu.cn; tao.jiang@ieee.org).

D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: w.k.ng@unsw.edu.au).

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/TCOMM.2019.2893276

by employing a two-dimensional ESPRIT to separate and estimate the DOAs and delays. However, the computational complexity of ESPRIT is also high in massive MIMO systems as it increases cubically with the number of antennas.

Moreover, estimating only two parameters, i.e., delay and DOA, is not enough to accurately model the overall channel. In practice, an inaccurate channel estimation causes a loss in system performance. In particular, fading coefficient is also an important parameter for efficient representation of a channel matrix. There are several research efforts on the investigation of the fading coefficient estimation problem in *conventional* MIMO systems, e.g. [16], [17]. For example, Chakraborty and Sen [16] proposed two iterative algorithms based on the maximum likelihood (ML) for joint estimation of time, frequency offsets, and fading coefficients in multi-relay distributed MIMO-orthogonal frequency division multiplexing (OFDM) system. With the ML algorithm, priori knowledge of the correlation of massive MIMO channels cannot be utilized, which causes a performance loss. Moreover, fading coefficient and DOA estimation for millimeter-wave MIMO is solved via *Parallel Factor* (PARAFAC) analysis in [17]. Recently, it is also employed for mmWave MIMO-OFDM system in [18]. PARAFAC analysis is a highly efficient algorithm for joint parameter estimation method for multi-way signals with low-rank property. It has been widely adopted in wireless communication as it enjoys a uniqueness property under simple conditions summarized by Kruskal [19]. Motivated by this, in [20], we investigated the joint estimation of three channel parameters for massive MIMO system via PARAFAC analysis. However, pilot contamination is not considered in [20]. In this case, the scheme of [20] suffers from a severe performance loss. Besides, in [20], a large number of computation steps is required when the number BS antennas is large, which degrades the system performance as it consumes an exceedingly large portion of the coherence time.

In this paper, we address the above issues. To this end, the joint estimation of channel fading coefficients, DOAs, and delays in the presence of pilot contamination is explored for massive MIMO systems. We first develop a comprehensive channel model for massive MIMO systems comprising all three relevant channel parameters. Thereby, we study the joint estimation problem of massive MIMO system in the presence of pilot contamination based on the aforework of PARAFAC analysis. Besides, in our previous work [21], the computational complexity reduction utilizing the correlation of massive MIMO channels across different antennas was investigated. Thereby, the correlation across both different antennas and times is used to further reduce the computational complexity of the proposed scheme. The main contributions of this paper are summarized as follows:

- Using the low-rank property of channel, we parameterize the massive MIMO channel in terms of fading coefficients, DOAs, and delays, thereby the channel is characterized via three equivalent PARAFAC models. Therein, the *identifiability* of the three parameters is proved to mitigate pilot contamination, which guarantees that the three channel parameters can be uniquely determined.
- The joint estimation of three parameters in the presence of pilot contamination is formulated as an optimization problem with three coupled variables, which is solved by an alternating least squares (ALS) algorithm. Subsequently, we state the convergence of the ALS algorithm, which shows that the proposed estimates are close-to-optimal.
- The computational complexity of the proposed scheme is analyzed in terms of the required number of floating-point operations (FLOPs). Then, we further develop two computationally efficient schemes by antenna selection and reducing the estimation frequency of DOAs and delays, respectively. These two schemes obtain nearly the same accuracy performance as PARAFAC-based scheme but have much lower computational complexities.
- The Cramer-Rao Bound (CRB) of all three channel parameters for the considered massive MIMO system is derived for comparison, which is the lower bound of parameter estimation for a given set of observations. Moreover, simulation results reveal that the performance of the proposed schemes nearly obtains the CRB.

The remainder of this paper is organized as follows. In Section II, the multi-cell massive MIMO system model is presented. The PARAFAC-based joint estimation scheme for channel parameters in the presence of pilot contamination is proposed in Section III, and an exact computational complexity analysis and two schemes for complexity reduction are provided in Section IV. Simulation results are presented in Section V, and conclusions are drawn in Section VI.

Notations: In this paper, we use boldface capital and lower case letters to denote matrices and vectors, respectively. $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\dagger}$ represent the transpose and pseudo-inverse of matrix $\mathbf{A}$, respectively; $||\mathbf{A}||_{\mathrm{F}}$ and $\mathrm{tr}(\mathbf{A})$ denote the Frobenius norm and trace of matrix $\mathbf{A}$, respectively; $A \rightarrow B$ indicates that A converges to B in the limit; $\mathbf{G}_1(\mathbf{a})$ and $\mathbf{G}_{n_b}(\mathbf{B})$ denote diagonal matrices with the diagonal elements given by vector $\mathbf{a}$ and row n_b of matrix $\mathbf{B}$, respectively; $a \in \mathbb{C}$ means that a is a complex number; $\mathbf{A} \in \mathbb{C}^{N_r \times N_b}$ is the set of complex $N_r \times N_b$ matrices; $\mathcal{CN}(0, \sigma^2)$ denotes the complex Gaussian distribution with mean 0 and variance σ^2; $\mathbf{A} \odot \mathbf{B}$ denotes the Khatri-Rao product of $\mathbf{A} \in \mathbb{C}^{N_a \times N}$ and $\mathbf{B} \in \mathbb{C}^{N_b \times N}$; $\mathbf{I}_N$ is N-dimensional identity matrix. For convenience, the important symbols and notations used in this paper are defined in Table I.

II. System Model

In this section, we present the uplink multi-cell massive MIMO system model. Then, the multipath channel is parameterized in terms of fading coefficients, DOAs, and delays.

A. Uplink Massive MIMO System

We consider an uplink massive MIMO system of N_c cells with a N_a-antenna BS and N_u single-antenna users per cell, where the considered transmission format is shown in Fig. 1. We assume time-division multiple access (TDMA) between N_s-length pilot transmissions of intra-cell users with a guard interval of length N_g between adjacent transmission, such that intra-cell interference can be avoided [22]. Thus, a single user in each cell, $N_u = 1$, is assumed to simplify the notations.

TABLE I
SYMBOLS AND NOTATIONS

Symbol	Definition
N_c	Number of cells
N_a	Number of antennas
N_s	Length of pilot signals
N_u	Number of antennas
N_b	Number of channel blocks
N_g	Guard interval in TDMA
N_f	Number of taps of the overall channel
N_p	Number of resolvable propagation paths between user and BS
T	Symbol period
b_{n_c,n_b,n_p}	Fading coefficient associated with path n_p in block n_b
θ_{n_c,n_p}	DOA of path n_p at the BS in cell n_c
τ_{n_c,n_p}	Delay of path n_p to the cell n_c
$d(t)$	Pulse shaping function
$s(n)$	Pilot symbol in time slot n
$\mathbf{v}_{n_b}(n_s)$	White Gaussian noise in time slot n_s of block b
$\mathbf{x}_{n_b}(n_s)$	Received signals in time slot n_s of block b
$\mathbf{h}_{n_c,n_b}(n_fT)$	Channel impulse response between the user in cell n_c and the target-cell BS
$\mathbf{a}(\theta_{n_c,n_p})$	Steering vector of path p
$\mathbf{H}$	Matrix of channels
$\mathbf{A}$	Matrix of DOAs
$\mathbf{B}$	Matrix of fading coefficients
$\mathbf{D}$	Matrix of delays
$\mathbf{S}$	Matrix of pilot signals
$\mathbf{V}$	Matrix of noise
$\mathbf{X}$	Received signals of all blocks

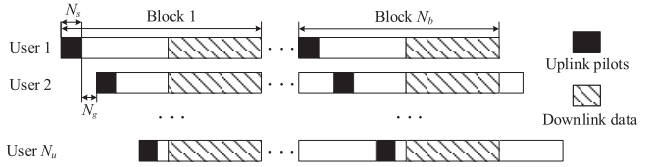


Fig. 1. The considered transmission format in each cell, where each user sends pilots in the uplink in N_s time slots, N_b is the number of blocks, and the guard interval spanning N_g symbol interval is used to avoid inter-user interference.

On the other hand, users of different cells may send the same pilot signals to their respective BSs in a same time slot, which gives rise to the inter-cell interference, i.e., pilot contamination problem, which is the focus of this paper.

Without loss of generality, let the first cell be the target cell, and other N_c-1 cells are the interfering cells. In the uplink training phase, the pilot signals of each user are filtered by the transmit filter, sent through the communication channel, and filtered by the receive filter. Thereby, the uplink received signals $\mathbf{x}_{n_b}(n_s) \in \mathbb{C}^{N_a\times 1}$ at the target cell in time slot $n_s \in \{1,\cdots,N_s\}$ of block $n_b \in \{1,\cdots,N_b\}$ is written as [23]

$$\mathbf{x}_{n_b}(n_s)=\sum_{n_c=1}^{N_c}\sum_{n_f=0}^{N_f-1}\mathbf{h}_{n_c,n_b}(n_fT)s(n_s-n_f)+\mathbf{v}_{n_b}(n_s), \quad (1)$$

where T is the symbol period, N_f is the number of taps of the overall channel, and N_b is the number of blocks. $\mathbf{h}_{n_c,n_b}(n_fT) \in \mathbb{C}^{N_a\times 1}$ is the channel impulse response between the user in cell n_c and N_a BS antennas in the target cell, which has finite support and encompasses the effects of the transmit filter, receive filter, and channel. $s(n)$, $n \in \{1-N_f,\cdots,N_s\}$, are the pilot symbols, which is periodically extended over multiple time-slots, and if $n \le 0$, we have $s(n)=N_s+n$. $\mathbf{v}_{n_b}(n_s) \in \mathbb{C}^{N_a\times 1}$ is additive white Gaussian noise with $\mathcal{CN}(0,\sigma^2)$, whose elements are independent and identically distributed (i.i.d.) variables.

B. Multipath Channel Model

We consider a realistic channel model assuming a physical environment with finite scattering, cf. [24]. Therein, the angle domain is partitioned into a finite number of directions causing N_p i.i.d. resolvable propagation paths between each user and the target BS. N_p is generally independent of the number of antennas, N_a, and the number of users, N_u, and determined by the number of scatterers in the propagation environment only. Hence, the channel impulse response of a user in cell n_c, $\mathbf{h}_{n_c,n_b}(n_fT)$, can be expressed as [23]

$$\mathbf{h}_{n_c,n_b}(n_fT)=\sum_{n_p=1}^{N_p} b_{n_c,n_b,n_p}\mathbf{a}(\theta_{n_c,n_p})d(n_fT-\tau_{n_c,n_p}), \quad (2)$$

where $b_{n_c,n_b,n_p} \in \mathbb{C}$ is the fading coefficient associated with path n_p in block n_b, $\theta_{n_c,n_p} \in \mathbb{R}$ denotes the DOA of path n_p at the BS in cell n_c, and $\tau_{n_c,n_p} \in \mathbb{R}$ is the delay of path n_p to the cell n_c. Therein, the DOA θ_{n_c,n_p} is assumed to not overlap with other DOAs to guarantee the independence of different DOAs. $d(t) \in \mathbb{C}$ is the known pulse shaping function,[1] and it is typically the convolution of transmit filter and receive filter. Furthermore, $\mathbf{a}(\theta_{n_c,n_p}) \in \mathbb{C}^{N_a\times 1}$ is the steering vector, which denotes the correlation of channels across different BS antennas caused by that the BS antennas are densely placed and not well separated in massive MIMO systems. For a uniform linear array, the steering vector is given by [26]

$$\mathbf{a}(\theta_{n_c,n_p})=\left[1,e^{-j\omega_{n_c,n_p}},\cdots,e^{-j\omega_{n_c,n_p}(N_a-1)}\right]^{\mathrm{T}}, \quad (3)$$

where $\omega_{n_c,n_p}=\frac{2\pi l}{\lambda}\cos(\theta_{n_c,n_p})$, l is the antenna spacing at the BS, and λ is the wavelength of the center frequency. In the context of this multipath channel model, the guard interval satisfying $N_g \ge N_f+\frac{\tau_{N_p}}{T}-1$ is necessary to guarantee that the pilot sequences sent by different users in each cell do not overlap at the BS.

In this model, each path of the signals from the user in cell n_c is parameterized by $(b_{n_c,n_b,n_p},\ \theta_{n_c,n_p},\tau_{n_c,n_p})$. Fading coefficient b_{n_c,n_b,n_p} is generally modeled as a complex Gaussian random variable[2] following the distribution of $\mathcal{CN}(0,g_{n_c}^2)$, where g_{n_c} is the pathloss between users in cell n_c and the target BS. For a block-fading channel model,

[1]We assume that $d(t)$ is bandlimited and sampled at or above the Nyquist rate. In this case, the transmit filter can be a windowed filter, and the receive filter can be a matched filter [25].

[2]Although we assume that fading coefficients of different blocks are independent to facilitate concise and insightful results, it is not a necessary assumption. The proposed scheme and the following analytical results can be easily extended to the correlated case such as the scenario with the first order autoregressive model or the Jakes' model.

fading coefficient is assumed to remain constant over a block of symbols, while the fading coefficients of different blocks are modeled as independent random variables. Moreover, the variations of the DOAs and delays across different blocks are generally much smaller than that of the fading coefficients, which means that the DOAs and delays remain unchanged for a relatively long period of time, e.g., N_b adjacent blocks is assumed in this paper. In mobile communication systems, this assumption is reasonable if N_b is chosen in accordance with the mobile speed and multipath geometry. As a consequence, channels across different times are correlated.

III. PARAFAC-Based Joint Parameter Estimation Scheme

In this section, we derive the PARAFAC models of massive MIMO channels in matrix form to facilitate the PARAFAC analysis. Thereby, we realize the PARAFAC-based joint estimation using the ALS algorithm.

A. PARAFAC Model in Matrix Form

The channel model in (2) includes a sum of triple terms, which is known as the PARAFAC model [27]. Therein, each term is characterized by $(b_{n_c,n_b,n_p}, \theta_{n_c,n_p}, \tau_{n_c,n_p})$, respectively. We represent this PARAFAC model in matrix form.

1) Least Squares Estimate: Let us define the $N_aN_b \times N_s$ received signal matrix for all N_b blocks in the target cell as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1(1) & \mathbf{x}_1(2) & \cdots & \mathbf{x}_1(N_s) \\ \mathbf{x}_2(1) & \mathbf{x}_2(2) & \cdots & \mathbf{x}_2(N_s) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N_b}(1) & \mathbf{x}_{N_b}(2) & \cdots & \mathbf{x}_{N_b}(N_s) \end{bmatrix}.$$

Then, based on (1) and (2), $\mathbf{X}$ can be decomposed into

$$\mathbf{X} = \sum_{n_c=1}^{N_c} \mathbf{H}_{n_c}\mathbf{S} + \mathbf{V} = \sum_{n_c=1}^{N_c} \begin{bmatrix} \mathbf{A}_{n_c}\mathbf{G}_1(\mathbf{B}_{n_c}) \\ \mathbf{A}_{n_c}\mathbf{G}_2(\mathbf{B}_{n_c}) \\ \vdots \\ \mathbf{A}_{n_c}\mathbf{G}_{N_b}(\mathbf{B}_{n_c}) \end{bmatrix} \mathbf{D}_{n_c}^{\mathrm{T}}\mathbf{S} + \mathbf{V}$$

$$= \begin{bmatrix} \mathbf{A}\mathbf{G}_1(\mathbf{B}) \\ \mathbf{A}\mathbf{G}_2(\mathbf{B}) \\ \vdots \\ \mathbf{A}\mathbf{G}_{N_b}(\mathbf{B}) \end{bmatrix} \mathbf{D}^{\mathrm{T}}\mathbf{S} + \mathbf{V}, \tag{4}$$

where $\mathbf{G}_{n_b}(\mathbf{B})$ denotes the diagonal matrix with the diagonal elements given by the row n_b of matrix $\mathbf{B}$, $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_{N_c}] \in \mathbb{C}^{N_a \times N_cN_p}$, $\mathbf{B} = [\mathbf{B}_1, \cdots, \mathbf{B}_{N_c}] \in \mathbb{C}^{N_b \times N_cN_p}$,

$$\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_{N_c}] \in \mathbb{C}^{N_f \times N_cN_p},$$

$$\mathbf{S} = \begin{bmatrix} s(1) & s(2) & \cdots & s(N_s) \\ s(N_s) & s(1) & \cdots & s(N_s-1) \\ \vdots & \vdots & \ddots & \vdots \\ s(N_s-N_f+2) & s(3) & \cdots & s(N_s-N_f+1) \end{bmatrix},$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1(1) & \mathbf{v}_1(2) & \cdots & \mathbf{v}_1(N_s) \\ \mathbf{v}_2(1) & \mathbf{v}_2(2) & \cdots & \mathbf{v}_2(N_s) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_{N_b}(1) & \mathbf{v}_{N_b}(2) & \cdots & \mathbf{v}_{N_b}(N_s) \end{bmatrix},$$

$$\mathbf{H}_{n_c} = \begin{bmatrix} \mathbf{h}_{n_c,1}(T) & \mathbf{h}_{n_c,1}(2T) & \cdots & \mathbf{h}_{n_c,1}(N_fT) \\ \mathbf{h}_{n_c,2}(T) & \mathbf{h}_{n_c,2}(2T) & \cdots & \mathbf{h}_{n_c,2}(N_fT) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{n_c,N_b}(T) & \mathbf{h}_{n_c,N_b}(2T) & \cdots & \mathbf{h}_{n_c,N_b}(N_fT) \end{bmatrix}.$$

Therein, $\mathbf{S}$ is a Toeplitz matrix, $\mathbf{A}_{n_c} = [\mathbf{a}(\theta_{n_c,n_1}), \mathbf{a}(\theta_{n_c,n_2}), \cdots, \mathbf{a}(\theta_{n_c,N_p})] \in \mathbb{C}^{N_a \times N_p}$,

$$\mathbf{B}_{n_c} = \begin{bmatrix} b_{n_c,1,1} & b_{n_c,1,2} & \cdots & b_{n_c,1,N_p} \\ b_{n_c,2,1} & b_{n_c,2,2} & \cdots & b_{n_c,2,N_p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n_c,N_b,1} & b_{n_c,N_b,2} & \cdots & b_{n_c,N_b,N_p} \end{bmatrix},$$

$$\mathbf{D}_{n_c} = \begin{bmatrix} d(-\tau_{n_c,1}) & \cdots & d(-\tau_{n_c,N_p}) \\ d(T-\tau_{n_c,1}) & \cdots & d(T-\tau_{n_c,N_p}) \\ \vdots & \cdots & \vdots \\ d((N_f-1)T-\tau_{n_c,1}) & \cdots & d((N_f-1)T-\tau_{n_c,N_p}) \end{bmatrix}.$$

The received signal matrix $\mathbf{X}$ in (4) can be written in a compact form using the Khatri-Rao product, i.e.,

$$\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{V} = [\mathbf{B} \odot \mathbf{A}]\mathbf{D}^{\mathrm{T}}\mathbf{S} + \mathbf{V}, \tag{5}$$

where $\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_{N_c}] \in \mathbb{C}^{N_aN_b \times N_cN_f}$. Then, the least squares (LS) estimates of all channels are given by

$$\hat{\mathbf{H}} = \mathbf{X}\mathbf{S}^{\dagger} = [\mathbf{B} \odot \mathbf{A}]\mathbf{D}^{\mathrm{T}} + \mathbf{W}, \tag{6}$$

where $\mathbf{W} = \mathbf{V}\mathbf{S}^{\dagger} \in \mathbb{C}^{N_aN_b \times N_cN_f}$ follows $\mathcal{CN}(0, \sigma^2)$.

2) Channel Model Transformation: To ensure that the three parameters are *identifiable*, we use the fact that the Fourier transform maps a constant delay to a constant phase shift. This technique has been used before in delay estimation algorithms, e.g. [23]. Collect the samples of the known waveform $d(t)$ into a column vector $\mathbf{d}_0 = [d(0), d(T), \cdots, d((N_f-1)T)]^{\mathrm{T}} \in \mathbb{C}^{N_f \times 1}$. Then, the discrete Fourier transform (DFT) is applied to $d(t)$ yielding $\widetilde{\mathbf{d}}_0 = \mathbf{F}\mathbf{d}_0$, where $\mathbf{F} \in \mathbb{C}^{N_f \times N_f}$ is the DFT matrix given by

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \phi & \cdots & \phi^{N_f-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi^{N_f-1} & \cdots & \phi^{(N_f-1)(N_f-1)} \end{bmatrix},$$

where $\phi = e^{-j2\pi/N_f}$. The column $(n_c-1)N_p+n_p$ of matrix $\mathbf{D}$ (i.e., the column n_p of matrix $\mathbf{D}_{n_c}$) contains the samples of $d(t-\tau_{n_c,n_p})$, and it is denoted by $\mathbf{d}_{n_c,n_p} = [d(-\tau_{n_c,n_p}), d(T-\tau_{n_c,n_p}), \cdots, d((N_f-1)T-\tau_{n_c,n_p})]^{\mathrm{T}}$. As we assumed

before, $d(t)$ is bandlimited and sampled at or above the Nyquist rate, then the DFT of $\mathbf{d}_{n_c,n_p}$ is given by [23]

$$\begin{aligned}\widetilde{\mathbf{d}}_{n_c,n_p} &= \mathbf{F}\mathbf{d}_{n_c n_p} \\ &= \mathbf{G}(\widetilde{\mathbf{d}}_0)[1, \phi^{\tau_{n_c,n_p}}, \phi^{2\tau_{n_c,n_p}}, \cdots, \phi^{(N_f-1)\tau_{n_c,n_p}}]^{\mathrm{T}}. \quad (7)\end{aligned}$$

With the above notations, we write N_f-point DFT of $\mathbf{D}$ as

$$\widetilde{\mathbf{D}} = \mathbf{F}\mathbf{D} = [\widetilde{\mathbf{d}}_{1,1}, \cdots, \widetilde{\mathbf{d}}_{N_c,N_p}] = \mathbf{G}(\widetilde{\mathbf{d}}_0)\mathbf{C}, \quad (8)$$

where $\widetilde{\mathbf{D}} \in \mathbb{C}^{N_f \times N_c N_p}$, and $\mathbf{C}$ is a Vandermonde matrix as

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \phi^{\tau_{1,1}} & \phi^{\tau_{1,2}} & \cdots & \phi^{\tau_{N_c,N_p}} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{(N_f-1)\tau_{1,1}} & \phi^{(N_f-1)\tau_{1,2}} & \cdots & \phi^{(N_f-1)\tau_{N_c,N_p}} \end{bmatrix}.$$

Then, we define $\widetilde{\mathbf{H}}$ as [23]

$$\begin{aligned}\widetilde{\mathbf{H}} &= \hat{\mathbf{H}}\mathbf{F}\mathbf{G}(\widetilde{\mathbf{d}}_0)^{\dagger} = [\mathbf{B} \odot \mathbf{A}]\mathbf{D}^{\mathrm{T}}\mathbf{F}\mathbf{G}(\widetilde{\mathbf{d}}_0)^{\dagger} + \mathbf{W}\mathbf{F}\mathbf{G}(\widetilde{\mathbf{d}}_0)^{\dagger} \\ &= [\mathbf{B} \odot \mathbf{A}]\left(\mathbf{G}(\widetilde{\mathbf{d}}_0)^{\dagger}\mathbf{F}\mathbf{D}\right)^{\mathrm{T}} + \mathbf{W}_1 = [\mathbf{B} \odot \mathbf{A}]\mathbf{C}^{\mathrm{T}} + \mathbf{W}_1, \quad (9)\end{aligned}$$

where $\mathbf{W}_1 = \mathbf{W}\mathbf{F}\mathbf{G}(\widetilde{\mathbf{d}}_0)^{\dagger} \in \mathbb{C}^{N_a N_b \times N_f}$ following the distribution of $\mathcal{CN}(0, \sigma^2)$. From (9), we can observe that the matrix of the useful signals in $\widetilde{\mathbf{H}}$ is low-rank since $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are correlated, respectively. Thus, (9) is referred to as the PARAFAC model in matrix form [32]. The perfect symmetry of PARAFAC model in (9) allows two more equivalent representations of channel, which is interpreted as stacking three-way array along different dimensions. In particular,

$$\widetilde{\mathbf{Y}} = (\mathbf{A} \odot \mathbf{C})\mathbf{B}^{\mathrm{T}} + \mathbf{W}_2 \in \mathbb{C}^{N_f N_a \times N_b}, \quad (10)$$

$$\widetilde{\mathbf{Z}} = (\mathbf{C} \odot \mathbf{B})\mathbf{A}^{\mathrm{T}} + \mathbf{W}_3 \in \mathbb{C}^{N_b N_f \times N_a}, \quad (11)$$

where $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are alternative versions of $\widetilde{\mathbf{H}}$, and $\mathbf{W}_2$ and $\mathbf{W}_3$ are alternative versions of $\mathbf{W}_1$.

B. Identifiability of Channel Parameters

In the aforementioned three equivalent PARAFAC models, we decompose the channel matrix into three matrices corresponding to three parameters respectively. In order to obtain a unique solution of three parameters and mitigate the impairment of pilot contamination, the *uncertainty* of $\mathbf{A}_1$, $\mathbf{B}_1$, and $\mathbf{C}_1$ in matrix decomposition should be reduced. Note that Kruskal [19] provides some uniqueness results for PARAFAC analysis. Therein, a distinguishing feature of PARAFAC models is that their decompositions are often unique due to low-rank property, which means that this is the only possible combination of rank-N_f matrices that sums to $\mathbf{H}$, whereas matrix decompositions are not. Based on this, we first discuss the uniqueness of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in the considered system. To this end, the following proposition is presented:

Proposition 1: The channel matrix decomposition in the considered PARAFAC model is unique up to permutation of columns of the matrices $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$*.*

Proof: Please refer to Appendix A. ∎

Remark 1: Proposition 1 suggests that only a permutation uncertainty for matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ exists for the channel matrix decomposition in the considered system. However, this permutation uncertainty does not have to be solved with regard to the parameter estimation problem, since the ordering of channel parameters is not important.

Remark 2: Note that the distance between adjacent-cell users and the target BS is generally much longer than that of the distance between target-cell users and the target BS, which means that the pathloss of adjacent-cell channels is much larger than that of target-cell channels. Thus, the variance of the $\mathbf{B}_1$ is much larger than that of the $\mathbf{B}_{n_c}, n_c = 2, \cdots, N_c$. As a consequence, the N_p columns with most significant variance of $\mathbf{B}$ are the N_p columns of $\mathbf{B}_1$. The corresponding N_p columns of $\mathbf{A}$ and $\mathbf{C}$ are the N_p columns of $\mathbf{A}_1$ and $\mathbf{C}_1$, respectively. The larger N_a, the more accurate the estimate of pathloss, and thereby the impairment of the pilot contamination vanishes asymptotically for large N_a. As a consequence, the *identifiability* of channel parameters $\mathbf{A}_1$, $\mathbf{B}_1$, and $\mathbf{C}_1$ is guaranteed, which means a unique solution for the three parameters can be obtained from PARAFAC models.

C. Parameter Estimation

The ALS algorithm is firstly adopted to the equivalent PARAFAC models in (9)-(11) to obtain estimates of channel matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and the convergence of the ALS algorithm is also stated for close-to-optimal estimates. Then, based on the proof of Proposition 1, we mitigate the permutation uncertainty and pilot contamination and obtain the *identifiable* estimates of the three channel parameters.

1) ALS Solution: In order to jointly estimate the three parameters, we first solve the following optimization problem to obtain matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$,

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \parallel \widetilde{\mathbf{H}} - [\mathbf{B} \odot \mathbf{A}]\,\mathbf{C}^{\mathrm{T}} \parallel_{\mathrm{F}}^2 . \quad (12)$$

It is evident that, (12) is a nonconvex optimization problem due to the Kronecker product of three coupled variables [28], and it is an NP-hard problem. Additionally, another challenge is that the inaccuracy of one variable may be propagated to the other variables due to couplings, and the optimal solution is determined by all three variables jointly. In general, it is not feasible to optimize the three variables independently. Fortunately, the PARAFAC models in (9)-(11) reveal the hidden decomposability of this problem in the considered system. Thereby, the typical ALS algorithm is considered, which subdivides the coupled problem into three sub-problems according to the PARAFAC models, and then obtains LS estimate for one of the parameter matrices under the assumption that the other two parameter matrices are already known [18]. The estimation is then alternately repeated until all three matrices have been estimated. This entire algorithm is then iterated until convergence (which is proved in the proof of Proposition 2) is obtained. Therein, ALS solution of (12) in iteration n_r is

$$\begin{cases} \hat{\mathbf{C}}^{\mathrm{T}}(n_r) = \left[\hat{\mathbf{B}}(n_r-1) \odot \hat{\mathbf{A}}(n_r-1)\right]^{\dagger}\widetilde{\mathbf{H}} & (13a) \\ \hat{\mathbf{B}}^{\mathrm{T}}(n_r) = \left[\hat{\mathbf{A}}(n_r-1) \odot \hat{\mathbf{C}}(n_r)\right]^{\dagger}\widetilde{\mathbf{Y}} & (13b) \\ \hat{\mathbf{A}}^{\mathrm{T}}(n_r) = \left[\hat{\mathbf{C}}(n_r) \odot \hat{\mathbf{B}}(n_r)\right]^{\dagger}\widetilde{\mathbf{Z}}, & (13c) \end{cases}$$

where $n_r \in \{1, 2, \cdots, N_r\}$, N_r is required number of iterations of ALS algorithm, and $\hat{\mathbf{A}}(n_r)$, $\hat{\mathbf{B}}(n_r)$, and $\hat{\mathbf{C}}(n_r)$ are ALS estimates after iteration n_r of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively.

The ALS algorithm has been widely used in such fields as psychometrics, chemometrics, and signal processing for fitting the PARAFAC model. For example, Zhou *et al.* [18] achieved an improved estimation accuracy by the ALS in mmWave MIMO-OFDM system. The reason for the popularity of this algorithm lies in the fact that it is simple to understand and implement. Additionally, Faber compared ALS with six different methods for PARAFAC models, none of which is better than ALS in terms of quality of solution [29]. Note that the ALS algorithm cannot guarantee to converge to a global minimum. Fortunately, the ALS algorithm in the context of the considered massive MIMO system is proved close-to-optimal, which is shown later.

2) Identifiable Estimates: The ALS yields the estimates $\hat{\mathbf{A}}(N_r)$, $\hat{\mathbf{B}}(N_r)$, and $\hat{\mathbf{C}}(N_r)$. Referring to the proof of Proposition 1, the *identifiable* estimates of the three unique matrices denoted by $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$ are given by

$$\hat{\mathbf{A}} = \hat{\mathbf{A}}(N_r)\hat{\mathbf{\Delta}}_1, \quad \hat{\mathbf{B}} = \hat{\mathbf{B}}(N_r)\hat{\mathbf{\Delta}}_2, \quad \hat{\mathbf{C}} = \hat{\mathbf{C}}(N_r)\hat{\mathbf{\Delta}}_3, \quad (14)$$

where $\hat{\mathbf{\Delta}}_1$, $\hat{\mathbf{\Delta}}_2$, and $\hat{\mathbf{\Delta}}_3$ are diagonal matrices to eliminate the scaling uncertainty. Moreover, considering the proof of Proposition 1, these matrices are given by

$$\hat{\mathbf{\Delta}}_1 = \mathbf{G}_1(\hat{\mathbf{a}}_1(N_r)), \quad \hat{\mathbf{\Delta}}_3 = \mathbf{G}_1(\hat{\mathbf{c}}_1(N_r)), \quad \hat{\mathbf{\Delta}}_2 = (\hat{\mathbf{\Delta}}_3\hat{\mathbf{\Delta}}_1)^{\dagger}, \quad (15)$$

where $\hat{\mathbf{a}}_1(N_r)$ and $\hat{\mathbf{c}}_1(N_r)$ are the first rows of matrices $\hat{\mathbf{A}}(N_r)$ and $\hat{\mathbf{C}}(N_r)$, respectively. Then, the N_p columns with most significant variance of $\hat{\mathbf{B}}$ are the estimates of N_p columns of $\mathbf{B}_1$ denoted by $\hat{\mathbf{B}}_1$. The corresponding N_p columns of $\hat{\mathbf{A}}$ and $\hat{\mathbf{C}}$ are the estimates of N_p columns of $\mathbf{A}_1$ and $\mathbf{C}_1$, which are denoted by $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{C}}_1$, respectively. In this way, the pilot contamination can be practically mitigated for large N_a. Furthermore, we determine the DOAs and delays by utilizing the structures of matrices $\mathbf{A}_1$ and $\mathbf{C}_1$ in Section III-A, respectively. The estimates of the fading coefficients are the elements of $\hat{\mathbf{B}}_1$. As a consequence, the identifiable estimates of the three channel parameters are obtained.

3) Convergence of the ALS Algorithm: We shall present that the ALS solution is close-to-optimal for the optimization problem in (12). First, the convergence of the ALS algorithm is analyzed, which is presented as the following proposition:

Proposition 2: With the ALS algorithm, the objective value decreases over the iterations and converges to a fixed point with sufficient number of iterations.

Proof: Please refer to Appendix B. ∎

Note that the objective function of ALS is also the objective function of (12). Thus, based on Proposition 2, the ALS algorithm converge to a solution where the objective value of (12) decreases and converges to a fixed point. Then, to show the performance of this fixed point, the optimal objective value is given as the following proposition:

Proposition 3: The optimal objective value is close to the noise power for massive MIMO systems.

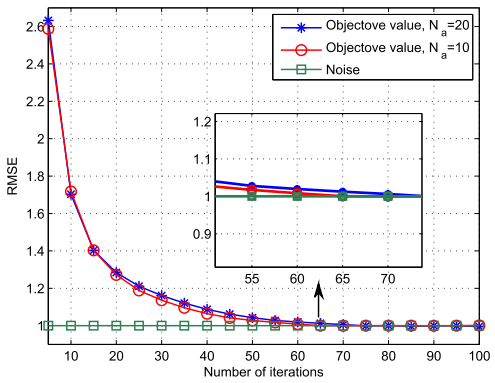


Fig. 2. RMSEs versus the number of iterations of the proposed ALS algorithm, $N_p = 5$, $N_b = 5$, $N_s = 5$, $N_f = 5$, and $\sigma^2 = 1$.

Proof: The objective function of (12) is actually the residual error between the estimates of channel parameters and $\widetilde{\mathbf{H}}$. According to (9), the globally optimal magnitude of the objective function value is [30]

$$\frac{1}{N_f N_a N_b} \parallel \widetilde{\mathbf{H}} - [\mathbf{B} \odot \mathbf{A}]\, \mathbf{C}^{\mathrm{T}} \parallel_{\mathrm{F}}^2 = \frac{1}{N_f N_a N_b} \parallel \mathbf{W}_1 \parallel_{\mathrm{F}}^2 \rightarrow \frac{1}{N_a N_b} \mathrm{tr}(\sigma^2 \mathbf{I}_{N_a N_b}) = \sigma^2, \quad (16)$$

where the second operation $\rightarrow$ holds for a large number of antennas in massive MIMO systems. ∎

Remark 3: Proposition 3 provides a necessary condition to check if the estimates are global optimal or suboptimal. It is that if the fixed point of the objective value in Proposition 2 reaches the noise power, the corresponding solution is globally optimal. Note that the fixed point of the objective value achieved by the ALS algorithm can be heavily dependent on the initial guess [31]. Thus, there is a probability that the converged fixed point is not the noise power, then we have reached a suboptimal solution. In this case, we restart the ALS algorithm with another initial guess. We proceed by this way until the fixed point reaches the noise power, i.e., global minimum is reached. For a wide range of parameters, usually one or two re-initializations are sufficient to yield the global minimum in our simulations. Hence, the success of the ALS algorithm is not hindered by the presence of additional fixed points. Additionally, our simulations in Fig. 2 show that the fixed point of the objective value in Proposition 2 in the context of the considered system is close to noise power. Thus, ALS solution is indeed close-to-optimal.

IV. Computational Complexity Reduction

A. Computational Complexity

Computational complexity is a critical factor in determining the merit of channel parameter estimation schemes especially in massive MIMO systems as the significantly large number of antennas leads much computation. We analyze the computational complexity of the proposed scheme and compare it with the existing schemes such as ESPRIT [14], as ESPRIT is a well-known and efficient DOA estimation approach. Note that the existing approaches in the literature focus on only

$$C_{\text{PARAFAC}} = 8N_a(2N_bN_rN_c^2N_p^2 + 2N_fN_rN_c^2N_p^2 + 3N_bN_fN_rN_cN_p + N_bN_rN_cN_p + N_fN_rN_cN_p + N_bN_fN_s + N_bN_f) + 8N_bN_fN_rN_cN_p + 16N_bN_fN_rN_c^2N_p^2\upsilon + 24N_rN_c^2N_p^2 + 12N_rN_c^3N_p^3 + 18N_rN_cN_p + 5N_cN_pN_f\log_2 N_f + 6N_f. \quad (17)$$

$$C_{\text{PARAFAC}}^{\text{DOA}} = 8N_a(2N_bN_rN_c^2N_p^2 + N_bN_fN_rN_cN_p + N_bN_rN_cN_p + \frac{1}{3}N_bN_f) + 4N_rN_c^3N_p^3 + 8N_rN_c^2N_p^2 + 6N_rN_cN_p + \frac{5}{3}N_cN_pN_f\log_2 N_f + 2N_f, \quad (18)$$

DOA or delay estimation in massive MIMO systems. Hence, for a fair comparison, the computational complexity of the proposed scheme for DOA estimation is also derived. In fact, the computational complexities for estimation of all three parameters have the same format due to the perfect symmetry of PARAFAC models. Here, the computational complexity is expressed in terms of the required number of FLOPs corresponding to the number of complex-valued multiplications and additions, where one complex-valued multiplication and one complex-valued addition require 6 and 2 FLOPs, respectively. We are interested in the complexity required for one user.

First, we determine the computational complexity of the proposed PARAFAC-based scheme. We neglect the required number of FLOPs for taking the pseudo-inverse of matrix $\mathbf{S}$, since it can be precomputed once and then be used subsequently for all estimates. Multiplying the received signal matrix requires $8N_aN_bN_fN_s$ FLOPs. The N_f-point DFT for N_cN_p paths requires $5N_cN_pN_f\log_2 N_f$ FLOPs. Subsequently, taking the inverse of diagonal matrix $\mathbf{G}(\widetilde{\mathbf{d}}_0)$ and multiplying the channel matrix require $6N_f$ and $8N_aN_bN_f$ FLOPs, respectively. Moreover, the ALS is an iterative algorithm, where each iteration includes three steps. The required number of FLOPs for performing the first step in (13a) is $16N_aN_bN_c^2N_p^2 + 8N_aN_bN_fN_cN_p + 8N_aN_bN_cN_p + 4N_c^3N_p^3 + 8N_c^2N_p^2 + 6N_cN_p$. Similarly, $16N_aN_fN_c^2N_p^2 + 8N_aN_bN_fN_cN_p + 8N_aN_fN_cN_p + 4N_c^3N_p^3 + 8N_c^2N_p^2 + 6N_cN_p$ and $16N_bN_fN_c^2N_p^2 + 8N_aN_bN_fN_cN_p + 8N_bN_fN_cN_p + 4N_c^3N_p^3 + 8N_c^2N_p^2 + 6N_cN_p$ FLOPs are needed for performing the next two steps (13b) and (13c), respectively. This leads to the total number of FLOPs for the proposed scheme in (17), shown at the top of this page. Therein, the required number of FLOPs for DOA estimation is given in (18), shown at the top of this page, which increases linearly with N_a. We represent the computational complexity as a function of the number of antennas, N_a, in (17) and (18). It can be observed that the computational complexity of estimating the three channel parameters is low since it increases linearly with N_a. The N_a irrelevant parts can be ignored as N_a is considerably larger compared to the other parameters in massive MIMO systems.

Next, we calculate the computational complexity of ESPRIT for comparison, see [14, Algorithm 2] in detail. For ESPRIT, in the first step, N_t matrix-matrix multiplications are performed to obtain an estimate of the covariance matrix of received signals requiring $8N_a^2N_sN_t$ FLOPs, where N_t is the number of received signal samples. Then, the required number of FLOPs for performing eigenvalue decomposition of the covariance matrix in step 2 is $\frac{16}{3}N_a^3$ FLOPs. After that, in order to calculate the selected signal subspaces, step 3 requires $2\frac{16}{3}N_a(6N_u)^3 + (4+8)N_a(3N_u)^3 + 8N_a(3N_u)^2 + 6\ N_a(3N_u)$ FLOPs for N_u users. Hence, each user requires $2628N_aN_u^2 + 72N_aN_u + 18N_a$ FLOPs. Finally, performing the eigenvalue matching algorithm in step 4 requires $144N_aN_u^2$ FLOPs. Thus, the total number of FLOPs for ESPRIT is

$$C_{\text{ESPRIT}}^{\text{DOA}} = \frac{16}{3}N_a^3 + 8N_a^2N_sN_t + 2772N_aN_u^2 + 72N_aN_u + 18N_a, \quad (19)$$

where $N_u = 1$. Thus, $C_{\text{ESPRIT}}^{\text{DOA}}$ increases cubically with N_a.

Note that the number of antennas, N_a, is much larger relative to the other parameters in massive MIMO systems, thus it is the main factor to determine the computational complexity. Thereby, we have

$$\frac{C_{\text{ESPRIT}}^{\text{DOA}}}{C_{\text{PARAFAC}}^{\text{DOA}}} \to \frac{2}{N_b(6N_rN_c^2N_p^2 + 3N_fN_rN_cN_p + 3N_rN_cN_p + N_f)}. \quad (20)$$

when $N_a^2(N_a \to \infty)$.

Remark 4: When N_a is sufficiently large, the first term in (18) and (19) is dominant over the other terms, respectively. As shown in (20), the ratio of the required number of FLOPs in ESPRIT to the required number of FLOPs for the proposed scheme approaches $\mathcal{O}(N_a^2)$. Therefore, the computational complexity of ESPRIT is considerably larger than that of the proposed scheme in massive MIMO systems.

B. Approaches for Computational Complexity Reduction

Although the computational complexity of the proposed scheme increases only linear with N_a, it is also high as N_a is very large in massive MIMO systems. Thus, utilizing the correlation of channels across antennas and times, we develop two schemes to further reduce the computational complexity.

1) PARAFAC-AS Scheme: High estimation performance can be obtained with a large number of antennas as illustrated in the simulation results, but then one must address the computational complexity that scales with the number of antennas N_a. Note that the numbers of estimated DOAs are much smaller than that of the received signals due to the low-rank property of massive MIMO channels caused by the channel correlation across different BS antennas refer to Section II. Thus, the largely redundant received signals for DOA estimation are existed in (13c), which can be exploited to alleviate the computational complexity and at the same time achieve a high estimation accuracy by a technique known as antenna selection (AS).

Considering low complexity, we adopt a simple random AS in this paper. It is to select the first adjacent N_{as} antennas as a subset for parameter estimation, where N_{as} is much smaller than the number of BS antennas. With PARAFAC-AS scheme, the ALS solution of (12) in iteration n_r is given by

$$\begin{cases} \hat{\mathbf{C}}^{\mathrm{T}}(n_r) = \left[\hat{\mathbf{B}}(n_r-1) \odot \dot{\mathbf{A}}(n_r-1)\right]^{\dagger} \dot{\mathbf{H}} & \text{(21a)} \\ \hat{\mathbf{B}}^{\mathrm{T}}(n_r) = \left[\dot{\mathbf{A}}(n_r-1) \odot \hat{\mathbf{C}}(n_r)\right]^{\dagger} \dot{\mathbf{Y}} & \text{(21b)} \\ \dot{\mathbf{A}}^{\mathrm{T}}(n_r) = \left[\hat{\mathbf{C}}(n_r) \odot \hat{\mathbf{B}}(n_r)\right]^{\dagger} \dot{\mathbf{Z}}, & \text{(21c)} \end{cases}$$

where $\dot{\mathbf{A}}(n_r) \in \mathbb{C}^{N_{as} \times N_c N_p}$ is the ALS estimate after iteration n_r of the first N_{as} rows of $\mathbf{A}$, and $\dot{\mathbf{Z}} \in \mathbb{C}^{N_b N_f \times N_{as}}$ is the first N_{as} columns of $\widetilde{\mathbf{Z}}$. Besides, $\dot{\mathbf{H}} \in \mathbb{C}^{N_{as} N_b \times N_f}$ and $\dot{\mathbf{Y}} \in \mathbb{C}^{N_{as} N_f \times N_b}$ are alternative versions of $\dot{\mathbf{Z}}$, which are interpreted as stacking three-way array along different dimensions.

Based on (18), the computational complexity of the proposed PARAFAC-AS for the DOA estimation is given in (22), shown at the bottom of this page. Thus, with PARAFAC-AS scheme, the computational complexity changes from linear with N_a to N_{as}, which reduces the computational complexity substantially for large N_a but small N_{as}. Moreover, this computational complexity reduction is traded by the estimation performance, which means that the AS may lead performance loss. In fact, our simulation results show that the performance improvement caused by increasing the number of antennas has an upper bound, which brings valuable insights into the performance of AS. Thus, there is a best trade-off with a N_{as} in massive MIMO systems, where we almost achieve the best performance with an alleviative computational complexity shown in the simulation results detailed.

2) PARAFAC-LS Scheme: The channel correlation across different times refer to Section II is also facilitative to complexity reduction. In particular, the fading coefficients of different blocks are independent, while the DOAs and delays remain unchanged for N_b adjacent blocks, which enables that different parameters can be estimated by different frequency.

We propose to reduce the frequency of the DOA and delay estimation, which can be carried out once and then be used subsequently for N_b blocks. Then, the computational complexity can be further reduced in this way. Specifically, two steps are included. First, we estimate the parameter matrices $\mathbf{A}_1$, $\mathbf{B}_1$, and $\mathbf{C}_1$ using the proposed PARAFAC scheme, which has been detailed described in Section III. Then, we use the estimates $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{C}}_1$ to estimate the fading coefficients matrix $\mathbf{B}^n$ in the next N_b-1 blocks by the LS method given by

$$\hat{\mathbf{B}}^n = \underset{\mathbf{B}}{\operatorname{argmin}} \parallel \widetilde{\mathbf{Y}}^n - \left[\hat{\mathbf{A}}_1 \odot \mathbf{S}\hat{\mathbf{C}}_1\right] \mathbf{B}^{\mathrm{T}} \parallel_{\mathrm{F}}^2, \qquad (23)$$

where $\widetilde{\mathbf{Y}}^n$ denotes the received signals in the next $N_b - 1$ blocks and they has the same form with $\widetilde{\mathbf{Y}}$. Then, the multipath fading coefficient matrix could be obtained by

$$\hat{\mathbf{B}}^n = \left[\left(\hat{\mathbf{A}}_1 \odot \mathbf{S}\hat{\mathbf{C}}_1\right)^{\dagger} \widetilde{\mathbf{Y}}^n\right]^{\mathrm{T}}. \qquad (24)$$

It is denoted by the joint PARAFAC and LS (PARAFA-LS) scheme. Compared with the PARAFAC, the PARAFAC-LS makes the parameter estimation more computationally efficient in multiple blocks. For example, the computational complexity of the proposed PARAFAC-LS for the DOA estimation of every block, $\mathcal{O}_{\text{PARAFAC-LS}}$, is given in (25), shown at the bottom of this page, in the case of using the PARAFAC in block 1 and using the LS in the next $N_b - 1$ blocks.

Note that we can use PARAFAC-AS to estimate the parameters in block 1 and use PARAFAC-LS to estimate the parameters in next N_b-1 blocks. The joint PARAFAC-AS and PARAFAC-LS could largely reduce computational complexity.

V. Simulation Results

In this section, we illustrate the performance and the complexity of the proposed schemes through simulations. Here, the signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 10 \log_{10} \frac{g_1^2}{\sigma^2}. \qquad (26)$$

The root mean square error (RMSE) of fading coefficients, DOAs, and delays are defined as $\frac{1}{N_b N_p} \sum_{n_b=1}^{N_b} \sum_{n_p=1}^{N_p} |\hat{b}_{1,n_b,n_p} - b_{1,n_b,n_p}|$, $\frac{1}{N_p} \sum_{n_p=1}^{N_p} |\hat{\theta}_{1,n_p} - \theta_{1,n_p}|$, and $\frac{1}{N_p} \sum_{n_p=1}^{N_p} |\hat{\tau}_{1,n_p} - \tau_{1,n_p}|$, respectively. In the simulations, the pathloss between users in cell n_c and the target BS is defined as $g_{n_c} = \frac{1}{1+l_{n_c}^{\gamma}}$, where $\gamma = 3.8$, and l_{n_c} is the distance between the user in cell n_c and target BS. Moreover, we normalize the cell radius such that $l_1 < 1$ for the target-cell users and $l_{n_c} = 1 + U, n_c \geq 2$ for the adjacent-cell users, where $U \in (0, 1)$ is random. Additionally, both the pilots and the initialization of the ALS algorithm for each run are random, and we assume a symbol period of $T = 1$. The DOA $\theta_{n_c,n_p} \in [-10°, 20°]$[3] and delay $\tau_{n_c,n_p} \in [0, 3T]$ are

[3]To show the DOA estimation accuracy in the case where users are in similar directions, we set a limited range of the DOA values as an example in this section. For interested readers, the simulation can be directly extended to the general case of $\theta_{n_c,n_p} \in [-180°, 180°]$.

$$C_{\text{PARAFAC-AS}}^{\text{DOA}} = 8N_{as}(2N_b N_r N_c^2 N_p^2 + N_b N_f N_r N_c N_p + N_b N_r N_c N_p + \frac{1}{3} N_b N_f) + 4N_r N_c^3 N_p^3 + 8N_r N_c^2 N_p^2 + 6N_r N_c N_p + \frac{5}{3} N_c N_p N_f \log_2 N_f + 2N_f. \qquad (22)$$

$$C_{\text{PARAFAC-LS}}^{\text{DOA}} = 8N_a(2N_r N_c^2 N_p^2 + N_f N_r N_c N_p + N_r N_c N_p) + \frac{8}{3} N_a N_f + \frac{1}{N_b}(4N_r N_c^3 N_p^3 + 8N_r N_c^2 N_p^2 + 6N_r N_c N_p + \frac{5}{3} N_c N_p N_f \log_2 N_f + 2N_f). \qquad (25)$$

random variables, and fading coefficient b_{n_c,n_b,n_p} is complex Gaussian random variable following distribution $\mathcal{CN}(0, g_{n_c}^2)$. The other adopted system parameters are provided in the captions of the figures.

A. Properties of Proposed Schemes

We first provide some discussions to understand the properties of the proposed schemes. Specifically, our aim is to show the convergence of the ALS algorithm and the trade-off between computational complexity and estimation performance in the PARAFAC-AS scheme. We investigate a simple case of not having pilot contamination, i.e., $N_c = 1$.

The RMSE versus the number of iterations is shown in Fig. 2. As can be observed, the objective value decreases monotonically towards a steady-state value for several iterations. This is due to the fact that the ALS converges to a fixed point with sufficient iterations (cf. Proposition 2). On the other hand, the steady-state value is very close to the noise power, which means that the fixed point achieved by the ALS is close to the global optimal objective value (cf. Proposition 3). Thus, the ALS solution is close-to-optimal for (12). Moreover, it also can be observed that the required number of iterations increases as the number of antennas increases. Fortunately, the increasing of the required number of iterations is small as shown in Fig. 2.

To show the trade-off between computational complexity and estimation performance in the PARAFAC-AS scheme, in Fig. 3, we show the RMSEs for the three channel parameters versus N_{as}. As can be observed, all RMSEs decrease monotonically with N_{as}. This is due to the fact that the correlation of channels across different antennas guarantees that the number of estimated parameters remains constant as N_{as} increases, while for larger N_{as} more received signals can be exploited for estimation. Moreover, it also can be observed from Fig. 3 that the estimation performance saturates at a constant value when the number of antennas is large, indicating that the full benefits of massive MIMO systems are exploited. These two features can be used to alleviate the computational complexity and at the same time achieve a high estimation accuracy by AS. It means that there is a best trade-off, where we almost achieve the best performance with an alleviative computational complexity, for example, the $N_{as} = 60$ for the considered set of parameters.

B. Accuracy Performance

The existing approaches focus on only DOA or delay estimation. Therefore, the efficient ESPRIT-based DOA and delay estimation algorithm reported in [14] is considered as a baseline scheme for comparison. Moreover, the average CRB derived in Appendix C of all three channel parameters is then used as a performance measure.

In Figs. 4, we show the RMSEs for the three channel parameters versus the SNR without pilot contamination. As expected, for the PARAFAC-based scheme, all the estimation errors decrease monotonically with increasing SNR. From Fig. 4(a), we also observe that the estimation performance of the PARAFAC-LS for fading coefficient estimation is close to

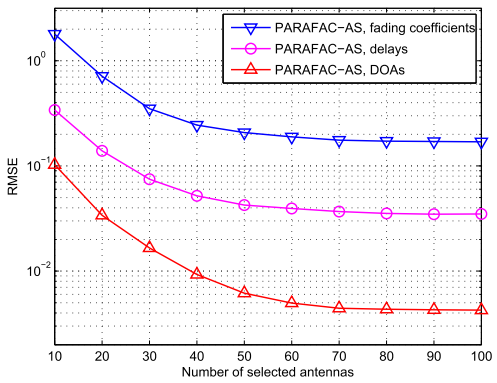


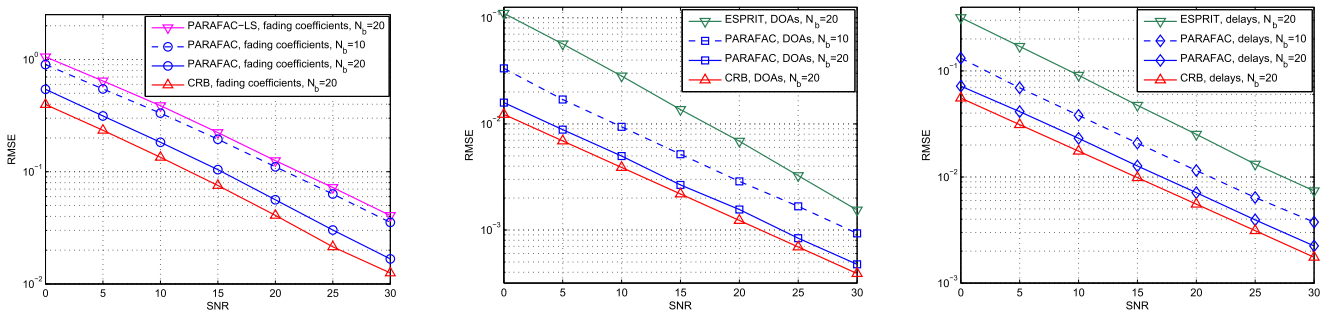
Fig. 3. RMSEs of the proposed PARAFAC-based scheme versus the number of selected antennas, $N_a = 100$, $N_p = 8$, $N_b = 15$, $N_f = 5$, $N_s = 5$, and SNR $= 10$ dB.

the PARAFAC-based scheme due to the fact that the DOAs and delays remain unchanged for several blocks. Compared with the PARAFAC-LS, the performance gain of the PARAFAC-based scheme is provided by joint estimation of all three parameters. Moreover, the estimation errors decrease as the number of blocks increases. This is because the larger the number of blocks, the more samples of the received signal are available for estimation.

Additionally, for DOA and delay estimations in Fig. 4(b) and Fig. 4(c),[4] we observe that the proposed PARAFAC-based scheme outperforms ESPRIT by a margin of more than 5 dB for the considered set of parameters. This is because that the joint estimation of all three parameters and the close-to-optimal ALS algorithm enable the accurate estimates. Moreover, the gap between the estimation errors of the proposed scheme and those of ESPRIT increases with decreasing SNR. These results suggest that the proposed scheme is more efficient in reducing the noise and interference relative to ESPRIT. More importantly, the RMSEs of the PARAFAC-based scheme for the three channel parameters are close to CRBs, which means that the proposed scheme stays close to what is theoretically achievable performance-wise.

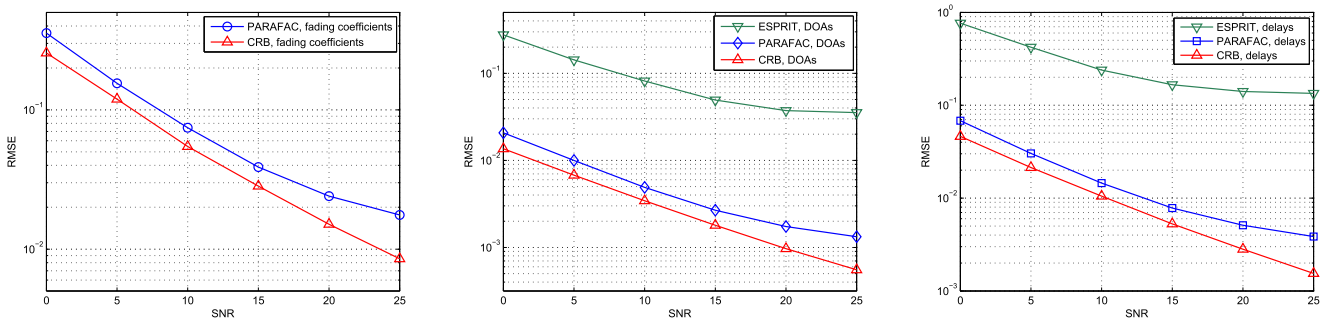
In Figs. 5, we show the RMSE versus the SNR in the presence of pilot contamination. As can be observed, the estimation accuracy of the proposed PARAFAC-based scheme is close to the CRB in low SNR, and the proposed scheme also suffers a performance loss in high SNR due to the pilot contamination. However, the proposed estimates achieve much better performance than the ESPRIT estimate in all the cases illustrating a substantial improvement. This is because the ESPRIT does not take the pilot contamination into consideration. Moreover, increasing SNR generally can improve this reduction due to that the stronger interference from adjacent cells are also addressed in PARAFAC-based scheme.

[4]Note that the estimates of the DOAs and delays of the PARAFAC-LS scheme in the next $N_b - 1$ blocks are the same with that of the PARAFAC scheme in block 1. Thus, for clarity, the simulation results and the corresponding analysis of the DOAs and delays with the PARAFAC-LS scheme are not included in Fig. 4(b) and Fig. 4(c).



(a) Fading coefficients (b) DOAs (c) Delays

Fig. 4. RMSEs of the proposed PARAFAC-based scheme, ESPRIT and CRB versus SNR, $N_a = 30$, $N_p = 5$, $N_f = 5$, and $N_s = 5$.



(a) Fading coefficients (b) DOAs (c) Delays

Fig. 5. RMSEs of the proposed PARAFAC-based scheme, ESPRIT and CRB versus SNR, $N_c = 3$, $N_a = 50$, $N_p = 5$, $N_b = 15$, $N_f = 10$, and $N_s = 8$.

C. Computational Complexity

In Fig. 6, we show the number of Mega FLOPs (MFLOPs) of the proposed three schemes and ESPRIT in [14] for DOA estimation versus N_a. We get that the computational complexity of the PARAFAC-based scheme is lower than that of ESPRIT especially when the number of antennas is larger than 60. This is due to the fact that, as shown in Section III, the computational complexity of ESPRIT increases cubically with N_a, while the computational complexity of the proposed scheme increases only linearly with N_a. Since N_a is considerably larger than the numbers of the other parameters in massive MIMO systems, the proposed PARAFAC-based scheme has a significantly lower computational complexity compared to ESPRIT. As expected, the PARAFAC-AS and PARAFAC-LS schemes have much lower computational complexity than the PARAFAC-based scheme. For example, the complexity of PARAFAC-AS scheme decreases by 50% compared to the PARAFAC scheme, and more than 50% of the complexity of PARAFAC scheme can be saved by the PARAFAC-LS scheme when $N_a = 70$ and $N_{as} = 30$. More importantly, the complexity of the RAPAFAC-LS scheme is much lower than that of the ESPRIT for all the considered set of parameters.

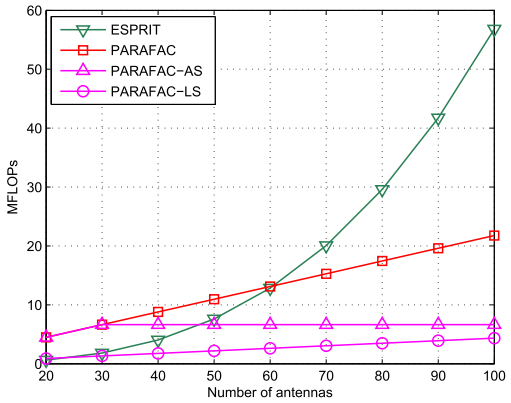


Fig. 6. MFLOPs of the proposed PARAFAC-based scheme and ESPRIT versus the number of antennas, $N_c = 3$, $N_u = 1$, $N_t = 20$, $N_b = 5$, $N_s = 10$, $N_p = 5$, and $N_f = 5$.

VI. Conclusions

In this paper, we developed a PARAFAC-based scheme for joint estimation of the fading coefficients, DOAs, and delays of a massive MIMO channel. Specifically, we derived three equivalent representations of the channel based on PARAFAC model. We further proved the *identifiability* of three relevant channel parameters in the PARAFAC model. Thereby, we formulated the joint estimation of these parameters as an optimization problem with three coupled variables, and achieved its optimal and unique solution through an ALS algorithm. Moreover, we analyzed the computational complexity of the proposed scheme in terms of the required number of FLOPs, and subsequently two schemes were developed

to further reduce the computational complexity. Simulation results revealed that our schemes shows a significant improvement in estimation performance compared to the conventional ESPRIT and stays close to what is theoretically achievable performance-wise. In addition, the computational complexity of the three proposed schemes were shown to be significantly smaller than that of the ESPRIT, especially when the number of antennas is large.

APPENDIX

A. Proof of Proposition 1

To prove this proposition, we first give the definition of Kruskal rank.

Definition 1 (Kruskal Rank): If a matrix $\mathbf{A}$ contains every set of $k_{\mathbf{A}}$ linearly independent columns but does not contain every set of $k_{\mathbf{A}}+1$ linearly independent columns, then the Kruskal rank (or k-rank) of matrix $\mathbf{A}$ is $k_{\mathbf{A}}$.

According to Definition 1, if a matrix is full-rank, it is also full k-rank. In the considered system, matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are full-rank as both matrices $\mathbf{A}$ and $\mathbf{C}$ in (9) are Vandermonde matrices, and the elements of $\mathbf{B}$ are random variables. Thus, these three matrices are also all full k-rank. As a result, the k-ranks of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $k_{\mathbf{A}}=\min(N_a,N_cN_p)$, $k_{\mathbf{B}}=\min(N_b,N_cN_p)$, and $k_{\mathbf{C}}=\min(N_f,N_cN_p)$, respectively.

In a PARAFAC model, $\mathbf{U}_{n_b}=\mathbf{A}\mathbf{G}_{n_b}(\mathbf{B})\,\mathbf{C}^{\mathrm{T}}, n_b=1,2,\cdots,N_b$, if

$$k_{\mathbf{A}}+k_{\mathbf{B}}+k_{\mathbf{C}}\geq 2(N_cN_p+1), \tag{27}$$

is satisfied, matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are unique up to permutation and scaling of columns [32]. Note that we focus on the physical finite scattering channel model. On the one hand, in a poor scattering propagation environment (i.e., non-line-of-sight and low angle spread), the number of paths is largely limited, since some of the radiation branches are too weak to support a diversity branch [33]. On the other hand, when the number of paths is large, the scatterers appear in groups (called clusters) with similar delays and AoAs, limiting the effective number of active paths [34]. As a result, N_cN_p is usually small relative to N_a and N_b.[5] Besides, N_cN_p and N_f are usually larger than or equal to 2. Thus, we have

$$\begin{aligned}&k_{\mathbf{A}}+k_{\mathbf{B}}+k_{\mathbf{C}}\\&\quad=\min(N_a,N_cN_p)+\min(N_b,N_cN_p)+\min(N_f,N_cN_p)\\&\quad=2N_cN_p+\min(N_f,N_cN_p)\geq 2N_cN_p+2.\end{aligned} \tag{28}$$

Hence, in the considered massive MIMO systems, matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are unique up to permutation and scaling of columns. Based on this, any matrices $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$, and $\bar{\mathbf{C}}$ satisfying the PARAFAC model in (9) can be expressed as

$$\bar{\mathbf{A}}=\mathbf{A}\mathbf{\Pi}\bar{\mathbf{\Delta}}_1,\quad \bar{\mathbf{B}}=\mathbf{B}\mathbf{\Pi}\bar{\mathbf{\Delta}}_2,\quad \bar{\mathbf{C}}=\mathbf{C}\mathbf{\Pi}\bar{\mathbf{\Delta}}_3. \tag{29}$$

Therein, $\mathbf{\Pi}\in\mathbb{C}^{N_cN_p\times N_cN_p}$ is a permutation matrix denoting the permutation uncertainty, and $\bar{\mathbf{\Delta}}_1$, $\bar{\mathbf{\Delta}}_2$, and $\bar{\mathbf{\Delta}}_3$ are diagonal matrices denoting the scaling uncertainty, which satisfies the condition that

$$\bar{\mathbf{\Delta}}_1\bar{\mathbf{\Delta}}_2\bar{\mathbf{\Delta}}_3=\mathbf{I}. \tag{30}$$

In the considered system, the scaling uncertainty can be eliminated thanks to the Vandermonde structure of $\mathbf{A}$ and $\mathbf{C}$. In other words, the first rows of both $\mathbf{A}$ and $\mathbf{C}$ have unity entries. Then, the diagonal matrices, $\bar{\mathbf{\Delta}}_1$, $\bar{\mathbf{\Delta}}_2$, and $\bar{\mathbf{\Delta}}_3$ are given by

$$\bar{\mathbf{\Delta}}_1=\mathbf{G}_1(\bar{\mathbf{a}}_1),\quad \bar{\mathbf{\Delta}}_3=\mathbf{G}_1(\bar{\mathbf{c}}_2),\quad \bar{\mathbf{\Delta}}_2=(\bar{\mathbf{\Delta}}_3\bar{\mathbf{\Delta}}_1)^{\dagger}, \tag{31}$$

where $\bar{\mathbf{a}}_1$ and $\bar{\mathbf{c}}_1$ are the first rows of matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{C}}$, respectively. As a consequence, the channel matrix decomposition in (9) is unique up to a permutation of the columns of matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. This completes the proof.

B. Proof of Proposition 2

We first prove that the objective value sequence, $[\chi(1),\chi(2),\cdots,\chi(N_r)]$, is decreasing. Therein, $\chi(n_r)$ is the objective function of the ALS after iteration n_r given by

$$\chi(n_r)=\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r)\right]\hat{\mathbf{C}}^{\mathrm{T}}(n_r)\,\|_{\mathrm{F}}^2\,. \tag{32}$$

Note that $\chi(n_r)$ is also the objective function of (12), and it is non-negative and continuous. In (13a), we obtain the $\hat{\mathbf{C}}(n_r)$ in an LS sense, thereby we have

$$\begin{aligned}&\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r-1)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r-1)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r-1)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\,.\end{aligned} \tag{33}$$

Thus, after the first step of the iteration n_r in (13a), the objective value is decreasing. Similarly, after the other two steps of iteration n_r in (13b) and (13c), we have

$$\begin{aligned}&\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\quad\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r-1)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2,\end{aligned} \tag{34}$$

$$\begin{aligned}&\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\quad\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\,.\end{aligned} \tag{35}$$

Thus, the relationship between $\chi(n_r-1)$ and $\chi(n_r)$ is

$$\begin{aligned}\chi(n_r)&=\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r-1)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&\leq\|\,\widetilde{\mathbf{H}}-\left[\hat{\mathbf{B}}(n_r-1)\odot\hat{\mathbf{A}}(n_r-1)\right]\hat{\mathbf{C}}(n_r-1)^{\mathrm{T}}\,\|_{\mathrm{F}}^2\\&=\chi(n_r-1).\end{aligned} \tag{36}$$

Moreover, based on the ALS, the iteration in (36) will stop once the next iteration cannot reduce the objective value compared to the current iteration, i.e., $\chi(N_r)=\chi(N_r+1)$ in the proposed scheme. Hence, we get that the objective value sequence, $[\chi(1),\chi(2),\cdots,\chi(N_r)]$, is decreasing.

Then, we give the definition of a fixed point.

[5] In the case of a small N_b, the uniqueness can be guaranteed by increasing the number of the channel taps so that $N_f=N_cN_p$ and the condition changes to $N_b\geq 2$.

Definition 2 (Fixed Point): Let $\gamma : V \to V$ be a mapping from a point in space V to a point also in space V. There is an iterative algorithm, $v_k = \gamma(v_{k-1})$, and point v_0 in V is given. A point v^* is a fixed point of γ if $v^* = \gamma(v^*)$.

We consider the ALS algorithm as a mapping, and its objective value sequence, $[\chi(1), \chi(2), \cdots, \chi(N_r)]$, is decreasing and non-negative. From the monotone convergence theorem [35], every bounded monotone sequence is convergent. Therefore, we can draw the conclusion that the objective value of the ALS algorithm converges to a fixed point $\chi(N_r)$.

C. CRB for Multipath Channel Model in Massive MIMO Systems

Consider the three equivalent forms of channel estimate shown in (9), (10) and (11). Therein, it assumes that the noise $\mathbf{W}_1$ follows complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$. Thus, the probability density function (PDF) of the noise is

$$f(w(n_a n_b, n_f), \sigma^2) = \frac{1}{\sigma\sqrt{\pi}} \exp\left\{-\frac{1}{\sigma^2} w(n_a n_b, n_f)^2\right\}, \quad (37)$$

where $w_1(n_a n_b, n_f)$ is the element of $\mathbf{W}_1$. Then, we write the likelihood function of channel estimate in three equivalent ways

$$\begin{aligned} &L(\mathbf{W}_1) \\ &= \frac{1}{(\pi\sigma^2)^{N_a N_b N_f}} \exp\left\{\frac{1}{\sigma^2} \sum_{n_f=1}^{N_f} ||\widetilde{\mathbf{h}}_{n_f} - (\mathbf{B} \odot \mathbf{A})\mathbf{c}_{n_f}^T||_2^2\right\} \\ &= \frac{1}{(\pi\sigma^2)^{N_a N_b N_f}} \exp\left\{\frac{1}{\sigma^2} \sum_{n_b=1}^{N_b} ||\widetilde{\mathbf{y}}_{n_b} - (\mathbf{A} \odot \mathbf{C})\mathbf{b}_{n_b}^T||_2^2\right\} \\ &= \frac{1}{(\pi\sigma^2)^{N_a N_b N_f}} \exp\left\{\frac{1}{\sigma^2} \sum_{n_a=1}^{N_a} ||\widetilde{\mathbf{z}}_{n_a} - (\mathbf{C} \odot \mathbf{B})\mathbf{a}_{n_a}^T||_2^2\right\}, \quad (38) \end{aligned}$$

where $\widetilde{\mathbf{h}}_{n_f}$, $\widetilde{\mathbf{y}}_{n_b}$, and $\widetilde{\mathbf{z}}_{n_a}$ represent the row n_f of $\widetilde{\mathbf{H}}$, row n_b of $\widetilde{\mathbf{Y}}$, row n_a of $\widetilde{\mathbf{Z}}$, respectively. Additionally, $\mathbf{a}_{n_a}$, $\mathbf{b}_{n_b}$, and $\mathbf{c}_{n_f}$ represent the row n_a of $\mathbf{A}$, row n_b of $\mathbf{B}$, row n_f of $\mathbf{C}$, respectively. Thereby, the log-likelihood functions of channel estimate are given by

$$\begin{aligned} &\ln L(\mathbf{W}_1) \\ &= -N_a N_b N_f \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{n_f=1}^{N_f} ||\widetilde{\mathbf{h}}_{n_f} - (\mathbf{B} \odot \mathbf{A})\mathbf{c}_{n_f}^T||_2^2 \\ &- N_a N_b N_f \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{n_b=1}^{N_b} ||\widetilde{\mathbf{y}}_{n_b} - (\mathbf{A} \odot \mathbf{C})\mathbf{b}_{n_b}^T||_2^2 \\ &- N_a N_b N_f \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{n_a=1}^{N_a} ||\widetilde{\mathbf{z}}_{n_a} - (\mathbf{C} \odot \mathbf{B})\mathbf{a}_{n_a}^T||_2^2. \quad (39) \end{aligned}$$

Note that $\mathbf{A}$ is Vandermonde matrix. Thus, the number of unknown complex parameters of $\mathbf{A}$ is $N_c N_p$ instead of $N_a N_c N_p$. Similarly, the number of unknown complex parameters of Vandermonde matrix $\mathbf{C}$ is also $N_c N_p$. While the number of unknown complex parameters of $\mathbf{B}$ is $N_f N_c N_p$. As a result, the total number of unknown complex parameters is $2N_c N_p + N_f N_c N_p$. We define a vector $[\mathbf{a}_2, \mathbf{b}_1, \cdots, \mathbf{b}_{N_b}]$, where the complex parameters to be estimated are included. To simplify the CRB derivation, it is useful to introduce the equivalent complex parameter vector $\boldsymbol{\varepsilon}$ is

$$\boldsymbol{\varepsilon} = [\mathbf{a}_2, \mathbf{b}_1, \cdots, \mathbf{b}_{N_b}, \mathbf{c}_2^*, \mathbf{a}_2^*, \cdots, \mathbf{c}_2^*]. \quad (40)$$

Hence, we have $\ln L(\mathbf{W}_1) = \ln L(\boldsymbol{\varepsilon})$. Thereby, the complex fisher information matrix (FIM) $\Omega(\boldsymbol{\varepsilon})$ is given by

$$\Omega(\boldsymbol{\varepsilon}) = \mathrm{E}\left\{\left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial\boldsymbol{\varepsilon}}\right)^{\mathrm{H}} \left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial\boldsymbol{\varepsilon}}\right)\right\}. \quad (41)$$

Taking partial derivative of $\ln L(\boldsymbol{\varepsilon})$ with respect to the unknown parameters,

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\varepsilon})}{\partial a_{2,p}} &= \sum_{n_a=1}^{N_a} \frac{n_a - 1}{\sigma^2} a_{2,p}^{n_a - 2} \left(\widetilde{\mathbf{z}}_{n_a} - (\mathbf{C} \odot \mathbf{B})\mathbf{a}_{n_a}^T\right)^{\mathrm{H}} \\ &\quad \times (\mathbf{C} \odot \mathbf{B})\mathbf{e}_p, \\ \frac{\partial \ln L(\boldsymbol{\varepsilon})}{\partial b_{n_b,p}} &= \frac{1}{\sigma^2} \left(\widetilde{\mathbf{y}}_{n_b} - (\mathbf{A} \odot \mathbf{C})\mathbf{b}_{n_b}^T\right)^{\mathrm{H}} (\mathbf{A} \odot \mathbf{C})\mathbf{e}_p, \\ \frac{\partial \ln L(\boldsymbol{\varepsilon})}{c_{2,p}} &= \sum_{n_f=1}^{N_f} \frac{n_f - 1}{\sigma^2} c_{2,p}^{n_f - 2} \left(\widetilde{\mathbf{h}}_{n_f} - (\mathbf{B} \odot \mathbf{A})\mathbf{c}_{n_f}^T\right)^{\mathrm{H}} \\ &\quad \times (\mathbf{B} \odot \mathbf{A})\mathbf{e}_p, \\ \frac{\partial \ln L(\boldsymbol{\varepsilon})}{\partial a_{2,p}^*} &= \left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial a_{2,p}}\right)^*, \\ \frac{\partial \ln L(\boldsymbol{\varepsilon})}{\partial \widetilde{d}_{2,p}^*} &= \left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial \widetilde{d}_{2,p}}\right)^*, \\ \frac{\partial \ln L(\boldsymbol{\varepsilon})}{\partial b_{n_b,p}^*} &= \left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial b_{n_b,p}}\right)^*, \quad (42) \end{aligned}$$

where $p = (n_c - 1)N_p + n_p$, $\mathbf{e}_p$ is the p^{th} unit coordinate vector, and $\boldsymbol{\Phi}_1 = \mathbf{B} \odot \mathbf{A}$, $\boldsymbol{\Phi}_2 = \mathbf{A} \odot \mathbf{C}$ and $\boldsymbol{\Phi}_3 = \mathbf{C} \odot \mathbf{B}$. Then, we can obtain second derivatives by (43), shown at the top of this page, where matrix $\mathrm{E}(n_a, n_b; N_f)$ is given by

$$\mathrm{E}(n_a, n_b; N_f) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.$$

$\mathrm{E}(n_a, n_f; N_b)$ and $\mathrm{E}(n_f, n_b; N_a)$ can be obtained similarly.

After that, according to (42), the FIM $\Omega(\boldsymbol{\varepsilon})$ in (41) could be rewritten as

$$\Omega(\boldsymbol{\varepsilon}) = \mathrm{E}\left\{\left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial\boldsymbol{\varepsilon}}\right)^{\mathrm{H}} \left(\frac{\partial\eta(\boldsymbol{\varepsilon})}{\partial\boldsymbol{\varepsilon}}\right)\right\} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^* \end{bmatrix}, \quad (44)$$

where $\boldsymbol{\Psi}$ is of size $(N_b + 2)N_c N_p \times (N_b + 2)N_c N_p$, and

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ab} & \boldsymbol{\Psi}_{ac} \\ \boldsymbol{\Psi}_{ab^{\mathrm{T}}} & \boldsymbol{\Psi}_{bb} & \boldsymbol{\Psi}_{bc} \\ \boldsymbol{\Psi}_{ac^{\mathrm{T}}} & \boldsymbol{\Psi}_{bc^{\mathrm{T}}} & \boldsymbol{\Psi}_{cc} \end{bmatrix}, \quad (45)$$

$$
\begin{aligned}
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial a_{2,p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial a_{2,p_2}}\right\} &= \sum_{n_{a_1}=1}^{N_a}\sum_{n_{a_2}=1}^{N_a}\frac{(n_{a_1}-1)(n_{a_2}-1)}{\sigma^2}\left(a_{2,p_1}^{n_{a_1}-2}\right)^* a_{2,p_2}^{n_{a_2}-2}\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{C}\odot\mathbf{B})^{\mathrm{H}}(\mathbf{C}\odot\mathbf{B})\mathbf{e}_{p_2}\delta(n_{a_1},n_{a_2}),\\
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial b_{n_{b_1},p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial b_{n_{b_2},p_2}}\right\} &= \frac{1}{\sigma^2}\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{A}\odot\mathbf{C})^{\mathrm{H}}(\mathbf{A}\odot\mathbf{C})\mathbf{e}_{p_2}\delta(n_{b_1},n_{b_2}),\\
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial c_{2,p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial c_{2,p_2}}\right\} &= \sum_{n_{f_1}=1}^{N_f}\sum_{n_{f_2}=1}^{N_f}\frac{(n_{f_1}-1)(n_{f_2}-1)}{\sigma^2}\left(c_{2,p_1}^{n_{f_1}-2}\right)^* c_{2,p_2}^{n_{f_2}-2}\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{A})^{\mathrm{H}}(\mathbf{B}\odot\mathbf{A})\mathbf{e}_{p_2}\delta(n_{f_1},n_{f_2}),\\
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial a_{2,p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial b_{n_b,p_2}}\right\} &= \sum_{n_a=1}^{N_a}\frac{n_a-1}{\sigma^2}a_{2,p_1}^{m-2}\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{C}\odot\mathbf{B})^{\mathrm{H}}\mathrm{E}(n_a,n_b;N_f)(\mathbf{A}\odot\mathbf{C})\mathbf{e}_{p_2},\\
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial a_{2,p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial c_{2,p_2}}\right\} &= \sum_{n_a=1}^{N_a}\sum_{n_f=1}^{N_f}\frac{(n_a-1)(n_f-1)}{\sigma^2}a_{2,p_1}^{n_a-2}c_{2,p_2}^{n_f-2}\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{C}\odot\mathbf{B})^{\mathrm{H}}\mathrm{E}(n_a,n_f;N_b)(\mathbf{B}\odot\mathbf{A})\mathbf{e}_{p_2},\\
\mathrm{E}\left\{\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial c_{2,p_1}^*}\frac{\partial \mathrm{In}L(\boldsymbol{\varepsilon})}{\partial b_{n_b,p_2}}\right\} &= \sum_{n_f=1}^{N_f}\frac{n_f-1}{\sigma^2}\left(c_{2,p_1}^{n_f-2}\right)^*\mathbf{e}_{p_1}^{\mathrm{H}}(\mathbf{B}\odot\mathbf{A})^{\mathrm{H}}\mathrm{E}(n_f,n_b;N_a)(\mathbf{A}\odot\mathbf{C})\mathbf{e}_{p_2},
\end{aligned}
\tag{43}
$$

where $\boldsymbol{\Psi}_{aa}$ represents the second derivative of $\mathbf{a}$ and similarly for others. Finally, we could obtain the FIM $\Omega(\boldsymbol{\varepsilon})$ by (45).

The CRB is given by the inverse of the $\boldsymbol{\Psi}$, which is

$$
\begin{bmatrix}
\mathbf{CRB}_{aa} & \mathbf{CRB}_{ab} & \mathbf{CRB}_{ac}\\
\mathbf{CRB}_{ab} & \mathbf{CRB}_{bb} & \mathbf{CRB}_{bc}\\
\mathbf{CRB}_{ac} & \mathbf{CRB}_{bc} & \mathbf{CRB}_{cc}
\end{bmatrix} = \boldsymbol{\Psi}^{\dagger}. \tag{46}
$$

Then, the CRB of fading coefficients, DOAs, and delays are the mean of $\mathbf{CRB}_{aa}$, $\mathbf{CRB}_{bb}$, and $\mathbf{CRB}_{cc}$, respectively. Thereby, the CRB of the three channel parameters in the target-cell can be obtained from $\mathbf{CRB}_{aa}$, $\mathbf{CRB}_{bb}$, and $\mathbf{CRB}_{cc}$.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Robert Schober, from Friedrich-Alexander-University Erlangen-Nürnberg, for his generous help in improving this paper.

REFERENCES

[1] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[2] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, Feb. 2016.

[3] S. Jin, X. Liang, K.-K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1444, Apr. 2013.

[5] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[6] 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.

[7] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1826–1829, Oct. 2015.

[8] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.

[9] Z. Chen and C. Yang, "Pilot decontamination in massive MIMO systems: Exploiting channel sparsity with pilot assignment," in *Proc. IEEE GlobalSIP*, Atlanta, GA, USA, Dec. 2014, pp. 637–641.

[10] U. Ugurlu, R. Wichman, C. B. Ribeiro, and C. Wijting, "A multipath extraction-based CSI acquisition method for FDD cellular networks with massive antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2940–2953, Apr. 2016.

[11] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.

[12] K. T. Wong and M. D. Zoltowski, "Self-initiating MUSIC-based direction finding and polarization estimation in spatio-polarizational beamspace," *IEEE Trans. Antennas Propag.*, vol. 48, no. 8, pp. 1235–1245, Aug. 2000.

[13] K.-Y. Yang, J.-Y. Wu, and W.-H. Li, "A low-complexity direction-of-arrival estimation algorithm for full-dimension massive MIMO systems," in *Proc. IEEE ICCS*, Shenzhen, China, Nov. 2014, pp. 472–476.

[14] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, "An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[15] Y. Zhu, L. Liu, and J. Zhang, "Joint angle and delay estimation for 2D active broadband MIMO-OFDM systems," in *Proc. IEEE Globecom*, Atlanta, GA, USA, Dec. 2013, pp. 3300–3305.

[16] S. Chakraborty and D. Sen, "Joint estimation of time, frequency offsets, and channel gains with ICIs in EF multi-relay DMIMO-OFDM system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5822–5838, Jul. 2017.

[17] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[18] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[19] 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.

[20] X. Wei, W. Peng, D. W. K. Ng, R. Schober, and T. Jiang, "Joint estimation of channel parameters in massive MIMO systems via PARAFAC analysis," in *Proc. IEEE ICNC*, Maui, HI, USA, Mar. 2018, pp. 496–502.

[21] X. Wei, W. Peng, D. Chen, R. Schober, and T. Jiang, "Uplink channel estimation in massive MIMO systems using factor analysis," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1620–1623, Aug. 2018.

[22] T. Ue, S. Sampei, N. Morinaga, and K. Hamaguchi, "Symbol rate and modulation level-controlled adaptive modulation/TDMA/TDD system for high-bit-rate wireless data transmission," *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1134–1147, Nov. 1998.

[23] A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
[24] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE WCNC*, Shanghai, China, Apr. 2013, pp. 2890–2895.
[25] G. C. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.
[26] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
[27] R. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," *UCLA Work. Papers Phonetics*, vol. 16, pp. 1–84, Dec. 1970.
[28] D. P. Bertsekas, *Nonlinear Programming*, Belmont, CA, USA: Athena Scientific, 1999.
[29] N. K. M. Faber, R. Bro, and P. K. Hopke, "Recent developments in CANDECOMP/PARAFAC algorithms: A critical review," *Chemometrics Intell. Lab. Syst.*, vol. 65, no. 1, pp. 119–137, Jan. 2003.
[30] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. I. Algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1184–1197, May 1996.
[31] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[32] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics Intell. Lab. Syst.*, vol. 38, no. 2, pp. 149–171, Oct. 1997.
[33] C. B. Chae, A. Forenza, R. W. Heath, M. R. McKay, and I. B. Collings, "Adaptive MIMO transmission techniques for broadband wireless communication systems," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 112–118, May 2010.
[34] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[35] J. Bibby, "Axiomatisations of the average and a further generalisation of monotonic sequences," *Glasgow Math. J.*, vol. 15, no. 11, pp. 63–65, 1974.

Authors' photographs and biographies not available at the time of publication.