

Compressed CPD-Based Channel Estimation and Joint Beamforming for RIS-Assisted Millimeter Wave Communications

Xi Zheng, Jun Fang [✉], *Senior Member, IEEE*, Hongwei Wang [✉], *Member, IEEE*, Peilan Wang [✉],
and Hongbin Li [✉], *Fellow, IEEE*

Abstract—We consider the problem of channel estimation and joint active and passive beamforming for reconfigurable intelligent surface (RIS) assisted millimeter wave (mmWave) multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems. We show that, with a well-designed frame-based training protocol, the received pilot signal can be organized into a low-rank third-order tensor that admits a canonical polyadic decomposition (CPD). Based on this observation, we propose a CPD-based method for estimating the cascade channels associated with different subcarriers. The proposed method exploits the intrinsic low-rankness of the CPD formulation, which is a result of the sparse scattering characteristics of mmWave channels, and thus has the potential to achieve a significant training overhead reduction. Specifically, our analysis shows that the proposed method has a sample complexity that scales quadratically with the sparsity of the cascade channel. Also, by utilizing the singular value decomposition-like structure of the effective channel, this paper develops a joint active and passive beamforming method based on the estimated cascade channels. Simulation results show that the proposed CPD-based channel estimation method attains mean square errors that are close to the Cramér-Rao bound (CRB) and present a clear advantage over the compressed sensing-based methods. In addition, the proposed joint beamforming method can effectively utilize the estimated channel parameters to achieve superior beamforming performance.

Index Terms—Channel estimation, joint active and passive beamforming, millimeter wave communications, reconfigurable intelligent surface.

Manuscript received 6 March 2024; accepted 3 June 2024. Date of publication 7 June 2024; date of current version 17 October 2024. The work of Jun Fang was supported by Sichuan Science and Technology Program under Grant 2023ZYD0146. The work of Hongwei Wang was supported by the Fundamental Research Funds for the Central University under Grant ZYGX2022J029. The work of Hongbin Li was supported by National Science Foundation under Grant CCF-2316865, Grant ECCS-2212940, and Grant ECCS-2332534. The review of this article was coordinated by Prof. Yue Gao. (*Corresponding author: Jun Fang.*)

Xi Zheng and Jun Fang are with the Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China, and also with the National Key Laboratory of Wireless Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: junfang@uestc.edu.cn).

Hongwei Wang and Peilan Wang are with the National Key Laboratory of Wireless Communications, University of Electronic Science and Technology of China, Chengdu 611731, China.

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

Digital Object Identifier 10.1109/TVT.2024.3411069

I. INTRODUCTION

MILLIMETER wave (mmWave) and terahertz (THz) communications are able to support extremely high data rate transmissions [1], [2], [3], [4]. Nevertheless, due to the reduced diffraction effect and high penetration loss, mmWave/THz systems require more antennas and active nodes such as access points (APs) and relays to improve the signal coverage [5], [6]. Deploying active nodes incurs an additional energy consumption and meanwhile presents a serious network interference issue. It is thus of practical significance to develop innovative technologies to address the coverage issue of future mmWave/THz wireless networks with a low cost and complexity.

Recently, reconfigurable intelligent surface (RIS) has emerged as an energy-efficient and cost-effective solution to tackle the above challenges. Generally speaking, RIS intelligently adjusts the signal reflection through a large number of low-cost passive reflection elements, which can dynamically reshape the wireless propagation environment and thereby improve the system performance [7], [8], [9]. An important advantage of RIS is that it does not require any active circuits such as radio frequency (RF) chains for signal transmission/reception, which reduces hardware complexity as well as energy consumptions compared to traditional active transceivers/relays. Furthermore, RIS can be easily attached to different objects (such as walls and ceilings), thus showing great flexibility and compatibility in practical deployment [10].

Channel state information (CSI) acquisition is a prerequisite to realize the full potential of RIS-assisted mmWave systems. Nevertheless, since RIS is usually composed of a large number of passive elements, CSI acquisition for RIS-assisted mmWave systems faces the difficulty of requiring a large amount of training overhead. In order to reduce the training overhead, the inherent sparse structure of the mmWave channel is exploited and the cascade channel estimation is cast into a compressed sensing framework [11], [12], [13], [14]. Specifically, the work [11] assumed that the base station (BS)-RIS and RIS-user channels are LOS-dominated. The work [12] considered a more general scenario where there are multiple paths between the BS (RIS) and the RIS (user). Nevertheless, due to the multidimensional structure of the cascade channel, the sensing matrix is excessively large even with low grid resolutions, especially in MIMO-OFDM systems. To solve this problem, [13] proposed

a two-stage RIS-aided channel estimation (TRICE) framework to obtain the BS/RIS/user angle information separately in each stage, thus decoupling the multidimensionality of the cascade channel. The work [14] combined machine learning into compressed sensing methods to improve the estimation accuracy in wideband systems. Yet, it needs extra training during offline training phases.

Meanwhile, most of these compressed sensing techniques require to discretize the continuous parameter space into a finite set of grid points, leading to grid discretization errors. Recently, tensor decomposition-based methods have been developed by exploiting the intrinsic multi-dimensional structure of the received signal [15], [16], [17], [18]. The tensor-based approach is gridless and therefore is free from grid discretization errors. This helps achieve a performance improvement over compressed sensing-based methods. However, most of the work [15], [16], [17] did not utilize the sparse scattering characteristics of mmWave channels and had a CP rank that is equal to the number of reflecting elements. As a result, these methods required a training overhead proportional to the number of reflecting elements, which is usually large in practice. In [18], a hybrid-RIS with active sensors was employed to separately estimate the BS-RIS and RIS-user channels. The received signals by these active sensors were formulated into an incomplete fourth-order tensor with missing entries, based on which the BS-RIS or the RIS-user channel can be estimated via a tensor completion approach.

In addition to channel estimation, joint beamforming is another important topic for RIS-assisted systems that has been extensively studied. Joint beamforming refers to jointly design the reflection coefficients at the RIS and the active precoding/combining matrix at the BS/user to optimize the system spectral efficiency. In recent years, there have been many works on joint beamforming [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Among them, a lot of works focused on single-input-single-output (SISO) or multiple-input-single-output (MISO) systems, e.g. [19], [20], [21], [22], [23], which cannot be straightforwardly extended to MIMO systems. For RIS-assisted MIMO systems, [24], [25], [26], [27] mainly utilised the idea of alternating optimization for the joint beamforming design. The renowned alternating optimization (AO) algorithm, originally introduced in [27], was designed to enhance the capacity of IRS-enhanced point-to-point MIMO systems. Despite its performance superiority, the AO method which alternatively optimizes variables suffers a prohibitively high computational complexity, rendering it less practical for practical systems. The work [28] optimized the spectral efficiency through maximizing the Frobenius-norm of the effective channel. However, this typically yields a large condition number, which limits its performance improvement. In [29], a joint beamforming method was proposed for RIS-assisted MIMO-OFDM systems. Its idea is to align the passive beamforming vector with the most prominent path of the reflected channel. Nevertheless, as this method exclusively relies on the dominant path for data transmission, it fails to exploit the spatial diversity inherent in the cascade channel.

In this paper, we study the problem of channel estimation and joint beamforming for RIS-assisted mmWave MIMO-OFDM systems. We show that by exploring the sparse scattering characteristics and the intrinsic multi-dimensional structure of the mmWave cascade channel, the received signal can be formulated into a low-rank third-order tensor that admits a canonical polyadic decomposition (CPD). Based on this formulation, an alternating least squares (ALS) method is developed for channel estimation. Theoretical analysis shows that the proposed methods have a sample complexity of $\mathcal{O}(U^2)$. Here U denotes the sparsity of the cascade channel. Since U is usually small relative to the dimension of the cascade channel, the proposed method can achieve a substantial training overhead reduction.

In addition to channel estimation, we also consider the problem of joint active and passive beamforming design. Our proposed joint beamforming design method directly optimizes the reflection coefficients by obtaining a relationship between the reflection coefficients and the singular values of the effective channel. In doing this way, a favorable propagation environment can be realized by directly manipulating the reflection coefficients of the RIS, without resorting to the complex alternating optimization procedure. Simulation results show that our proposed method presents a clear performance advantage over existing state-of-the-art joint beamforming methods.

The current work is an extension of our previous work [30] which developed a CPD-based channel estimation method for RIS-assisted mmWave MISO systems. The contribution of the current work beyond [30] consists of the following two aspects. First, we extend the CPD-based method to the MIMO scenarios. Note that such an extension is nontrivial. In fact, for the MISO case, as pointed out in [30], the Kruskals condition which is essential to the uniqueness of the CPD does not hold and hence the classical ALS method cannot be applied. We show that due to the diversity brought by multi-antenna at the receiver, the Kruskals condition can be satisfied for MIMO scenarios. Such a fact enables us to develop an ALS-based channel estimation method that performs better than [30], especially in the low SNR regime. Second, besides channel estimation, the current work also considers how to optimize the active and passive beamforming coefficients based on the estimated cascade channel, which is a challenging problem (particularly for MIMO systems) and has not studied in [30].

The rest of the paper is organized as follows. In Section II, the system model and the formulation of the channel estimation problem are discussed. CPD-based channel estimation method is developed in Section III. In Section IV, the problem of joint active and passive beamforming design is studied. Finally, simulation results are presented in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a point-to-point RIS-aided mmWave MIMO-OFDM system, where an RIS is deployed to assist the downlink data transmission of N_s data streams from the BS to the user (see Fig. 1). For simplicity, we assume that the direct link between the BS and the user is blocked due to poor propagation conditions.

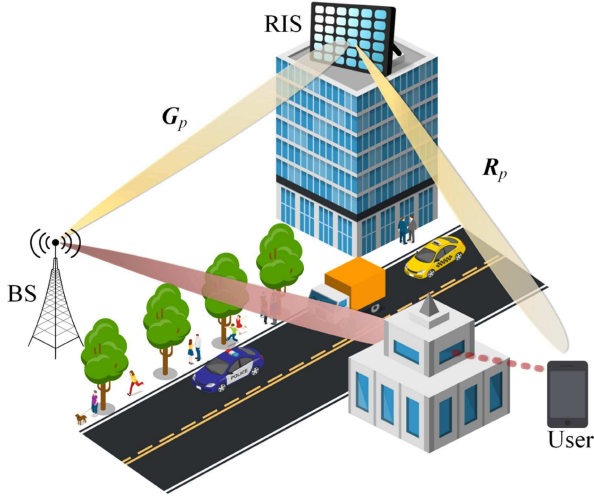


Fig. 1. RIS-assisted mmWave MIMO-OFDM systems.

The BS is equipped with a uniform linear array (ULA) with N_t antennas and R_t radio frequency (RF) chains, and the user is equipped with a ULA with N_r antennas and R_r RF chains, where $R_t \ll N_t$ and $R_r \ll N_r$. The RIS is a uniform planar array (UPA) with $M = M_y \times M_z$ passive reflecting elements. Each element, say the m th element, can independently reflect the incident signal with a reconfigurable phase shift $e^{j\varsigma_m}$. For notational simplicity, let $\mathbf{v} \triangleq [e^{j\varsigma_1} \dots e^{j\varsigma_M}]^H$ denote the reflection coefficient vector, and $\Phi \triangleq \text{diag}(\mathbf{v}^H)$ denote the reflection matrix.

A. Channel Model

In this paper, we adopt a geometric wideband mmWave channel model [31] to characterize the channel between the BS (RIS) and the IRS (user). Specifically, the BS-RIS channel in the delay domain can be expressed as

$$\mathbf{G}(\tau) = \sum_{l=1}^L \alpha_l \mathbf{a}_{\text{IRS}}(\vartheta_l^r, \chi_l^r) \mathbf{a}_{\text{BS}}^H(\phi_l) \delta(\tau - \tau_l), \quad (1)$$

where L is the total number of paths between the BS and the RIS, α_l is the complex gain associated with the l th path, ϕ_l represents the spatial angle of departure (AoD), $\{\vartheta_l^r, \chi_l^r\}$ denote the spatial azimuth and elevation angles of arrival (AoAs), τ_l denotes the time delay, $\delta(\tau)$ denotes the Dirac-delta function, $\mathbf{a}_{\text{IRS}}(\vartheta, \chi)$ and $\mathbf{a}_{\text{BS}}(\phi)$ denote the normalized array response vectors associated with the IRS and the BS, respectively. Similarly, the IRS-user channel in the delay domain is modeled as

$$\mathbf{R}(\tau) = \sum_{l=1}^{L_r} \beta_l \mathbf{a}_{\text{UE}}(\theta_l) \mathbf{a}_{\text{IRS}}^H(\vartheta_l^t, \chi_l^t) \delta(\tau - \kappa_l), \quad (2)$$

where L_r is the number of paths between the RIS and the user, β_l denotes the associated complex path gain, θ_l represents the spatial AoA, $\{\vartheta_l^t, \chi_l^t\}$ denote the spatial azimuth and elevation AoDs, κ_l is the time delay, and $\mathbf{a}_{\text{UE}}(\theta)$ denotes the normalized array response vectors associated with the user. The normalized array response vectors $\mathbf{a}_{\text{BS}}(\phi)$, $\mathbf{a}_{\text{UE}}(\theta)$ and $\mathbf{a}_{\text{IRS}}(\vartheta, \chi)$ are

respectively defined as

$$\mathbf{a}_{\text{BS}}(\phi) \triangleq \frac{1}{\sqrt{N_t}} [1 \ e^{j\phi} \dots e^{j(N_t-1)\phi}]^T, \quad (3)$$

$$\mathbf{a}_{\text{UE}}(\theta) \triangleq \frac{1}{\sqrt{N_r}} [1 \ e^{j\theta} \dots e^{j(N_r-1)\theta}]^T, \quad (4)$$

$$\mathbf{a}_{\text{IRS}}(\vartheta, \chi) \triangleq \mathbf{a}_y(\vartheta) \otimes \mathbf{a}_z(\chi), \quad (5)$$

$$= \frac{1}{\sqrt{M_y}} [1 \ e^{j\vartheta} \dots e^{j(M_y-1)\vartheta}]^T \quad (6)$$

$$\otimes \frac{1}{\sqrt{M_z}} [1 \ e^{j\chi} \dots e^{j(M_z-1)\chi}]^T, \quad (7)$$

where $\otimes$ denotes the Kronecker product, $\phi \triangleq \frac{2\pi d}{\lambda} \sin(\eta)$, $\theta \triangleq \frac{2\pi d}{\lambda} \sin(\gamma)$, $\vartheta \triangleq \cos(\varpi) \sin(\psi)$, and $\chi \triangleq \sin(\varpi)$. Here ψ and ϖ denote the azimuth and elevation angles associated with the IRS, γ (η) represents the angle associated with the user (BS), d and λ denote the adjacent spacing and the signal wavelength, respectively.

Accordingly, the frequency-domain BS-RIS and RIS-user channels associated with the p th subcarrier can be respectively expressed as

$$\mathbf{G}_p = \sum_{l=1}^L \alpha_l e^{-j2\pi f_s \tau_l \frac{P}{P_0}} \mathbf{a}_{\text{IRS}}(\vartheta_l^r, \chi_l^r) \mathbf{a}_{\text{BS}}^H(\phi_l), \quad (8)$$

$$\mathbf{R}_p = \sum_{l=1}^{L_r} \beta_l e^{-j2\pi f_s \kappa_l \frac{P}{P_0}} \mathbf{a}_{\text{UE}}(\theta_l) \mathbf{a}_{\text{IRS}}^H(\vartheta_l^t, \chi_l^t), \quad (9)$$

where f_s is the sample frequency, and P_0 denotes the total number of OFDM tones.

B. Downlink Training and Signal Model

To facilitate the algorithm development, we employ a frame-based downlink training protocol (see Fig. 2). Specifically, the training period is divided into T time frames, where the BS (user) employs different beamforming (combining) vectors at different time frames. Each time frame is further divided into Q time slots. At the q th time slot, the RIS adopts an individual phase-shift matrix Φ_q to reflect the impinging signal. Suppose the total number of OFDM tones is P_0 , among which P , say $\{1, 2, \dots, P\}$, subcarriers are selected for training. The transmitted signal associated with the p th subcarrier at the t th time frame can be expressed as

$$\mathbf{f}_{t,p} = \mathbf{F}_{\text{RF},t} \mathbf{F}_{\text{BB},t,p} \mathbf{s}_{t,p} \in \mathbb{C}^{N_t}, \quad (10)$$

where $\mathbf{s}_{t,p} \in \mathbb{C}^{N_s}$ denotes the p th subcarrier's pilot symbol vector, $\mathbf{F}_{\text{BB},t,p} \in \mathbb{C}^{R_t \times N_s}$ is the baseband precoding matrix associated with the p th subcarrier, and $\mathbf{F}_{\text{RF},t} \in \mathbb{C}^{N_t \times R_t}$ is a radio frequency (RF) precoder common to all subcarriers. The transmitted signal arrives at the user via propagation through the BS-RIS-user channel. The user employs the hybrid combiner $\mathbf{W}_{t,p} = \mathbf{W}_{\text{RF},t} \mathbf{W}_{\text{BB},t,p} \in \mathbb{C}^{N_r \times N_s}$ to combine the received signal, where $\mathbf{W}_{\text{BB},t,p} \in \mathbb{C}^{R_r \times N_s}$ denotes the baseband combiner associated with the p th subcarrier and $\mathbf{W}_{\text{RF},t} \in \mathbb{C}^{N_r \times R_r}$ is a RF combiner common to all subcarriers. Hence, the received

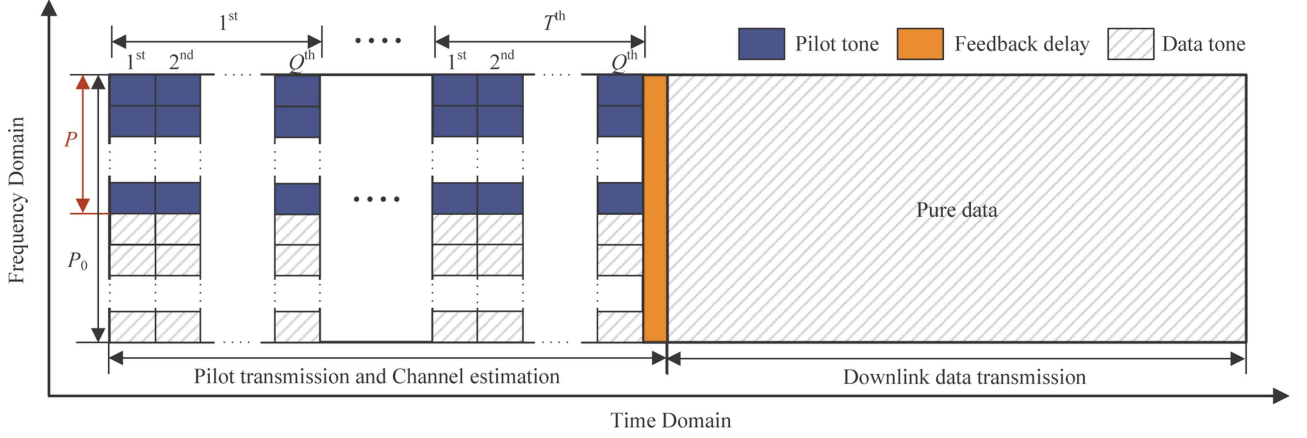


Fig. 2. Illustration of the proposed transmission protocol.

signal $\mathbf{y}_{q,t,p} \in \mathbb{C}^{N_s}$ associated with the p th sub-carrier at the q th time slot of the t th time frame can be written as

$$\mathbf{y}_{q,t,p} = \mathbf{W}_{\text{BB},t,p}^H \mathbf{W}_{\text{RF},t}^H \mathbf{R}_p \Phi_q \mathbf{G}_p \mathbf{F}_{\text{RF},t} \mathbf{F}_{\text{BB},t,p} \mathbf{s}_{t,p} + \mathbf{n}_{q,t,p}. \quad (11)$$

In the channel estimation stage, for simplicity, we assume $\mathbf{F}_{\text{BB},t,p} = \mathbf{F}_{\text{BB},t}$ and $\mathbf{s}_{t,p} = \mathbf{s}_t$, in which case we have $\mathbf{f}_{t,p} = \mathbf{f}_t \triangleq \mathbf{F}_{\text{RF},t} \mathbf{F}_{\text{BB},t} \mathbf{s}_t$. Similarly, let $\mathbf{W}_{\text{BB},t,p} = \mathbf{W}_{\text{BB},t}$. We have $\mathbf{W}_{t,p} = \mathbf{W}_t \triangleq \mathbf{W}_{\text{RF},t} \mathbf{W}_{\text{BB},t}$. The received signal can thus be expressed as

$$\begin{aligned} \mathbf{y}_{q,t,p} &= \mathbf{W}_t^H \mathbf{R}_p \Phi_q \mathbf{G}_p \mathbf{f}_t + \mathbf{n}_{q,t,p}, \\ &= (\mathbf{f}_t^T \otimes \mathbf{W}_t^H) \text{vec}(\mathbf{R}_p \Phi_q \mathbf{G}_p) + \mathbf{n}_{q,t,p}, \\ &= \mathbf{X}_t^T (\mathbf{G}_p^T \odot \mathbf{R}_p) \mathbf{v}_q^* + \mathbf{n}_{q,t,p}, \end{aligned} \quad (12)$$

where we define $\mathbf{X}_t \triangleq \mathbf{f}_t \otimes \mathbf{W}_t^* \in \mathbb{C}^{N_t N_r \times N_s}$, $\odot$ denotes the Khatri-Rao product, and $\mathbf{H}_p \triangleq \mathbf{G}_p^T \odot \mathbf{R}_p \in \mathbb{C}^{N_t N_r \times M}$ denotes the cascade channel associated with the p th subcarrier. Recalling (8) and (9), the cascade channel $\mathbf{H}_p$ can be further expressed as

$$\begin{aligned} \mathbf{H}_p &= \mathbf{G}_p^T \odot \mathbf{R}_p, \\ &= \left(\sum_{m=1}^L \alpha_m e^{-j2\pi f_s \tau_m \frac{P}{P_0}} \mathbf{a}_{\text{BS}}^*(\phi_m) \mathbf{a}_{\text{IRS}}^T(\vartheta_m^r, \chi_m^r) \right) \\ &\quad \odot \left(\sum_{n=1}^{L_r} \beta_n e^{-j2\pi f_s \kappa_n \frac{P}{P_0}} \mathbf{a}_{\text{UE}}(\theta_n) \mathbf{a}_{\text{IRS}}^H(\vartheta_n^t, \chi_n^t) \right), \\ &= \sum_{m=1}^L \sum_{n=1}^{L_r} \alpha_m \beta_n e^{-j2\pi f_s (\tau_m + \kappa_n) \frac{P}{P_0}} (\mathbf{a}_{\text{BS}}^*(\phi_m) \\ &\quad \otimes \mathbf{a}_{\text{UE}}(\theta_n)) (\mathbf{a}_{\text{IRS}}^T(\vartheta_m^r, \chi_m^r) \odot \mathbf{a}_{\text{IRS}}^H(\vartheta_n^t, \chi_n^t)), \\ &= \sum_{m=1}^L \sum_{n=1}^{L_r} \alpha_m \beta_n e^{-j2\pi f_s (\tau_m + \kappa_n) \frac{P}{P_0}} \mathbf{a}_{\text{S}}(\phi_m, \theta_n) \\ &\quad \times \mathbf{a}_{\text{IRS}}^T(\vartheta_m^r - \vartheta_n^t, \chi_m^r - \chi_n^t), \end{aligned}$$

$$\stackrel{(a)}{=} \sum_{u=1}^{LL_r} \varrho_u e^{-j2\pi f_s \iota_u \frac{P}{P_0}} \mathbf{a}_{\text{S}}(\phi_u, \theta_u) \mathbf{a}_{\text{IRS}}^T(\zeta_u, \xi_u), \quad (13)$$

where $\mathbf{a}_{\text{S}}(\phi_m, \theta_n) \triangleq \mathbf{a}_{\text{BS}}^*(\phi_m) \otimes \mathbf{a}_{\text{UE}}(\theta_n)$, and the mapping process (a) is defined as

$$\begin{aligned} (m-1)L_r + n &\mapsto u, u = 1, \dots, LL_r, \\ \alpha_m \beta_n &\mapsto \varrho_u, u = 1, \dots, LL_r, \\ \tau_m + \kappa_n &\mapsto \iota_u, u = 1, \dots, LL_r, \\ \vartheta_m^r - \vartheta_n^t &\mapsto \zeta_u, u = 1, \dots, LL_r, \\ \chi_m^r - \chi_n^t &\mapsto \xi_u, u = 1, \dots, LL_r, \\ \phi_m &\mapsto \phi_u, m = \left\lceil \frac{u}{L_r} \right\rceil, u = 1, \dots, LL_r, \\ \theta_n &\mapsto \theta_u, n = u - \left(\left\lceil \frac{u}{L_r} \right\rceil - 1 \right) L_r, u = 1, \dots, LL_r, \end{aligned} \quad (14)$$

where $\lceil x \rceil$ denotes the ceiling function which gives the least integer greater than or equal to x .

Our objective of this paper is two-fold. First, we wish to develop a method to estimate the cascade channel matrices $\{\mathbf{H}_p\}$ from the received measurements $\{\mathbf{y}_{q,t,p}\}$. In particular, the proposed method is expected to provide a reliable channel estimate by using as few pilot symbols as possible. After the cascade channel matrices are obtained, another purpose of this work is to develop a joint active and passive beamforming method which aims to maximize the spectral efficiency by exploiting the knowledge of the cascade channels.

III. PROPOSED CPD-BASED CHANNEL ESTIMATION METHOD

A. Low-Rank Tensor Representation

Substituting (13) into (12), we have

$$\begin{aligned} \mathbf{y}_{q,t,p} &= \sum_{u=1}^{LL_r} \varrho_u e^{-j2\pi f_s \iota_u \frac{P}{P_0}} \mathbf{X}_t^T \mathbf{a}_{\text{S}}(\phi_u, \theta_u) \\ &\quad \times \mathbf{a}_{\text{IRS}}^T(\zeta_u, \xi_u) \mathbf{v}_q^* + \mathbf{n}_{q,t,p}. \end{aligned} \quad (15)$$

Define $\mathbf{Y}_{t,p} \triangleq [\mathbf{y}_{1,t,p} \ \cdots \ \mathbf{y}_{Q,t,p}]^T \in \mathbb{C}^{Q \times N_s}$. The received signal at the t th time frame can be written as

$$\mathbf{Y}_{t,p} = \sum_{u=1}^{LL_r} \varrho_u e^{-j2\pi f_s \iota_u \frac{p}{P_0}} \mathbf{V}^T \mathbf{a}_{\text{IRS}}(\zeta_u, \xi_u) \times \mathbf{a}_S^T(\phi_u, \theta_u) \mathbf{X}_t + \mathbf{N}_{t,p}, \quad (16)$$

where

$$\mathbf{V} \triangleq [\mathbf{v}_1^* \ \cdots \ \mathbf{v}_Q^*] \in \mathbb{C}^{M \times Q}, \quad (17)$$

$$\mathbf{N}_{t,p} \triangleq [\mathbf{n}_{1,t,p} \ \cdots \ \mathbf{n}_{Q,t,p}]^T \in \mathbb{C}^{Q \times N_s}. \quad (18)$$

Collecting the received signals from all time frames and defining $\mathbf{Y}_p \triangleq [\mathbf{Y}_{1,p} \ \cdots \ \mathbf{Y}_{T,p}] \in \mathbb{C}^{Q \times TN_s}$, we have

$$\begin{aligned} \mathbf{Y}_p &= \sum_{u=1}^{LL_r} \varrho_u e^{-j2\pi f_s \iota_u \frac{p}{P_0}} \mathbf{V}^T \mathbf{a}_{\text{IRS}}(\zeta_u, \xi_u) \\ &\quad \times \mathbf{a}_S^T(\phi_u, \theta_u) \mathbf{F} + \mathbf{N}_p, \\ &= \sum_{u=1}^{LL_r} \varrho_u e^{-j2\pi f_s \iota_u \frac{p}{P_0}} \tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u) \tilde{\mathbf{a}}_S^T(\phi_u, \theta_u) + \mathbf{N}_p, \end{aligned} \quad (19)$$

where $\tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u) \triangleq \mathbf{V}^T \mathbf{a}_{\text{IRS}}(\zeta_u, \xi_u) \in \mathbb{C}^Q$, $\tilde{\mathbf{a}}_S(\phi_u, \theta_u) \triangleq \mathbf{F}^T \mathbf{a}_S(\phi_u, \theta_u) \in \mathbb{C}^{TN_s}$, and

$$\mathbf{F} \triangleq [\mathbf{X}_1 \ \cdots \ \mathbf{X}_T] \in \mathbb{C}^{N_t N_r \times TN_s}, \quad (20)$$

$$\mathbf{N}_p \triangleq [\mathbf{N}_{1,p} \ \cdots \ \mathbf{N}_{T,p}] \in \mathbb{C}^{Q \times TN_s}. \quad (21)$$

Recall that $\mathbf{Y}_p$ is the received signal associated with the p th subcarrier. As we have multiple subcarriers, the entire received signal $\{\mathbf{Y}_p\}_p$ can be expressed as a third-order tensor $\mathcal{Y} \in \mathbb{C}^{Q \times TN_s \times P}$. Each slice of the tensor $\mathcal{Y}$ is $\mathbf{Y}_p$. It can be easily verified that this third-order tensor $\mathcal{Y}$ can be expressed as a sum of rank-one component tensors, a.k.a. the CPD, i.e.,

$$\mathcal{Y} = \sum_{u=1}^U \tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u) \circ (\varrho_u \tilde{\mathbf{a}}_S(\phi_u, \theta_u)) \circ \mathbf{g}(\iota_u) + \mathcal{N}, \quad (22)$$

where $U \triangleq LL_r$, $\mathcal{N} \in \mathbb{C}^{Q \times TN_s \times P}$ is the noise tensor, and

$$\mathbf{g}(\iota_u) \triangleq [e^{-j2\pi f_s \iota_u \frac{1}{P_0}} \ \cdots \ e^{-j2\pi f_s \iota_u \frac{P}{P_0}}]^T. \quad (23)$$

Define

$$\mathbf{A} \triangleq [\tilde{\mathbf{a}}_{\text{IRS}}(\zeta_1, \xi_1) \ \cdots \ \tilde{\mathbf{a}}_{\text{IRS}}(\zeta_U, \xi_U)] \in \mathbb{C}^{Q \times U}, \quad (24)$$

$$\mathbf{B} \triangleq [\varrho_1 \tilde{\mathbf{a}}_S(\phi_1, \theta_1) \ \cdots \ \varrho_U \tilde{\mathbf{a}}_S(\phi_U, \theta_U)] \in \mathbb{C}^{TN_s \times U}, \quad (25)$$

$$\mathbf{C} \triangleq [\mathbf{g}(\iota_1) \ \cdots \ \mathbf{g}(\iota_U)] \in \mathbb{C}^{P \times U}. \quad (26)$$

Here $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ are the factor matrices of the tensor $\mathcal{Y}$. We see that the three factor matrices are characterized by different channel parameters, and it is convenient to extract the respective parameters from each factor matrix.

B. CP Decomposition

For generic CPD problems, an alternating least squares (ALS) method is usually employed to search for the factor matrices.

Specifically, assume that the CP rank U , is known a priori. The CP decomposition of $\mathcal{Y}$ can be accomplished by solving

$$\min_{\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}} \left\| \mathcal{Y} - \sum_{u=1}^U \hat{\mathbf{a}}_u \circ \hat{\mathbf{b}}_u \circ \hat{\mathbf{c}}_u \right\|_F^2, \quad (27)$$

where we define $\hat{\mathbf{A}} \triangleq [\hat{\mathbf{a}}_1 \ \cdots \ \hat{\mathbf{a}}_U]$, $\hat{\mathbf{B}} \triangleq [\hat{\mathbf{b}}_1 \ \cdots \ \hat{\mathbf{b}}_U]$ and $\hat{\mathbf{C}} \triangleq [\hat{\mathbf{c}}_1 \ \cdots \ \hat{\mathbf{c}}_U]$. The above optimization can be solved by an ALS procedure which alternatively solves the least squares problems [32].

For the case where the CP rank U is unknown a priori, we can either employ the minimum description length (MDL) criterion [33] to estimate the CP rank or resort to some sophisticated CPD techniques, e.g. [34], which jointly estimate the CP rank and the factor matrices. Interested readers can refer [34] for more details.

C. Channel Estimation

From CPD theories, we know that there exist a scaling ambiguity and a permutation ambiguity between the estimated factor matrices and the true factor matrices. The permutation ambiguity is common to all estimated factor matrices, and thus can be ignored. We only need to consider the scaling ambiguity in the channel estimation stage. Mathematically, the estimated factor matrices can be expressed as

$$\hat{\mathbf{A}} = \mathbf{A} \Psi_1 + \mathbf{E}_1, \quad (28)$$

$$\hat{\mathbf{B}} = \mathbf{B} \Psi_2 + \mathbf{E}_2, \quad (29)$$

$$\hat{\mathbf{C}} = \mathbf{C} \Psi_3 + \mathbf{E}_3, \quad (30)$$

where $\{\Psi_i\}_{i=1}^3$ denote the scaling ambiguity matrices, and we have $\Psi_1 \Psi_2 \Psi_3 = \mathbf{I}$ according to CPD theories.

We now discuss how to extract channel parameters from each estimated factor matrix. Recall that each column of the factor matrix $\mathbf{A}$ is characterized by a pair of angular parameters $\{\zeta_u, \xi_u\}$. Hence a correlation-based technique whose objective is to maximize the correlation between the estimated vector and the parametric vector can be used to search for this pair of parameters:

$$\{\hat{\zeta}_u, \hat{\xi}_u\} = \arg \max_{\zeta_u, \xi_u} \frac{|\hat{\mathbf{a}}_u^H \tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u)|}{\|\hat{\mathbf{a}}_u\|_2 \|\tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u)\|_2}, \quad (31)$$

where $\hat{\mathbf{a}}_u$ represents the u th column of $\hat{\mathbf{A}}$. For the factor matrices $\mathbf{B}$ and $\mathbf{C}$, a similar correlation-based technique can be employed to respectively estimate $\{\hat{\phi}_u, \hat{\theta}_u\}$ and $\hat{\iota}_u$.

Next, based on the estimated channel parameters $\{\zeta_u, \xi_u, \phi_u, \theta_u, \iota_u\}$, the composite path gains $\{\varrho_u\}$ can be recovered. After obtaining $\{\hat{\zeta}_u, \hat{\xi}_u\}$, the factor matrix $\mathbf{A}$ can be accordingly estimated as $\hat{\mathbf{A}} = [\tilde{\mathbf{a}}_{\text{IRS}}(\hat{\zeta}_1, \hat{\xi}_1) \ \cdots \ \tilde{\mathbf{a}}_{\text{IRS}}(\hat{\zeta}_U, \hat{\xi}_U)]$. Recall that $\hat{\mathbf{A}}$ and $\mathbf{A}$ are related as $\hat{\mathbf{A}} = \mathbf{A} \Psi_1$. Hence the nonsingular diagonal matrix Ψ_1 can be estimated as $\Psi_1 = \hat{\mathbf{A}}^\dagger \hat{\mathbf{A}}$. Similarly, we have $\Psi_3 = \hat{\mathbf{C}}^\dagger \hat{\mathbf{C}}$, in which $\hat{\mathbf{C}} = [\mathbf{g}(\hat{\iota}_1) \ \cdots \ \mathbf{g}(\hat{\iota}_U)]$. Since $\Psi_1 \Psi_2 \Psi_3 = \mathbf{I}$, Ψ_2 can be obtained as $\Psi_2 = \Psi_1^{-1} \Psi_3^{-1}$.

Notice that we have

$$\hat{\mathbf{B}} = \begin{bmatrix} \tilde{\mathbf{a}}_S(\hat{\phi}_1, \hat{\theta}_1) & \cdots & \tilde{\mathbf{a}}_S(\hat{\phi}_U, \hat{\theta}_U) \end{bmatrix} \text{diag}(\varrho_1, \dots, \varrho_U) \Psi_2. \quad (32)$$

Thus, $\mathbf{D} = \text{diag}\{\varrho_1, \dots, \varrho_U\}$ can be estimated as $\hat{\mathbf{D}} = \tilde{\mathbf{B}}^\dagger \hat{\mathbf{B}} \Psi_2^{-1}$, where we have $\tilde{\mathbf{B}} = [\tilde{\mathbf{a}}_S(\hat{\phi}_1, \hat{\theta}_1) \cdots \tilde{\mathbf{a}}_S(\hat{\phi}_U, \hat{\theta}_U)]$.

D. Uniqueness Condition and Sample Complexity

We now study the uniqueness condition of CPD for our considered problem. This condition also sheds light on the sample complexity of the proposed method, i.e. the amount of training overhead required to reliably estimate the channel. A well-known condition for the uniqueness of CPD is Kruskal's condition, which is summarized as follows.

Theorem 1: Let $\mathcal{X} \in \mathbb{C}^{I \times J \times K}$ be a third-order tensor decomposed of three factor matrices $\mathbf{A}^{(1)} \in \mathbb{C}^{I \times R}$, $\mathbf{A}^{(2)} \in \mathbb{C}^{J \times R}$ and $\mathbf{A}^{(3)} \in \mathbb{C}^{K \times R}$. If

$$k_{\mathbf{A}^{(1)}} + k_{\mathbf{A}^{(2)}} + k_{\mathbf{A}^{(3)}} \geq 2R + 2, \quad (33)$$

then the CPD of $\mathcal{X}$ is unique up to scaling and permutation ambiguities.

Here $k_{\mathbf{A}}$ denotes the k-rank of $\mathbf{A}$, which is defined as the largest value of $k_{\mathbf{A}}$ such that every subset of $k_{\mathbf{A}}$ columns of $\mathbf{A}$ is linearly independent. To obtain the uniqueness condition for our problem, we analyze the k-rank of each factor matrix. Consider the factor matrix $\mathbf{A}$. Recall that

$$\mathbf{A} \triangleq \mathbf{V}^T [\mathbf{a}_{\text{IRS}}(\zeta_1, \xi_1) \cdots \mathbf{a}_{\text{IRS}}(\zeta_U, \xi_U)]. \quad (34)$$

Assume that each entry of $\mathbf{V} \in \mathbb{C}^{M \times Q}$ is randomly generated from the unit circle, i.e. $v_{m,n} = e^{j\sigma_{m,n}}$, where $\sigma_{m,n} \in [-\pi, \pi]$ is drawn from a uniform distribution. Let $a_{m,i} \triangleq \mathbf{v}_m^T \mathbf{a}_{\text{IRS}}(\zeta_i, \xi_i)$ denote the (m, i) th entry of $\mathbf{A}$, where $\mathbf{v}_m$ is the m th column of $\mathbf{V}$. It can be easily verified that $E[a_{m,i}] = 0, \forall m, i$ and

$$E[a_{m,i}^H a_{n,j}] = \begin{cases} 0, & m \neq n, \\ \mathbf{a}_{\text{IRS}}^H(\zeta_j, \xi_j) \mathbf{a}_{\text{IRS}}(\zeta_i, \xi_i), & m = n. \end{cases} \quad (35)$$

Recall that $\mathbf{a}_{\text{IRS}}(\zeta_i, \xi_i)$ is a Kronecker product of two steering vectors. Hence we have $\mathbf{a}_{\text{IRS}}^H(\zeta_j, \xi_j) \mathbf{a}_{\text{IRS}}(\zeta_i, \xi_i) \approx 0$, when $\zeta_i \neq \zeta_j$ or $\xi_i \neq \xi_j$ [35]. In reality, due to the random nature of the channel parameters, the angles $\{\zeta_u\}_{u=1}^U$ are mutually distinct with probability one, and so are the angles $\{\xi_u\}_{u=1}^U$. We thus have $E[a_{m,i}^H a_{n,j}] \approx 0$ even for the case $m = n$. On the other hand, according to the central limit theorem, $a_{m,i}$ approximately follows a Gaussian distribution. Therefore entries of $\mathbf{A}$ can be considered as i.i.d. Gaussian variables with zero mean and unit variance. As a result, we have

$$k_{\mathbf{A}} = \min\{Q, U\}. \quad (36)$$

By following a similar derivation, we have

$$k_{\mathbf{B}} = \min\{TN_s, U\}. \quad (37)$$

Note that due to the mapping process, the set $\{\phi_u\}$ only contains L distinct elements, and $\{\theta_u\}$ only contains L_r distinct elements. Nevertheless, for $i \neq j$, we still have $\mathbf{a}_S(\phi_i, \theta_i) \neq \mathbf{a}_S(\phi_j, \theta_j)$

since each pair of (ϕ_u, θ_u) is unique according to the mapping rule defined in (14).

As for the factor matrix $\mathbf{C}$, it is a Vandermonde matrix with distinct generators, i.e. $\iota_i \neq \iota_j, \forall i \neq j$. Thus we have

$$k_{\mathbf{C}} = \min\{P, U\}. \quad (38)$$

Based on the above results, we know that Kruskal's condition is equivalent to

$$\min\{Q, U\} + \min\{TN_s, U\} + \min\{P, U\} \geq 2U + 2. \quad (39)$$

Note that the total number of pilot signals for downlink training is QTP . To meet condition (39), we can set $Q \geq U, TN_s \geq U$ and $P \geq 2$, in which case the amount of training overhead is in the order of $\mathcal{O}(2U^2/N_s)$. We see that the sample complexity of the proposed method only depends on the sparsity of the cascade channel U . As U is usually small relative to the dimension of the cascade channel, a substantial training overhead reduction can be achieved.

IV. JOINT ACTIVE AND PASSIVE BEAMFORMING DESIGN

In this section, we consider the problem of joint active and passive beamforming design based on the estimated channel parameters. Specifically, we aim to maximize the spectral efficiency by jointly optimizing the precoding matrices at the transmitter, the combining matrices at the receiver, and the reflection coefficients at the RIS. Such an optimization problem can be formulated as

$$\begin{aligned} & \max_{\{\mathbf{F}_p\}_{p=1}^P, \{\mathbf{W}_p\}_{p=1}^P, \Phi} \frac{1}{P} \sum_{p=1}^P \log_2 \det \left(\mathbf{I}_{N_s} + \frac{1}{\sigma^2} \mathbf{W}_p^\dagger \bar{\mathbf{H}}_p \right. \\ & \quad \left. \times \mathbf{F}_p \mathbf{F}_p^H \bar{\mathbf{H}}_p^H \mathbf{W}_p \right), \\ & \text{s.t. } \|\mathbf{F}_p\|_F^2 \leq \rho, \forall p = 1, \dots, P, \\ & \quad \bar{\mathbf{H}}_p = \mathbf{R}_p \Phi \mathbf{G}_p, \forall p = 1, \dots, P, \\ & \quad \Phi = \text{diag}(e^{j\varsigma_1}, e^{j\varsigma_2}, \dots, e^{j\varsigma_M}), \end{aligned} \quad (40)$$

where $\dagger$ represents the pseudo-inverse operation, $\|\mathbf{F}_p\|_F^2 \leq \rho$ denotes a transmit power constraint for each subcarrier, $\bar{\mathbf{H}}_p \triangleq \mathbf{R}_p \Phi \mathbf{G}_p$ denotes the effective channel associated with the p th subcarrier. Note that the definition of $\bar{\mathbf{H}}_p$ is different from that of the cascade channel $\mathbf{H}_p$. $\mathbf{F}_p \in \mathbb{C}^{N_t \times N_s}$ and $\mathbf{W}_p \in \mathbb{C}^{N_r \times N_s}$ respectively denote a fully digital precoder and a fully digital combiner associated with the p th subcarrier. Once an optimal fully digital precoder/combiner is found, we can use the manifold optimization method [36] to search for a hybrid precoder/combiner to approximate the optimal fully digital precoder/combiner.

Given a fixed Φ , the optimal $\mathbf{F}_p$ and $\mathbf{W}_p$ can be obtained via the singular value decomposition (SVD) of $\bar{\mathbf{H}}_p$. Substituting the optimal fully digital precoder and combiner into (40), we can arrive at a problem which concerns only the optimization of the passive beamforming matrix Φ :

$$\max_{\Phi} \frac{1}{P} \sum_{p=1}^P \log_2 \det \left(\mathbf{I}_{N_s} + \frac{\rho}{N_s \sigma^2} (\Sigma_{1,p})^2 \right),$$

$$\text{s.t. } \Phi = \text{diag}(e^{j\varsigma_1}, e^{j\varsigma_2}, \dots, e^{j\varsigma_M}), \quad (41)$$

where $\Sigma_{1,p} \in \mathbb{C}^{N_s \times N_s}$ is the submatrix of Σ_p , and Σ_p is obtained via the SVD of $\bar{\mathbf{H}}_p$:

$$\begin{aligned} \bar{\mathbf{H}}_p &= \mathbf{U}_p \Sigma_p \mathbf{V}_p^H, \\ &= [\mathbf{U}_{1,p} \quad \mathbf{U}_{2,p}] \begin{bmatrix} \Sigma_{1,p} & \mathbf{0} \\ \mathbf{0} & \Sigma_{2,p} \end{bmatrix} [\mathbf{V}_{1,p} \quad \mathbf{V}_{2,p}]^H. \end{aligned} \quad (42)$$

Here $\mathbf{U}_{1,p} \in \mathbb{C}^{N_r \times N_s}$, $\mathbf{V}_{1,p} \in \mathbb{C}^{N_t \times N_s}$, Σ_p is an $r_p \times r_p$ diagonal matrix, and $r_p \triangleq \text{rank}(\bar{\mathbf{H}}_p)$.

The main difficulty of solving (41) is that the singular values of the effective channel $\bar{\mathbf{H}}_p$ cannot be explicitly expressed in terms of the passive reflecting coefficient matrix Φ . To address this difficulty, [37] proposes to exploit the inherent structure of the effective channel $\bar{\mathbf{H}}_p$. Specifically, substituting (8)–(9) into the definition of $\bar{\mathbf{H}}_p$, the effective channel can be written as

$$\bar{\mathbf{H}}_p = \mathbf{A}_{\text{UE}} \mathbf{D}_p \mathbf{A}_{\text{BS}}^H, \quad (43)$$

where $\mathbf{A}_{\text{UE}} \triangleq [\mathbf{a}_{\text{UE}}(\theta_1) \quad \dots \quad \mathbf{a}_{\text{UE}}(\theta_{L_r})]$, $\mathbf{A}_{\text{BS}} \triangleq [\mathbf{a}_{\text{BS}}(\phi_1) \quad \dots \quad \mathbf{a}_{\text{BS}}(\phi_L)]$, and the (m, n) th entry of $\mathbf{D}_p$ is given by

$$\mathbf{D}_p(m, n) \triangleq \alpha_m \beta_n e^{-j2\pi f_s(\tau_m + \kappa_n) \frac{p}{F_0}} d_{mn}, \quad (44)$$

in which

$$\begin{aligned} d_{mn} &\triangleq \mathbf{a}_{\text{IRS}}^H(\vartheta_n^t, \chi_n^t) \Phi \mathbf{a}_{\text{IRS}}(\vartheta_m^r, \chi_m^r), \\ &= \mathbf{v}^H (\mathbf{a}_{\text{IRS}}^*(\vartheta_n^t, \chi_n^t) \circ \mathbf{a}_{\text{IRS}}(\vartheta_m^r, \chi_m^r)), \\ &= \mathbf{v}^H \mathbf{a}_{\text{IRS}}(\vartheta_m^r - \vartheta_n^t, \chi_m^r - \chi_n^t). \end{aligned} \quad (45)$$

Based on the above expression, we propose to use $\bar{\mathbf{H}}_p = \mathbf{A}_{\text{UE}} \mathbf{D}_p \mathbf{A}_{\text{BS}}^H$ to approximate a truncated SVD of $\bar{\mathbf{H}}_p$. Based on this approximation, the optimization problem (41) can be rewritten as

$$\begin{aligned} \max_{\mathbf{v}} \quad & \frac{1}{P} \sum_{p=1}^P \sum_{i=1}^{N_s} \log_2 \left(1 + \frac{\rho}{N_s \sigma^2} |\alpha_i \beta_i|^2 \mathbf{v}^H \mathbf{P}_{ii} \mathbf{v} \right), \\ \text{s.t.} \quad & \mathbf{v} = [e^{j\varsigma_1} \quad e^{j\varsigma_2} \quad \dots \quad e^{j\varsigma_M}]^H, \end{aligned} \quad (46)$$

where $\mathbf{P}_{ii} \triangleq \mathbf{a}_{\text{IRS}}(\vartheta_i^r - \vartheta_i^t, \chi_i^r - \chi_i^t) \mathbf{a}_{\text{IRS}}^H(\vartheta_i^r - \vartheta_i^t, \chi_i^r - \chi_i^t)$. Such an optimization can be efficiently solved via the manifold optimization technique.

From (46), we see that to optimize the reflection coefficients, we only need the knowledge of the composite gains $\{\alpha_i \beta_i\}_{i=1}^{N_s}$ and the composite RIS angles $\{\vartheta_i^r - \vartheta_i^t, \chi_i^r - \chi_i^t\}_{i=1}^{N_s}$. Recall that, in the channel estimation stage, the following channel parameters $\{\hat{\varsigma}_u, \hat{\xi}_u, \hat{\phi}_u, \hat{\theta}_u, \hat{\iota}_u, \hat{\rho}_u\}_{u=1}^U$ are obtained, in which we have

$$\begin{aligned} \alpha_m \beta_n &\mapsto \varrho_u, \quad u = 1, \dots, LL_r, \\ \vartheta_m^r - \vartheta_n^t &\mapsto \zeta_u, \quad u = 1, \dots, LL_r, \\ \chi_m^r - \chi_n^t &\mapsto \xi_u, \quad u = 1, \dots, LL_r. \end{aligned} \quad (47)$$

We see that our proposed channel estimator can provide an estimate of the composite gains as well as the composite RIS angles associated with all U composite paths. The problem now

is how to appropriately choose N_s composite paths from these U composite paths. Randomly choosing N_s composite paths certainly does not work. In fact, it is easy to know that the composite paths corresponding to the diagonal entries of $\mathbf{D}_p$ must have mutually distinct AoDs at the BS and mutually distinct AoAs at the user. Also, to improve the spectral efficiency, clearly we should choose those composite paths whose composite gains are as large as possible. Based on the above considerations, the N_s composite paths can be selected based on the following criterion:

$$\begin{aligned} \max_{\mathcal{I}} \quad & \sum_{i \in \mathcal{I}} |\hat{\rho}_i|^2, \\ \text{s.t.} \quad & \mathcal{I} \subset \{1, \dots, U\}, \quad |\mathcal{I}| = N_s, \\ & |\mathbf{a}_{\text{BS}}^H(\hat{\phi}_i) \mathbf{a}_{\text{BS}}(\hat{\phi}_j)| < \delta_{\text{BS}}, \quad i \neq j, \forall i, j \in \mathcal{I}, \\ & |\mathbf{a}_{\text{UE}}^H(\hat{\theta}_i) \mathbf{a}_{\text{UE}}(\hat{\theta}_j)| < \delta_{\text{UE}}, \quad i \neq j, \forall i, j \in \mathcal{I}, \end{aligned} \quad (48)$$

where the last two constraints are imposed to ensure that the selected composite paths have mutually distinct AoDs at the BS and mutually distinct AoAs at the user, in which δ_{BS} and δ_{UE} are small positive parameters of user's choice. Based on the selected N_s composite paths, the optimization problem (46) can be further written as

$$\begin{aligned} \max_{\mathbf{v}} \quad & \sum_{i \in \mathcal{I}} \log_2 \left(1 + \frac{\rho \hat{\rho}_i^2}{N_s \sigma^2} \mathbf{v}^H \mathbf{P}_{ii} \mathbf{v} \right), \\ \text{s.t.} \quad & \mathbf{P}_{ii} = \mathbf{a}_{\text{IRS}}(\hat{\zeta}_i, \hat{\xi}_i) \mathbf{a}_{\text{IRS}}^H(\hat{\zeta}_i, \hat{\xi}_i), \quad \forall i \in \mathcal{I}, \\ & \mathbf{v} = [e^{j\varsigma_1}, e^{j\varsigma_2}, \dots, e^{j\varsigma_M}]^H. \end{aligned} \quad (49)$$

The above optimization problem can be efficiently solved via the manifold optimization method.

After the passive beamforming vector $\mathbf{v}$ is determined, the p th subcarrier's equivalent channel $\hat{\mathbf{H}}_p$ can be estimated as

$$\begin{aligned} \hat{\mathbf{H}}_p &= \sum_{u=1}^U \hat{\rho}_u e^{-j2\pi f_s \hat{\iota}_u \frac{p}{F_0}} \mathbf{v}^H \mathbf{a}_{\text{IRS}}(\hat{\zeta}_u, \hat{\xi}_u) \\ &\quad \times \mathbf{a}_{\text{UE}}(\hat{\theta}_u) \mathbf{a}_{\text{BS}}^H(\hat{\phi}_u), \\ &= [\hat{\mathbf{U}}_{1,p} \quad \hat{\mathbf{U}}_{2,p}] \begin{bmatrix} \hat{\Sigma}_{1,p} & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_{2,p} \end{bmatrix} [\hat{\mathbf{V}}_{1,p} \quad \hat{\mathbf{V}}_{2,p}]^H, \end{aligned} \quad (50)$$

where the second equality is a truncated SVD of $\hat{\mathbf{H}}_p$. According to [37], the optimal fully digital precoder/combiner can be obtained as

$$\mathbf{F}_{\text{opt},p}^* = \sqrt{\rho/N_s} \hat{\mathbf{V}}_{1,p}, \quad \mathbf{W}_{\text{opt},p}^* = \hat{\mathbf{U}}_{1,p}. \quad (51)$$

After the optimal fully digital precoder/combiner is obtained, we search for a common analog precoding (combining) matrix $\mathbf{F}_{\text{RF}}$ ($\mathbf{W}_{\text{RF}}$) and a set of baseband precoding (combining) matrices $\{\mathbf{F}_{\text{BB},p}\}$ ($\{\mathbf{W}_{\text{BB},p}\}$) to approximate the optimal precoder (combiner) $\{\mathbf{F}_{\text{opt},p}\}$ ($\{\mathbf{W}_{\text{opt},p}\}$). The problem can be solved via the manifold optimization technique.

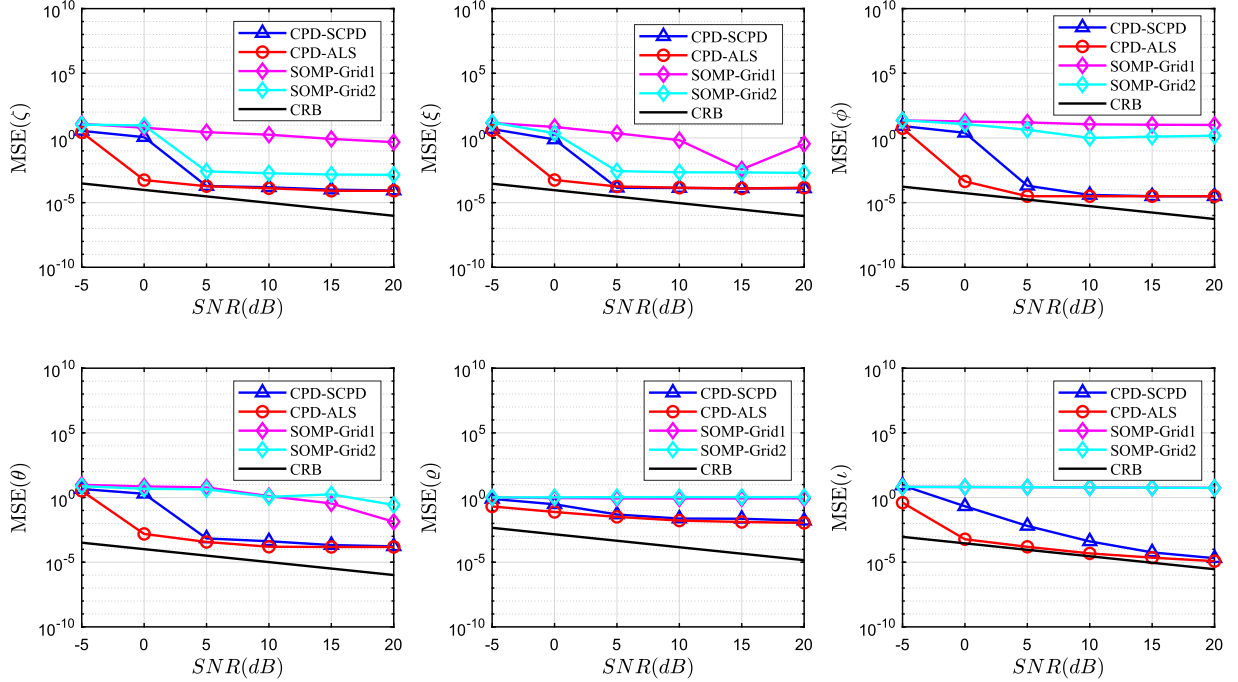


Fig. 3. MSEs of respective methods vs. SNR, where $Q = 32$, $T = 16$, and $P = 16$.

V. SIMULATION RESULTS

We present simulation results to evaluate the performance of the proposed ALS-based CPD channel estimation method and the proposed joint beamforming scheme.

In our simulations, we assume that the BS employs a ULA with $N_t = 32$ antennas and $R_t = 3$ RF chains, the IRS is equipped with $M = 16 \times 16$ passive reflecting elements, and the user employs a ULA with $N_r = 32$ antennas and $R_r = 3$ RF chains. The angular parameters $\{\vartheta_l^r, \chi_l^r, \phi_l\}_{l=1}^L$, $\{\vartheta_l^t, \chi_l^t, \theta_l\}_{l=1}^{L_r}$ are randomly generated from $[0, 2\pi]$, where we set $L = 3$ and $L_r = 3$. The delay spreads $\{\tau_l\}_{l=1}^L$, $\{\kappa_l\}_{l=1}^{L_r}$ are drawn from a uniform distribution $\mathcal{U}(0, 100 \text{ ns})$. The complex gains $\{\alpha_l\}_{l=1}^L$ and $\{\beta_l\}_{l=1}^{L_r}$ follow a circularly symmetric Gaussian distribution $\mathcal{CN}(0, 1)$. The number of data streams is set to $N_s = 2$. The total number of subcarriers is set to $P_0 = 128$, among which P subcarriers are used for training. The sampling rate is set to $f_s = 0.32 \text{ GHz}$. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} \triangleq \frac{\|\mathcal{Y} - \mathcal{N}\|_F^2}{\|\mathcal{N}\|_F^2}. \quad (52)$$

We first examine the estimation accuracy of the channel parameters $\{\zeta_u, \xi_u, \phi_u, \theta_u, \iota_u, \varrho_u\}_{u=1}^U$ and the overall channel estimation performance. Note that the channel estimation problem being considered in this work can be cast as a multi-measurement vector (MMV) compressed sensing problem, and the simultaneous-OMP method (SOMP) [38] can be used to estimate the cascade channel. The TRICE [13] and the Structured-CPD (SCPD) [30] methods are also included for comparison. For the SOMP method, two different grids are employed to discretize the continuous parameter space: the first grid discretizes the

multi-dimensional parameter space into $128 \times (128 \times 128) \times 128 \times 128$ points, and the second grid discretizes the continuous parameter space into $256 \times (256 \times 256) \times 256 \times 256$ points. For the TRICE method, the 4D angular space is discretized into $256 \times 256 \times 256 \times 256$ points. The CRB results are also included to provide a benchmark for evaluating the performance of our proposed method. The calculation of CRB can be found in Appendix VII.

In Fig. 3, we plot the mean square errors (MSEs) of the estimated channel parameters as a function of the SNR, where we set $P = 16$, $T = 16$, and $Q = 32$. Note that the parameters $\{\iota_u, \varrho_u\}_{u=1}^U$ of the cascaded channel cannot be separated by TRICE, thus TRICE is not included in this figure. From Fig. 3, we see that both of the tensor-based approach ALS and SCPD can achieve an estimation accuracy close to the theoretical lower bound, and present a substantial advantage over the SOMP method. Meanwhile, the proposed ALS method is superior to the SCPD method in the low SNR regime. This is probably because the SCPD method which depends critically on the structure of the factor matrix is more sensitive to noise.

In Fig. 4(a), we plot the estimation performance of respective methods as a function of the SNR. The performance is evaluated via the normalized mean squared error (NMSE) of the cascaded channel, which is defined as $\sum_{p=1}^P \|\hat{\mathbf{H}}_p - \mathbf{H}_p\|_F^2 / \sum_{p=1}^P \|\mathbf{H}_p\|_F^2$. Again, we see that our proposed method ALS present a significant performance improvement over other methods. In particular, the ALS method achieves a decent estimation performance even in a low SNR regime, say $\text{SNR} = 0 \text{ dB}$. Note that in mmWave communications, due to the severe path loss, the SNR for channel estimation is usually low, with the range of interest from 0 dB to 10 dB at most. Thus the ability of delivering an accurate channel estimate

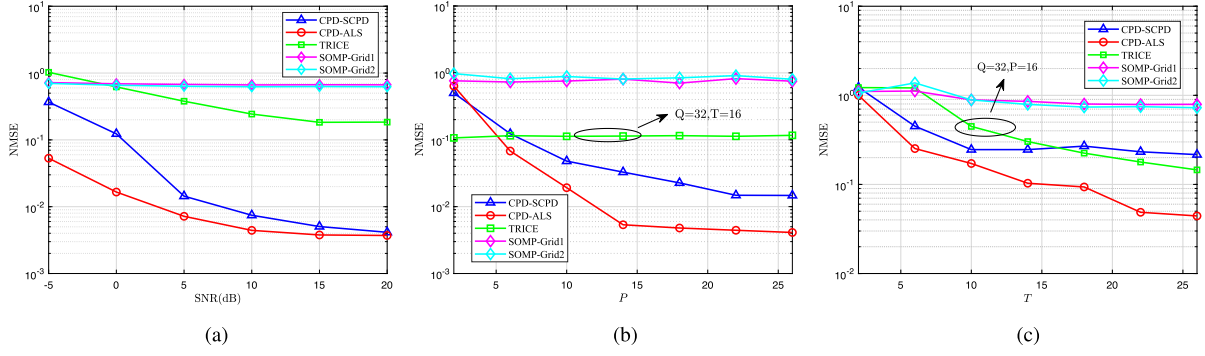


Fig. 4. (a) NMSEs of respective methods vs. SNR, where $Q = 32$, $T = 16$, and $P = 16$; (b) NMSEs of respective methods vs. P , where $Q = 10$, $T = 5$, and SNR = 20 dB; and (c) NMSEs of respective methods vs. T , where $Q = 10$, $P = 10$, and SNR = 20 dB.

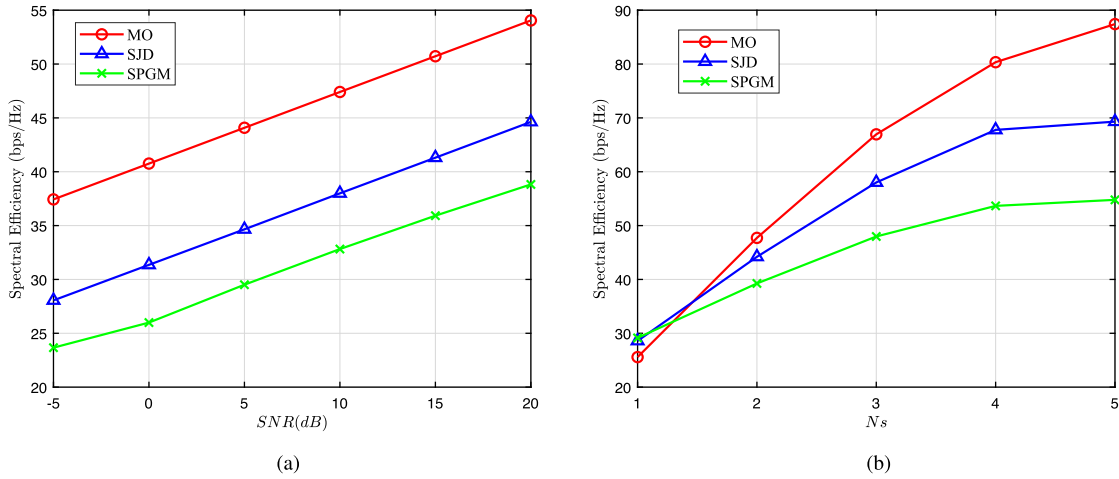


Fig. 5. (a) Spectral efficiency of respective methods (full CSI) vs. SNR, where $N_s = 2$; (b) Spectral efficiency of respective methods (full CSI) vs. N_s , where SNR = 20 dB.

in the low SNR regime is highly desirable and has important practical implications.

In Fig. 4(b), we plot the estimation performance of respective methods as a function of the number of subcarriers P , where we set $Q = 10$ and $T = 5$. It can be seen that the proposed method provides a reliable channel estimate when $P \geq 10$, which corresponds to a total number of 500 measurements for training. As a comparison, note that the cascade channel $\mathbf{H}_p$ to be estimated has a size of $N_t N_r \times M = 1024 \times 256$, which has more than 2.6×10^5 parameters. This result indicates that the proposed method can achieve a substantial training overhead reduction. Fig. 4(c) plots the estimation performance versus the number of time slots T , where we set $Q = 10$ and $P = 10$. This result, again, demonstrates the superiority of the proposed method over the compressed sensing-based method SOMP and TRICE.

Next, we examine the performance of joint beamforming scheme proposed in Section IV. we compare our proposed method (named MO) with state-of-the-art methods SPGM [28] and SJD [29]. Fig. 5(a) depicts the spectral efficiency achieved

by respective methods assuming perfect knowledge of the cascade channel, where we set the number of data stream $N_s = 2$. It can be observed that three algorithms show a linear increase in spectral efficiency as SNR increases. Our method MO has an obvious advantage over the other methods SPGM and SJD. This is due to the fact that the designed method MO has the ability of fully exploiting the spatial diversity inherent in the cascade channel. In Fig. 5(b), we plot the spectral efficiency of respective methods as a function of the number of data stream N_s , where we set SNR = 20 dB. It can be observed that our proposed method presents a clear performance advantage over the other two methods.

Finally, we examine the beamforming performance attained by our proposed joint beamforming method based on the estimated cascade channel. To illustrate the effectiveness of the proposed channel estimator, we include the beamforming performance attained by assuming the perfect knowledge of the CSI, which serves as an upper bound on the beamforming performance attained by using the estimated CSI. Fig. 6 plots the spectral efficiency of the proposed joint beamforming scheme

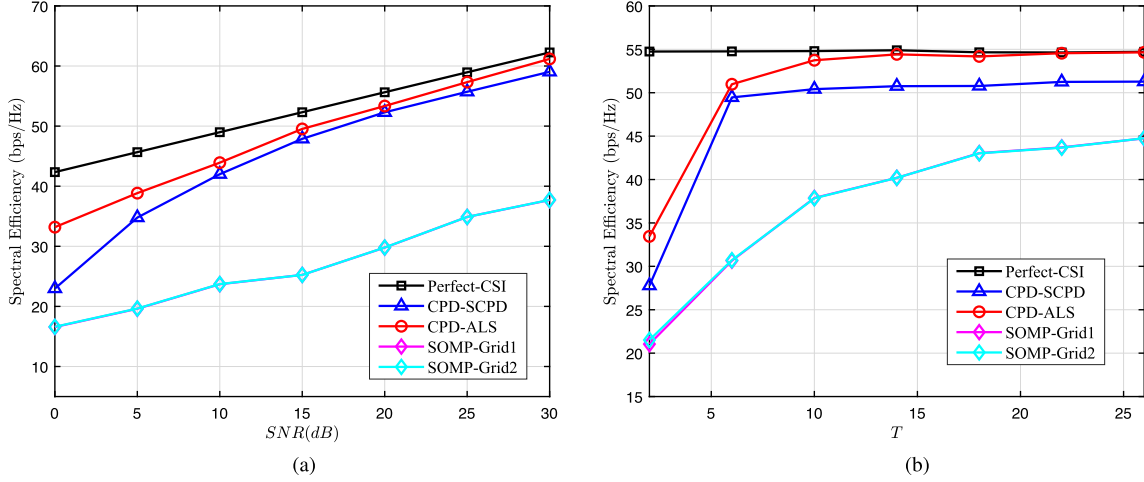


Fig. 6. (a) Spectral efficiency of respective methods (estimated CSI) vs. SNR, where $Q = 16$, $P = 16$, $T = 16$; (b) Spectral efficiency of respective methods (estimated CSI) vs. T , where $Q = 16$, $P = 16$, and SNR = 20 dB.

as a function of SNR and the number of time frames T , respectively. We see that our proposed CPD-based estimator ALS achieves performance close to that attained by assuming perfect CSI knowledge even in the low SNR regime, which verifies the effectiveness of the proposed estimation method. Also, our proposed method presents a clear performance improvement over the compressed sensing-based methods. Particularly, when the training overhead is low, say, $P = Q = 16$ and $T = 6$, the proposed CPD-based method can still achieve decent beamforming performance, whereas the compressed sensing-based methods incurs a significant performance loss.

VI. CONCLUSION

In this paper, by exploiting the intrinsic multi-dimensional structure as well as the sparse scattering characteristics of the mmWave channels, we developed a CPD-aided channel estimation method, namely, an ALS-based CPD method, for RIS-assisted mmWave MIMO-OFDM systems. The proposed method effectively utilizes the low-rankness of the CPD formulation and can achieve a substantial training overhead reduction. We also developed a joint beamforming scheme that utilizes the estimated cascade channel parameters for optimizing the system's active and passive variables. Simulation results show that our proposed method presents a significant performance advantage over the compressed sensing methods, and can achieves superior channel estimation and beamforming performance with a low training overhead.

VII. DERIVATION OF CRAMÉR-RAO LOWER BOUND

Consider the $Q \times TN_s \times P$ observation tensor $\mathcal{Y}$ in (22)

$$\mathcal{Y} = \sum_{u=1}^U \tilde{\mathbf{a}}_{\text{IRS}}(\zeta_u, \xi_u) \circ (\varrho_u \tilde{\mathbf{a}}_{\text{S}}(\phi_u, \theta_u)) \circ \mathbf{g}(\iota_u) + \mathcal{N}, \quad (53)$$

where $\mathcal{N}(q, t, p) \sim \mathcal{CN}(0, \sigma^2)$, $\{\zeta_u, \xi_u, \phi_u, \theta_u, \varrho_u, \iota_u\}$ are the unknown channel parameters to be estimated. Let $\mathbf{p} \triangleq$

$[\zeta^T, \xi^T, \phi^T, \theta^T, \varrho^T, \iota^T]$, where

$$\begin{aligned} \zeta &\triangleq [\zeta_1 \ \cdots \ \zeta_U]^T, \ \xi \triangleq [\xi_1 \ \cdots \ \xi_U]^T, \\ \phi &\triangleq [\phi_1 \ \cdots \ \phi_U]^T, \ \theta \triangleq [\theta_1 \ \cdots \ \theta_U]^T, \\ \varrho &\triangleq [\varrho_1 \ \cdots \ \varrho_U]^T, \ \iota \triangleq [\iota_1 \ \cdots \ \iota_U]^T. \end{aligned}$$

Thus the log-likelihood function of $\mathbf{p}$ can be expressed as

$$\begin{aligned} L(\mathbf{p}) &= f(\mathcal{Y}; \mathbf{A}, \mathbf{B}, \mathbf{C}), \\ &= -QTN_s P \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right\|_F^2, \\ &= -QTN_s P \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \mathbf{Y}_{(2)}^T - (\mathbf{C} \odot \mathbf{A}) \mathbf{B}^T \right\|_F^2, \\ &= -QTN_s P \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \left\| \mathbf{Y}_{(3)}^T - (\mathbf{B} \odot \mathbf{A}) \mathbf{C}^T \right\|_F^2. \end{aligned} \quad (54)$$

The complex Fisher information matrix (FIM) for $\mathbf{p}$ is given by

$$\Omega(\mathbf{p}) = \mathbb{E} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \mathbf{p}} \right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \mathbf{p}} \right) \right\}. \quad (55)$$

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

A. Partial Derivative of $L(\mathbf{p})$ W.r.t $\mathbf{p}$

For simplicity, we consider the partial derivative of $L(\mathbf{p})$ with respect to ζ_u . Partial derivations of $L(\mathbf{p})$ with respect to other parameters can be deduced in a similar way and thus omitted. We have

$$\frac{\partial L(\mathbf{p})}{\partial \zeta_u} = \text{tr} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \mathbf{A}} \right)^T \frac{\partial \mathbf{A}}{\partial \zeta_u} + \left(\frac{\partial L(\mathbf{p})}{\partial \mathbf{A}^*} \right)^T \frac{\partial \mathbf{A}^*}{\partial \zeta_u} \right\}, \quad (56)$$

where

$$\frac{\partial L(\mathbf{p})}{\partial \mathbf{A}} = \frac{1}{\sigma^2} \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right)^H (\mathbf{C} \odot \mathbf{B}), \quad (57)$$

$$\frac{\partial L(\mathbf{p})}{\partial \mathbf{A}^*} = \frac{1}{\sigma^2} \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right)^T (\mathbf{C} \odot \mathbf{B})^*, \quad (58)$$

$$\frac{\partial \mathbf{A}}{\partial \zeta_u} = [\mathbf{0} \ \cdots \ \tilde{\mathbf{a}}_{a,u} \ \cdots \ \mathbf{0}], \quad (59)$$

in which $\tilde{\mathbf{a}}_{a,u} \triangleq j \mathbf{V}^T \mathbf{D}_1 \mathbf{a}_{\text{IRS}}(\zeta_u, \xi_u)$ and

$$\mathbf{D}_1 \triangleq \text{diag}(\underbrace{0, \dots, 0}_{M_z}, \underbrace{1, \dots, 1}_{M_z}, \dots, \underbrace{M_y - 1, \dots, M_y - 1}_{M_z}). \quad (60)$$

Thus we have

$$\begin{aligned} \frac{\partial L(\mathbf{p})}{\partial \zeta_u} &= \mathbf{e}_u^T \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^T \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right)^* \tilde{\mathbf{a}}_{a,u} \\ &+ \mathbf{e}_u^T \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^H \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right) \tilde{\mathbf{a}}_{a,u}^*, \\ &= 2 \text{Re} \left\{ \mathbf{e}_u^T \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^T \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right)^* \tilde{\mathbf{A}}_a \mathbf{e}_u \right\}, \end{aligned} \quad (61)$$

where $\text{Re}\{\cdot\}$ represents the real part of a complex number, $\mathbf{e}_u$ is a unit vector whose u th entry equals to one and all other entries equal to zeros, and $\tilde{\mathbf{A}}_a \triangleq [\tilde{\mathbf{a}}_{a,1} \ \cdots \ \tilde{\mathbf{a}}_{a,U}]$. Similarly, we have

$$\begin{aligned} \frac{\partial L(\mathbf{p})}{\partial \xi_u} &= 2 \text{Re} \left\{ \mathbf{e}_u^T \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^T \left(\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \right)^* \tilde{\mathbf{A}}_e \mathbf{e}_u \right\}, \end{aligned} \quad (62)$$

where $\tilde{\mathbf{A}}_e \triangleq [\tilde{\mathbf{a}}_{e,1} \ \cdots \ \tilde{\mathbf{a}}_{e,U}]$, $\tilde{\mathbf{a}}_{e,u} = j \mathbf{V}^T \mathbf{D}_2 \mathbf{a}_{\text{IRS}}(\zeta_u, \xi_u)$, and

$$\mathbf{D}_2 \triangleq \text{diag}(\underbrace{0, \dots, M_z - 1}_{M_y}, \underbrace{0, \dots, M_z - 1}_{M_y}, \dots, \underbrace{0, \dots, M_z - 1}_{M_y}). \quad (63)$$

The calculations of $\frac{\partial L(\mathbf{p})}{\partial \phi_u}$, $\frac{\partial L(\mathbf{p})}{\partial \theta_u}$, $\frac{\partial L(\mathbf{p})}{\partial \varrho_u}$ and $\frac{\partial L(\mathbf{p})}{\partial \nu_u}$ are similar, which is omitted here.

B. Calculation of Fisher Information Matrix $\Omega(\mathbf{p})$

We first calculate the entries in the principal minors of $\Omega(\mathbf{p})$. For instance, the (u_1, u_2) th entry of $\mathbb{E}\left\{\left(\frac{\partial L(\mathbf{p})}{\partial \zeta}\right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \zeta}\right)\right\}$ is given by

$$\begin{aligned} &\mathbb{E} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \zeta_{u_1}} \right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \zeta_{u_2}} \right) \right\} \\ &= 4 \mathbb{E} \left\{ \text{Re} \left\{ \mathbf{e}_{u_1}^T \mathbf{W}_1 \mathbf{e}_{u_2} \right\} \text{Re} \left\{ \mathbf{e}_{u_2}^T \mathbf{W}_1 \mathbf{e}_{u_1} \right\} \right\}, \\ &= \mathbb{E} \left\{ (\mathbf{W}_1(u_1, u_1) + \mathbf{W}_1(u_1, u_1)^*) \right. \\ &\quad \left. (\mathbf{W}_1(u_2, u_2) + \mathbf{W}_1(u_2, u_2)^*) \right\}, \end{aligned} \quad (64)$$

where $\mathbf{W}_1 \triangleq \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^T (\mathbf{Y}_{(1)}^T - (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T)^* \tilde{\mathbf{A}}_a \triangleq \frac{1}{\sigma^2} (\mathbf{C} \odot \mathbf{B})^T (\mathbf{N}_{(1)}^H) \tilde{\mathbf{A}}_a$. Let $\mathbf{w}_1 = \text{vec}(\mathbf{W}_1)$. We have

$$\mathbf{w}_1 = \frac{1}{\sigma^2} \left(\tilde{\mathbf{A}}_a^T \otimes (\mathbf{C} \odot \mathbf{B})^T \right) \text{vec}(\mathbf{N}_{(1)}^H), \quad (65)$$

where $\mathbf{N}_{(1)}$ is the mode-1 unfolding of $\mathcal{N}$, $\text{vec}(\mathbf{N}_{(1)}^H) \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$. Since $\mathbf{w}_1$ is a linear transformation of $\text{vec}(\mathbf{N}_{(1)}^H)$, it also follows a circularly symmetric complex Gaussian distribution. Its covariance matrix $\mathbf{C}_{\mathbf{w}_1} \in \mathbb{C}^{U^2 \times U^2}$ and second-order moments $\mathbf{M}_{\mathbf{w}_1} \in \mathbb{C}^{U^2 \times U^2}$ are respectively given by

$$\begin{aligned} \mathbf{C}_{\mathbf{w}_1} &= \mathbb{E} \{ \mathbf{w}_1 \mathbf{w}_1^H \}, \\ &= \left(\frac{1}{\sigma^2} \right)^2 \left(\tilde{\mathbf{A}}_a^T \otimes (\mathbf{C} \odot \mathbf{B})^T \right) \\ &\quad \times \mathbb{E} \left\{ \text{vec}(\mathbf{N}_{(1)}^H) \text{vec}(\mathbf{N}_{(1)}^H)^H \right\} \left(\tilde{\mathbf{A}}_a^T \otimes (\mathbf{C} \odot \mathbf{B})^T \right)^H, \\ &= \frac{1}{\sigma^2} \left(\tilde{\mathbf{A}}_a^T \tilde{\mathbf{A}}_a^* \right) \otimes \left((\mathbf{C} \odot \mathbf{B})^T (\mathbf{C} \odot \mathbf{B})^* \right), \end{aligned} \quad (66)$$

and

$$\mathbf{M}_{\mathbf{w}_1} = \mathbb{E} \{ \mathbf{w}_1 \mathbf{w}_1^T \} = \mathbf{0}. \quad (67)$$

Thus we have

$$\mathbb{E} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \zeta_{u_1}} \right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \zeta_{u_2}} \right) \right\} = 2 \text{Re} \{ \mathbf{C}_{\mathbf{w}_1}(m, n) \}, \quad (68)$$

where $m \triangleq U(u_1 - 1) + u_1$, $n \triangleq U(u_2 - 1) + u_2$. Similarly, we have

$$\mathbb{E} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \xi_{u_1}} \right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \xi_{u_2}} \right) \right\} = 2 \text{Re} \{ \mathbf{C}_{\mathbf{w}_2}(m, n) \}, \quad (69)$$

where $\mathbf{C}_{\mathbf{w}_2} = \frac{1}{\sigma^2} (\tilde{\mathbf{A}}_e^T \tilde{\mathbf{A}}_e^*) \otimes ((\mathbf{C} \odot \mathbf{B})^T (\mathbf{C} \odot \mathbf{B})^*)$. The derivations of other entries in the principal minors of $\Omega(\mathbf{p})$ are similar and thus omitted here.

For elements in the off-principal minors of $\Omega(\mathbf{p})$, such as the (l_1, l_2) th entry of $\mathbb{E}\left\{\left(\frac{\partial L(\mathbf{p})}{\partial \zeta}\right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \xi}\right)\right\}$, we have

$$\mathbb{E} \left\{ \left(\frac{\partial L(\mathbf{p})}{\partial \zeta_{u_1}} \right)^H \left(\frac{\partial L(\mathbf{p})}{\partial \xi_{u_2}} \right) \right\} = 2 \text{Re} \{ \mathbf{C}_{\mathbf{w}_1, \mathbf{w}_2}(m, n) \}, \quad (70)$$

where

$$\begin{aligned} \mathbf{C}_{\mathbf{w}_1, \mathbf{w}_2} &= \mathbb{E} \{ \mathbf{w}_1 \mathbf{w}_2^H \}, \\ &= \left(\frac{1}{\sigma^2} \right)^2 \left(\tilde{\mathbf{A}}_a^T \otimes (\mathbf{C} \odot \mathbf{B})^T \right) \mathbb{E} \left\{ \text{vec}(\mathbf{N}_{(1)}^H) \right. \\ &\quad \left. \times \text{vec}(\mathbf{N}_{(1)}^H)^H \right\} \left(\tilde{\mathbf{A}}_b^T \otimes (\mathbf{C} \odot \mathbf{B})^T \right)^H, \\ &= \frac{1}{\sigma^2} \left(\tilde{\mathbf{A}}_a^T \tilde{\mathbf{A}}_b^* \right) \otimes \left((\mathbf{C} \odot \mathbf{B})^T (\mathbf{C} \odot \mathbf{B})^* \right). \end{aligned} \quad (71)$$

Other entries in the off-principal minors of $\Omega(\mathbf{p})$ can be similarly calculated.

C. Cramér-Rao Bound

After obtaining the FIM, the CRB for the parameters $\mathbf{p}$ can be calculated as

$$\text{CRB}(\mathbf{p}) = \Omega^{-1}(\mathbf{p}). \quad (72)$$

REFERENCES

- [1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [2] S. Sun, T. S. Rappaport, M. Shafi, P. Tang, J. Zhang, and P. J. Smith, "Propagation models and performance evaluation for 5G millimeter-wave bands," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8422–8439, Sep. 2018.
- [3] M. Majumder, H. Saxena, S. Srivastava, and A. K. Jagannatham, "Optimal bit allocation-based hybrid precoder-combiner design techniques for mmWave MIMO-OFDM systems," *IEEE Access*, vol. 9, pp. 54109–54125, 2021.
- [4] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse, group-sparse, and online bayesian learning aided channel estimation for doubly-selective mmWave hybrid MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5843–5858, Sep. 2021.
- [5] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [6] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.
- [7] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.
- [8] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 270–278.
- [9] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [10] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.
- [11] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.
- [12] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020.
- [13] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Process. Lett.*, vol. 28, pp. 513–517, 2021.
- [14] S. Liu, Z. Gao, J. Zhang, M. Di Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmwave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.
- [15] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.
- [16] G. T. de Araújo, A. L. De Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, Apr. 2021.
- [17] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Channel estimation and user localization for IRS-assisted MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2320–2335, Apr. 2022.
- [18] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based algebraic channel estimation for hybrid IRS-assisted MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3770–3784, Jun. 2021.
- [19] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.
- [20] J. Tang, Z. Peng, D. K. C. So, X. Zhang, K.-K. Wong, and J. A. Chambers, "Energy efficiency optimization for a multiuser IRS-aided MISO system with SWIPT," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5950–5962, Oct. 2023.
- [21] W. Yan, X. Yuan, and X. Cao, "Frequency reflection modulation for reconfigurable intelligent surface aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9381–9393, Nov. 2022.
- [22] Y. Wang, W. Shi, M. Huang, F. Shu, and J. Wang, "Intelligent reflecting surface aided secure transmission with colluding eavesdroppers," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10155–10160, Sep. 2022.
- [23] D. Zhao, H. Lu, Y. Wang, H. Sun, and Y. Gui, "Joint power allocation and user association optimization for IRS-assisted mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 577–590, Jan. 2022.
- [24] F. Shu et al., "Enhanced secrecy rate maximization for directional modulation networks via IRS," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8388–8401, Dec. 2021.
- [25] X. Zhao, K. Xu, S. Ma, S. Gong, G. Yang, and C. Xing, "Joint transceiver optimization for IRS-aided MIMO communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3467–3482, May 2022.
- [26] W. Jiang, B. Chen, J. Zhao, Z. Xiong, and Z. Ding, "Joint active and passive beamforming design for the IRS-assisted MIMOME-OFDM secure communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10369–10381, Oct. 2021.
- [27] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.
- [28] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted MIMO: A sum-path-gain maximization approach," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1105–1109, Jul. 2020.
- [29] S. H. Hong, J. Park, S.-J. Kim, and J. Choi, "Hybrid beamforming for intelligent reflecting surface aided millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7343–7357, Sep. 2022.
- [30] X. Zheng, P. Wang, J. Fang, and H. Li, "Compressed channel estimation for IRS-assisted millimeter wave OFDM systems: A low-rank tensor decomposition-based approach," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1258–1262, Jun. 2022.
- [31] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [32] P.-A. Absil, R. Mahony, and R. Sepulchre, "Optimization Algorithms on Matrix Manifolds," in *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [33] K. Liu, J. P. C. da Costa, H. C. So, L. Huang, and J. Ye, "Detection of number of components in CANDECOMP/PARAFAC models via minimum description length," *Digit. Signal Process.*, vol. 51, pp. 110–123, 2016.
- [34] 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.
- [35] J. Chen, "When does asymptotic orthogonality exist for very large arrays?," in *Proc. IEEE Glob. Commun. Conf.*, 2013, pp. 4146–4150.
- [36] H. Kasai, "Fast optimization algorithm on complex oblique manifold for hybrid precoding in millimeter wave MIMO systems," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2018, pp. 1266–1270.
- [37] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, Feb. 2021.
- [38] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.



Xi Zheng received the B.Eng. degree in electronic and information engineering and the M.Eng. degree in information and communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2017 and 2024, respectively. Her research interests include millimeter wave/THz communications, signal processing, and intelligent reflecting surface.



Jun Fang (Senior Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in electrical engineering from the Xidian University, Xi'an, China in 1998 and 2001, respectively, and the Ph.D. degree in electrical engineering from the National University of Singapore, Singapore, in 2006. In 2006, he was a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering,

Stevens Institute of Technology, Hoboken, NJ, USA. Since 2011, he has been with the University of Electronic Science and Technology of China, where he is currently a Professor. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference. Dr. Fang was the recipient of the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was an Associate Technical Editor for *IEEE Communications Magazine* during 2012–2020, and a Senior Associate Editor for IEEE SIGNAL PROCESSING LETTERS during 2018–2022. He is currently a Member of the IEEE SPS Sensor Array and Multichannel TC.



Hongbin Li (Fellow, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1991 and 1994, respectively, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 1999. From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently the Charles and Rosanna Batchelor Memorial Chair Professor. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory in the summers of 2003, 2004, and 2009. His research interests include statistical signal processing, wireless communications, and radars. Dr. Li was the recipient of the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Society, Outstanding Paper Award from the IEEE AFICON Conference in 2011, Provost's Award for Research Excellence in 2019, Harvey N. Davis Teaching Award in 2003, and Jess H. Davis Memorial Research Award in 2001 from Stevens Institute of Technology, and Sigma Xi Graduate Research Award from the University of Florida in 1999. He is a Member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for *Signal Processing* (Elsevier), IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as the General Co-Chair of the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, June 17–20, 2012. He is a Member of Tau Beta Pi and Phi Kappa Phi.

stitute of Technology, Hoboken, NJ, USA, where he is currently the Charles and Rosanna Batchelor Memorial Chair Professor. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory in the summers of 2003, 2004, and 2009. His research interests include statistical signal processing, wireless communications, and radars. Dr. Li was the recipient of the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Society, Outstanding Paper Award from the IEEE AFICON Conference in 2011, Provost's Award for Research Excellence in 2019, Harvey N. Davis Teaching Award in 2003, and Jess H. Davis Memorial Research Award in 2001 from Stevens Institute of Technology, and Sigma Xi Graduate Research Award from the University of Florida in 1999. He is a Member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for *Signal Processing* (Elsevier), IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as the General Co-Chair of the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, June 17–20, 2012. He is a Member of Tau Beta Pi and Phi Kappa Phi.



Hongwei Wang (Member, IEEE) received the B.S. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2019, respectively. He is currently a Postdoctoral Researcher with the University of Electronic Science and Technology of China, Chengdu, China. His research interests include statistical signal processing, compressed sensing and sparse theory, millimeter wave communications, and wideband spectrum sensing.



Peilan Wang received the B.Eng. and the Ph.D. degree from the University of Electronic Science and Technology, Chengdu, China, in 2018 and 2023, respectively. From 2022 to 2023, she was a Visiting Research Scholar with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. Her research interests include compressed sensing, millimeter wave/THz communications, intelligent reflecting surface, signal processing, and integrated sensing and communications.