

Spatial Channel Covariance Estimation for Hybrid Architectures Based on Tensor Decompositions

Sungwoo Park, *Student Member, IEEE*, Anum Ali, *Member, IEEE*, Nuria González-Prelcic, *Senior Member, IEEE*, and Robert W. Heath Jr., *Fellow, IEEE*

Abstract—Spatial channel covariance information can replace full instantaneous channel state information for the analog precoder design in hybrid analog/digital architectures. Obtaining spatial channel covariance estimation, however, is challenging in the hybrid structure due to the use of fewer radio frequency (RF) chains than the number of antennas. In this paper, we propose a spatial channel covariance estimation method based on higher-order tensor decomposition for spatially sparse time-varying frequency-selective channels. The proposed method leverages the fact that the channel can be represented as a low-rank higher-order tensor. We also derive the Cramér-Rao lower bound on the estimation accuracy of the proposed method. Numerical results and theoretical analysis show that the proposed tensor-based approach achieves higher estimation accuracy in comparison with prior compressive-sensing-based approaches or conventional angle-of-arrival estimation approaches.

I. INTRODUCTION

Hybrid analog/digital precoding uses a smaller number of RF chains to reduce the number of power-consuming devices like analog-to-digital converters (ADCs) or digital-to-analog converters (DACs). Consequently, the hybrid approach can reduce power consumption and implementation complexity in millimeter wave multiple-input-multiple-output (MIMO) systems [1]–[4] and massive MIMO systems [5]–[7]. The rate loss incurred by the hybrid architecture is insignificant for spatially sparse channels such as in millimeter wave systems or in suburban/rural areas in sub-6 GHz systems [1]–[4], [7].

A main challenge in the hybrid architecture is to configure the analog and digital precoding stages. Many previous methods accomplish this task based on full CSI [1], [4], [8]. These approaches require frequent estimation of the channel, obtained for example via sparse recovery techniques. An alternative is to use only long-term statistical knowledge such as that contained in spatial channel covariance matrices, for the analog precoder design [7], [9]–[11]. Once the analog precoder is determined based on the spatial channel covariance, the digital precoder, of a much smaller dimension is designed by using instantaneous full CSI of the low-dimensional effective channel, i.e., the propagation channel combined with the analog precoder. While the accurate estimation of the full CSI

is difficult for time-varying frequency-selective channels, the long-term statistical CSI can be efficiently estimated. It was shown in [7], [9]–[11] that the hybrid precoding methods based on spatial channel covariance achieve spectral efficiency close to that of the hybrid precoding obtained from full CSI when the channels are spatially sparse.

Although the use of the spatial channel covariance matrix helps the hybrid precoding design to be simpler and more practical, the hybrid architecture makes it difficult to estimate the covariance matrix. Since there is no digital access to the outputs of every antenna, and only the signals combined in an analog way are available at baseband, it is difficult to estimate the spatial covariance of the high dimension channel. Different approaches have been suggested to solve the spatial channel covariance estimation problem under such an environment. In [6], a least-squares-based covariance estimation method was proposed by using time-varying analog beamforming matrices. Since the method does not exploit the sparse channel property, it is not an efficient method for sparse channels, which are of our interest in this paper. The sparse ruler array in [12]–[14] and the coprime sensor array [15] can omit measurements on some antenna elements by leveraging the fact that correlations between antenna elements are wide-sense stationary in spatial domain. Although these so-called compressive covariance sensing (CCS) methods can reduce the number of RF chains, the methods have a limitation on the configuration of the number of RF chains and antennas.

The CCS methods were initially developed by using only a subset of antennas. It is, however, possible to extend the work to general hybrid architecture where the analog part is composed of phase shifters and thus is represented as a dense matrix [14]. In this dense sensing matrix case, the CCS methods become closely related to typical compressive sensing (CS) techniques. For example, in [11], the spatial channel estimation method was developed by adopting the time-varying analog combiners used in [6], which are dense matrices. Instead of the least-squares method, one of the well-known conventional vector-type CS techniques, orthogonal matching pursuit (OMP), was adopted to exploit the sparse channel property. It is worthwhile to note that conventional vector-type CS techniques were typically developed for channel estimation [16]–[18] but can be extended to spatial channel covariance estimation as well. The vector-type CS techniques for covariance estimation, however, are outperformed by matrix-type CS techniques developed for so-called multiple measurement vector (MMV) problems [19]–[21] that enable the joint sparse recovery. An advanced spatial channel covariance estimation

S. Park, N. González-Prelcic, and R. W. Heath Jr. are with the Wireless Networking and Communication Group (WNCG), Department of Electrical and Computer Engineering, The University of Texas at Austin, TX, 78701 USA. (e-mail: {swpark96,ngprelcic,rheath}@utexas.edu).

Anum Ali was with The University of Texas at Austin, Austin TX, 78701. He is now with Samsung Research America, Plano, TX, USA (email: anum.ali@samsung.com).

This work is supported in part by the National Science Foundation under Grant No. 1514275, and by a gift from Huawei Technologies.

method based on the MMV approach was proposed in [22] by applying time-varying sensing matrices and exploiting the Hermitian property of the covariance matrix. The subspace estimation work based on MMV in [23], [24] was also proposed for the hybrid architecture. It was shown that the proposed MMV based methods have performance close to approximate maximum-likelihood (AML). The estimation methods in [22]–[24], however, assumed narrowband channels, i.e., the frequency-selective wideband channels were not considered.

Besides the CS-based work, conventional AoA estimation methods such as the multiple signal classification (MUSIC) [25] and the estimation of signal parameters via rotational invariance technique (ESPRIT) [26] can also be applied to the spatial channel covariance estimation problem via some modification. In the conventional AoA estimation methods for fully-digital architectures, the spatial channel covariance is directly calculated from the received signal vectors, and the AoA is estimated from the obtained covariance matrix. A covariance estimation based on AoA estimation for the hybrid architecture requires the estimation process in the opposite direction. First, the covariance matrix of the low-dimensional baseband received signal is calculated. Second, the AoAs are estimated from the covariance of the baseband received signal vectors. Finally, the covariance of the actual channel is reconstructed from the estimated AoAs. This basic approach has been adopted for different scenarios with some variations [27], [28]. This approach, however, has a weak point: the estimation accuracy rapidly decreases as the number of channel paths increases toward the number of RF chains. In addition, the methods based on this approach do not work properly when the number of channel paths exceeds that of RF chains.

In [29], higher-order tensor decomposition techniques were applied to channel estimation for millimeter wave MIMO-OFDM systems. Leveraging the fact that the frequency selective channel in MIMO-OFDM systems can be represented as a higher-order tensor, the canonical polyadic decomposition (CPD), which is called the CANDECOMP/PARAFAC [30], [31], was adopted to estimate the sparse channel parameters. The method, however, does not fully exploit multiple RF chains and requires that the channel must be fixed during the estimation process that takes many subframes/frames in the time domain. In [32], [33], another type of higher-order decomposition called the higher-order SVD (HOSVD), was applied to the subspace estimation problem. The methods, however, did not consider hybrid architecture or frequency-selective wideband channels.

In this paper, we propose a spatial channel covariance estimation method for the hybrid architecture over uplink time-varying frequency-selective channels. We consider a time division duplex (TDD) system where the estimated covariance over uplink channels can be used for the downlink precoding design at a base station (BS). We represent the channel and the received baseband signal as higher-order tensors. Considering spatially sparse channels, we use the fact that these higher-order tensors have a low tensor rank and their CPD forms are unique up to a common permutation and scaling of columns under some mild conditions [34]–[36]. We also analyze the

theoretical performance by adopting the performance metric in [22], [23] that is associated with the dominant eigenvalues and their eigenspaces of the spatial channel covariance matrix. We will call this performance metric the relative precoding efficiency (RPE) in this paper. After showing that the RPE is closely related to the mean squared error (MSE) of the AoA estimation, we derive Cramér-Rao lower bound (CRLB) for the AoA estimation and its associated bound for the performance metric. Using numerical results, we first show that the performance of the proposed method approaches the performance bound as SNR increases. We also show that the lower bound of the tensor-based method is lower than that of the MUSIC-based method, which provides insight into the benefits of using tensor-based methods. Simulations show that the proposed tensor-based method is promising compared to CS-based methods and MUSIC-based methods.

The rest of the paper is organized as follows. Section II briefly introduces the basics of high-order tensor algebra. Section III provides a system and channel model by using tensor representations. Section IV describes the proposed spatial channel covariance method. The performance metric is analyzed in Section V, and the CRLB is derived in Section VI. In Section VII, the proposed tensor-based work is compared with prior work based on CS or MUSIC. Simulation results are presented in Section VIII, and conclusions are drawn in Section IX. This paper is the journal version of [37] with theoretical analysis added.

Notation: We use the following notation throughout this paper: a is a scalar, $\mathbf{a}$ is a vector, $\mathbf{A}$ is a matrix, and $\mathcal{A}$ is a tensor. $\mathbf{A}^T$, $\mathbf{A}^C$, $\mathbf{A}^*$, and $\mathbf{A}^\dagger$ are transpose, conjugate, conjugate transpose, and Moore-Penrose pseudoinverse. $[\mathbf{A}]_{i,:}$ and $[\mathbf{A}]_{:,j}$ are the i -th row and the j -th column of the matrix $\mathbf{A}$. $\mathbf{A} \otimes \mathbf{B}$, $\mathbf{A} \odot \mathbf{B}$, and $\mathbf{A} \circ \mathbf{B}$ denote the Kronecker product, the Hadamard product, and the column-wise Khatri-Rao product. $\mathbf{a} \circ \mathbf{b}$ denotes the outer product, which is also known as the tensor product. $\text{Re}(\mathcal{A})$ and $\text{Im}(\mathcal{A})$ denote the real part and the imaginary part of $\mathcal{A}$. $\text{diag}(\mathbf{A})$ is a column vector whose elements are composed of the diagonal elements of $\mathbf{A}$.

II. PRELIMINARIES: OVERVIEW OF TENSOR ALGEBRA AND CANONICAL POLYADIC DECOMPOSITION

In this section, we review the basics of tensor algebra that will be used in this paper. Readers who are interested in more details about tensors can refer to [34]–[36] and the references therein. A tensor denotes a multi-dimensional (a.k.a. multi-way or multi-mode) array. The *order* of a tensor is defined as the number of dimensions of the tensor. A vector and a matrix are special cases of a tensor, i.e., a vector is a tensor of order one, and a matrix is a tensor of order two.

Given an N -th order tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, let its $(i_1, i_2, \dots, i_N)$ -th element be denoted by $x_{i_1 i_2 \dots i_N} = \mathcal{X}(i_1, i_2, \dots, i_N)$. The *mode- n fibers* of $\mathcal{X}$ are defined as vector-valued sub-tensors obtained by fixing all but one index associated with mode- n , i.e., $\mathcal{X}(i_1, \dots, i_{n-1}, :, i_{n+1}, \dots, i_N)$. The number of mode- n fibers in $\mathcal{X}$ is $\prod_{m=1, m \neq n}^N I_m$.

The *mode- n matricization* (a.k.a. *unfolding*) is a process that transforms a tensor into a matrix whose columns are composed

of mode- n fibers of the tensor. The *mode- n unfolding matrix* of $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $\mathbf{X}_{(n)} \in \mathbb{C}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N}$. The tensor $\mathcal{X}(i_1, i_2, \dots, i_N)$ maps to $\mathbf{X}_{(n)}(i_n, j)$ such that $j = 1 + \sum_{k=1, k \neq n}^N ((i_k - 1) \prod_{m=1, m \neq n}^{k-1} I_m)$.

The *mode- n product* of a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and a matrix $\mathbf{A} \in \mathbb{C}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n \mathbf{A}$. Let $\mathcal{Y} = \mathcal{X} \times_n \mathbf{A}$. Then, the elements of the tensor $\mathcal{Y} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$ are given by $y_{i_1 i_2 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} a_{j i_n}$. The mode- n product representation $\mathcal{Y} = \mathcal{X} \times_n \mathbf{A}$ can also be expressed by using the mode- n matricization as

$$\mathbf{Y}_{(n)} = \mathbf{A} \mathbf{X}_{(n)}. \quad (1)$$

Given a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and matrices $\mathbf{A}^{(n)} \in \mathbb{C}^{J_n \times I_n}$ for $n = 1, \dots, N$, their *full multilinear product* is defined as

$$\llbracket \mathcal{X}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket = \mathcal{X} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_N \mathbf{A}^{(N)}. \quad (2)$$

For a special case where $I_1 = \dots = I_N = R$ and $\mathcal{X}$ is a diagonal tensor $\mathcal{I} \in \mathbb{C}^{R \times R \times \dots \times R}$ that has zero off-diagonal elements and unit diagonal elements, there exists a simplified notation of the full multilinear product as

$$\llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_N \mathbf{A}^{(N)}. \quad (3)$$

The *norm* of a tensor is defined as

$$\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} |x_{i_1 i_2 \dots i_N}|^2}, \quad (4)$$

which is analogous to the Frobenius norm in the matrix case.

Let $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and $\mathcal{Y} \in \mathbb{C}^{J_1 \times J_2 \times \dots \times J_M}$. Then, the *outer product* (a.k.a. *tensor product*) of $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $\mathcal{X} \circ \mathcal{Y}$. Let $\mathcal{Z} = \mathcal{X} \circ \mathcal{Y}$. Then, the elements of the tensor $\mathcal{Z} \in \mathbb{C}^{I_1 \times \dots \times I_N \times J_1 \times \dots \times J_M}$ are given by $z_{i_1 \dots i_N j_1 \dots j_M} = x_{i_1 \dots i_N} y_{j_1 \dots j_M}$, $\forall i_1, \dots, i_N, j_1, \dots, j_M$.

A tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is called a *rank-one tensor* if it can be written as the outer product of vectors as

$$\mathcal{X} = \mathbf{x}^{(1)} \circ \mathbf{x}^{(2)} \circ \dots \circ \mathbf{x}^{(N)}, \quad (5)$$

where $\mathbf{x}^{(n)} \in \mathbb{C}^{I_n \times 1}$, $\forall n$.

The canonical polyadic decomposition (CPD), which is also known as CANDECOMP/PARAFAC decomposition, factorizes a tensor into a sum of component rank-one tensors. The CPD of $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ has a form

$$\mathcal{X} = \sum_{r=1}^R \mathbf{x}_r^{(1)} \circ \mathbf{x}_r^{(2)} \circ \dots \circ \mathbf{x}_r^{(N)}, \quad (6)$$

where $\mathbf{x}_r^{(n)} \in \mathbb{C}^{I_n \times 1}$ for $r = 1, \dots, R$. The minimum possible value of the number of rank-one tensors that constitute $\mathcal{X}$, which is R in (6), is called the *rank* of $\mathcal{X}$.

III. CHANNEL MODEL AND SYSTEM MODEL

Consider a TDD system where a base station with N_{ant} antennas and $M_{\text{RF}} (\leq N_{\text{ant}})$ RF chains communicates with a mobile station that has a single antenna.

A. Channel model

We consider a spatially sparse channel that has L_{ch} paths between the BS and mobile user. Let τ_ℓ and ϕ_ℓ denote the path

delay and AoA of the ℓ^{th} path. Let $g_{t,\ell}$ denote the short-term fading complex path gain of the ℓ^{th} path at frame t . Let $p_{\text{PS}}(\tau)$ denote the low pass filter including pulse shaping and analog filters. We assume a uniform linear array (ULA) with antenna element spacing d_a and signal wavelength λ . It is possible to extend the proposed method to a uniform planar array (UPA). The array response vector associated with the ℓ^{th} AoA ϕ_ℓ is expressed as

$$\mathbf{a}(\phi_\ell) = \left[1 \quad e^{\frac{j2\pi d_a \sin(\phi_\ell)}{\lambda}} \quad \dots \quad e^{\frac{j2\pi d_a (N_{\text{ant}}-1) \sin(\phi_\ell)}{\lambda}} \right]^T. \quad (7)$$

Let T_s and N_{CP} be the sampling duration and the cyclic prefix length. We assume that the large-scale fading parameters, τ_ℓ 's and ϕ_ℓ 's, are constant during the estimation process while the channel path gains of different paths $g_{t,\ell}$ at every time t are uncorrelated random variables according to the wide sense stationary uncorrelated scattering (WSSUS) model [24], [38]. By using the delay- d channel model [39]–[41], the uplink channel at frame t can be represented as

$$\mathbf{h}_t[d] = \sum_{\ell=1}^{L_{\text{ch}}} g_{t,\ell} p_{\text{PS}}(dT_s - \tau_\ell) \mathbf{a}(\phi_\ell) \quad \text{for } d = 0, \dots, N_{\text{CP}} - 1. \quad (8)$$

By letting $c_{k,\ell} = \sum_{d=0}^{N_{\text{CP}}-1} p_{\text{PS}}(dT_s - \tau_\ell) e^{-\frac{j2\pi(k-1)d}{K_{\text{schr}}}}$, the channel frequency response vector can be expressed as

$$\mathbf{h}_{t,k} = \sum_{\ell=1}^{L_{\text{ch}}} g_{t,\ell} c_{k,\ell} \mathbf{a}(\phi_\ell), \quad (9)$$

at frame t and subcarrier k .

B. System model

Let $s_{t,k}$ be an uplink training symbol at frame t and subcarrier k with $|s_{t,k}| = 1$, and $\mathbf{z}_{t,k} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ be a circularly symmetric Gaussian noise. The $N_{\text{ant}} \times 1$ received signal vector at each frame and subcarrier can be represented as

$$\mathbf{r}_{t,k} = \mathbf{h}_{t,k} s_{t,k} + \mathbf{z}_{t,k}. \quad (10)$$

Let $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_{\text{ant}} \times M_{\text{RF}}}$ and $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{M_{\text{RF}} \times M_{\text{RF}}}$ be an analog combining matrix and a digital baseband processing matrix. Similar to a sensing matrix in prior CS-based channel estimation work [42], [43], we assume that the elements of $\mathbf{W}_{\text{RF}}$ have random phases with a unit amplitude. Let $\mathbf{W}$ denote the hybrid combining matrix as $\mathbf{W} = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$. After combining with the hybrid combiner and multiplying by $s_{t,k}^*$, the $M_{\text{RF}} \times 1$ baseband received signal vector becomes

$$\mathbf{y}_{t,k} = s_{t,k}^* \mathbf{W}^* \mathbf{r}_{t,k} = \mathbf{x}_{t,k} + \mathbf{n}_{t,k}, \quad (11)$$

where $\mathbf{x}_{t,k} = \mathbf{W}^* \mathbf{h}_{t,k}$ denotes the signal part of $\mathbf{y}_{t,k}$, and $\mathbf{n}_{t,k} = s_{t,k}^* \mathbf{W}^* \mathbf{z}_{t,k} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{W}^* \mathbf{W})$ denotes the noise part. From the viewpoint of the estimator at baseband, the effective noise $\mathbf{n}_{t,k}$ becomes colored for an arbitrary hybrid combiner $\mathbf{W}$. It is possible to whiten the effective noise by using the baseband combiner $\mathbf{W}_{\text{BB}} = (\mathbf{W}_{\text{RF}}^* \mathbf{W}_{\text{RF}})^{-\frac{1}{2}}$. With this choice of $\mathbf{W}_{\text{BB}}$, the hybrid combiner $\mathbf{W}$ satisfies $\mathbf{W}^* \mathbf{W} = \mathbf{I}$ for any $\mathbf{W}_{\text{RF}}$. We will use this unitary hybrid combiner $\mathbf{W}$ throughout this paper.

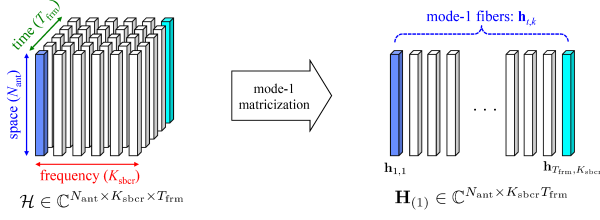


Fig. 1. Tensor representation of a time-varying frequency-selective channel: the left figure shows the third-order tensor representation $\mathcal{H} \in \mathbb{C}^{N_{\text{ant}} \times K_{\text{sbc}} \times T_{\text{frm}}}$, and the right figure shows its mode-1 unfolding matrix $\mathbf{H}_{(1)} \in \mathbb{C}^{N_{\text{ant}} \times K_{\text{sbc}} T_{\text{frm}}}$.

C. Tensor representation of channels and received signals

In this subsection, we show that the time-varying frequency-selective channel in (9) can be represented as a low-rank third-order tensor. Let $\mathbf{a}_\ell = \mathbf{a}(\phi_\ell)$, $\mathbf{c}_\ell = [c_{1,\ell} \ \cdots \ c_{K_{\text{sbc}},\ell}]^\top$, and $\mathbf{g}_\ell = [g_{1,\ell} \ \cdots \ g_{T_{\text{frm}},\ell}]^\top$ for $\ell = 1, \dots, L_{\text{ch}}$. With $\mathbf{a}_\ell$, $\mathbf{c}_\ell$, and $\mathbf{g}_\ell$, let us define $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{G}$ as $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{L_{\text{ch}}}]$, $\mathbf{C} = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_{L_{\text{ch}}}]$, and $\mathbf{G} = [\mathbf{g}_1 \ \cdots \ \mathbf{g}_{L_{\text{ch}}}]$.

The $K_{\text{sbc}} T_{\text{frm}}$ channel frequency response vectors $\mathbf{h}_{t,k} \in \mathbb{C}^{N_{\text{ant}} \times 1}$ for $t = 1, \dots, T_{\text{frm}}$ and $k = 1, \dots, K_{\text{sbc}}$ in (9) can be regarded as $T_{\text{frm}} K_{\text{sbc}}$ mode-1 fibers in a third-order tensor $\mathcal{H} \in \mathbb{C}^{N_{\text{ant}} \times K_{\text{sbc}} \times T_{\text{frm}}}$ as shown in Fig. 1. In CPD form, the rank- L_{ch} third-order tensor is

$$\mathcal{H} = \llbracket \mathbf{A}, \mathbf{C}, \mathbf{G} \rrbracket = \mathcal{I} \times_1 \mathbf{A} \times_2 \mathbf{C} \times_3 \mathbf{G} = \sum_{\ell=1}^{L_{\text{ch}}} \mathbf{a}_\ell \circ \mathbf{c}_\ell \circ \mathbf{g}_\ell. \quad (12)$$

In this CPD form in (12), the channel tensor $\mathcal{H}$ is factorized into the three matrices, $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{G}$, which are called factor matrices. The mode-1 factor matrix $\mathbf{A}$ is associated with antennas in the space domain, the mode-2 factor matrix $\mathbf{C}$ with subcarriers in the frequency domain, and the mode-3 factor matrix $\mathbf{G}$ with frames in the time domain.

From the channel frequency response tensor model, the received signal at baseband, $\mathbf{y}_{t,k}$, $\forall t, k$ in (11) can also be represented as a third-order tensor as

$$\mathcal{Y} = \mathcal{X} + \mathcal{N}, \quad (13)$$

where $\mathcal{N}$ is the noise tensor whose mode-1 fibers are IID Gaussian vectors with $CN(\mathbf{0}, \sigma^2 \mathbf{I})$. The signal tensor $\mathcal{X}$ is given by

$$\mathcal{X} = \mathcal{H} \times_1 \mathbf{W}^* \in \mathbb{C}^{M_{\text{RF}} \times K_{\text{sbc}} \times T_{\text{frm}}}. \quad (14)$$

The tensor representation in (14) can also be expressed by using a mode-1 matricization as $\mathbf{X}_{(1)} = \mathbf{W}^* \mathbf{H}_{(1)}$ where $\mathbf{H}_{(1)} = \mathbf{A}(\mathbf{G} \odot \mathbf{C})^\top$ is the mode-1 matricization of $\mathcal{H}$ shown in Fig. 1. Let $\mathbf{B} = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_{L_{\text{ch}}}] = \mathbf{W}^* \mathbf{A} \in \mathbb{C}^{M_{\text{RF}} \times L_{\text{ch}}}$, which can be regarded as the effective array response matrices from the viewpoint of the baseband estimator. The CPD of $\mathcal{X}$ is given by

$$\mathcal{X} = \llbracket \mathbf{B}, \mathbf{C}, \mathbf{G} \rrbracket. \quad (15)$$

Note that both the time-varying frequency-selective channel $\mathcal{H}$ and its associated baseband received signal part $\mathcal{X}$ are third-order tensors with rank L_{ch} .

IV. SPATIAL CHANNEL COVARIANCE ESTIMATION BASED ON TENSOR DECOMPOSITION

The spatial channel covariance matrix can be estimated from the sample covariance of $K_{\text{sbc}} T_{\text{frm}}$ mode-1 fibers in $\mathcal{H}$. Since we assume that the elements in $\mathcal{H}$ have zero mean, the sample covariance of the mode-1 fibers becomes

$$\begin{aligned} \mathbf{R}_h &= \frac{1}{K_{\text{sbc}} T_{\text{frm}}} \mathbf{H}_{(1)} \mathbf{H}_{(1)}^* \\ &= \frac{1}{K_{\text{sbc}} T_{\text{frm}}} \mathbf{A} (\mathbf{G}^* \mathbf{G} \odot \mathbf{C}^* \mathbf{C}) \mathbf{A}^*. \end{aligned} \quad (16)$$

The goal of the spatial channel covariance estimation for the hybrid architecture is to calculate $\mathbf{R}_h$ of $\mathcal{H} \in \mathbb{C}^{N_{\text{ant}} \times K_{\text{sbc}} \times T_{\text{frm}}}$ from the baseband received signal $\mathcal{Y} \in \mathbb{C}^{M_{\text{RF}} \times K_{\text{sbc}} \times T_{\text{frm}}}$, which has smaller dimensions than $\mathcal{H}$. In this section, we propose an estimation method that has three steps. In the first step, the measurement tensor $\mathcal{Y}$ is decomposed into three factor matrices: $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{G}}$ in a CPD form. Note that the three factor matrices obtained in the first step are strongly related to the actual factor matrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$, but are not identical. The relationship between these matrices is further explained in Section IV-A. In the second step, the estimate of the actual factor matrix $\mathbf{A}$ of the channel tensor $\mathcal{H}$ is obtained from $\hat{\mathbf{B}}$, which is denoted by $\hat{\mathbf{A}}$. The spatial channel covariance matrix is calculated from $\hat{\mathbf{A}}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{G}}$ in the last step. Each step is explained in detail in the following subsections.

A. First step: factorization of $\mathcal{Y}$ in a CPD form

If the factor matrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ are given, we can exactly calculate the signal part $\mathcal{X}$ of the measurement tensor $\mathcal{Y}$. The reverse process does not hold in general; the perfect reconstruction of the original $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ from any $\mathcal{X}$ are not guaranteed. There is, however, a special case where the factor matrices can be reconstructed from $\mathcal{X}$. If the tensor rank of a higher-order (more than second-order) tensor is low, its CPD is unique under some mild constraints [35]. The uniqueness of the CPD means that there exists only one possible combination of rank-one tensors that sum to the given tensor subject to two types of indeterminacy: scaling and permutation. The scaling indeterminacy means that the columns in each factor matrix can be scaled arbitrarily, i.e., the CPD form in (15) can be rewritten as

$$\mathcal{X} = \sum_{\ell=1}^{L_{\text{ch}}} (\delta_{\mathbf{b},\ell} \mathbf{b}_\ell) \circ (\delta_{\mathbf{c},\ell} \mathbf{c}_\ell) \circ (\delta_{\mathbf{g},\ell} \mathbf{g}_\ell), \quad (17)$$

as long as $\delta_{\mathbf{b},\ell} \delta_{\mathbf{c},\ell} \delta_{\mathbf{g},\ell} = 1$ for $\ell = 1, \dots, L_{\text{ch}}$. The CPD form in (17) can be expressed in a multilinear product format as $\mathcal{X} = \llbracket \mathbf{B} \Delta_{\mathbf{B}}, \mathbf{C} \Delta_{\mathbf{C}}, \mathbf{G} \Delta_{\mathbf{G}} \rrbracket$, where $\Delta_{\mathbf{B}}$, $\Delta_{\mathbf{C}}$, and $\Delta_{\mathbf{G}}$ are any diagonal matrices satisfying $\Delta_{\mathbf{B}} \Delta_{\mathbf{C}} \Delta_{\mathbf{G}} = \mathbf{I}$. The permutation indeterminacy means that the column vectors in each factor matrix can be reordered with a permutation matrix that is common for all the modes, i.e., the CPD form in (15) can also be represented as $\mathcal{X} = \llbracket \mathbf{B} \mathbf{\Pi}, \mathbf{C} \mathbf{\Pi}, \mathbf{G} \mathbf{\Pi} \rrbracket$ for any permutation matrix $\mathbf{\Pi}$. Considering the indeterminacy, the general form of the CPD of $\mathcal{X}$ becomes

$$\mathcal{X} = \llbracket \mathbf{B} \mathbf{\Pi} \Delta_{\mathbf{B}}, \mathbf{C} \mathbf{\Pi} \Delta_{\mathbf{C}}, \mathbf{G} \mathbf{\Pi} \Delta_{\mathbf{G}} \rrbracket. \quad (18)$$

Let k_A denote the Kruskal-rank of a matrix $\mathbf{A}$, which is defined as the maximum value k such that any k columns in $\mathbf{A}$ are linearly independent. Then, the sufficient condition for the uniqueness of the CPD [34]–[36] is

$$k_B + k_C + k_G \geq 2L_{\text{ch}} + 2. \quad (19)$$

Since the Kruskal-rank can be approximated to be the minimum between the number of rows and columns, the sufficient condition in (19) can be rewritten as

$$L_{\text{ch}} \leq \frac{M_{\text{RF}} + K_{\text{sbc}} + T_{\text{frm}}}{2} + 1, \quad (20)$$

if M_{RF} , K_{sbc} , and T_{frm} are smaller than L_{ch} .

If we set the indeterminacy issue aside, we can exactly reconstruct its factor matrices by leveraging the uniqueness of the CPD. In this subsection, we focus on how to find a CPD solution. We will discuss how to deal with the indeterminacy in the following subsections.

Given a received signal tensor $\mathcal{Y}$, the problem of finding its CPD form is formulated as

$$\{\hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{G}}\} = \arg \min_{\mathbf{B}, \mathbf{C}, \mathbf{G}} \|\mathcal{Y} - [\![\mathbf{B}, \mathbf{C}, \mathbf{G}]\!]\|. \quad (21)$$

There are many known algorithms to solve (21) for CPD. One algorithm to compute the CPD problem in (21) is the alternating least squares (ALS) [34]. By rewriting the objective function in (21) in matrix form as

$$\begin{aligned} \|\mathcal{Y} - [\![\mathbf{B}, \mathbf{C}, \mathbf{G}]\!]\| &= \|\mathbf{Y}_{(1)} - \hat{\mathbf{B}}(\hat{\mathbf{G}} \odot \hat{\mathbf{C}})^T\|_F \\ &= \|\mathbf{Y}_{(2)} - \hat{\mathbf{C}}(\hat{\mathbf{G}} \odot \hat{\mathbf{B}})^T\|_F \\ &= \|\mathbf{Y}_{(3)} - \hat{\mathbf{G}}(\hat{\mathbf{C}} \odot \hat{\mathbf{B}})^T\|_F, \end{aligned} \quad (22)$$

the ALS algorithm first finds the mode-1 factor matrix $\hat{\mathbf{B}}$ assuming that the mode-2 and mode-3 factor matrices, $\hat{\mathbf{C}}$ and $\hat{\mathbf{G}}$ are fixed. This subproblem is formulated as

$$\min_{\mathbf{B}} \|\mathbf{Y}_{(1)} - \hat{\mathbf{B}}(\hat{\mathbf{G}} \odot \hat{\mathbf{C}})^T\|_F. \quad (23)$$

The solution to (23) can be found by using the least squares algorithm as

$$\hat{\mathbf{B}} = \mathbf{Y}_{(1)} \left((\hat{\mathbf{G}} \odot \hat{\mathbf{C}}) (\hat{\mathbf{G}}^* \hat{\mathbf{G}} \odot \hat{\mathbf{C}}^* \hat{\mathbf{C}})^{\dagger} \right)^C. \quad (24)$$

Similar to (23), the mode-2 factor matrix $\hat{\mathbf{C}}$ and the mode-3 factor matrix $\hat{\mathbf{G}}$ can be calculated by fixing other factor matrices except for its own factor matrix. The ALS algorithm iterates the three steps until the objective function converges. The convergence is guaranteed although the converged solution may not be a global optimum.

The solution after convergence provides an estimate of the CPD form of $\mathcal{X}$ as

$$\hat{\mathcal{X}} = [\![\hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{G}}]\!]. \quad (25)$$

Note that $\hat{\mathcal{X}}$ is an estimate of the actual $\mathcal{X}$ in (18). Due to the scaling and permutation indeterminacy, the estimated factor matrices $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{G}}$ obtained in the first step are related to

the actual factor matrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ as

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{B} \mathbf{\Pi} \mathbf{\Delta}_B + \mathbf{\Omega}_B, \\ \hat{\mathbf{C}} &= \mathbf{C} \mathbf{\Pi} \mathbf{\Delta}_C + \mathbf{\Omega}_C, \\ \hat{\mathbf{G}} &= \mathbf{G} \mathbf{\Pi} \mathbf{\Delta}_G + \mathbf{\Omega}_G, \end{aligned} \quad (26)$$

where $\mathbf{\Delta}_B$, $\mathbf{\Delta}_C$, and $\mathbf{\Delta}_G$ are complex-valued diagonal matrices that satisfy $\mathbf{\Delta}_B \mathbf{\Delta}_C \mathbf{\Delta}_G = \mathbf{I}$, and $\mathbf{\Omega}_B$, $\mathbf{\Omega}_C$, and $\mathbf{\Omega}_G$ denote the estimation errors caused by CPD.

We assumed that the tensor rank, which is the number of channel paths L_{ch} , is known to the estimator. There are some known methods to estimate the tensor rank (see [29] and references therein), which can be used in our application. In practice, the actual number of channel paths may not be so small because the channel can be composed of multiple clusters each with multiple subpaths. Even in this case, the tensor can be approximated by a low-rank tensor. We showed in Section VIII-B that the proposed method works well even when we use the number of RF chains as the tensor rank instead of the actual number of channel paths.

B. Second step: estimation of $\mathbf{A}$ of the channel tensor $\mathcal{H}$

The goal of the second step is to estimate $\mathbf{A} \mathbf{\Pi}$ from $\hat{\mathbf{B}}$ that is obtained in the first step as described in Section IV-A. Let $\hat{\mathbf{A}} = \mathbf{A} \mathbf{\Pi}$. We will show in Section IV-C that we do not need to obtain $\mathbf{\Pi}$ explicitly to estimate the spatial channel covariance. Let $\mathbf{a}(\hat{\phi}_\ell)$ denote the ℓ -th column in the estimate of $\hat{\mathbf{A}}$. The problem of finding $\hat{\phi}_\ell$ that minimizes the angle between $\hat{\mathbf{b}}_\ell$ and $\mathbf{W}^* \mathbf{a}(\hat{\phi}_\ell)$ is represented as

$$\hat{\phi}_\ell = \arg \min_{\phi} \left(1 - \frac{|\hat{\mathbf{b}}_\ell^* \mathbf{W}^* \mathbf{a}(\phi)|^2}{\|\hat{\mathbf{b}}_\ell\|^2 \|\mathbf{W}^* \mathbf{a}(\phi)\|^2} \right), \quad (27)$$

and its solution can be found by one-dimensional search methods with respect to ϕ , which is a continuous variable. Since we assume ULA, the solution can be obtained more efficiently by using a polynomial equation similar to the Root-MUSIC algorithm [44]. By letting $z = e^{\frac{j2\pi d_a \sin(\phi)}{\lambda}}$, the array response vector $\mathbf{a}(\phi)$ can be denoted by $\mathbf{a}(z) = [1 \ z \ \dots \ z^{N_{\text{ant}}-1}]^T$. Then, the optimization problem in (27) is rewritten as

$$\hat{z}_\ell = \arg \min_z \left(\frac{\mathbf{a}^*(z) \mathbf{W} \left(\|\hat{\mathbf{b}}_\ell\|^2 \mathbf{I} - \hat{\mathbf{b}}_\ell \hat{\mathbf{b}}_\ell^* \right) \mathbf{W}^* \mathbf{a}(z)}{\|\hat{\mathbf{b}}_\ell\|^2 \mathbf{a}^*(z) \mathbf{W} \mathbf{W}^* \mathbf{a}(z)} \right). \quad (28)$$

Let $\mathbf{Q}_\ell = \mathbf{W} \left(\|\hat{\mathbf{b}}_\ell\|^2 \mathbf{I} - \hat{\mathbf{b}}_\ell \hat{\mathbf{b}}_\ell^* \right) \mathbf{W}^*$. The numerator in (28) is represented as a polynomial with respect to z and becomes zero in the noiseless case, i.e.,

$$\mathbf{a}^*(z) \mathbf{Q}_\ell \mathbf{a}(z) = \sum_{m=-N_{\text{ant}}+1}^{N_{\text{ant}}-1} \left(\sum_{n_1=n_2=m} [\mathbf{Q}_\ell]_{n_1, n_2} \right) z^m = 0. \quad (29)$$

Note that if ω is a root of (29), then $1/\omega^*$ is also its root, and there are $(N_{\text{ant}} - 1)$ roots within the unit circle. Let $\omega_1, \dots, \omega_{N_{\text{ant}}-1}$ denote the $(N_{\text{ant}} - 1)$ roots normalized by their amplitudes. Then, the solution to (28) can be obtained by searching over $z \in \{\omega_1, \dots, \omega_{N_{\text{ant}}-1}\}$ that has discrete $(N_{\text{ant}} - 1)$

elements, i.e., $\hat{z}_\ell = \arg \max_{z \in \{\omega_1, \dots, \omega_{N_{\text{ant}}-1}\}} \frac{|\hat{\mathbf{b}}_\ell^* \mathbf{W}^* \mathbf{a}(z)|^2}{\|\mathbf{W}^* \mathbf{a}(z)\|^2}$. After obtaining $\hat{z}_\ell$, the diagonal elements in $\hat{\Delta}_{\mathbf{B}}$ can be estimated as

$$\hat{\Delta}_{\mathbf{B},\ell} = \frac{\mathbf{a}^*(\hat{z}_\ell) \mathbf{W} \hat{\mathbf{b}}_\ell}{\|\mathbf{W}^* \mathbf{a}(\hat{z}_\ell)\|^2}. \quad (30)$$

Let $\hat{\Delta}_{\mathbf{B}}$ and $\hat{\mathbf{A}}$ denote the estimate of $\Delta_{\mathbf{B}}$ and $\tilde{\mathbf{A}}$. Then, $\hat{\Delta}_{\mathbf{B}}$ and $\hat{\mathbf{A}}$ can be represented as $\hat{\Delta}_{\mathbf{B}} = \text{diag}([\hat{\Delta}_{\mathbf{B},1} \ \dots \ \hat{\Delta}_{\mathbf{B},L_{\text{ch}}}]$ and $\hat{\mathbf{A}} = [\mathbf{a}(\hat{z}_1) \ \dots \ \mathbf{a}(\hat{z}_{L_{\text{ch}}})]$. Note that $\hat{\mathbf{A}}$ is the estimate of $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}$, in which the permutation matrix $\mathbf{\Pi}$ is not known.

C. Third step: estimation of the spatial channel covariance matrix from $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{G}}$

While the estimate of $\mathbf{A}$ is given by $\hat{\mathbf{A}}$ with only permutation indeterminacy at the second step in Section IV-B, the estimated factor matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{C}}$ at the first step in Section IV-A still contain both scaling and permutation indeterminacy. Consequently, it is impossible to simply replace $\mathbf{G}$ and $\mathbf{C}$ by $\hat{\mathbf{G}}$ and $\hat{\mathbf{C}}$ in (16) without considering $\Delta_{\mathbf{C}}$, $\Delta_{\mathbf{G}}$, and $\mathbf{\Pi}$.

Let $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{G}}$ denote the estimate of the actual $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{G}$ without any indeterminacy, which are defined as $\tilde{\mathbf{A}} = \hat{\mathbf{A}}\mathbf{\Pi}^*$, $\tilde{\mathbf{C}} = \hat{\mathbf{C}}\Delta_{\mathbf{C}}^{-1}\mathbf{\Pi}^*$ and $\tilde{\mathbf{G}} = \hat{\mathbf{G}}\Delta_{\mathbf{G}}^{-1}\mathbf{\Pi}^*$. Note that $\mathbf{\Pi}^{-1} = \mathbf{\Pi}^*$ for any permutation matrix $\mathbf{\Pi}$. Then, the estimate of the sample spatial channel covariance matrix in (16) can be calculated from $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{G}}$ as

$$\begin{aligned} \tilde{\mathbf{R}}_{\mathbf{h}} &= \frac{1}{K_{\text{sbc}} T_{\text{frm}}} \tilde{\mathbf{A}} (\tilde{\mathbf{G}}^* \tilde{\mathbf{G}} \odot \tilde{\mathbf{C}}^* \tilde{\mathbf{C}}) \tilde{\mathbf{A}}^* \\ &\stackrel{(a)}{\approx} \frac{1}{K_{\text{sbc}} T_{\text{frm}}} \hat{\mathbf{A}} \hat{\Delta}_{\mathbf{B}}^* (\hat{\mathbf{G}}^* \hat{\mathbf{G}} \odot \hat{\mathbf{C}}^* \hat{\mathbf{C}}) \hat{\Delta}_{\mathbf{B}} \hat{\mathbf{A}}^*, \end{aligned} \quad (31)$$

where (a) comes from the fact that $\Delta_{\mathbf{B}} \Delta_{\mathbf{C}} \Delta_{\mathbf{G}} = \mathbf{I}$, and $\hat{\Delta}_{\mathbf{B}}$ is the estimate of $\Delta_{\mathbf{B}}$.

V. RELATIVE PRECODING EFFICIENCY FOR SPATIAL CHANNEL COVARIANCE ESTIMATION

The MSE or normalized MSE (NMSE) is typically used as a performance metric for channel estimation methods. Other metrics, though, are more relevant for spatial channel covariance estimation. This is because the dominant eigenvalues and their eigenspaces are more useful for hybrid precoding rather than each element in the covariance matrix. In this regard, we adopt the performance metric used in [22], [23], which we call relative precoding efficiency (RPE). Let $\mathbf{R}_{\mathbf{h}} \in \mathbb{C}^{N_{\text{ant}} \times N_{\text{ant}}}$ and $\tilde{\mathbf{R}}_{\mathbf{h}} \in \mathbb{C}^{N_{\text{ant}} \times N_{\text{ant}}}$ be the spatial channel covariance and its estimate. Let $\mathbf{U}$ and $\tilde{\mathbf{U}}$ denote the matrices composed of the dominant M_{RF} eigenvectors of $\mathbf{R}_{\mathbf{h}}$ and $\tilde{\mathbf{R}}_{\mathbf{h}}$. The RPE is defined as

$$\eta = \frac{\text{Tr}(\tilde{\mathbf{U}}^* \mathbf{R}_{\mathbf{h}} \tilde{\mathbf{U}})}{\text{Tr}(\mathbf{U}^* \mathbf{R}_{\mathbf{h}} \mathbf{U})}. \quad (32)$$

This metric lies between zero and one, i.e., $0 \leq \eta \leq 1$, and higher η indicates more accurate estimation. The RPE η in (32) is closely related to the relative spectral efficiency of the hybrid beamforming based on $\tilde{\mathbf{R}}_{\mathbf{h}}$ compared to that of the hybrid beamforming based on $\mathbf{R}_{\mathbf{h}}$. Consider a hybrid beamforming system where the analog part is composed of $\mathbf{U}$ or $\tilde{\mathbf{U}}$ with M_{RF} RF chains as in [9], [10]. For analytical tractability, we ignore

the fact that phase shifter are typically used for the analog part. At low SNR region, the achievable rate ratio approximates to

$$\begin{aligned} \frac{\text{rate}_{\text{hybrid}}^{(\text{est.})}}{\text{rate}_{\text{hybrid}}^{(\text{ideal})}} &= \frac{\mathbb{E} \left[\log(1 + \frac{1}{\sigma^2} \mathbf{h}^* \tilde{\mathbf{U}} \tilde{\mathbf{U}}^* \mathbf{h}) \right]}{\mathbb{E} \left[\log(1 + \frac{1}{\sigma^2} \mathbf{h}^* \mathbf{U} \mathbf{U}^* \mathbf{h}) \right]} \\ &\stackrel{(a)}{\approx} \frac{\mathbb{E} \left[\frac{1}{\sigma^2} \mathbf{h}^* \tilde{\mathbf{U}} \tilde{\mathbf{U}}^* \mathbf{h} \right]}{\mathbb{E} \left[\frac{1}{\sigma^2} \mathbf{h}^* \mathbf{U} \mathbf{U}^* \mathbf{h} \right]} \\ &\stackrel{(b)}{=} \frac{\text{Tr}(\tilde{\mathbf{U}}^* \mathbf{R}_{\mathbf{h}} \tilde{\mathbf{U}})}{\text{Tr}(\mathbf{U}^* \mathbf{R}_{\mathbf{h}} \mathbf{U})}, \end{aligned} \quad (33)$$

where (a) comes from the fact that $\ln(1+x) \approx x$ for $x \approx 0$, and (b) comes from $\text{Tr}(\mathbf{A}\mathbf{B}) = \text{Tr}(\mathbf{B}\mathbf{A})$ and $\mathbf{R}_{\mathbf{h}} = \mathbb{E}[\mathbf{h}\mathbf{h}^*]$. This result shows that the metric η allows us to anticipate how much relative loss will be caused by the estimation error in terms of achievable rate at low SNR region.

The RPE η defined in (32) can be analyzed approximately in large antenna array regimes. Let $\mathbf{A} \in \mathbb{C}^{N_{\text{ant}} \times L_{\text{ch}}}$ be a matrix composed of array response vectors and $\mathbf{R}_{\mathbf{g}} = \mathbb{E}[\mathbf{g}\mathbf{g}^*] \in \mathbb{C}^{L_{\text{ch}} \times L_{\text{ch}}}$ be the covariance of channel path gains. The spatial channel covariance matrix is represented as

$$\mathbf{R}_{\mathbf{h}} = \mathbf{A} \mathbf{R}_{\mathbf{g}} \mathbf{A}^*. \quad (34)$$

As N_{ant} becomes large, $\mathbf{A}^* \mathbf{A} \approx N_{\text{ant}} \mathbf{I}$, i.e., $\frac{1}{\sqrt{N_{\text{ant}}}} \mathbf{A}$ becomes semi-unitary asymptotically [45]. If we assume that $g_{\ell,k}$ are IID complex random variables with zero mean and variance $1/L_{\text{ch}}$ for analytical tractability, then (34) can be regarded as the approximate eigenvalue decomposition of $\mathbf{R}_{\mathbf{h}}$, i.e., $\mathbf{U} \approx \frac{1}{\sqrt{N_{\text{ant}}}} \mathbf{A}$. Let e_ℓ be the AoA estimation error and $\tilde{\phi}_\ell = \phi_\ell + e_\ell$ be the estimated AoA for the ℓ -th path. Similar to $\mathbf{U}$, we assume that $\tilde{\mathbf{U}}$ approximates to $\frac{1}{\sqrt{N_{\text{ant}}}} \tilde{\mathbf{A}}$ as N_{ant} increases. Then, the RPE η becomes

$$\begin{aligned} \eta &\approx \frac{\text{Tr}(\tilde{\mathbf{U}}^* \mathbf{U} \left(\frac{N_{\text{ant}}}{L_{\text{ch}}} \mathbf{I}_{L_{\text{ch}}} \right) \mathbf{U}^* \tilde{\mathbf{U}})}{\text{Tr} \left(\frac{N_{\text{ant}}}{L_{\text{ch}}} \mathbf{I}_{L_{\text{ch}}} \right)} \\ &\approx \frac{1}{N_{\text{ant}}^2 L_{\text{ch}}} \|\mathbf{A}^* \tilde{\mathbf{A}}\|_{\text{F}}^2 \\ &\approx \frac{1}{N_{\text{ant}}^2 L_{\text{ch}}} \sum_{\ell=1}^{L_{\text{ch}}} |\mathbf{a}^*(\phi_\ell) \mathbf{a}(\phi_\ell + e_\ell)|^2, \end{aligned} \quad (35)$$

where we assume that e_ℓ is small and $\frac{1}{N} \mathbf{a}^*(\phi_{\ell_1}) \mathbf{a}(\phi_{\ell_2}) \approx 0$ for $\ell_1 \neq \ell_2$.

Let κ_ℓ be defined as $\kappa_\ell = \frac{\pi d_a}{\lambda} (\sin(\phi_\ell + e_\ell) - \sin(\phi_\ell))$ which approximates to $\kappa_\ell \approx \frac{\pi d_a}{\lambda} \cos(\phi_\ell) e_\ell$ for small e_ℓ . In the ULA case, η in (35) is given by

$$\begin{aligned} \eta &\approx \frac{1}{N_{\text{ant}}^2 L_{\text{ch}}} \sum_{\ell=1}^{L_{\text{ch}}} \frac{\sin^2(N_{\text{ant}} \kappa_\ell)}{\sin^2(\kappa_\ell)} \\ &\stackrel{(a)}{\approx} \frac{1}{L_{\text{ch}}} \sum_{\ell=1}^{L_{\text{ch}}} \left(1 - \frac{N_{\text{ant}}^2 \kappa_\ell^2}{3} \right) \\ &\approx \frac{1}{L_{\text{ch}}} \sum_{\ell=1}^{L_{\text{ch}}} \left(1 - \frac{N_{\text{ant}}^2 \pi^2 d_a^2 \cos^2(\phi_\ell)}{3 \lambda^2} e_\ell^2 \right), \end{aligned} \quad (36)$$

where (a) comes from the second-order approximation of Maclaurin series for small κ_ℓ . Consequently, $1 - \mathbb{E}[\eta]$ approximately becomes

$$\begin{aligned} 1 - \mathbb{E}[\eta] &\approx \frac{N_{\text{ant}}^2 \pi^2 d_a^2}{3L_{\text{ch}} \lambda^2} \sum_{\ell=1}^{L_{\text{ch}}} \cos^2(\phi_\ell) \mathbb{E}[(\phi_\ell - \tilde{\phi}_\ell)^2] \\ &\geq \frac{N_{\text{ant}}^2 \pi^2 d_a^2}{3L_{\text{ch}} \lambda^2} \sum_{\ell=1}^{L_{\text{ch}}} \cos^2(\phi_\ell) \text{CRLB}(\phi_\ell), \end{aligned} \quad (37)$$

where the CRLB of the ϕ_ℓ estimation $\text{CRLB}(\phi_\ell)$ will be derived in the following section.

VI. CRAMÉR-RAO LOWER BOUND FOR THE AOA ESTIMATION

In this section, we derive the CRLB of the MSE of AoAs. The basic tool to derive CRLB is based on the method in [46] and [29]. In [46], all the factor matrices are non-structure matrices, i.e., a factor matrix $\mathbf{F} \in \mathbb{C}^{N_{\text{ant}} \times M}$ is determined by $N_{\text{ant}} M_{\text{RF}}$ complex-valued variables and has no specific structure. In [29], all the factor matrices have a special structure and are determined by only L_{ch} real-valued variables. In this paper, we derive CRLB in the case where the tensor has the combination of two structured factor matrices, $\mathbf{A}$ and $\mathbf{C}$, and one unstructured factor matrix $\mathbf{G}$. We also simplify the complicated CRLB expression to a more compact form.

Focusing on the fact that the channel tensor is determined by L_{ch} AoAs, L_{ch} path delays, and $L_{\text{ch}} T_{\text{frm}}$ time-varying channel path gains, we define three parameter vectors as $\boldsymbol{\phi} = [\phi_1 \ \cdots \ \phi_{L_{\text{ch}}}]^T$, $\boldsymbol{\tau} = [\tau_1 \ \cdots \ \tau_{L_{\text{ch}}}]^T$, and $\mathbf{g} = \text{vec}(\mathbf{G}) = [g_{1,1} \ \cdots \ g_{T,L_{\text{ch}}}]^T$. Let $\boldsymbol{\theta}$ denote a column vector that includes all the parameters such that $\boldsymbol{\theta} = [\boldsymbol{\phi}^T \ \boldsymbol{\tau}^T \ \mathbf{g}^T \ \mathbf{g}^*]^T$. Note that $\mathbf{g}$ is a complex vector while $\boldsymbol{\phi}$ and $\boldsymbol{\tau}$ are real vectors.

Since the analog combining matrix combined with the baseband post-processing matrix is unitary, the elements in the noise tensor $\mathcal{N}$ become IID circularly symmetric Gaussian with $\mathcal{CN}(0, \sigma^2)$. Consequently, the log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} f(\boldsymbol{\theta}) &= -M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \ln(\pi \sigma^2) - \frac{1}{\sigma^2} \|\mathbf{Y}_{(1)} - \mathbf{B}(\mathbf{G} \odot \mathbf{C})^T\|_F^2 \\ &= -M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \ln(\pi \sigma^2) - \frac{1}{\sigma^2} \|\mathbf{Y}_{(2)} - \mathbf{C}(\mathbf{G} \odot \mathbf{B})^T\|_F^2 \\ &= -M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \ln(\pi \sigma^2) - \frac{1}{\sigma^2} \|\mathbf{Y}_{(3)} - \mathbf{G}(\mathbf{C} \odot \mathbf{B})^T\|_F^2. \end{aligned} \quad (38)$$

Then, the CRLB with respect to the parameter set $\boldsymbol{\theta}$ can be obtained as

$$\text{CRLB}(\boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}), \quad (39)$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta}) \in \mathbb{C}^{2L_{\text{ch}}(T_{\text{frm}}+1) \times 2L_{\text{ch}}(T_{\text{frm}}+1)}$ is the complex Fisher information matrix (FIM) defined as

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^* \right]. \quad (40)$$

The FIM $\boldsymbol{\Omega}(\boldsymbol{\theta})$ in (40) is divided into submatrices as

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Omega}_{\phi\phi} & \boldsymbol{\Omega}_{\phi\tau} & \boldsymbol{\Omega}_{\phi\mathbf{g}} & \boldsymbol{\Omega}_{\phi\mathbf{g}^C} \\ \boldsymbol{\Omega}_{\phi\tau}^* & \boldsymbol{\Omega}_{\tau\tau} & \boldsymbol{\Omega}_{\tau\mathbf{g}} & \boldsymbol{\Omega}_{\tau\mathbf{g}^C} \\ \boldsymbol{\Omega}_{\phi\mathbf{g}}^* & \boldsymbol{\Omega}_{\tau\mathbf{g}}^* & \boldsymbol{\Omega}_{\mathbf{g}\mathbf{g}} & \boldsymbol{\Omega}_{\mathbf{g}\mathbf{g}^C} \\ \boldsymbol{\Omega}_{\phi\mathbf{g}^C}^* & \boldsymbol{\Omega}_{\tau\mathbf{g}^C}^* & \boldsymbol{\Omega}_{\mathbf{g}\mathbf{g}^C}^* & \boldsymbol{\Omega}_{\mathbf{g}^C\mathbf{g}^C} \end{bmatrix}. \quad (41)$$

Each submatrix in (41) is calculated in the following subsections.

A. Calculation of $\boldsymbol{\Omega}_{\phi\phi} \in \mathbb{C}^{L_{\text{ch}} \times L_{\text{ch}}}$

The partial derivative of $f(\boldsymbol{\theta})$ with respect to ϕ_ℓ is given by

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \phi_\ell} = \text{Tr} \left(\left(\frac{\partial f(\boldsymbol{\theta})}{\partial \mathbf{B}} \right)^T \frac{\partial \mathbf{B}}{\partial \phi_\ell} + \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \mathbf{B}^C} \right)^T \frac{\partial \mathbf{B}^C}{\partial \phi_\ell} \right). \quad (42)$$

Let $\dot{\mathbf{B}} = [\dot{\mathbf{b}}_1(\phi_1) \ \cdots \ \dot{\mathbf{b}}_{L_{\text{ch}}}(\phi_{L_{\text{ch}}})]$ be defined as

$$\begin{aligned} \dot{\mathbf{B}} &= \frac{j2\pi d_a}{\lambda} \mathbf{W}^* \text{diag}([0 \ 1 \ \cdots \ N_{\text{ant}} - 1]) \\ &\quad \cdot [\cos(\phi_1) \mathbf{a}(\phi_1) \ \cdots \ \cos(\phi_{L_{\text{ch}}}) \mathbf{a}(\phi_{L_{\text{ch}}})], \end{aligned} \quad (43)$$

and let $\mathbf{N}_{(1)} = \mathbf{Y}_{(1)} - \mathbf{B}(\mathbf{G} \odot \mathbf{C})^T$ be the mode-1 unfolding matrix of $\mathcal{N}$.

By using (42) and (43), the partial derivative of $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\phi}$ can be represented as

$$\begin{aligned} \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= \text{diag} \left(\left(\frac{\partial f(\boldsymbol{\theta})}{\partial \mathbf{B}} \right)^T \dot{\mathbf{B}} + \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \mathbf{B}^C} \right)^* \dot{\mathbf{B}}^C \right) \\ &= \frac{2}{\sigma^2} \text{Re} \left(\text{diag} \left((\mathbf{G} \odot \mathbf{C})^T \mathbf{N}_{(1)}^* \dot{\mathbf{B}} \right) \right). \end{aligned} \quad (44)$$

Let $\mathbf{V}_{\mathbf{B}} = \frac{1}{\sigma^2} (\mathbf{G} \odot \mathbf{C})^T \mathbf{N}_{(1)}^* \dot{\mathbf{B}}$ and $\mathbf{d}_{\mathbf{B}} = \text{diag}(\mathbf{V}_{\mathbf{B}})$. Then, $\mathbf{d}_{\mathbf{B}}$ can be represented as

$$\mathbf{d}_{\mathbf{B}} = \frac{1}{\sigma^2} \left(\dot{\mathbf{B}} \odot \mathbf{G} \odot \mathbf{C} \right)^T \text{vec}(\mathbf{N}_{(1)}^*). \quad (45)$$

Since we consider IID zero mean circularly symmetric complex Gaussian noise, the covariance matrix and the pseudo-covariance matrix of $\text{vec}(\mathbf{N}_{(1)}^*)$ in (45) become $\mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(1)}^*) \right)^* \right] = \sigma^2 \mathbf{I}$ and $\mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(1)}^*) \right)^T \right] = \mathbf{0}$. Then, the covariance matrix of $\mathbf{d}_{\mathbf{B}}$ becomes

$$\mathbf{C}_{\mathbf{d}_{\mathbf{B}}} = \mathbb{E} [\mathbf{d}_{\mathbf{B}} \mathbf{d}_{\mathbf{B}}^*] = \frac{1}{\sigma^2} \left(\dot{\mathbf{B}}^* \dot{\mathbf{B}} \odot \mathbf{G}^* \mathbf{G} \odot \mathbf{C}^* \mathbf{C} \right)^T, \quad (46)$$

and the pseudo-covariance becomes $\tilde{\mathbf{C}}_{\mathbf{d}_{\mathbf{B}}} = \mathbb{E} [\mathbf{d}_{\mathbf{B}} \mathbf{d}_{\mathbf{B}}^T] = \mathbf{0}$. Then, the submatrix $\boldsymbol{\Omega}_{\phi\phi}$ becomes

$$\begin{aligned} \boldsymbol{\Omega}_{\phi\phi} &= \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \right)^* \right] \\ &= \frac{2}{\sigma^2} \text{Re} \left(\left(\dot{\mathbf{B}}^* \dot{\mathbf{B}} \odot \mathbf{G}^* \mathbf{G} \odot \mathbf{C}^* \mathbf{C} \right)^T \right). \end{aligned} \quad (47)$$

B. Calculation of $\boldsymbol{\Omega}_{\tau\tau} \in \mathbb{C}^{L_{\text{ch}} \times L_{\text{ch}}}$

Let $\dot{\boldsymbol{\epsilon}}_\ell(\tau_\ell) \in \mathbb{C}^{K_{\text{sbc}} \times 1}$ be a vector whose k -th element is defined as

$$[\dot{\boldsymbol{\epsilon}}_\ell(\tau_\ell)]_k = - \sum_{d=0}^{N_{\text{CP}}-1} p'_{\text{PS}}(dT_s - \tau_\ell) e^{-\frac{j2\pi(k-1)d}{K_{\text{sbc}}}}, \quad (48)$$

where $p'_{\text{PS}}(x)$ is the first derivative of $p_{\text{PS}}(x)$. Then, the partial derivative of $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\tau}$ can be obtained as

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} = \frac{2}{\sigma^2} \text{Re} \left(\text{diag} \left((\mathbf{G} \odot \mathbf{B})^\top (\mathbf{Y}_{(2)}^* - (\mathbf{G} \odot \mathbf{B})^c \mathbf{C}^*) \dot{\mathbf{C}} \right) \right), \quad (49)$$

where $\dot{\mathbf{C}} = [\dot{\mathbf{c}}_1(\tau_1) \ \cdots \ \dot{\mathbf{c}}_{L_{\text{ch}}}(\tau_{L_{\text{ch}}})]$. Let $\mathbf{N}_{(2)} = \mathbf{Y}_{(2)} - \mathbf{C}(\mathbf{G} \odot \mathbf{B})^\top$ be the mode-2 unfolding matrix of $\mathcal{N}$, and $\mathbf{V}_{\mathbf{C}} = \frac{1}{\sigma^2} (\mathbf{G} \odot \mathbf{B})^\top \mathbf{N}_{(2)}^* \dot{\mathbf{C}}$. Let $\mathbf{d}_{\mathbf{C}}$ denote

$$\mathbf{d}_{\mathbf{C}} = \text{diag}(\mathbf{V}_{\mathbf{C}}) = \frac{1}{\sigma^2} \left(\dot{\mathbf{C}} \odot \mathbf{G} \odot \mathbf{B} \right)^\top \text{vec}(\mathbf{N}_{(2)}^*). \quad (50)$$

Similarly to Section VI-A, the submatrix $\boldsymbol{\Omega}_{\boldsymbol{\tau}\boldsymbol{\tau}}$ can be calculated as

$$\boldsymbol{\Omega}_{\boldsymbol{\tau}\boldsymbol{\tau}} = \frac{2}{\sigma^2} \text{Re} \left(\left(\dot{\mathbf{C}}^* \dot{\mathbf{C}} \odot \mathbf{G}^* \mathbf{G} \odot \mathbf{B}^* \mathbf{B} \right)^\top \right). \quad (51)$$

C. Calculation of $\boldsymbol{\Omega}_{\text{gg}}$ $\in \mathbb{C}^{T_{\text{frm}} L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$, $\boldsymbol{\Omega}_{\text{gg}^c} \in \mathbb{C}^{T_{\text{frm}} L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$, and $\boldsymbol{\Omega}_{\text{g}^c \text{g}^c} \in \mathbb{C}^{T_{\text{frm}} L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$

From the fact that $\mathbb{E} \left[\text{vec}(\mathbf{N}_{(3)}^c) \left(\text{vec}(\mathbf{N}_{(3)}^c) \right)^* \right] = \sigma^2 \mathbf{I}$, the submatrix $\boldsymbol{\Omega}_{\text{gg}}$ becomes

$$\begin{aligned} \boldsymbol{\Omega}_{\text{gg}} &= \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \text{vec}(\mathbf{G})} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \text{vec}(\mathbf{G})} \right)^* \right] \\ &= \frac{1}{\sigma^4} \left((\mathbf{C} \odot \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}} \right) \\ &\quad \cdot \mathbb{E} \left[\text{vec}(\mathbf{N}_{(3)}^c) \left(\text{vec}(\mathbf{N}_{(3)}^c) \right)^* \right] \left((\mathbf{C} \odot \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}} \right)^* \\ &= \frac{1}{\sigma^2} (\mathbf{C}^* \mathbf{C} \odot \mathbf{B}^* \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}}. \end{aligned} \quad (52)$$

Since $\mathbb{E} \left[\text{vec}(\mathbf{N}_{(3)}^c) \left(\text{vec}(\mathbf{N}_{(3)}^c) \right)^\top \right] = \mathbf{0}_{M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \times M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}}}$, the submatrix $\boldsymbol{\Omega}_{\text{gg}^c}$ is given by

$$\begin{aligned} \boldsymbol{\Omega}_{\text{gg}^c} &= \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \text{vec}(\mathbf{G})} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \text{vec}(\mathbf{G}^c)} \right)^* \right] \\ &= \frac{1}{\sigma^4} \left((\mathbf{C} \odot \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}} \right) \\ &\quad \cdot \mathbb{E} \left[\text{vec}(\mathbf{N}_{(3)}^c) \left(\text{vec}(\mathbf{N}_{(3)}^c) \right)^\top \right] \left((\mathbf{C} \odot \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}} \right)^\top \\ &= \mathbf{0}_{T_{\text{frm}} L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}. \end{aligned} \quad (53)$$

The submatrix $\boldsymbol{\Omega}_{\text{g}^c \text{g}^c}$ can be obtained from $\boldsymbol{\Omega}_{\text{gg}}$ such that $\boldsymbol{\Omega}_{\text{g}^c \text{g}^c} = \boldsymbol{\Omega}_{\text{gg}}^c$.

D. Calculation of $\boldsymbol{\Omega}_{\boldsymbol{\phi}\boldsymbol{\tau}}$ $\in \mathbb{C}^{L_{\text{ch}} \times L_{\text{ch}}}$

The submatrix $\boldsymbol{\Omega}_{\boldsymbol{\phi}\boldsymbol{\tau}}$ is given by

$$\begin{aligned} \boldsymbol{\Omega}_{\boldsymbol{\phi}\boldsymbol{\tau}} &= \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \right)^* \right] \\ &= 2\text{Re}(\mathbf{C}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}}) + 2\text{Re}(\tilde{\mathbf{C}}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}}), \end{aligned} \quad (54)$$

where $\mathbf{C}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}} = \mathbb{E}[\mathbf{d}_{\mathbf{B}} \mathbf{d}_{\mathbf{C}}^*]$ and $\tilde{\mathbf{C}}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}} = \mathbb{E}[\mathbf{d}_{\mathbf{B}} \mathbf{d}_{\mathbf{C}}^\top]$.

To calculate $\mathbf{C}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}}$, let us first start with calculating the cross-covariance matrix of $\text{vec}(\mathbf{N}_{(1)}^*)$ and $\text{vec}(\mathbf{N}_{(2)}^*)$, which are associated with the mode-1 and mode-2 unfolding matrix of

$\mathcal{N}$. Let $\mathbf{e}_i \in \mathbb{C}^{M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \times 1}$ be the i -th unit coordinate vector and $\mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}} = \mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(2)}^*) \right)^* \right]$. Using the fact that $[\mathcal{N}]_{m,k,t}$ is expressed in different ways as

$$\begin{aligned} [\mathcal{N}]_{m,k,t} &= [\text{vec}(\mathbf{N}_{(1)})]_{k+(t-1)K_{\text{sbc}}+(m-1)K_{\text{sbc}}T_{\text{frm}}} \\ &= [\text{vec}(\mathbf{N}_{(2)})]_{m+(t-1)M_{\text{RF}}+(k-1)M_{\text{RF}}T_{\text{frm}}}, \end{aligned} \quad (55)$$

the cross-covariance matrix $\mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}}$ can be represented as

$$\begin{aligned} \mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}} &= \mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(2)}^*) \right)^* \right] \\ &= \sigma^2 \sum_{m=1}^{M_{\text{RF}}} \sum_{k=1}^{K_{\text{sbc}}} \sum_{t=1}^{T_{\text{frm}}} \mathbf{e}_{k+(t-1)K_{\text{sbc}}+(m-1)K_{\text{sbc}}T_{\text{frm}}} \\ &\quad \cdot \mathbf{e}_{m+(t-1)M_{\text{RF}}+(k-1)M_{\text{RF}}T_{\text{frm}}}^\top. \end{aligned} \quad (56)$$

Consequently, $\mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}} \in \mathbb{C}^{M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}} \times M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}}}$ is a matrix that has only $M_{\text{RF}} K_{\text{sbc}} T_{\text{frm}}$ nonzero elements whose amplitudes are equal to σ^2 . From (56), the cross-covariance matrix of $\mathbf{d}_{\mathbf{B}}$ and $\mathbf{d}_{\mathbf{C}}$ can be expressed as

$$\begin{aligned} \mathbf{C}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}} &= \frac{1}{\sigma^4} \left(\dot{\mathbf{B}} \odot \mathbf{G} \odot \mathbf{C} \right)^\top \mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}} \left(\dot{\mathbf{C}} \odot \mathbf{G} \odot \mathbf{B} \right)^c \\ &= \frac{1}{\sigma^2} \left(\mathbf{B}^* \dot{\mathbf{B}} \odot \mathbf{G}^* \mathbf{G} \odot \dot{\mathbf{C}}^* \mathbf{C} \right)^\top. \end{aligned} \quad (57)$$

Since $\tilde{\mathbf{C}}_{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}} = \mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(2)}^*) \right)^\top \right] = \mathbf{0}$, the pseudo-cross-covariance matrix of $\mathbf{d}_{\mathbf{B}}$ and $\mathbf{d}_{\mathbf{C}}$ becomes

$$\tilde{\mathbf{C}}_{\mathbf{d}_{\mathbf{B}}, \mathbf{d}_{\mathbf{C}}} = \mathbb{E}[\mathbf{d}_{\mathbf{B}} \mathbf{d}_{\mathbf{C}}^\top] = \mathbf{0}. \quad (58)$$

From (57) and (58), the submatrix $\boldsymbol{\Omega}_{\boldsymbol{\phi}\boldsymbol{\tau}}$ in (54) can be rewritten as

$$\boldsymbol{\Omega}_{\boldsymbol{\phi}\boldsymbol{\tau}} = \frac{2}{\sigma^2} \text{Re} \left(\left(\mathbf{B}^* \dot{\mathbf{B}} \odot \mathbf{G}^* \mathbf{G} \odot \dot{\mathbf{C}}^* \mathbf{C} \right)^\top \right). \quad (59)$$

E. Calculation of $\boldsymbol{\Omega}_{\boldsymbol{\phi}\mathbf{g}}$ $\in \mathbb{C}^{L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\phi}\mathbf{g}^c} \in \mathbb{C}^{L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$

The submatrix $\boldsymbol{\Omega}_{\boldsymbol{\phi}\mathbf{g}}$ is expressed as

$$\begin{aligned} \boldsymbol{\Omega}_{\boldsymbol{\phi}\mathbf{g}} &= \mathbb{E} \left[\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \text{vec}(\mathbf{G})} \right)^* \right] \\ &= \frac{1}{\sigma^2} \mathbb{E} \left[\left(\mathbf{d}_{\mathbf{B}} + \mathbf{d}_{\mathbf{B}}^c \right) \left(\text{vec}(\mathbf{N}_{(3)}^c) \right)^* \left((\mathbf{C} \odot \mathbf{B})^\top \otimes \mathbf{I}_{T_{\text{frm}}} \right)^* \right] \\ &= \frac{1}{\sigma^4} \left(\left(\dot{\mathbf{B}} \odot \mathbf{G} \odot \mathbf{C} \right)^\top \mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(3)}^c} + \left(\dot{\mathbf{B}} \odot \mathbf{G} \odot \mathbf{C} \right)^* \tilde{\mathbf{C}}_{\mathbf{n}_{(1)}, \mathbf{n}_{(3)}^c} \right) \\ &\quad \cdot \left((\mathbf{C} \odot \mathbf{B})^c \otimes \mathbf{I}_{T_{\text{frm}}} \right), \end{aligned} \quad (60)$$

where $\tilde{\mathbf{C}}_{\mathbf{n}_{(1)}, \mathbf{n}_{(3)}^c} = \mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}) \left(\text{vec}(\mathbf{N}_{(3)}) \right)^\top \right] = \mathbf{0}$ and

$$\begin{aligned} \mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(3)}^c} &= \mathbb{E} \left[\text{vec}(\mathbf{N}_{(1)}^*) \left(\text{vec}(\mathbf{N}_{(3)}) \right)^\top \right] \\ &= \sigma^2 \sum_{m=1}^{M_{\text{RF}}} \sum_{k=1}^{K_{\text{sbc}}} \sum_{t=1}^{T_{\text{frm}}} \mathbf{e}_{k+(t-1)K_{\text{sbc}}+(m-1)K_{\text{sbc}}T_{\text{frm}}} \\ &\quad \cdot \mathbf{e}_{t+(m+(k-1)M_{\text{RF}}-1)T_{\text{frm}}}^\top. \end{aligned} \quad (61)$$

By using (61), we can further simplify $\mathbf{\Omega}_{\phi\mathbf{g}}$ in (60) as

$$\begin{aligned}\mathbf{\Omega}_{\phi\mathbf{g}} &= \frac{1}{\sigma^4} \left(\mathbf{\dot{B}} \odot \mathbf{G} \odot \mathbf{C} \right)^\top \mathbf{C}_{\mathbf{n}_{(1)}, \mathbf{n}_{(3)}^{\mathbf{C}}} \left((\mathbf{C} \odot \mathbf{B})^{\mathbf{C}} \otimes \mathbf{I}_{T_{\text{frm}}} \right) \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{T_{\text{frm}}} \left(\left((\mathbf{B}^* \mathbf{\dot{B}}) \odot (\mathbf{C}^* \mathbf{C}) \odot (\mathbf{1}_{L_{\text{ch}}}[\mathbf{G}]_{t,:}) \right) \otimes \mathbf{e}_t \right)^\top \\ &= \frac{1}{\sigma^2} \left((\mathbf{B}^* \mathbf{\dot{B}} \odot \mathbf{C}^* \mathbf{C}) \odot \mathbf{G} \right)^\top.\end{aligned}\quad (62)$$

The submatrix $\mathbf{\Omega}_{\phi\mathbf{g}\mathbf{C}}$ is expressed as $\mathbf{\Omega}_{\phi\mathbf{g}\mathbf{C}} = \mathbf{\Omega}_{\phi\mathbf{g}}^{\mathbf{C}}$.

F. Calculation of $\mathbf{\Omega}_{\tau\mathbf{g}} \in \mathbb{C}^{L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$ and $\mathbf{\Omega}_{\tau\mathbf{g}\mathbf{C}} \in \mathbb{C}^{L_{\text{ch}} \times T_{\text{frm}} L_{\text{ch}}}$

Similar to Section VI-E, we can obtain $\mathbf{\Omega}_{\tau\mathbf{g}}$ and $\mathbf{\Omega}_{\tau\mathbf{g}\mathbf{C}}$ as

$$\begin{aligned}\mathbf{\Omega}_{\tau\mathbf{g}} &= \mathbb{E} \left[\frac{\partial f(\theta)}{\partial \boldsymbol{\tau}} \left(\frac{\partial f(\theta)}{\partial \text{vec}(\mathbf{G})} \right)^* \right] \\ &= \frac{1}{\sigma^2} \left((\mathbf{C}^* \mathbf{\dot{C}} \odot \mathbf{B}^* \mathbf{B}) \odot \mathbf{G} \right)^\top,\end{aligned}\quad (63)$$

and $\mathbf{\Omega}_{\tau\mathbf{g}\mathbf{C}} = \mathbf{\Omega}_{\tau\mathbf{g}}^{\mathbf{C}}$.

G. CRLB for the ϕ estimation

The results of the preceding subsections are summarized as

$$\begin{aligned}\mathbf{\Omega}_{\phi\phi} &= \frac{2}{\sigma^2} \text{Re} \left(\left(\mathbf{\dot{B}}^* \mathbf{\dot{B}} \odot \mathbf{C}^* \mathbf{C} \odot \mathbf{G}^* \mathbf{G} \right)^\top \right), \\ \mathbf{\Omega}_{\tau\tau} &= \frac{2}{\sigma^2} \text{Re} \left(\left(\mathbf{B}^* \mathbf{B} \odot \mathbf{\dot{C}}^* \mathbf{\dot{C}} \odot \mathbf{G}^* \mathbf{G} \right)^\top \right), \\ \mathbf{\Omega}_{\phi\tau} &= \frac{2}{\sigma^2} \text{Re} \left(\left(\mathbf{B}^* \mathbf{\dot{B}} \odot \mathbf{\dot{C}}^* \mathbf{C} \odot \mathbf{G}^* \mathbf{G} \right)^\top \right), \\ \mathbf{\Omega}_{\mathbf{g}\mathbf{g}} &= \frac{1}{\sigma^2} (\mathbf{B}^* \mathbf{B} \odot \mathbf{C}^* \mathbf{C})^\top \otimes \mathbf{I}_{T_{\text{frm}}}, \\ \mathbf{\Omega}_{\phi\mathbf{g}} &= \frac{1}{\sigma^2} \left((\mathbf{B}^* \mathbf{\dot{B}} \odot \mathbf{C}^* \mathbf{C}) \odot \mathbf{G} \right)^\top, \\ \mathbf{\Omega}_{\tau\mathbf{g}} &= \frac{1}{\sigma^2} \left((\mathbf{B}^* \mathbf{B} \odot \mathbf{C}^* \mathbf{\dot{C}}) \odot \mathbf{G} \right)^\top.\end{aligned}\quad (64)$$

Letting $\mathbf{\Omega}_1 = \begin{bmatrix} \mathbf{\Omega}_{\phi\tau} & \mathbf{\Omega}_{\phi\mathbf{g}} & \mathbf{\Omega}_{\phi\mathbf{g}}^{\mathbf{C}} \end{bmatrix}$ and

$$\mathbf{\Omega}_2 = \begin{bmatrix} \mathbf{\Omega}_{\tau\tau} & \mathbf{\Omega}_{\tau\mathbf{g}} & \mathbf{\Omega}_{\tau\mathbf{g}}^{\mathbf{C}} \\ \mathbf{\Omega}_{\tau\mathbf{g}}^* & \mathbf{\Omega}_{\mathbf{g}\mathbf{g}} & \mathbf{0} \\ \mathbf{\Omega}_{\tau\mathbf{g}}^{\mathbf{C}*} & \mathbf{0} & \mathbf{\Omega}_{\mathbf{g}\mathbf{g}}^{\mathbf{C}} \end{bmatrix}, \quad (65)$$

the CRLB for the ϕ_ℓ estimation can be expressed in a compact form as

$$\text{CRLB}(\phi_\ell) = \left[\left(\mathbf{\Omega}_{\phi\phi} - \mathbf{\Omega}_1^* \mathbf{\Omega}_2^{-1} \mathbf{\Omega}_1 \right)^{-1} \right]_{\ell, \ell}, \quad (66)$$

by using the Schur complement and the matrix inversion lemma.

VII. COMPARISON WITH CS-BASED OR MUSIC-BASED METHODS

In this section, we explain two other approaches that estimate spatial channel covariance or subspace for comparison: 1) CS-based methods and 2) MUSIC-based methods.

A. Prior work based on CS

There has been some work on CS-based spatial channel covariance estimation for narrowband channels [22]–[24]. We can extend these methods to the wideband channel case via two different approaches.

The first approach is to exploit the fact that all the $T_{\text{frm}} K_{\text{sbc}} \mathbf{x}_{t,k} \in \mathbb{C}^{M_{\text{RF}} \times 1}$ for $t = 1, \dots, T_{\text{frm}}$ and $k = 1, \dots, K_{\text{sbc}}$ can be stacked such that the problem becomes an MMV problem with $T_{\text{frm}} K_{\text{sbc}}$ measurements. The mode-1 fibers of the signal tensor $\mathcal{X}$ in (14) can be represented as

$$\mathbf{x}_{t,k} = \mathbf{W}^* \mathbf{A} \left([\mathbf{G}^\top]_{:,t} \odot [\mathbf{C}^\top]_{:,k} \right). \quad (67)$$

Let $\mathbf{A}_{\text{D}} \in \mathbb{C}^{N_{\text{ant}} \times N_{\text{grid}}}$ be a dictionary matrix whose N_{grid} columns are composed of the array response vectors associated with a predefined set of AoAs. In the CS framework, the signal model in (67) is rewritten as

$$\mathbf{x}_{t,k} \approx \mathbf{W}^* \mathbf{A}_{\text{D}} (\hat{\mathbf{g}}_{\text{D},t} \odot \hat{\mathbf{c}}_{\text{D},k}), \quad (68)$$

where $\hat{\mathbf{g}}_{\text{D},t} \in \mathbb{C}^{N_{\text{grid}} \times 1}$ and $\hat{\mathbf{c}}_{\text{D},k} \in \mathbb{C}^{N_{\text{grid}} \times 1}$ are sparse column vectors with L_{ch} nonzero elements of $\hat{\mathbf{g}}_t$ and $\hat{\mathbf{c}}_k$ in the space domain. The positions of the L_{ch} nonzero elements indicate AoAs, and thus $\hat{\mathbf{g}}_{\text{D},t}$ and $\hat{\mathbf{c}}_{\text{D},k}$ share the same support for all t and k , which means that this is still the MMV problem with an increased number of measurement vectors. Compared to the narrowband case which is the MMV problem with T_{frm} measurement vectors, the wideband case can also be represented as an extended MMV problem that has $T_{\text{frm}} K_{\text{sbc}}$ measurement vectors. It is worthwhile to note that the dictionary matrix size is the same as $M_{\text{RF}} \times N_{\text{grid}}$ for both cases. While simultaneous orthogonal matching pursuit (SOMP) is an adequate algorithm for the general MMV problems, a more advanced CS algorithm, so-called WB-DCOMP, was proposed for the spatial channel covariance estimation problem in [22]. We will compare WB-DCOMP with the proposed method in Section VIII-B.

The second approach is to use the fact that the channel is sparse in the delay domain as well as the spatial domain. The t -th frontal slice of $\mathcal{X}$ can be represented in a matrix form as

$$\mathbf{X}_t = \mathbf{W}^* \mathbf{A} \text{diag}([\mathbf{G}]_{t,:}) \mathbf{C}^\top. \quad (69)$$

Similar to the spatial domain dictionary matrix $\mathbf{A}_{\text{D}}$, we can define a delay-domain dictionary matrix $\mathbf{C}_{\text{D}} \in \mathbb{C}^{K_{\text{sbc}} \times K_{\text{grid}}}$ such that the (k, q) -th element in $\mathbf{C}_{\text{D}}$ is $c_{k,q} = e^{-\frac{j2\pi(k-1)\tau_{\text{grid},q}}{K_{\text{sbc}} T_s}}$ where $\{\tau_{\text{grid},1}, \dots, \tau_{\text{grid},K_{\text{grid}}}\}$ is a predefined grid set of path delays in $[0, N_{\text{CP}} T_s]$. Using the two dictionary matrices, the frontal slice in (69) can be rewritten in the CS framework as

$$\mathbf{X}_t \approx \mathbf{W}^* \mathbf{A}_{\text{D}} \mathbf{G}_{\text{D},t} \mathbf{C}_{\text{D}}^\top, \quad (70)$$

where $\mathbf{G}_{\text{D}} \in \mathbb{C}^{N_{\text{grid}} \times K_{\text{grid}}}$ is a sparse matrix with only L_{ch} nonzero elements. Using vectorization, the frontal slice in (70) can be rewritten in a vector form as

$$\begin{aligned}\text{vec}(\mathbf{X}_t) &= (\mathbf{C} \odot \mathbf{W}^* \mathbf{A}) [\mathbf{G}]_{t,:}^\top \\ &\approx (\mathbf{C}_{\text{D}} \otimes \mathbf{W}^* \mathbf{A}_{\text{D}}) \text{vec}(\mathbf{G}_{\text{D},t}).\end{aligned}\quad (71)$$

Since $\text{vec}(\mathbf{G}_{D,t})$'s for $t = 1, \dots, T_{\text{frm}}$ are sparse nonzero vectors sharing a common support set, this also becomes an MMV problem with an increased dictionary size.

As we will show in Section VIII-B, the second approach outperforms the first approach because the former exploits sparsity in both spatial and temporal domain while the latter exploits sparsity only in the spatial domain. The first approach, however, has a considerably high complexity due to the Kronecker structure of the dictionary matrix; the dictionary matrix size increases from $M_{\text{RF}} \times N_{\text{grid}}$ to $M_{\text{RF}} K_{\text{sbc}} \times N_{\text{grid}} K_{\text{grid}}$.

B. Prior work based on MUSIC

In conventional fully-digital architectures, the goal of the MUSIC algorithm is to estimate AoAs from the spatial channel covariance matrix. In other words, the covariance must be known prior to applying the MUSIC algorithms. Note that the spatial channel covariance can be estimated from the covariance of the received signal vectors in fully-digital architectures. Although the spatial channel covariance estimation is not straightforward in the hybrid architectures, the MUSIC algorithm can be applied to the subspace estimation problem for hybrid architectures. It is worthwhile to note that only the subspace can be estimated and the covariance cannot be estimated by using the MUSIC-based approach. Since the subspace is enough for the hybrid precoder design in some cases as in SU-MIMO systems, we will compare our proposed work with the MUSIC-based method in terms of subspace estimation. The overall process for the MUSIC-based method is composed of three steps. First, the sample covariance of the baseband received signal vectors $\mathbf{y}_{t,k}$ in (11) is estimated for all t and k as

$$\begin{aligned} \mathbf{R}_y &= \frac{1}{T_{\text{frm}} K_{\text{sbc}}} \sum_{t=1}^{T_{\text{frm}}} \sum_{k=1}^{K_{\text{sbc}}} \mathbf{y}_{t,k} \mathbf{y}_{t,k}^* \\ &= \frac{1}{T_{\text{frm}} K_{\text{sbc}}} \mathbf{W}^* \mathbf{A} (\mathbf{G}^* \mathbf{G} \odot \mathbf{C}^* \mathbf{C}) \mathbf{A}^* \mathbf{W} + \mathbf{R}_n, \end{aligned} \quad (72)$$

where $\mathbf{R}_n = \frac{1}{T_{\text{frm}} K_{\text{sbc}}} \left(\sum_{t=1}^{T_{\text{frm}}} \sum_{k=1}^{K_{\text{sbc}}} \mathbf{n}_{t,k} \mathbf{n}_{t,k}^* \right)$. Let the SVD of $\mathbf{R}_y$ be

$$\mathbf{R}_y = \mathbf{U}_x \Sigma_x \mathbf{U}_x^* + \mathbf{U}_n \Sigma_n \mathbf{U}_n^*, \quad (73)$$

where $\mathbf{U}_x \in \mathbb{C}^{M_{\text{RF}} \times L_{\text{ch}}}$ is the signal subspace and $\mathbf{U}_n \in \mathbb{C}^{M_{\text{RF}} \times (M_{\text{RF}} - L_{\text{ch}})}$ is the subspace orthogonal to the signal subspace. Let $\mathbf{b}_w(\phi) = \mathbf{W}^* \mathbf{a}(\phi)$. Since the RF chains can be regarded as the effective antennas from the viewpoint of the estimator at baseband, the vector $\mathbf{b}_w(\phi)$ can be considered as the *effective* array response vector with a reduced size. The second step is to find the L_{ch} highest peaks of the function of ϕ defined as

$$f_{\text{MUSIC}}(\phi) = \frac{1}{\mathbf{b}_w^*(\phi) \mathbf{U}_n \mathbf{U}_n^* \mathbf{b}_w(\phi)}. \quad (74)$$

The final step is to reconstruct the subspace of the channel by using the ϕ_ℓ 's for $\ell = 1, \dots, L_{\text{ch}}$ that are obtained from the subspace of the baseband received signals. The subspace of the channel is given by the subspace of $[\mathbf{a}(\phi_1) \ \dots \ \mathbf{a}(\phi_{L_{\text{ch}}})]$.

The CRLB for ϕ_ℓ 's in the MUSIC-based method case is given by [47]

$$\text{CRLB}_{\text{MUSIC}}(\boldsymbol{\phi}) = \frac{\sigma^2}{2} \left(\sum_{t=1}^{T_{\text{frm}}} \sum_{k=1}^{K_{\text{sbc}}} \text{Re} \left(\mathbf{Z}_{t,k}^* \hat{\mathbf{B}}^* \left(\mathbf{I} - \mathbf{B} (\mathbf{B}^* \mathbf{B})^{-1} \mathbf{B}^* \right) \hat{\mathbf{B}} \mathbf{Z}_{t,k} \right) \right)^{-1}, \quad (75)$$

where $\hat{\mathbf{B}}$ is defined in (43) and $\mathbf{Z}_{t,k}$ is defined as $\mathbf{Z}_{t,k} = \text{diag}([\mathbf{G}^T]_{:,t} \odot [\mathbf{C}^T]_{:,k})$. The bound of the RPE in the MUSIC-based method case can also be obtained from (37) as in the tensor-based method case.

VIII. SIMULATION RESULTS

In this section, we numerically evaluate the CRLB analysis in Section VI. We also present simulation results to demonstrate the performance of the proposed spatial channel covariance estimation algorithms based on CPD of higher-order tensors.

A. Analytical results on CRLB

In Fig. 2, we show the MSE of the estimation of $\boldsymbol{\phi}$. We also compare the MSE results to the CRLB derived in Section VI for $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $L_{\text{ch}} = 6$, $T_{\text{frm}} = 20$, and $K_{\text{sbc}} = 128$. We assume that the path gains $g_{t,\ell}$'s are generated from $\mathcal{CN}(0, 1/L_{\text{ch}})$ and $p_{\text{PS}}(\tau) = \text{sinc}(\tau/T_s)$. Since CRLB depends on the deterministic value of AoA and path delays, we set the values as $\boldsymbol{\phi} = [-66, 13, 49, -7, 81, 62]$ in degrees and $\boldsymbol{\tau}/T_s = [0, 4.34, 7.13, 17.05, 21.08, 25.73]$ for the purpose of reproduction. We can see that the proposed method achieves the MSE that is close to its theoretical lower bound at moderate and high SNR region.

Fig. 2(a) compares the proposed tensor-based method with the MUSIC-based method in terms of the $\text{MSE}(\boldsymbol{\phi})$. In addition to numerical results, the analytical results indicate the superiority of the tensor-based method over the MUSIC-based method. The metric $1 - \mathbb{E}[\eta]$ is plotted in Fig. 2(b) with the lower bound of its approximation derived in Section V. As shown in Section V, the RPE is closely related to $\text{MSE}(\boldsymbol{\phi})$ and its CRLB.

B. Performance evaluation of the spatial channel covariance estimation methods

In this subsection, we evaluate the performance of the spatial channel covariance estimation in terms of the RPE for $N_{\text{ant}} = 64$ and $K_{\text{sbc}} = 128$. Unlike Section VIII-A, the AoA ϕ_ℓ 's are uniformly distributed in $[-180^\circ, 180^\circ]$, and the normalized delay τ_ℓ/T_s 's are uniformly distributed in $[0, N_{\text{CP}}]$, where the cyclic prefix length N_{CP} is set to $K_{\text{sbc}}/4$. The path gains $g_{t,\ell}$'s are IID complex Gaussian random variables as $g_{t,\ell} \sim \mathcal{CN}(0, 1/L_{\text{ch}})$ and $p_{\text{PS}}(\tau) = \text{sinc}(\tau/T_s)$.

Fig. 3 shows the comparison among different initialization strategies for ALS. The first option is to use random matrices. For comparison, we also include the case where multiple random initializations are used and the best result is kept. The second option is to use the direct trilinear decomposition (DTLD) method [48]. Finally, we also tested the cpd

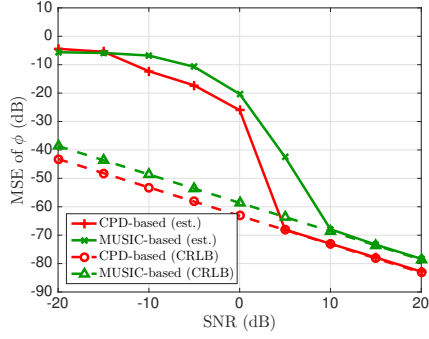
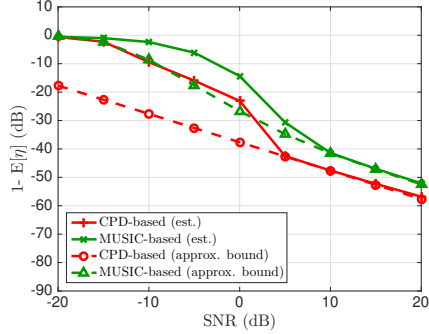
(a) MSE of ϕ (b) $1 - \mathbb{E}[\eta]$

Fig. 2. Comparison between the proposed CPD-based method and the MUSIC-based method in terms of the MSE of ϕ and $1 - \mathbb{E}[\eta]$ when $T_{\text{frm}} = 20$, $K_{\text{sbc}} = 128$, $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, and $L_{\text{ch}} = 6$.

function included in Tensorlab MATLAB package [49]. The function automatically computes an initialization and uses a more sophisticated algorithm than the ALS. Simulation results show that all initialization strategies have similar performance except a single random initialization. Fig. 3(b) shows the computation time for different methods when simulations are conducted with MATLAB R2018b in 2.6 GHz Intel Core i7. Although the actual computation time depends on code optimization, Fig. 3(b) shows the relative computational complexity. Fig. 3(a) and Fig. 3(b) show the DTLD-based initialization can be a reasonable solution considering the trade-off between performance and complexity, and we used the DTLD-based initialization for experiments in the paper.

Fig. 4(a) shows a snapshot of the convergence of the ALS with the DTLD-based initialization. The normalized mean squared error (NMSE) in dB is defined as $10 \log_{10} \left(\frac{\|\mathcal{Y} - \hat{\mathcal{X}}\|^2}{\|\mathcal{Y}\|^2} \right)$ where $\mathcal{Y}$ is the input of the ALS and $\hat{\mathcal{X}}$ is the tensor reconstructed from the estimated factor matrices. It can be seen that the convergence rate of the DTLD initialization is faster than that of the random initialization. To speed up simulations, the ALS algorithm terminates if the difference between the NMSE of the current iteration and that of the previous iteration is less than 0.001 dB.

In Fig. 4(b), different options for a sensing matrix $\mathbf{W}$ are compared. As described in Section III-B, we used the hybrid combiner $\mathbf{W} = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$ where $\mathbf{W}_{\text{RF}}$ is an analog combiner composed of random phase shifters and $\mathbf{W}_{\text{BB}}$ is a baseband compensation matrix defined as $\mathbf{W}_{\text{BB}} = (\mathbf{W}_{\text{RF}}^* \mathbf{W}_{\text{RF}})^{-\frac{1}{2}}$. The

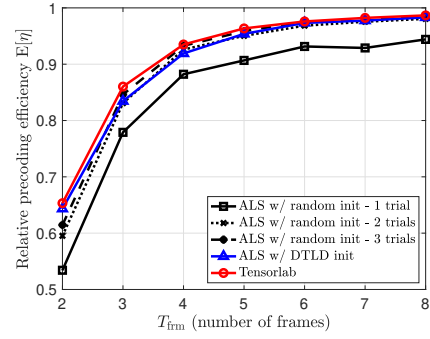
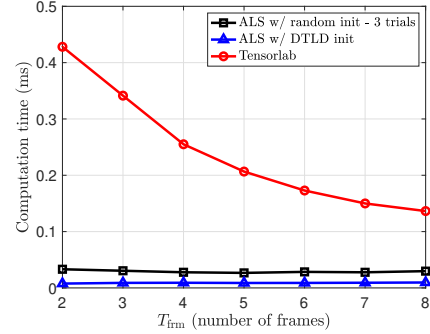
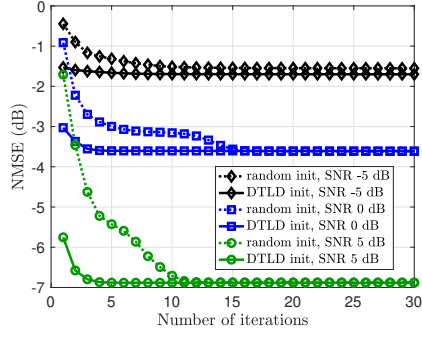
(a) RPE vs. T_{frm} (b) Computation time vs. T_{frm}

Fig. 3. Comparison among different initialization methods when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 4$, $L_{\text{ch}} = 3$, SNR = 0 dB, and $K_{\text{sbc}} = 128$.

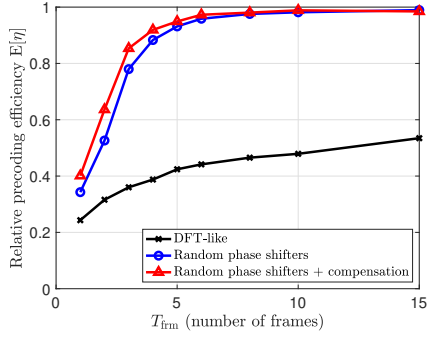
figure shows that the use of the baseband compensation matrix $\mathbf{W}_{\text{BB}}$ yield some gain compared to using only the analog combiner $\mathbf{W}_{\text{RF}}$. For comparison, we also add the case where the M_{RF} columns of the sensing matrix are randomly selected in an $N_{\text{ant}} \times N_{\text{ant}}$ DFT matrix. Since this DFT-based sensing matrix focuses only on some directions and disregards other directions, this DFT-based method is not a good option as shown in the figure.

Fig. 5 compares the proposed method with other methods explained in Section VII in terms of the RPE when $M_{\text{RF}} = 8$ and SNR 0 dB. In the figure, we compare the methods for different L_{ch} values: 6 (solid) and 7 (dashed). It is worthwhile to note that the MUSIC-based method does not work properly if $L_{\text{ch}} \geq M_{\text{RF}}$ because $\mathbf{U}_{\text{n}}$ in (74) must have at least one column. Even when $L_{\text{ch}} < M_{\text{RF}}$, the figure shows that the performance degradation of the MUSIC-based method is more severe than other two methods as L_{ch} approaches M_{RF} . The figure also shows several different types of CS-based methods. WB-DCOMP [22] is the extension of MMV by stacking more measurement vectors, and “FBS + Kron.” is the extension of the FBS with Nestrov’s update [24] by using Kronecker product of two dictionary matrices as described in Section VII-A. The “(fixed)” or “(varying)” denote that the combining matrix $\mathbf{W}$ are fixed or varying over time. The figures show that the proposed CPD-based method outperforms various types of the CS-based methods.

Fig. 6 shows the case of SNR -10 dB where other simulation parameters are the same as Fig. 5. It can be seen that the Kronecker-product-based extension of FBS is robust to noise



(a) Convergence of ALS



(b) Sensing matrix comparison when SNR=0 dB

Fig. 4. Convergence of ALS and comparison among different sensing matrices and when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $L_{\text{ch}} = 6$, and $K_{\text{sbc}} = 128$.

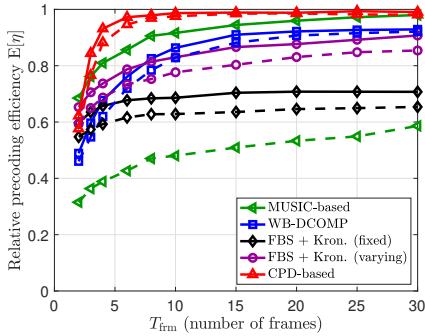


Fig. 5. Comparison with prior work in terms of RPE when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $K_{\text{sbc}} = 128$, $L_{\text{ch}} = 6$ or 7, and SNR 0 dB.

while the MUSIC-based method and WB-DCOMP suffer from high noise power. It is worthwhile to note that the use of a time-varying combiner yields considerable gain compared to the use of a fixed combiner. The figure shows that the extended CS-based method with a time-varying combiner outperforms the proposed CPD-based method at this very low SNR region. A promising point is that the proposed CPD-based method uses a fixed combining matrix, which implies that there is still room for improvement. Since the extension of the CPD-based method to the time-varying combining case is not straightforward, we leave this issue for future work.

Until now, we assumed that the channel has only L_{ch} channel paths as in (8). To evaluate performance for more realistic channels, we consider a clustered channel model [41]

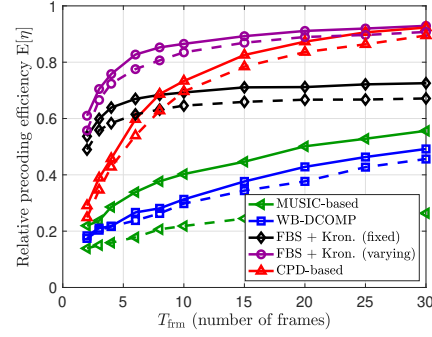


Fig. 6. Comparison with prior work in terms of RPE when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $K_{\text{sbc}} = 128$, $L_{\text{ch}} = 6$ or 7, and SNR -10 dB.

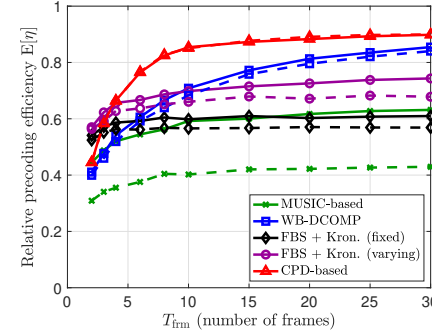


Fig. 7. RPE vs. the number of frames T_{fm} when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $K_{\text{sbc}} = 128$, and SNR 0 dB. The channel has $L_{\text{cluster}} = 6$ or 7 clusters, and each cluster has $L_{\text{subray}} = 10$ subrays with angular spread 2° .

that has multiple clusters with multiple subrays as $\mathbf{h}_t[d] = \sum_{\ell_c=1}^{L_{\text{cluster}}} \sum_{\ell_s=1}^{L_{\text{subray}}} g_{t,\ell_c,\ell_s} \text{PPS}(dT_s - \tau_{\ell_c,\ell_s}) \mathbf{a}(\phi_{\ell_c,\ell_s})$. The channel has L_{cluster} clusters whose AoAs are uniformly distributed in $[-180^\circ, 180^\circ]$, and each cluster has L_{subray} subrays whose AoA offsets are Laplacian distributed with angular spread 2° . Although the rank of the channel tensor $L_{\text{cluster}} L_{\text{subrays}}$ becomes high in general, we can approximate the channel tensor to be a low-rank tensor for spatially sparse channels. Fig. 7 shows the RPE results when $N_{\text{ant}} = 64$, $M_{\text{RF}} = 8$, $K_{\text{sbc}} = 128$, $L_{\text{subray}} = 10$, SNR= 0 dB, and $L_{\text{cluster}} = 6$ or 7. Instead of using the actual number of channel paths $L_{\text{ch}} = L_{\text{cluster}} L_{\text{subray}}$, the methods based on CPD and CS use M_{RF} for its low-rank (or sparse) approximation while the MUSIC-based method uses L_{cluster} due to its inherent limitation of using M_{RF} . The figure shows that, although the multiple subrays result in performance loss compared to the single subray case, the proposed method still works properly even in this case.

IX. CONCLUSIONS

In this paper, we proposed a spatial channel covariance estimation method for the hybrid analog/digital architecture over time-varying frequency-selective channels. Leveraging the fact that a low-rank higher-order tensor can be uniquely decomposed into factor matrices in each domain, we formulated the estimation problem by using high-order tensors and proposed a solution that achieves performance close to its

theoretical bound. We also derived the CRLB of the proposed method and showed that compared it is lower than the CRLB of MUSIC-based approach. Numerical results showed that the proposed method also work properly even though the actual number of channel paths is not so small and larger than the number of RF chains. The results also show the possibility of further improvement by using time-varying combiners, which we leave for future work.

REFERENCES

- [1] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [2] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [3] R. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [4] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [5] X. Zhang, A. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.
- [6] V. Venkateswaran and A. J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.
- [7] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [8] J. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmwave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.
- [9] S. Park, A. Alkhateeb, and R. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.
- [10] S. Park, J. Park, A. Yazdan, and R. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.
- [11] R. Méndez-Rial, N. González-Prelcic, and R. Heath, "Adaptive hybrid precoding and combining in mmwave multiuser MIMO systems based on compressed covariance estimation," in *Proc. IEEE Int. Workshop on Comp. Adv. in Multi-Sensor Adap. Proc. (CAMSAP)*, Dec. 2015, pp. 213–216.
- [12] D. Romero and G. Leus, "Compressive covariance sampling," in *Proc. Inf. Theory and Appl. Workshop (ITA)*, Feb 2013, pp. 1–8.
- [13] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, June 2012, pp. 525–529.
- [14] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, 2012.
- [15] P. P. Vaidyanathan and P. Pal, "Theory of sparse coprime sensing in multiple dimensions," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3592–3608, Aug 2011.
- [16] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [17] Y. Peng, Y. Li, and P. Wang, "An enhanced channel estimation method for millimeter wave systems with massive antenna arrays," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1592–1595, Sep. 2015.
- [18] J. Lee, G. Gil, and Y. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [19] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, July 2005.
- [20] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
- [21] J. Determe, J. Louveau, L. Jacques, and F. Horlin, "On the exact recovery condition of simultaneous orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 164–168, Jan 2016.
- [22] S. Park and R. Heath, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing-based approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8047–8062, Dec 2018.
- [23] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Jan 2017.
- [24] S. Haghighatshoar and G. Caire, "Low-complexity massive MIMO subspace estimation and tracking from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1832–1844, Apr. 2018.
- [25] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [26] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [27] S. Chuang, W. Wu, and Y. Liu, "High-resolution AoA estimation for hybrid antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 63, no. 7, pp. 2955–2968, July 2015.
- [28] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.
- [29] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. 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.
- [30] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an n -way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sep. 1970.
- [31] R. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Pap. Phonet.*, vol. 16, pp. 1–84, 1970.
- [32] F. Roemer, M. Haardt, and G. Del Galdo, "Analytical performance assessment of multi-dimensional matrix- and tensor-based ESPRIT-type algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2611–2625, May 2014.
- [33] L. Cheng, Y. Wu, J. Zhang, and L. Liu, "Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5897–5909, Nov 2015.
- [34] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [35] A. Cichocki, D. Mandic, A. Phan, C. Caiafa, G. Zhou, Q. Ahao, and L. Lathauwer, "Tensor decompositions for signal processing applications: from two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- [36] N. Sidiropoulos, L. Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [37] S. Park, A. Ali, N. González-Prelcic, and R. Heath, "Spatial channel covariance estimation for the hybrid architecture at a base station: A tensor-decomposition-based approach," in *Proc. IEEE Global Conf. on Signal and Info. Process (GlobalSIP)*, Nov. 2018, pp. 1008–1012.
- [38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [39] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. Asilomar Conf. Signals, Sys. and Comput.*, Nov. 2014, pp. 273–277.
- [40] A. Alkhateeb and R. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [41] P. Kyosit et al. IST-4-027756 WINNER II D1.1.2 V1.2 WINNER II Channel Models. [Online]. Available: www.ist-winner.org
- [42] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [43] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

- [44] B. D. Rao and K. V. S. Hari, "Performance analysis of Root-MUSIC," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
- [45] O. Ayach, R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2012, pp. 100–104.
- [46] X. Liu and N. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
- [47] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [48] E. Sanchez and B. Kowalski, "Tensor resolution: a direct trilinear decomposition," *J. Chemometrics*, vol. 4, no. 1, pp. 29–45, Jan. 1990.
- [49] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. (2016, Mar.) Tensorlab 3.0. Available online. [Online]. Available: <https://www.tensorlab.net>