Training Signal Design for Sparse Channel Estimation in Intelligent Reflecting Surface-Assisted Millimeter-Wave Communication

Song Noh, *Member, IEEE*, Heejung Yu, *Senior Member, IEEE*, and Youngchul Sung, *Senior Member, IEEE*

***Abstract*— In this paper, the problem of training signal design for intelligent reflecting surface (IRS)-assisted millimeter-wave (mmWave) communication under a sparse channel model is considered. The problem is approached based on the Cramér-Rao lower bound (CRB) on the mean-square error (MSE) of channel estimation. By exploiting the sparse structure of mmWave channels, the CRB for the channel parameter composed of path gains and path angles is derived in closed form under Bayesian and hybrid parameter assumptions. Based on the derivation and analysis, an IRS reflection pattern design method is proposed by minimizing the CRB as a function of design variables under constant modulus constraint on reflection coefficients. Extensions of the proposed design to a multi-antenna transceiver, a uniform planar array (UPA)-based IRS, and multi-user case are discussed. Numerical results validate the effectiveness of the proposed design method for sparse mmWave channel estimation.**

***Index Terms*— Millimeter wave (mmWave) communication, intelligent reflecting surface (IRS), signal design, sparse channel estimation, Cramer-Rao bounds.**

I. Introduction

IRSs have gained much attention as one of the potential technologies for 6G under various names such as reconfigurable intelligent surfaces (RISs), reflector-arrays or intelligent walls. The objective of communication system design up to now was to design optimal signal waveforms and encoding/decoding schemes for given wireless channels, but IRSs have changed the paradigm of this conventional communication system design. IRSs aim to realize intelligent wireless channels by controlling the radio propagation model rather than optimizing transmission-and-reception schemes under given channels [2], [3]. An IRS consists of an array of passive scattering elements and the phase and/or other characteristics of the signal reflected by each element is controlled. The controllable radio propagation model is beneficial for wireless communication to improve signal quality and coverage [4], [5]. In particular, IRSs can be very useful in mmWave channels since the propagation is directive and the propagation loss is large in this band. IRSs can provide diversity paths to improve the link quality to fully harness the large bandwidth available in this band [6], [7]. In this regard, the authors in [8] designed a transmit precoder and the IRS reflection pattern for data transmission in mmWave channels. For such design of transmit beamformer and IRS phase shifters, channel state information (CSI) is required and most previous works assume that perfect CSI of the IRS and the direct link is available. In practice, however, channels should be estimated. Channel estimation in IRS-assisted communication is not simple because an IRS is composed of passive reflecting elements which cannot send their own pilot signal and this fact makes the available channel estimation methods devised for relay channels not directly applicable. In this paper, we consider mmWave communication, which can potentially get much benefit from IRSs, and investigate training signal design for enhanced channel estimation in IRS-assisted mmWave communication.

A. Related Works and Contributions

We here focus on works on channel estimation in IRS-assisted communication relevant to our work. One line of approaches to channel estimation in IRS-assisted systems is to estimate individual channels, i.e., one from the transmitter to the IRS and the other from the IRS to the receiver [9], [10]. Another approach is to estimate the cascaded channel composed of the transmitter-to-IRS and IRS-to-receiver channels by treating the cascaded channel as a single effective channel [11]–[14] (this approach is adopted in this paper). In [11], the authors proposed an one-by-one channel estimation protocol in which only a single IRS element is activated while other elements remain off in each step. In [12], the authors investigated a cascaded channel framework, exploiting sparse matrix factorization and matrix completion techniques in a downlink massive multiple-input multiple-output (MIMO) system. Under the assumption of a rank-deficient channel, the pilot symbols and the IRS reflection pattern having on-off states are generated randomly by using Gaussian and Bernoulli distributions, respectively. In [13], the authors designed an

Manuscript received December 28, 2020; revised July 4, 2021; accepted September 7, 2021. Date of publication September 21, 2021; date of current version April 11, 2022. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) under Grant NRF-2019R1G1A1009021 and Grant NRF-2020R1F1A1072761. A preliminary version of this work was presented in ICC Workshops 2021 [1], in which only the hybrid CRB-based approach is considered for a single antenna receiver. In this paper, the Bayesian CRB-based approach and several extensions of the proposed design are included with extensive simulations. The associate editor coordinating the review of this article and approving it for publication was M. Payaró. *(Corresponding author: Youngchul Sung.)*

Song Noh is with the Department of Information and Telecommunication Engineering, Incheon National University, Incheon 22012, South Korea (e-mail: songnoh@inu.ac.kr).

Heejung Yu is with the Department of Electronics and Information Engineering, Korea University, Sejong 30019, South Korea (e-mail: heejungyu@korea.ac.kr).

Youngchul Sung is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: ysung@ee.kaist.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3112173.

Digital Object Identifier 10.1109/TWC.2021.3112173

optimal channel estimation scheme based on minimization of the Cramér-Rao lower bound (CRB) under a dense channel model and showed that an optimal IRS activation patterns is given by the columns of the discrete Fourier transform (DFT) matrix. In [14], a deep learning based method is proposed to extrapolate the CSI from a partial channel measurement.

The majority of the existing works assume that the number of training symbols for channel estimation is large enough compared to the number of (grouped) IRS reflecting elements [9]–[13]. However, it can be difficult to satisfy this assumption in practical systems because an IRS typically adopts a large number of reflecting elements to achieve high passive beamforming gain. In this paper, two approaches to CRB-based training signal design are investigated for enhanced sparse channel estimation. The Bayesian CRB provides a lower bound on the channel estimation MSE of any Bayesian channel estimator capable of exploiting the prior knowledge about the channel parameter [15] in which the path gains and angles are modeled as random parameters. Then, the problem of training signal design is formulated by minimizing the Bayesian CRB that is independent of parameter realizations. The hybrid CRB determines a lower bound on the MSE by taking into account the case that such prior knowledge is not available especially for the path angle. In this approach, the path gains are modeled as random parameters whereas the path angles are considered as unknown deterministic parameters. Thus, the hybrid CRB depends on specific realization of the path angle and enables training signal design relevant to a set of fixed target angles. Based on this result, the performance for sparse channel estimation can be improved by using the proposed design at the IRS during the training period. The contributions of this paper are summarized as follows.

- In the Bayesian parameter case, we derived a condition to minimize the Bayesian CRB and showed that the Fisher information "density" for angle estimation is given by a quadratic form composed of the derivative of the array response and the IRS reflection pattern matrix square, and this determines the quality of angle estimation across the angle domain.
- In the hybrid parameter case, treating the path gains as random nuisance parameters, we formulated the problem of optimal IRS reflection pattern matrix design by minimizing the CRB on angle estimation, and solved this optimization problem by using the projected gradient method (PGM) [16], [17].
- Extensions of the proposed design are discussed, including a multi-antenna transceiver, a uniform planar array (UPA)-based IRS, and multi-user case.
- With numerical evaluation, we demonstrated that the proposed design method exploiting channel sparsity yields noticeable gain in sparse mmWave channels when the number of training symbols is less than the number of the reflecting elements in the IRS.

B. Notations

Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^\dagger$ indicate the transpose, Hermitian transpose, and Moore-Penrose inverse of $\mathbf{A}$, respectively. $\mathbf{A} \succeq 0$ means that the matrix $\mathbf{A}$ is a Hermitian positive semidefinite matrix. $[\mathbf{A}]_{i,j}$ denotes the (i,j)-th element of $\mathbf{A}$. $[\mathbf{A}]_{i_1:i_2,j_1:j_2}$ denotes the submatrix of $\mathbf{A}$ given by the intersection of rows $i_1,\ldots,i_2$ and columns $j_1,\ldots,j_2$. For a vector $\mathbf{a}$, $[\mathbf{a}]_i$ denotes the i-th element of $\mathbf{a}$, and $\|\mathbf{a}\|$ represents vector ℓ_2-norm. $\mathbf{0}_n$ and $\mathbf{1}_n$ are the zero vector of length n and the vector of length n composed of all one elements, respectively. $\mathbf{e}_n(i)$ denotes the i-th column vector of the identity matrix of size $n \times n$. $\mathbf{I}_n$ is the identity matrix of size $n \times n$. $\mathrm{diag}(\mathbf{A})$ denotes a vector of the elements on the diagonal of $\mathbf{A}$ and $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix whose diagonal is $\mathbf{a}$. $\mathbb{E}\{\cdot\}$ denotes statistical expectation. $\mathbf{a} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that random vector $\mathbf{a}$ is complex circularly-symmetric Gaussian distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{a} \sim \mathrm{Unif}[a,b]$ means that the elements of $\mathbf{a}$ are randomly and uniformly distributed over the interval $[a,b]$. $\mathbb{C}$ denotes the set of complex numbers. The symbol $\odot$ and $\otimes$ denote the Hadamard product and the Kronecker product, respectively. $\mathrm{Re}(\cdot)$ $\mathrm{Im}(\cdot)$ denote the real part and the imaginary part, respectively. $\mathrm{tr}(\cdot)$ denotes the trace operator. $\mathrm{vec}(\cdot)$ denotes the vectorization to stack the columns of the input matrix. $\iota := \sqrt{-1}$.

II. System Model

Consider an IRS system composed of a transmitter, an IRS, and a receiver, as shown in Fig. 1. We assume that the transmitter and the receiver have a single antenna and the IRS has a uniform linear array (ULA) of N passive reflecting elements. In Fig. 1, there exists a direct channel path from the transmitter to the receiver, where the direct-path channel gain is denoted by $\alpha_0 \in \mathbb{C}$, and there exist channel links between the transmitter and the IRS $\mathbf{h}^{(t)} = [h_1^{(t)},\ldots,h_N^{(t)}]^T \in \mathbb{C}^N$ and between the IRS and the receiver $\mathbf{h}^{(r)} = [h_1^{(r)},\ldots,h_N^{(r)}]^T \in \mathbb{C}^N$. The considered IRS system operates in the mmWave band and hence adopt a geometry-based sparse channel model relevant to mmWave communication [7], [8], [18]–[20]. Thus, the channel vectors of $\mathbf{h}^{(t)}$ and $\mathbf{h}^{(r)}$ are modelled as $\mathbf{h}^{(t)} = \sum_{i=1}^{L^{(t)}} \alpha_i^{(t)} \mathbf{u}_N(\psi_i^{(t)})$ and $\mathbf{h}^{(r)} = \sum_{j=1}^{L^{(r)}} \alpha_j^{(r)} \mathbf{u}_N(\psi_j^{(r)})$, where $\alpha_i^{(t)}$ and $\psi_i^{(t)}$ are the gain and angle-of-arrival (AoA) of the i-th path of the channel from the transmitter to the IRS, $\alpha_j^{(r)}$ and $\psi_j^{(r)}$ are the gain and angle-of-departure (AoD) of the j-th path of the channel from the IRS to the receiver, $L^{(t)}$ and $L^{(r)}$ are the numbers of multi-paths of $\mathbf{h}^{(t)}$ and $\mathbf{h}^{(r)}$, respectively, and $\mathbf{u}_N(\psi)$ is the array response vector of an N-element ULA, given by $\mathbf{u}_N(\psi) = \left[1, e^{\iota 2\pi \frac{d}{\lambda}\psi}, e^{\iota 2\pi \frac{d}{\lambda} 2\psi}, \ldots, e^{\iota 2\pi \frac{d}{\lambda}(N-1)\psi}\right]^T$. Here, ψ is the normalized path angle given by $\psi = \sin(\phi)$, d is the spacing between adjacent reflecting elements at the IRS, λ is the carrier wavelength, and $\phi \in [-\pi/2, \pi/2]$ is the unnormalized physical path angle. In particular, we have $\psi \in [-1,1]$ with critical sampling $d = \frac{\lambda}{2}$.[1]

We consider training-based channel estimation with a training period composed of K symbols in the beginning of each

[1]With $d = \frac{\lambda}{2}$, the visible region equals to one Nyquist interval $\omega := 2\pi \frac{d}{\lambda}\psi \in [-\pi, \pi]$ covering the entire circle [21].

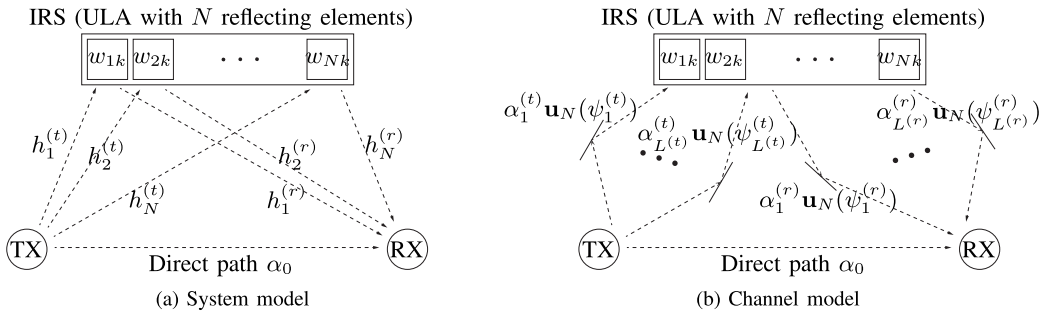


Fig. 1. The considered IRS-assisted point-to-point communication link.

channel coherence interval. At the k-th symbol time during the training period (i.e., $1 \le k \le K$), the transmitter sends a training symbol s_k with $|s_k|^2 = \rho$, and the N passive reflecting elements at the IRS reflect the incoming signal from the transmitter with a complex gain vector $\mathbf{w}_k = [w_{1k}, \ldots, w_{Nk}]^T$ with $|w_{nk}| = \sqrt{\gamma}$ for some $\gamma \in (0, 1]$. Then, the received signal y_k at the receiver at the k-th symbol time is given by [7], [8], [20]

$$\begin{aligned} y_k &= \left(\alpha_0 + \sum_{n=1}^{N} w_{nk}^* h_n^{(t)} h_n^{(r)}\right) s_k + n_k \\ &= \left(\alpha_0 + \mathbf{w}_k^H \left(\mathbf{h}^{(t)} \odot \mathbf{h}^{(r)}\right)\right) s_k + n_k, \end{aligned} \tag{1}$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is circularly-symmetric complex Gaussian noise. The IRS-assisted cascaded channel vector $\mathbf{h}$ from the transmitter to the receiver is defined as

$$\begin{aligned} \mathbf{h} := \mathbf{h}^{(t)} \odot \mathbf{h}^{(r)} &= \sum_{i=1}^{L^{(t)}} \sum_{j=1}^{L^{(r)}} \alpha_i^{(t)} \alpha_j^{(r)} \mathbf{u}_N\left(\psi_i^{(t)} + \psi_j^{(r)}\right) \\ &= \sum_{\ell=1}^{L} \alpha_\ell \mathbf{u}_N(\psi_\ell) = \mathbf{U}_{\boldsymbol{\psi}} [\alpha_1, \ldots, \alpha_L]^T, \end{aligned} \tag{2}$$

where $L = L^{(t)} L^{(r)}$ denotes the number of paths in the cascaded channel $\mathbf{h}$, α_ℓ $(= \alpha_i^{(t)} \alpha_j^{(r)})$ and ψ_ℓ $(= \psi_i^{(t)} + \psi_j^{(r)})$ denote the effective gain and angle of the ℓ-th path of the cascaded channel $\mathbf{h}$, respectively, and $\mathbf{U}_{\boldsymbol{\psi}} = [\mathbf{u}_N(\psi_1), \ldots, \mathbf{u}_N(\psi_L)] \in \mathbb{C}^{N \times L}$. Then, based on (1), the received signal during the entire training period can be written in vector form as

$$\begin{aligned} \mathbf{y} &= \mathbf{S} \underbrace{\begin{bmatrix} \mathbf{1}_K & \mathbf{W}^H \end{bmatrix}}_{=: \widetilde{\mathbf{W}}^H} \begin{bmatrix} \alpha_0 \\ \mathbf{h} \end{bmatrix} + \mathbf{n} \\ &= \mathbf{S} \widetilde{\mathbf{W}}^H \begin{bmatrix} 1 & \mathbf{0}_L^T \\ \mathbf{0}_N & \mathbf{U}_{\boldsymbol{\psi}} \end{bmatrix} \boldsymbol{\alpha} + \mathbf{n}, \end{aligned} \tag{3}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T$, $\mathbf{n} = [n_1, n_2, \ldots, n_K]^T$, $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_K)$, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{N \times K}$, and $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_L]^T$. We will refer to $\mathbf{W}$ as the *reflection pattern matrix* used at the IRS during the training period.

The goal of channel estimation is to obtain the channel parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_L]^T$ based on the received signal $\mathbf{y}$. Once the channel $\mathbf{h}$ is estimated during the training period, some control information can be sent from the receiver to the IRS to match the reflection coefficients $\{w_{ij}\}$ to $\mathbf{h}$ in order to maximize the data rate or other desired performance measure based on the received signal model (1) during the data transmission period.

Remark 1: Due to the existence of the direct path, the signal model (3) incorporating the channel sparsity in the mmWave band has a special structure. It is not a simple linear model in terms of the unknown parameters $\{\alpha_0, \alpha_1, \ldots, \alpha_L, \psi_1, \ldots, \psi_L\}$ or a signal model associated with simple AoA estimation. It is a mixed nonlinear model in terms of the channel path-gain parameters $\alpha_0, \alpha_1, \ldots, \alpha_L$ and the path-angle parameters $\psi_1, \ldots, \psi_L$.

For the rest of this paper, we consider the problem of optimal training signal design under the signal model (3). Here, we have two design variables $\mathbf{S}$ and $\mathbf{W}$, where $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_K)$ is the training symbol sequence at the transmitter and $\mathbf{W}$ is the IRS reflection pattern matrix during the training period. Then, the goal is to design the training sequence $\mathbf{S}$ and the IRS reflection pattern matrix $\mathbf{W}$ so that channel estimation based on the designed $\mathbf{S}$ and $\mathbf{W}$ yield best performance under a reasonable criterion.

Remark 2: Previous IRS reflection pattern designs under sparse scattering propagation include sparse random vector [12] for sparse representation of the indirect channel, any fixed unit-modulus vector [20] by approximating the sparse channel with virtual channel representation, and the IRS array manifold vectors [22] to maximize passive beamforming gain. Note that for sparse channel estimation, the error covariance matrix is not derived analytically in general [23] and the numerically obtained mean square error (MSE) is algorithm-specific. Therefore, it is not straightforward to establish the MSE-based training signal design.

Unlike the previous works in which the training signal is designed to facilitate an associated channel estimation algorithm, in this paper we investigate a systematic way to design training signal that is independent of any particular algorithm by employing the CRB as the optimality criterion. For the given parameter $\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\psi}^T]^T$, where $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_L]$ and $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_L]$, the CRB is analyzed as a function of training signal [24]:

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}} \{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}))^H\} \succeq \mathbf{J}(\boldsymbol{\theta})^{-1}, \tag{4}$$

where $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is any unbiased estimator of $\boldsymbol{\theta}$ and $\mathbf{J}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM) given by the covariance

matrix of the score function $\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, i.e., $\mathbf{J}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\{\big(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)\big(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)^H\}$. Then, the training signal is optimized by minimizing the CRB-based cost function, which is discussed in details in the following sections.

III. Analysis From Bayesian CRB

The derivation of FIM and corresponding CRB depends on the assumption on the parameter $\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\psi}^T]^T$. In this section, to gain insights into the IRS reflection pattern design with analytical tractability in the sparse channel case, we consider the Bayesian approach that assumes a prior distribution on $\boldsymbol{\theta}$ and considers the MSE performance averaged over the prior distribution on $\boldsymbol{\theta}$. The Bayesian CRB is obtained by taking expectation $\mathbb{E}_{\boldsymbol{\theta}}$ over $p(\boldsymbol{\theta})$ on both sides of (4) as

$$\mathbb{E}_{\boldsymbol{\theta}}\{\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{y}))(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{y}))^H\}\} \overset{(a)}{\succeq} \mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{J}(\boldsymbol{\theta})^{-1}\} \overset{(b)}{\succeq} \left(\mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{J}(\boldsymbol{\theta})\}\right)^{-1}, \quad (5)$$

where $\mathbf{J}(\boldsymbol{\theta})$ is defined in (4) and Step (b) is valid due to Jensen's inequality on positive-definite matrix inverse. Here, it is not simple to derive $\mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{J}(\boldsymbol{\theta})^{-1}\}$ analytically since the inverse of $\mathbf{J}(\boldsymbol{\theta})$ may not be simplified due to possible non-zero off-diagonal elements depending on the realization of $\boldsymbol{\theta}$. Instead, the inverse of the averaged FIM (i.e., Bayesian FIM) $\mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{J}(\boldsymbol{\theta})\}$ provides a tractable lower bound on the channel estimation MSE averaged over both $\mathbf{y}$ and $\boldsymbol{\theta}$.

A. Derivation of Bayesian CRB

In high-frequency channels such as mmWave channels, the gain and angle of a path are typically uncorrelated [25]–[27]. Hence, we assume that $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ are statistically independent with marginal distributions $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\psi})$, respectively. From the signal model (3) with additive Gaussian noise, the joint probability density function (pdf) is written as

$$\begin{aligned} p(\mathbf{y},\boldsymbol{\theta}) &= p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\psi}) = p(\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\psi})p(\boldsymbol{\alpha})p(\boldsymbol{\psi}) \\ &= \left(\pi\sigma_n^2\right)^{-K}\exp\left(-\left\|\mathbf{y}-\mathbf{m}\right\|^2/\sigma_n^2\right)p(\boldsymbol{\alpha})p(\boldsymbol{\psi}), \quad (6) \end{aligned}$$

where $\mathbf{m} := \mathbf{S}\widetilde{\mathbf{W}}^H\left[\alpha_0, \mathbf{h}^T\right]^T$, as seen in (3). Note that $\boldsymbol{\alpha}$ is complex-valued whereas $\boldsymbol{\psi}$ is real-valued. Hence, it is convenient to consider the real-valued version $\tilde{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$, defined as $\tilde{\boldsymbol{\theta}} := \left[\mathrm{Re}\{\boldsymbol{\alpha}^T\}, \mathrm{Im}\{\boldsymbol{\alpha}^T\}, \boldsymbol{\psi}^T\right]^T$. By using the property of the mapping $\mathbf{x} \mapsto \begin{bmatrix}\mathrm{Re}\{\mathbf{x}\}\\ \mathrm{Im}\{\mathbf{x}\}\end{bmatrix}$ for a complex vector $\mathbf{x}$, the Bayesian CRB for $\tilde{\boldsymbol{\theta}}$ is given by [28]

$$\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\left(\tilde{\boldsymbol{\theta}}-g(\mathbf{y})\right)\left(\tilde{\boldsymbol{\theta}}-g(\mathbf{y})\right)^H\right\} \succeq \mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}, \quad (7)$$

for any unbiased channel estimator $g(\mathbf{y})$ for $\tilde{\boldsymbol{\theta}}$, where $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ is the real-valued-version Bayesian FIM obtained by taking expectation over $\mathbf{y}$ and $\boldsymbol{\theta}$ [28]:

$$\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}} = \tilde{\mathbf{M}}\begin{bmatrix}\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}^*} & \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi}}\\ \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}} & \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*} & \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\psi}}\\ \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha}} & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha}^*} & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}\end{bmatrix}\tilde{\mathbf{M}}^H, \text{ where}$$

$$\tilde{\mathbf{M}} = \begin{bmatrix}\mathbf{M} & \mathbf{0}_{2(L+1)}\mathbf{0}_L^T\\ \mathbf{0}_L\mathbf{0}_{2(L+1)}^T & \mathbf{I}_L\end{bmatrix},$$

$$\mathbf{M} = \frac{1}{2}\begin{bmatrix}\mathbf{I}_{L+1} & \mathbf{I}_{L+1}\\ -\iota\mathbf{I}_{L+1} & \iota\mathbf{I}_{L+1}\end{bmatrix}, \quad (8)$$

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial}{\partial\boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\psi})}{\partial\boldsymbol{\alpha}^*}\right)^H\right\},$$

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}^*} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial}{\partial\boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\psi})}{\partial\boldsymbol{\alpha}}\right)^H\right\}, \quad (9)$$

and $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi}}$, $\mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\psi}}$, $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha}}$, $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha}^*}$ and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ are similarly defined by considering that $\boldsymbol{\psi}$ is a real-valued vector. Here, $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}$ denotes the expectation with respect to the joint pdf $p(\mathbf{y},\boldsymbol{\theta}) = p(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\psi})$, $\mathbf{M}$ in (8) is the complex-to-real conversion matrix [28], [29], and $\frac{\partial}{\partial\boldsymbol{\alpha}}$ and $\frac{\partial}{\partial\boldsymbol{\alpha}^*}$ in (9) are the Wirtinger complex derivative [30]. Note that the Bayesian FIM does not depend on the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, and the parameter subscript in $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ is for matrix partition purpose.

To enable derivation of Bayesian CRB with still incorporating many meaningful distributions [31], [32], we assume that the direct path gain α_0 has non-zero mean μ_0, the reflected path gains α_ℓ, $\ell = 1,\ldots,L$ have zero mean,[2] and all the path gains α_ℓ, $\ell = 0,1,\ldots,L$ have the same finite second-order central moment σ^2, and that the path angle is uniformly distributed over $[\Delta_1,\Delta_2] \subset [-1,1]$, i.e., $\psi_\ell \overset{i.i.d.}{\sim} \mathrm{Unif}[\Delta_1,\Delta_2]$, $\ell = 1,\ldots,L$. Under this prior assumption, the Bayesian CRB is given by the following theorem.

Theorem 1: For the system model in (3) with $\mathbb{E}\{\boldsymbol{\alpha}\} = \boldsymbol{\mu} = [\mu_0, \mathbf{0}_L^T]^T$, $\mathbb{E}\{(\boldsymbol{\alpha}-\boldsymbol{\mu})(\boldsymbol{\alpha}-\boldsymbol{\mu})^H\} = \sigma^2\mathbf{I}_{L+1}$, and $\psi_\ell \overset{i.i.d.}{\sim} \mathrm{Unif}[\Delta_1,\Delta_2]$, $\ell = 1,\ldots,L$, the Bayesian CRB for any unbiased estimator of $\tilde{\boldsymbol{\theta}}$ is given by $\mathrm{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$, where $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}} = \mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D} + \mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$. The likelihood part $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ and the prior part $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ are respectively given by

$$\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D} = \tilde{\mathbf{M}}\begin{bmatrix}\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} & \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{0}_{L+1}\mathbf{0}_L^T\\ \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D}^* & \mathbf{0}_{L+1}\mathbf{0}_L^T\\ \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\end{bmatrix}\tilde{\mathbf{M}}^H \text{ and}$$

$$\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P} = \tilde{\mathbf{M}}\begin{bmatrix}\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} & \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{0}_{L+1}\mathbf{0}_L^T\\ \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*,P} & \mathbf{0}_{L+1}\mathbf{0}_L^T\\ \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P}\end{bmatrix}\tilde{\mathbf{M}}^H,$$

where the corresponding submatrices are in (10)-(14), shown at the bottom of the next page.

Proof: See Appendix A.

Note that the Bayesian FIM in Theorem 1 is a function of $\mathbf{W}$ and ρ only because the terms involving $\mathbf{S}$ except the quadratic term $\mathbf{S}^H\mathbf{S}$ vanish in the derivation of the FIM due to taking derivative of the score function and expectation over the channel parameters, as seen in the proof of Theorem 1 in Appendix, but the quadratic term $\mathbf{S}^H\mathbf{S} = \rho\mathbf{I}_K$ under the constant-modulus training symbol assumption. This means that there is no loss in terms of CRB performance with any training symbol sequence satisfying the constant-modulus symbol constraint. Thus, we assume $s_k = \sqrt{\rho}$, $1 \le k \le K$ for ease of exposition.

[2]By treating the non-zero first-order moment of the path gain as a LoS component [33], [34], this can be estimated based on the sliding window versions of the average algorithms.

B. Analysis Under the Least Information Prior

Solving the optimization problem for the reflection pattern matrix by minimizing the Bayesian CRB in Theorem 1 is not easy in general cases. In this subsection, we consider the least information prior case under the assumption of Theorem 1 and investigate the reflection pattern matrix design in this case. To this end, we assume that the direct path gain is modelled by Rician fading (i.e., $\alpha_0 \sim \mathcal{CN}(\mu_0, \sigma^2)$), the indirect path gain is modelled by Rayleigh fading (i.e., $\alpha_\ell \sim \mathcal{CN}(0, \sigma^2)$ for $\ell = 1, \ldots, L$),[3] and the path angle is uniformly distributed over the entire angle domain (i.e., $\psi_\ell \sim \text{Unif}[-1, 1]$, $\ell = 1, \ldots, L$). These prior distributions of α_ℓ and ψ_ℓ have maximum entropy under the finite second moment and finite support assumptions, respectively. Based on this prior assumption, the Bayesian FIM is given by the following corollary.

Corollary 1: For the signal model (3) with $\boldsymbol{\alpha} \sim \mathcal{CN}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_{L+1})$, $\boldsymbol{\mu} = [\mu_0, \mathbf{0}_L^T]^T$, $\psi_\ell \overset{i.i.d.}{\sim} \text{Unif}[-1, 1]$, $\ell = 1, \ldots, L$, and $d = \frac{\lambda}{2}$, the diagonal submatrices of $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ and $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ in Theorem 1 are given by

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} = \frac{\rho}{\sigma_n^2}\begin{bmatrix} K & \bar{w}_1^*\mathbf{1}_L^T \\ \bar{w}_1\mathbf{1}_L & \text{tr}(\mathbf{Q})\mathbf{I}_L + [\mathbf{Q}]_{1,1}(\mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L) \end{bmatrix}, \tag{15}$$

$$\begin{aligned} \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} &= (2\rho\sigma^2/\sigma_n^2)\left(\pi^2\sum_{m=1}^{N-1} m^2\,[\mathbf{Q}]_{m+1,m+1}\right)\mathbf{I}_L, \\ \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} &= \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*,P} = (1/\sigma^2)\mathbf{I}_{L+1}, \text{ and } \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{0}_L\mathbf{0}_L^T. \end{aligned} \tag{16}$$

Proof: See Appendix B.

Due to the block diagonal structure of $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$, the inverse of the Bayesian FIM is given by (17), shown at the bottom of the page, where $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} + \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P}$, $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}} = \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} + \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$, and we used the fact that the inverse of $\mathbf{M}$ in (8) is $\mathbf{M}^{-1} = 2\mathbf{M}^H$.

[3]A multiplicative factor including the path-loss (PL) can be incorporated into the path gain part, e.g., $\alpha_\ell \sim \mathcal{CN}(0, \sigma^2 10^{-0.1\text{PL}})$ [35].

Now, let us consider the IRS reflection pattern design that minimizes the Bayesian CRB given by the trace of (17) while maintaining the constant modulus constraint on the reflection pattern matrix $|[\mathbf{W}]_{p,q}| = \sqrt{\gamma}$ for all p, q.

Lemma 1: The inverse of the matrix $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} + \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P}$ of size $(L+1)\times(L+1)$ in (17) is eigen-decomposed as

$$\begin{aligned} \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} = \frac{\sigma_n^2}{\rho}\mathbf{B} \times \text{diag}\Big(&K' + \frac{\tau-\kappa}{2}, K' + \frac{\tau+\kappa}{2}, \\ &(K' + \tau - L[\mathbf{Q}]_{1,1})\,\mathbf{1}_{L-1}^T\Big)^{-1} \times \mathbf{B}^H, \end{aligned} \tag{18}$$

where $K' = K + \frac{\sigma_n^2}{\rho\sigma^2}$, $\mathbf{B}$ is a unitary matrix defined in (57) in Appendix C,

$$\begin{aligned} \tau &= \text{tr}(\mathbf{Q}) + (L-1)[\mathbf{Q}]_{1,1} - K, \text{ and} \\ \kappa &= \sqrt{\tau^2 + 4L|\bar{w}_1|^2}. \end{aligned} \tag{19}$$

Here, $\mathbf{Q} = \mathbf{W}\mathbf{W}^H$ as defined in (13) and $\bar{w}_1 = \sum_{k=1}^K [\mathbf{W}]_{1,k}$.

Proof: See Appendix C.

Proposition 1: The IRS reflection pattern $\mathbf{W}$ minimizing the Bayesian CRB from Corollary 1 satisfies $\sum_{k=1}^K [\mathbf{W}]_{1,k} = 0$.

Proof: By Lemma 1, the Bayesian CRB from Corollary 1 can be rewritten as

$$\begin{aligned} \text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big) &\overset{(a)}{=} \text{tr}\left(\begin{bmatrix} 2\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} & \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & 2\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{*-1} & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} \end{bmatrix}\right) \\ &\overset{(b)}{=} 4\frac{\sigma_n^2}{\rho}\Bigg(\frac{1}{K' + \frac{\tau-\kappa}{2}} + \frac{1}{K' + \frac{\tau+\kappa}{2}} \\ &\quad + \frac{L-1}{K' + \tau - L[\mathbf{Q}]_{1,1}}\Bigg) \\ &\quad + \frac{\sigma_n^2}{\rho\sigma^2}\frac{L}{2\pi^2\sum_{m=1}^{N-1} m^2[\mathbf{Q}]_{m+1,m+1}} \end{aligned} \tag{20}$$

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} = \frac{\rho}{\sigma_n^2}\begin{bmatrix} K & (\boldsymbol{\varphi}_\psi^H\bar{\mathbf{w}})^*\mathbf{1}_L^T \\ (\boldsymbol{\varphi}_\psi^H\bar{\mathbf{w}})\mathbf{1}_L & \big(\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\varphi_\psi\big(2\pi\frac{d}{\lambda}(n-m)\big)\,[\mathbf{Q}]_{m+1,n+1}\big)\mathbf{I}_L + \boldsymbol{\varphi}_\psi^H\mathbf{Q}\boldsymbol{\varphi}_\psi(\mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L) \end{bmatrix}, \tag{10}$$

$$\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} = \frac{2\rho\sigma^2}{\sigma_n^2}\left(\Big(2\pi\frac{d}{\lambda}\Big)^2\sum_{m=1}^{N-1}\sum_{n=1}^{N-1} mn\varphi_\psi\Big(2\pi\frac{d}{\lambda}(n-m)\Big)\,[\mathbf{Q}]_{m+1,n+1}\right)\mathbf{I}_L, \tag{11}$$

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} = -\mathbb{E}\Big\{\frac{\partial}{\partial\boldsymbol{\alpha}^*}\Big(\frac{\partial\ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\alpha}^*}\Big)^H\Big\},\ \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*,P} = -\mathbb{E}\Big\{\frac{\partial}{\partial\boldsymbol{\alpha}}\Big(\frac{\partial\ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\alpha}}\Big)^H\Big\},\ \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{0}_L\mathbf{0}_L^T, \tag{12}$$

$$\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]^T = \sum_{k=1}^K \mathbf{w}_k, \qquad \mathbf{Q} = \mathbf{W}\mathbf{W}^H, \quad \text{and} \tag{13}$$

$$\boldsymbol{\varphi}_\psi = \begin{bmatrix} 1 \\ \varphi_\psi\big(2\pi\frac{d}{\lambda}\big) \\ \vdots \\ \varphi_\psi\big(2\pi\frac{d}{\lambda}(N-1)\big) \end{bmatrix} \quad \text{with} \quad \varphi_\psi\Big(2\pi\frac{d}{\lambda}n\Big) = \frac{e^{\iota 2\pi\frac{d}{\lambda}n\frac{(\Delta_2+\Delta_1)}{2}}}{\pi\frac{d}{\lambda}n(\Delta_2-\Delta_1)}\sin\Big(2\pi\frac{d}{\lambda}n\Big(\frac{\Delta_2-\Delta_1}{2}\Big)\Big). \tag{14}$$

$$\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} = \underbrace{\tilde{\mathbf{M}}\begin{bmatrix} 2\mathbf{I}_{2(L+1)} & \\ & \mathbf{I}_L \end{bmatrix}}_{=\tilde{\mathbf{M}}^{-H}}\left[\begin{array}{cc|c} \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} & \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{*-1} & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \hline \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} \end{array}\right]\underbrace{\begin{bmatrix} 2\mathbf{I}_{2(L+1)} & \\ & \mathbf{I}_L \end{bmatrix}\tilde{\mathbf{M}}^H}_{=\tilde{\mathbf{M}}^{-1}}, \tag{17}$$

$$\overset{(c)}{=} 4\frac{\sigma_n^2}{\rho}\left(\frac{1}{K'+\frac{\tau-\kappa}{2}}+\frac{1}{K'+\frac{\tau+\kappa}{2}}+\frac{L-1}{K'+\tau-LK\gamma}\right) + \frac{\sigma_n^2}{\rho\sigma^2}\frac{3L}{\pi^2K\gamma(N-1)N(2N-1)}, \quad (21)$$

where Step (a) is computation of (17) by using $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA})$ and $\mathbf{M}^H\mathbf{M} = \text{diag}(\frac{1}{2}\mathbf{I}_{2(L+1)}, \mathbf{I}_L)$ due to $\mathbf{M}^{-1} = 2\mathbf{M}^H$, for Step (b) we use the eigenvalues of $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ in Lemma 1 and the result of $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ in (16) from Corollary 1, and for Step (c) we use the constant modulus constraint on $\mathbf{W}$, i.e., $[\mathbf{Q}]_{m,m} = \sum_{k=1}^{K}|[\mathbf{W}]_{m,k}|^2 = K\gamma,\ \forall m$ and the series sum $\sum_{m=1}^{N-1} m^2 = (N-1)N(2N-1)/6$. Hence, the dependence of $\text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$ in (21) on $\mathbf{W}$ is only through τ and κ. As seen in (19), furthermore, the dependence of τ on $\mathbf{Q}$ disappears due to the constant modulus constraint on $\mathbf{W}$ because $\text{tr}(\mathbf{Q}) = NK\gamma$ and $[\mathbf{Q}]_{1,1} = K\gamma$ for all $\mathbf{W}$ satisfying the constant modulus constraint. Therefore, the remaining dependence of the CRB on $\mathbf{W}$ is only through κ, which is a function of $|\bar{w}_1|$ only, as seen in (19). The derivative of the CRB $\text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$ in (21) (as a function of κ) with respect to (w.r.t.) κ is given by $\frac{d}{d\kappa}\text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big) = \frac{\sigma_n^2}{\rho}\frac{2(2K'+\tau)\kappa}{(K+(\tau-\kappa)/2)^2(K+(\tau+\kappa)/2)^2} \geq 0$. Hence, the CRB $\text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$ in (21) is a monotone increasing function of κ. Therefore, the minimum CRB $\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$ is attained at the smallest value of κ and in turn the smallest κ is obtained with $|\bar{w}_1| = 0$ from (19). ■

Note that in this case, $\text{tr}(\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1})$ is given by a constant due to the symmetric prior assumption $\psi_\ell \overset{i.i.d.}{\sim} \text{Unif}[-1,1]$ as seen in (20) - (21). Hence, the overall CRB reduction by the condition in Proposition 1 is by improving the estimation performance for $\boldsymbol{\alpha}$. The received signal (3) during the training period (without noise term) can be rewritten as

$$\mathbf{y} = \sqrt{\rho}\left(\alpha_0\mathbf{1}_K + \mathbf{W}^H\mathbf{h}\right) = \sqrt{\rho}\Bigg(\alpha_0\mathbf{1}_K + \left(\sum_{\ell=1}^{L}\alpha_\ell\right) [\mathbf{W}]_{1,1:K}^H + [\mathbf{W}]_{2:N,1:K}^H[\mathbf{U}_{\boldsymbol{\psi}}]_{2:N,1:L}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_L\end{bmatrix}\Bigg), \quad (22)$$

since the first element of the ULA response vector $\mathbf{u}_N(\psi_\ell)$ is one. Due to the condition $\bar{w}_1 = \sum_{k=1}^{K}[\mathbf{W}]_{1,k} = 0$ of Proposition 1, the inner product between the first column $\alpha_0\mathbf{1}_K$ and the second column $(\sum_{\ell=1}^{L}\alpha_\ell)[\mathbf{W}]_{1,1:K}^H$ in the right-hand side (RHS) of (22) are orthogonal since $\mathbf{1}_K^T[\mathbf{W}]_{1:K,1}^H = \sum_{k=1}^{K}[\mathbf{W}]_{1,k}^* = \bar{w}_1^* = 0$. Hence, we have reduced interference from $\{\alpha_1,\ldots,\alpha_L\}$ to α_0, which can facilitate the estimation of $\boldsymbol{\alpha}$.

1) Quantitative Analysis: The sum condition $\sum_{k=1}^{K}[\mathbf{W}]_{1,k} = 0$ in Proposition 1 can be applied to previous designs including the DFT-based reflection patterns [13], [36]. This can be done by adjusting the first row elements of the reflection patterns so that the sum condition is satisfied. For example, such a $\mathbf{W}$ with $N = 8$ and $K = 4$ is given by $\mathbf{W} = [\mathbf{u}_8(-1+\frac{1}{8}), \mathbf{u}_8(-1+\frac{3}{8}), \mathbf{u}_8(-1+\frac{5}{8}), \mathbf{u}_8(-1+\frac{7}{N})]\text{diag}(1, e^{j\frac{2\pi}{4}}, e^{j2\frac{2\pi}{4}}, e^{j3\frac{2\pi}{4}})$, where this phase shifted $\mathbf{W}$ satisfies both the constant-modulus constraint and the sum condition $\sum_{k=1}^{K}[\mathbf{W}]_{1,k} = 0$. Fig. 2(a) shows the Bayesian CRB performance of several IRS reflection patterns for the estimation of the overall parameter $\alpha_0, \alpha_1, \ldots, \alpha_L, \psi_1, \ldots, \psi_L$ when the number of IRS elements $N = 8$, the number of training symbols $K = 4$, and the number of paths $L = 2$. Here, the signal-to-noise ratio (SNR) is defined as $\text{SNR} = \rho/\sigma_n^2$. The on-off method switching the on-off states of the reflecting elements [11] shows degraded performance over the entire range of SNR. This is because the on-off switching method does not exploit all reflecting elements simultaneously for channel estimation. The orthogonal reflection patterns including the first K columns [13], [36] and equi-spaced K columns of the $N \times N$ DFT matrix yield improved performance compared to the on-off method. We implemented the condition of Proposition 1 on the equi-spaced K columns on the DFT matrix by progressively shifting the phases of the K column vectors by $e^{j\frac{2\pi}{K}(k-1)}$, $k = 1, 2, \ldots, K$ so that $\sum_{k=1}^{K}[\mathbf{W}]_{1,k} = 0$ as already mentioned in the above example (note that the first row of a DFT matrix is an all-one vector). Indeed, the phase-shifted equi-spaced K columns of the DFT matrix shows further performance improvement, as seen in Fig. 2(a).

Fig. 2(b) shows the quantity $2\frac{\partial}{\partial\psi}\mathbf{u}_N^H(\psi)\mathbf{Q}\frac{\partial}{\partial\psi}\mathbf{u}_N(\psi)$ as a function of ψ for the IRS reflection patterns considered in Fig. 2(a). Note that $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ in (16) whose inverse constitutes the lower diagonal block of $\text{tr}\big(\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\big)$ in (20) determining the angle estimation performance is given by $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} = \frac{\rho\sigma^2}{\sigma_n^2}\mathbb{E}_\psi\{2\frac{\partial}{\partial\psi}\mathbf{u}_N^H(\psi)\mathbf{Q}\frac{\partial}{\partial\psi}\mathbf{u}_N(\psi)\}\mathbf{I}_L$, as seen in the proof of Theorem 1 in Appendix. Hence, $2\frac{\partial}{\partial\psi}\mathbf{u}_N^H(\psi)\mathbf{Q}\frac{\partial}{\partial\psi}\mathbf{u}_N(\psi)$ shows the density of Fisher information over the angle domain before taking the expectation over $\psi_\ell \sim \text{Unif}[-1,1]$. High Fisher information density at a certain angle means high estimation performance (i.e., low estimation error) at that angle. It is seen that the Fisher information density is angle-dependent for the considered IRS reflection patterns. The first 4 columns of the 8×8 DFT matrix has high Fisher information density for $\psi \in [-1, 0]$ and low Fisher information density for $\psi \in [0, 1]$. This is because the first 4 columns are equivalent to the orthogonal steering vectors for the look-angles in $[-1, 0]$. On the other hand, the equi-spaced 4 columns of the 8×8 DFT matrix spreads out Fisher information over the entire angle range with some ripples. The phase-shifted equi-spaced DFT columns satisfying the condition in Proposition 1 has the same value of the integrated Fisher information as the equi-spaced DFT columns. Note that all the reflection patterns other than the on-off method satisfy the constant modulus constraint and they have the same integrated Fisher information $\mathbb{E}_\psi\{2\frac{\partial}{\partial\psi}\mathbf{u}_N^H(\psi)\mathbf{Q}\frac{\partial}{\partial\psi}\mathbf{u}_N(\psi)\}$ with $\psi_\ell \sim \text{Unif}[-1,1]$. At a particular path angle, however, the angle estimation performance can be different for each reflection pattern because the Fisher information density $2\frac{\partial}{\partial\psi}\mathbf{u}_N^H(\psi)\mathbf{Q}\frac{\partial}{\partial\psi}\mathbf{u}_N(\psi)$ varies over the angle domain and has distinct peak-to-valley height, as seen in Fig. 2(b). This suggests that we can exploit the Fisher information density itself rather than its integrated value when we focus on the angle estimation performance considering the desired performance distribution over the angle domain. This will be investigated in the next section.

Remark 3: In the case of the indirect path gain with non-zero mean, the block diagonal structure of the FIM (i.e., $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$

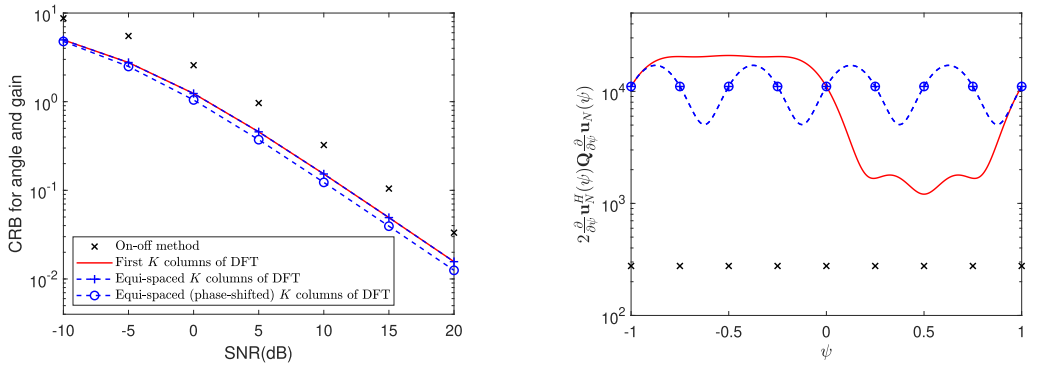


Fig. 2. Bayesian CRB performance: $N = 8$, $K = 4$, and $L = 2$: (a) Bayesian CRB for the overall parameter $\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\psi}^T]^T$ and (b) Density of $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ over the angle domain (the same legend as in (a)).

in Theorem 1) does not hold due to a possible non-zero off-diagonal submatrix. In particular, the off-diagonal submatrix $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi},D}$ is expressed as a function of the mean of the indirect path gain. Consequently, minimizing the CRB as a function of training signal is very difficult because the computation of the CRB requires the inverse of the whole FIM matrix $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ which is not easy to simplify for tractable optimization.

IV. Hybrid CRB-Based Reflection Pattern Design

To take advantage of the Fisher information density before taking expectation, we need deterministic parameter assumption. However, the full deterministic assumption for both path gains and path angles makes the derivation of CRB very difficult. We circumvent this difficulty by adopting a hybrid parameter assumption. Since in sparse mmWave channels, the estimation of path angles is crucial to collect the propagation power spread in the space with beamforming and yield high SNR [37], [38], we now take the path angle parameter $\boldsymbol{\psi}$ as a deterministic parameter (hence the expectation over $\boldsymbol{\psi}$ is not taken), whereas we treat the path gain parameter $\boldsymbol{\alpha}$ as a random nuisance parameter.

A. Derivation of Hybrid CRB

The FIM for the estimation of the random vector $\boldsymbol{\alpha}$ and the deterministic vector $\boldsymbol{\psi}$ can be obtained by averaging $\mathbf{y}$ and $\boldsymbol{\alpha}$ out from the covariance matrix of the score function conditioned on $\boldsymbol{\psi}$. Then, the derived CRB becomes a function of the path-angle parameter $\boldsymbol{\psi}$ and thus yields a lower bound on the MSE of estimation for given true underlying parameter $\boldsymbol{\psi}$. From the system model (3), the conditional pdf of $\mathbf{y}$ and $\boldsymbol{\alpha}$ given $\boldsymbol{\psi}$ is given by

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\psi}) &= p(\mathbf{y}|\boldsymbol{\alpha}; \boldsymbol{\psi}) p(\boldsymbol{\alpha}) \\ &= \left(\pi\sigma_n^2\right)^{-K} \exp\left(-\left\|\mathbf{y} - \mathbf{m}\right\|^2 / \sigma_n^2\right) p(\boldsymbol{\alpha}), \end{aligned} \quad (23)$$

where $\mathbf{m} = \sqrt{\rho}\widetilde{\mathbf{W}}^H \left[\alpha_0, \mathbf{h}^T\right]^T$ and $\mathbf{h} = \mathbf{U}_{\boldsymbol{\psi}}[\alpha_1, \ldots, \alpha_L]^T$. The CRB in this hybrid case is provided in the following theorem.

Theorem 2: Under the system model (3) with random path-gain parameter $\boldsymbol{\alpha} \sim \mathcal{CN}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_{L+1})$, $\boldsymbol{\mu} = [\mu_0, \mathbf{0}_L^T]^T$, and deterministic path-angle parameter $\boldsymbol{\psi}$, the hybrid CRB for any unbiased estimator of the overall parameter $\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\psi}^T]^T$ is given by $\mathrm{tr}\left(\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}\right) = \mathrm{tr}\left(\left[\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D} + \breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}\right]^{-1}\right)$, where $\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ is the hybrid FIM for which the path gain is averaged out but the path angle is given. The two matrices $\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ and $\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ are given by

$$\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D} = \tilde{\mathbf{M}} \begin{bmatrix} \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} & \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T & \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D}^* & \mathbf{0}_{L+1}\mathbf{0}_L^T \\ \mathbf{0}_L\mathbf{0}_{L+1}^T & \mathbf{0}_L\mathbf{0}_{L+1}^T & \breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D} \end{bmatrix} \tilde{\mathbf{M}}^H \text{ and}$$

$$\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P} = \mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P},$$

where

$$\begin{aligned} \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} &= \frac{\rho}{\sigma_n^2} \begin{bmatrix} K & \bar{\mathbf{w}}^H\mathbf{U}_{\boldsymbol{\psi}} \\ \mathbf{U}_{\boldsymbol{\psi}}^H\bar{\mathbf{w}} & \mathbf{U}_{\boldsymbol{\psi}}^H\mathbf{Q}\mathbf{U}_{\boldsymbol{\psi}} \end{bmatrix}; \\ [\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}]_{p,q} &= \begin{cases} 2\frac{\rho\sigma^2}{\sigma_n^2} \frac{\partial \mathbf{u}_N^H(\psi_p)}{\partial\psi_p} \mathbf{Q} \frac{\partial \mathbf{u}_N(\psi_q)}{\partial\psi_q}, & \text{if } p = q \\ 0, & \text{if } p \neq q; \end{cases} \end{aligned} \quad (24)$$

$\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$, $\bar{\mathbf{w}}$, and $\mathbf{Q}$ are given in Theorem 1; and $\mathbf{u}_N(\cdot)$ is the ULA response vector.

Proof: See Appendix D.

Note that in the hybrid case, the p-th diagonal element of $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ is nothing but the Fisher information density at ψ_p, mentioned in the previous section. Since $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ $(= \breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P})$ is block-diagonal, as seen in Theorem 1, with its constituent matrices $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} = \frac{1}{\sigma^2}\mathbf{I}_{L+1}$ and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{0}_L\mathbf{0}_L^T$, the hybrid FIM $\breve{\mathbf{J}}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ has a block-diagonal form and we have separate sub-FIMs for the path gains and the path angles, as seen in Theorem 2. Thus, the MSE lower bound for the estimation of the path-angle parameter $\boldsymbol{\psi}$ is determined by the inverse of the lower diagonal block $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ due to $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{0}_L\mathbf{0}_L^T$, whereas the MSE lower bound for the estimation of the path-gain parameter $\boldsymbol{\alpha}$ is determined by the upper diagonal element $\breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} + \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P}$. Both of the FIM submatrices are functions of the design variable $\mathbf{Q}$ $(= \mathbf{W}\mathbf{W}^H)$ and the true angle parameter $\boldsymbol{\psi}$. Focusing on the path-angle parameter while treating the path-gain parameter as a nuisance parameter, we can optimize the IRS reflection matrix $\mathbf{W}$ in order to yield the best angle-estimation CRB based on $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$. (Note that

in Theorem 2, the lower diagonal block of $\tilde{\mathbf{M}}$ multiplied to $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ is the identity matrix, as seen in (8).)

B. Hybrid CRB-Based Reflection Pattern

As seen in Theorem 2, the FIM submatrix $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ determining the performance of path angle estimation is a function of the design variable $\mathbf{Q}$ $(=\mathbf{W}\mathbf{W}^H)$ and the true underlying path angle $\boldsymbol{\psi}$. A key point to note is that the CRB $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big)$ determines the estimation performance when the true path angles are $\boldsymbol{\psi}=[\psi_1,\ldots,\psi_L]^T$. Hence, our design approach with the hybrid CRB is that we first design a set of targeted look-angles and then optimize $\mathbf{W}$ so that $\mathbf{W}$ yields best estimation performance when the true path angles are at the targeted look-angles. Here, we can exploit some side information about the path angles if any, e.g., its support information $\psi_\ell \in [\Delta_1,\Delta_2],\ \forall \ell$. With such information, we design the set of targeted look-angles as $\xi_{L_T,\ell}\in\big\{\Delta_1+\frac{\Delta_2-\Delta_1}{L_T}(\ell-1)\mid \ell=1,\ldots,L_T\big\}$, i.e., the targeted look-angles are evenly spaced over the entire angle range $[\Delta_1,\Delta_2]$ with L_T angles. Then, the optimization problem is formulated as

$$\min_{\mathbf{W}}\ \mathrm{tr}\left(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\,\Big|\boldsymbol{\psi}=\boldsymbol{\xi}\triangleq\ [\xi_{L_T,1},\ldots,\xi_{L_T,L_T}]^T\right) \quad (25)$$

$$\text{s.t.}\ \ |[\mathbf{W}]_{p,q}|=\sqrt{\gamma},\quad \forall p,q. \quad (26)$$

Remark 4: Using the Cauchy-Schwartz inequality $\big(\sum_{\ell=1}^{L}\sqrt{\lambda_\ell}\frac{1}{\sqrt{\lambda_\ell}}\big)^2 \le \sum_{\ell=1}^{L}\lambda_\ell\cdot\sum_{\ell=1}^{L}\frac{1}{\lambda_\ell}$, where $\lambda_\ell = 2\frac{\rho\sigma^2}{\sigma_n^2}\frac{\partial\mathbf{u}_N^H(\psi_\ell)}{\partial\psi_\ell}\mathbf{Q}\frac{\partial\mathbf{u}_N(\psi_\ell)}{\partial\psi_\ell}$, we have the following inequality regarding the trace of the inverse of $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$: $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big)\ge L^2\big[\mathrm{tr}\big(\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big)\big]^{-1}$ for any true $\boldsymbol{\psi}$. The equality is achieved if and only if $\lambda_1=\ldots=\lambda_L$. This condition can be achieved by $\mathbf{Q}=\mathbf{W}\mathbf{W}^H=K\gamma\mathbf{I}_N$ if $K\ge N$. That is, with $\mathbf{Q}=K\gamma\mathbf{I}_N$, $\lambda_\ell = 2K\gamma\frac{\rho\sigma^2}{\sigma_n^2}\frac{\partial\mathbf{u}_N^H(\psi_\ell)}{\partial\psi_\ell}\frac{\partial\mathbf{u}_N(\psi_\ell)}{\partial\psi_\ell} = 2K\gamma\frac{\rho\sigma^2}{\sigma_n^2}(2\pi\frac{d}{\lambda})^2\sum_{m=1}^{N-1}m^2$ does not depend on the index ℓ. (Please see in (41) in Appendix A regarding $\frac{\partial\mathbf{u}_N(\psi_\ell)}{\partial\psi_\ell}$.) So, we can see the optimality of $\mathbf{Q}=K\gamma\mathbf{I}_N$ obtained in the previous work [13], [36] in the case of $K\ge N$ in this sense too.

Since our main focus in this paper is sparse mmWave channels, the situation is $L<K<N$. We consider the optimization problem (25) - (26) in the case of $K<N$. The objective function (25) can be rewritten as

$$\begin{aligned}
&\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\,|\boldsymbol{\psi}=\boldsymbol{\xi}\big)\\
&=\frac{\sigma_n^2}{2\rho\sigma^2}\sum_{\ell=1}^{L_T}\frac{1}{\mathrm{tr}\big(\mathbf{W}^H\frac{\partial\mathbf{u}_N(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\mathbf{W}\big)} \quad (27)\\
&\overset{(a)}{=}\frac{\sigma_n^2}{2\rho\sigma^2}\sum_{\ell=1}^{L_T}\frac{1}{\mathrm{vec}\big(\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\mathbf{W}\big)^H\mathrm{vec}\big(\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\mathbf{W}\big)} \quad (28)\\
&\overset{(b)}{=}\frac{\sigma_n^2}{2\rho\sigma^2}\sum_{\ell=1}^{L_T}\frac{1}{\overrightarrow{\mathbf{w}}^H\big(\mathbf{I}_K\otimes\frac{\partial\mathbf{u}_N(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\big)\overrightarrow{\mathbf{w}}}\\
&\triangleq\ f(\overrightarrow{\mathbf{w}}), \quad (29)
\end{aligned}$$

where $\overrightarrow{\mathbf{w}}=\mathrm{vec}(\mathbf{W})$, Step (a) is valid due to $\mathrm{tr}(\mathbf{A}^H\mathbf{B})=\mathrm{vec}(\mathbf{A})^H\mathrm{vec}(\mathbf{B})$, and Step (b) is valid because $\mathrm{vec}(\mathbf{AXB})=(\mathbf{B}^T\otimes\mathbf{A})\mathrm{vec}(\mathbf{X})$ and $(\mathbf{A}_1\otimes\mathbf{B}_1)(\mathbf{A}_2\otimes\mathbf{B}_2)=(\mathbf{A}_1\mathbf{A}_2\otimes\mathbf{B}_1\mathbf{B}_2)$. Thus, the objective function is the sum of the inverses of Rayleigh quotients since $\overrightarrow{\mathbf{w}}^H\overrightarrow{\mathbf{w}}=NK\gamma$ under the constant modulus constraint $|[\mathbf{W}]_{p,q}|=\sqrt{\gamma},\forall p,q$. It is known that even the optimization of the sum of Rayleigh quotients is not a simple problem. In our case, we have the sum of the inverses of Rayleigh quotients and the element-wise constant modulus constraint in addition. To tackle this optimization problem, we adopt the PGM [16], [17]. PGM is an iterative method applying gradient descent and projection onto the constraint set in an alternating manner and is suited to complicated constraints. The overall algorithm is summarized in Algorithm 1. The objective function (29) is differentiable and its Wirtinger complex gradient is given by

$$\nabla_{\overrightarrow{\mathbf{w}}^*}f(\overrightarrow{\mathbf{w}})=-\sum_{\ell=1}^{L_T}\frac{\big(\mathbf{I}_K\otimes\frac{\partial\mathbf{u}_N(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\big)}{\big[\overrightarrow{\mathbf{w}}^H\big(\mathbf{I}_K\otimes\frac{\partial\mathbf{u}_N(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\frac{\partial\mathbf{u}_N^H(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}\big)\overrightarrow{\mathbf{w}}\big]^2}\overrightarrow{\mathbf{w}}. \quad (30)$$

We start with an initial point $\overrightarrow{\mathbf{w}}^{(0)}$. At the i-th iteration, we update the current $\overrightarrow{\mathbf{w}}^{(i)}$ by using gradient descent based on the gradient vector (30) and then project the updated vector onto the constraint set $\mathcal{C}:=\{\overrightarrow{\mathbf{w}}=[w_1,\ldots,w_{KN}]^T\in\mathbb{C}^{KN}:|w_i|=\sqrt{\gamma}\}$. We repeat this iteration. In the step size determination, we apply the technique in [39] to mitigate the slow crawling problem of gradient decent near a local minimum and help escape from local stationary points of the objective function. That is, we normalize the magnitude of each component of the gradient vector by multiplying the i-th component of the gradient vector by its magnitude inverse, as seen in Line 5 of Algorithm 1.

Note that the instantaneous CSI is not required for the proposed training signal design. Because the Bayesian FIM is derived by taking the expectation over the random gain and angle and the hybrid FIM is derived by taking the expectation over the random gain with a fixed set of target angles.

Algorithm 1 The Proposed PGM-Based IRS Reflection Pattern Design

Require: Stopping criterion ϵ, step size δ, and targeted look-angles $\{\xi_{L_T,\ell}\}$

1: Compute $\frac{\partial\mathbf{u}_N(\xi_{L_T,\ell})}{\partial\xi_{L_T,\ell}}$ for $\ell=1,\ldots,L_T$.
2: Set $i=0$ and initialize $\overrightarrow{\mathbf{w}}^{(0)}$.
3: **repeat**
4: Compute the gradient vector (30) at $\overrightarrow{\mathbf{w}}^{(i)}$.
5: Update $\overrightarrow{\mathbf{w}}^{(i+1)}=\Pi_{\mathcal{C}}\left(\overrightarrow{\mathbf{w}}^{(i)}-\mathbf{r}_i\odot\nabla_{\overrightarrow{\mathbf{w}}^*}f(\overrightarrow{\mathbf{w}}^{(i)})\right)$, where $\Pi_{\mathcal{C}}$ is the projection onto $\mathcal{C}$; $\mathbf{r}_i=\Big[\frac{\delta}{\big|\big[\nabla_{\overrightarrow{\mathbf{w}}^*}f(\overrightarrow{\mathbf{w}}^{(i)})\big]_1\big|},\ldots,\frac{\delta}{\big|\big[\nabla_{\overrightarrow{\mathbf{w}}^*}f(\overrightarrow{\mathbf{w}}^{(i)})\big]_{NK}\big|}\Big]^T$; and $\odot$ denotes elementwise multiplication.
6: $i=i+1$
7: **until** $\big\|\overrightarrow{\mathbf{w}}^{(i)}-\overrightarrow{\mathbf{w}}^{(i-1)}\big\|^2/\big\|\overrightarrow{\mathbf{w}}^{(i-1)}\big\|^2\le\epsilon$

V. Extensions of the Proposed Design

In this section, we consider several extensions relevant to practical implementation of the proposed design.

Extension to Multi-Antenna Transceiver: First, consider the case of an M-antenna ULA at the receiver. Then, the direct-path channel is modeled as $\mathbf{h}_D = \sum_{i=1}^{L_D} \alpha_i \mathbf{u}_M(\varrho_i)$, where α_i and ϱ_i are the gain and AoA of the i-th path of the channel from the transmitter to the receiver with L_D multiple paths. The indirect cascaded channel matrix $\mathbf{H}_I \in \mathbb{C}^{M\times N}$ is similarly defined as

$$\begin{aligned}\mathbf{H}_I &= \sum_{i=1}^{L^{(t)}}\sum_{j=1}^{L^{(r)}} \beta_i^{(t)}\beta_j^{(r)}\mathbf{u}_M(\xi_i^{(r)})\mathbf{u}_N^T(\psi_i^{(t)}+\psi_j^{(r)}) \\ &= \sum_{\ell=1}^{L_I} \beta_\ell \mathbf{u}_M(\xi_\ell)\mathbf{u}_N^T(\psi_\ell) = \mathbf{U}_{\boldsymbol{\xi}}\text{diag}(\boldsymbol{\beta})\mathbf{U}_{\boldsymbol{\psi}}, \end{aligned} \quad (31)$$

where $\beta_i^{(t)}$ and $\psi_i^{(t)}$ are the gain and AoA of the i-th path of the channel from the transmitter to the IRS; $\beta_j^{(r)}$, $\xi_j^{(r)}$, and $\psi_j^{(r)}$ are the gain, AoA, and AoD of the j-th path of the channel from the IRS to the receiver; $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{L_I}]^T$ with $L_I = L^{(t)}L^{(r)}$. $\beta_\ell := \beta_i^{(t)}\beta_j^{(r)}$, $\xi_\ell =: \xi_i^{(r)}$, and $\psi_\ell = \psi_i^{(t)} + \psi_j^{(r)}$ denote the effective gain, AoA and AoD of the ℓ-th path, respectively, and the array response matrices are defined as $\mathbf{U}_{\boldsymbol{\xi}} = [\mathbf{u}_M(\xi_1), \ldots, \mathbf{u}_M(\xi_{L_I})]$ and $\mathbf{U}_{\boldsymbol{\psi}} = [\mathbf{u}_N(\psi_1), \ldots, \mathbf{u}_N(\psi_{L_I})]$. Then, the received signal at the k-th symbol time corresponding to (1) is rewritten as $\mathbf{y}_k = \left(\mathbf{h}_D + \mathbf{H}_I\mathbf{w}_k^*\right)s_k + \mathbf{n}_k$ where $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_n^2\mathbf{I}_M)$. Accordingly, the received signal during the entire training period can be expressed as

$$\begin{aligned}\mathbf{Y} = [\mathbf{y}, \ldots, \mathbf{y}_K] = &\left(\mathbf{U}_{\boldsymbol{\varrho}}\boldsymbol{\alpha} + \mathbf{U}_{\boldsymbol{\xi}}\text{diag}(\boldsymbol{\beta})\mathbf{U}_{\boldsymbol{\psi}}\mathbf{W}^*\right)\mathbf{S} \\ &+ [\mathbf{n}_1, \ldots, \mathbf{n}_K], \end{aligned} \quad (32)$$

where $\mathbf{U}_{\boldsymbol{\varrho}} = [\mathbf{u}_M(\varrho_1), \ldots, \mathbf{u}_M(\varrho_{L_D})]$ and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{L_D}]^T$. Similarly to (3), the received signal (32) can be expressed in vector form as

$$\begin{aligned}\mathbf{y} = \text{vec}(\mathbf{Y}) = &\left(\mathbf{S}\otimes\mathbf{I}_M\right)\left[\mathbf{1}_K \otimes \mathbf{U}_{\boldsymbol{\varrho}}\ (\mathbf{W}^H\mathbf{U}_{\boldsymbol{\psi}}) * \mathbf{U}_{\boldsymbol{\xi}}\right] \\ &\times \begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix} + \mathbf{n}, \end{aligned} \quad (33)$$

where $\mathbf{n} = [\mathbf{n}_1^T, \ldots, \mathbf{n}_K^T]^T$ and the column-wise Khatri-Rao product is defined as $\mathbf{A} * \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_n \otimes \mathbf{b}_n]$. In (33), the IRS reflection pattern matrix $\mathbf{W}$ transforms only the array manifold $\mathbf{U}_{\boldsymbol{\psi}}$ of the IRS as in the single-antenna receiver case (3).[4] Given this, the results in Sections III and IV can directly be applied to the CRB derivation for the multi-antenna case. In particular, when $L_D = 1$, the minimum Bayesian CRB is achieved with $\bar{w}_1 = 0$ as in Proposition 1.[5] In the hybrid approach, the FIM relevant to $\boldsymbol{\psi}$ is derived as $[\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}]_{p,q} = \frac{2\rho\sigma^2}{\sigma_n^2} M \frac{\partial \mathbf{u}_N^H(\psi_p)}{\partial \psi_p}\mathbf{Q}\frac{\partial \mathbf{u}(\psi_q)}{\partial \psi_q}\delta_{p,q}$, which is the same as the corresponding FIM (24) up to a constant multiplicative factor. Therefore, the PGM-based algorithm for the single antenna receiver can be applied to the multi-antenna receiver case as well.

[4]The Bayesian and hybrid CRBs for the multi-antenna case can be derived in a similar way for the single antenna case. The detailed derivation is omitted due to space limitation.

[5]An extension of the Bayesian CRB-based design to the case of $L_D > 1$ remains for future work.

In the multi-antenna transmitter case, the transmit beam search is also required. For each transmit beam, the channel between the transmitter and the IRS becomes a single beam-weighted channel, which makes the multiple transmit antennas completely transparent to the receiver. When the transmitter sends K pilot symbols for each transmit beam, the corresponding system model is equivalent to that in (32) and the proposed K IRS reflection patterns can be jointly used for channel estimation. This procedure is repeated until all transmit beams are probed.

Remark 5: Note from (33) that when the path angles $\boldsymbol{\varrho}, \boldsymbol{\psi}, \boldsymbol{\xi}$ are known, we have a linear observation model for path gain estimation. The necessary condition for the system matrix to be invertible is $K \geq \frac{L_I+L_D}{M}$. On the other hand, the observation (33) is related to a non-linear measurement model from the angle estimation perspective due to the non-linear relation between $\mathbf{y}$ and $\boldsymbol{\varrho}, \boldsymbol{\psi}, \boldsymbol{\xi}$. Hence, analyzing the minimum training overhead for angle estimation is not straightforward and could be algorithm-specific as seen in [20], [22].

Extension to the UPA-based IRS: Second, the proposed design can be extended to UPA-based IRSs because a UPA is in essence the Kronecker product of two ULAs. Accordingly, the IRS reflection pattern is defined as the Kronecker product of the horizontal and vertical patterns as $\mathbf{w}_k = \mathbf{w}_k^{(h)} \otimes \mathbf{w}_k^{(v)}$. Then, the received signal matrix with a UPA-based IRS is expressed as

$$\begin{aligned}\mathbf{Y} = &\left(\mathbf{U}_{\boldsymbol{\varrho}}\boldsymbol{\alpha} + \mathbf{U}_{\boldsymbol{\xi}}\text{diag}(\boldsymbol{\beta})\mathbf{U}_{\boldsymbol{\psi}^{(h)}\boldsymbol{\psi}^{(v)}}\mathbf{W}^*\right)\mathbf{S} \\ &+ [\mathbf{n}_1, \ldots, \mathbf{n}_K], \end{aligned} \quad (34)$$

where $\mathbf{U}_{\boldsymbol{\psi}^{(h)}\boldsymbol{\psi}^{(v)}} = [\mathbf{u}_N(\psi_1^{(h)}, \psi_1^{(v)}), \ldots, \mathbf{u}_N(\psi_{L_I}^{(h)}, \psi_{L_I}^{(v)})]$ and the array manifold of the IRS is defined as $\mathbf{u}_N(\psi^{(h)}, \psi^{(v)}) = \mathbf{u}_{N_h}(\psi^{(h)}) \otimes \mathbf{u}_{N_v}(\psi^{(v)})$ with $\psi^{(h)} = \sin(\phi^{(h)})\cos(\phi^{(v)})$ and $\psi^{(v)} = \sin(\phi^{(v)})$ for the azimuth angle $\phi^{(h)}$ (rad) and the elevation angle $\phi^{(v)}$ (rad). Compared to (32), the only difference is that the IRS's array manifold is changed to $\mathbf{U}_{\boldsymbol{\psi}^{(h)}\boldsymbol{\psi}^{(v)}}$. Hence, the Bayesian and hybrid CRBs are similarly derived so that the proposed design based on CRB can be applied.

Extension to the Multi-User Case: Finally, consider the multi-user case. In the downlink, the pilot symbols can be shared across all users [40] and then the system model (3) or (33) can be regarded as the channel measurement at each user. In the uplink, it is required to transmit user-dedicated pilot symbols so that the channels of multiple users are separately estimated. Then, the pilot symbols of the multiple users can be multiplexed over time [41], frequency [42], or code [43] domain. Therefore, application of the proposed design to the multi-user case is possible by treating (3) or (33) as a system model for each multiplexed user.

VI. Numerical Results

In this section, we provide numerical results to evaluate the proposed IRS reflection pattern design. During the simulation, we set the magnitude of the reflection coefficient of each element at the IRS as $\gamma = 1$. We used the channel model (2), where the number of the IRS-assisted channel paths was set $L \in \{1, \ldots, 4\}$, $\alpha_0 \sim \mathcal{CN}(\mu_0, \sigma^2)$ and $\alpha_\ell \overset{i.i.d.}{\sim} \mathcal{CN}(0, \sigma^2)$,

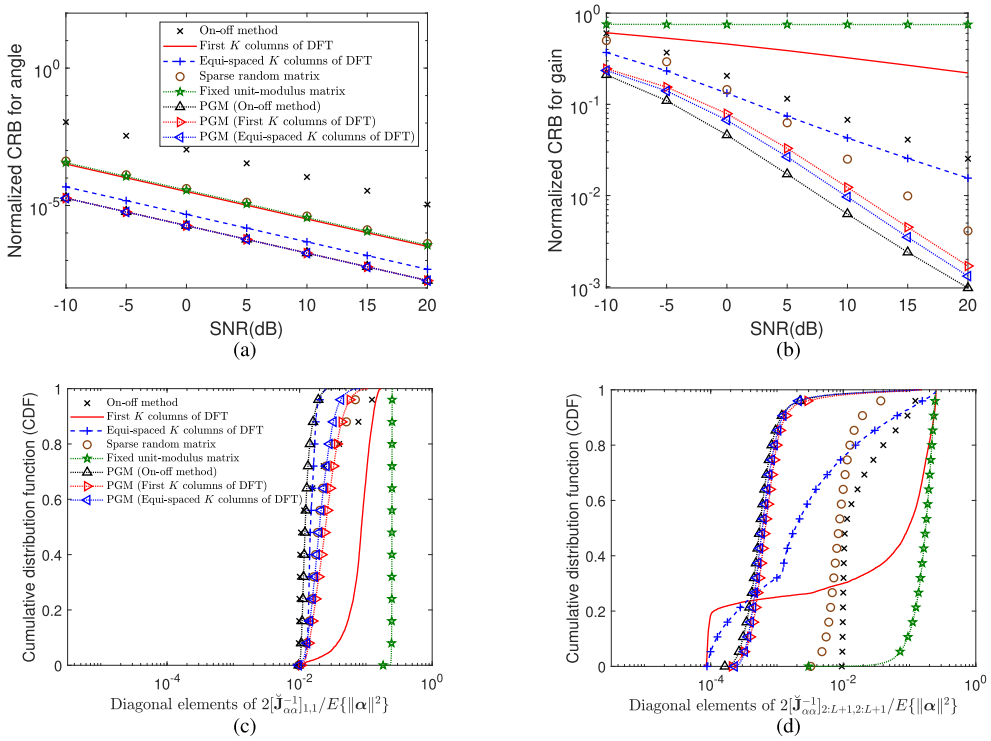
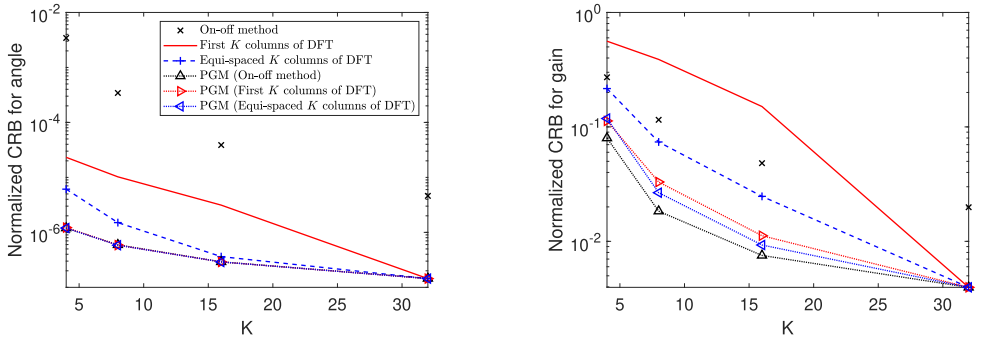


Fig. 3. Hybrid CRB comparision with several IRS reflection patterns versus SNR where $N = 32$, $K = 8$, and $L = 3$ (PGM represents the proposed design with the initial pattern for Algorithm 1 which is shown in the parenthesis.): (a) $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big)/\mathbb{E}\{\|\boldsymbol{\psi}\|^2\}$, (b) $\big(\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\big) + \mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*}\big)\big)/\mathbb{E}\{\|\boldsymbol{\alpha}\|^2\}$, (c) $\big([\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}]_{1,1} + [\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*}]_{1,1}\big)/\mathbb{E}\{\|\boldsymbol{\alpha}\|^2\}$ at SNR $= 5$dB, and (d) $\mathrm{diag}\big([\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}]_{2:L+1,2:L+1} + [\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*}]_{2:L+1,2:L+1}\big)/\mathbb{E}\{\|\boldsymbol{\alpha}\|^2\}$ at SNR $= 5$dB. (The subfigures (a)-(d) have the same legend.).

$\ell = 1, \ldots, L$ with $\mu_0 = \sigma^2 = 1$ and $d = \frac{\lambda}{2}$. The SNR is given by $\mathrm{SNR} = \rho/\sigma_n^2$.

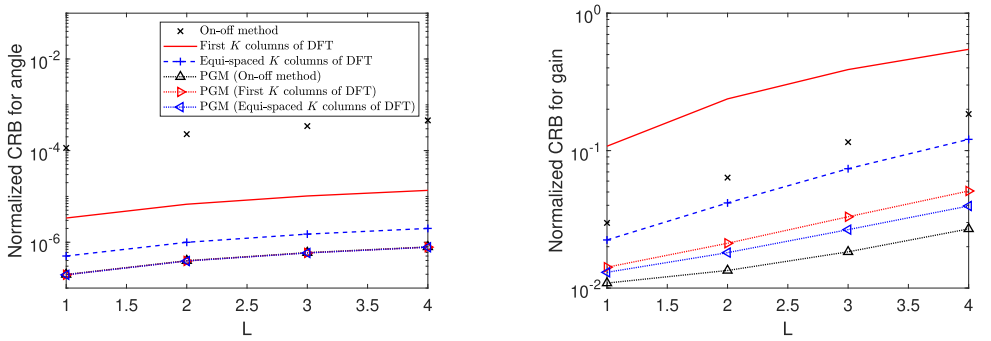
We considered the case of $N = 32$ and $K = 8$. We designed the IRS pattern matrix by running Algorithm 1 with $\epsilon = 10^{-10}$, $\delta = 1$, and $L_T = N$ targeted look-angles evenly spaced over $[-1, 1]$. Then, we generated $\psi_\ell \overset{i.i.d.}{\sim} \mathrm{Unif}[-1, 1]$, $\ell = 1, \ldots, L$ and computed $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big)$ and $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\big)$ with $\breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} + \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P}$ by using the designed $\mathbf{W}$ and the generated true angles $\psi_1, \ldots, \psi_L$. We repeated this procedure for 5000 random channel realizations and then took average over the 5000 Monte Carlo runs. Hence, this angle estimation performance average corresponds to $\mathbb{E}_{\boldsymbol{\psi}}\big\{\breve{\mathbf{J}}^{-1}_{\boldsymbol{\psi}\boldsymbol{\psi},D}\big\}$, i.e., the bound (a) not the bound (b) in (5). For other IRS pattern matrix designs, we applied the same Monte Carlo evaluation. The result is shown in Fig. 3. It is seen in Fig. 3(a) that the equi-spaced K DFT columns shows better performance than the on-off method [11], the first K DFT columns [13], a sparse Bernoulli random matrix with probability 0.1 [12], a fixed unit-modulus matrix [20] since ψ_ℓ is distributed over the angle range $[-1, 1]$. It is also seen that the proposed design yields noticeable gain over the equi-spaced K DFT columns and other designs. Note that the first K DFT columns and the equi-spaced K DFT columns show different performance here, whereas the two methods shows the same performance in Fig. 2(a). This is because Fig. 3(a) shows the tighter bound (a) in (5), whereas Fig. 2(a) shows the looser bound (b) in (5). Fig. 3(b) shows the corresponding path-gain CRB $\mathrm{tr}\big(\breve{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\big)$ averaged over the Monte Carlo runs. It is seen that the first K DFT columns and the fixed unit-modulus matrix show severe performance degradation in path gain estimation. This is because these methods have a heavy tail in the error distribution due to large gain estimation error occurring when the realized path angles are not within the coverage of the first K DFT columns or the fixed unit-modulus vector, as seen in Figs. 3(c) and (d). Due to the non-convexity in the objective function (29), there exists several local stationary points and the PGM-based algorithm can stop at some local minimum. Thus, the performance of the PGM-based algorithm slightly varies with distinct initial conditions especially for path gain estimation in Fig. 3(b).

Then, we evaluated the hybrid CRB performance of the IRS reflection patterns with different values of K. Fig. 4 shows that the designs using the orthogonal DFT matrix achieve the best performance when $K = N$ as expected, but there is noticeable gain by the proposed design method as K decreases from N. Finally, we examined the performance of the proposed method with respect to the number of channel paths L and the result is shown in Fig. 5. It is seen that the CRB performance degrades as L increases due to the increased number of parameters to estimate. It is also seen that the design by Algorithm 1 yields better performance than other methods. Note that the Bayesian CRB-based design does not require any numerical calculation because the optimal structure of $\mathbf{W}$ is derived analytically in Proposition 1. On the other hand, the hybrid CRB-based design is numerically obtained by Algorithm 1 for which the computational complexity can be



(a) Hybrid CRB for $\boldsymbol{\psi}$ (b) Hybrid CRB for $\boldsymbol{\alpha}$ (same legend as in (a))

Fig. 4. Hybrid CRB comparision with several IRS reflection patterns versus K where $N = 32$, $L = 3$, and SNR $= 5$dB.



(a) Hybrid CRB for $\boldsymbol{\psi}$ (b) Hybrid CRB for $\boldsymbol{\alpha}$ (same legend as in (a))

Fig. 5. Hybrid CRB comparision with several IRS reflection patterns versus L where $N = 32$, $K = 8$, and SNR $= 5$dB.

measured by the number of required complex multiplications, i.e., $N_{iter}(L_T(N+(KN)^2+KN+1)+2KN)$, where N_{iter} denotes the expected number of iterations for convergence.

VII. Conclusion

We have considered the problem of training signal design for IRS-assisted mmWave communication under a sparse channel model. We have approached the problem based on CRB for channel estimation. With the full Bayesian approach, we have shown that the main factor affecting the angle estimation performance is the Fisher information density given by a quadratic form composed of the derivative of the ULA response and the IRS reflection pattern matrix square $\mathbf{Q}$. Based on this fact, we have approached the IRS pattern matrix design problem with a hybrid CRB under the assumption of random path gains and unknown deterministic path angles. We have proposed a PGM-based algorithm to solve optimal IRS pattern matrix design focusing on the path angle estimation critical in mmWave communication. We have validated the effectiveness of the proposed design method with numerical evaluation. Furthermore, development of estimation algorithms achieving the CRB with the signal model (3) shall be studied as future work.

Appendix

A. Proof of Theorem 1

Step 1: To prove Theorem 1, we first provide the following lemma. Note that in the Bayesian case we need the expectation of the second order derivatives of the logarithm of $p(\mathbf{y},\boldsymbol{\theta})$ with $\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\psi}^T]^T$, as seen in (7) - (9). The logarithm of $p(\mathbf{y},\boldsymbol{\theta})$ is decomposed as $\ln p(\mathbf{y},\boldsymbol{\theta}) = \ln p(\boldsymbol{\theta}) + \ln p(\mathbf{y}|\boldsymbol{\theta})$ as seen in (6), and Lemma 2 provides the second order derivatives of $\ln p(\mathbf{y}|\boldsymbol{\theta})$.

Lemma 2: The second-order derivatives of the log-likelihood function $\ln p(\mathbf{y}|\boldsymbol{\theta})$ under the observation model (3) are given by

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H = -\frac{\rho}{\sigma_n^2}\begin{bmatrix} K & \bar{\mathbf{w}}^H\mathbf{U}_{\boldsymbol{\psi}} \\ \mathbf{U}_{\boldsymbol{\psi}}^H\bar{\mathbf{w}} & \mathbf{U}_{\boldsymbol{\psi}}^H\mathbf{Q}\mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix} = \left(\frac{\partial}{\partial \boldsymbol{\alpha}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right)^H\right)^*, \tag{35}$$

$$\frac{\partial}{\partial \psi_p}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_q}\right)^H = -\frac{1}{\sigma_n^2}\left(-(\mathbf{y}-\mathbf{m})^H\frac{\partial^2\mathbf{m}}{\partial\psi_p\partial\psi_q} + \frac{\partial\mathbf{m}^H}{\partial\psi_p}\frac{\partial\mathbf{m}}{\partial\psi_q} + \frac{\partial\mathbf{m}^H}{\partial\psi_q}\frac{\partial\mathbf{m}}{\partial\psi_p} - \frac{\partial^2\mathbf{m}^H}{\partial\psi_p\partial\psi_q}(\mathbf{y}-\mathbf{m})\right), \tag{36}$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_\ell}\right)^H = -\frac{1}{\sigma_n^2}\left(-\widetilde{\mathbf{U}}_{\psi_\ell}^H\widetilde{\mathbf{W}}\mathbf{S}^H\left(\mathbf{y} - \mathbf{S}\widetilde{\mathbf{W}}^H\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix}\boldsymbol{\alpha}\right) + \rho\begin{bmatrix} K & \bar{\mathbf{w}}^H \\ \mathbf{U}_{\boldsymbol{\psi}}^H\bar{\mathbf{w}} & \mathbf{U}_{\boldsymbol{\psi}}^H\mathbf{Q}\end{bmatrix}\widetilde{\mathbf{U}}_{\psi_\ell}\boldsymbol{\alpha}\right) \tag{37}$$

$$= \left(\frac{\partial}{\partial \boldsymbol{\alpha}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_\ell}\right)^H\right)^* = \left(\frac{\partial}{\partial \psi_\ell}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H\right)^H = \left(\frac{\partial}{\partial \psi_\ell}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right)^H\right)^T, \tag{38}$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right)^H = \frac{\partial}{\partial \boldsymbol{\alpha}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H = \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T, \tag{39}$$

where

$$\widetilde{\mathbf{U}}_{\psi_\ell} \triangleq \Big[\mathbf{0}_{N+1},\ [0,\ldots,0,\underbrace{1}_{\ell^{th}},0,\ldots,0]\otimes \Big[0,\ \frac{\partial \mathbf{u}_N^T(\psi_\ell)}{\partial \psi_\ell}\Big]^T\Big] \in \mathbb{C}^{(N+1)\times(L+1)} \tag{40}$$

$$\frac{\partial \mathbf{u}_N(\psi)}{\partial \psi} \triangleq \iota 2\pi \frac{d}{\lambda}\Big[0, e^{\iota 2\pi \frac{d}{\lambda}\psi}, 2e^{\iota 2\pi \frac{d}{\lambda}2\psi}, \ldots, (N-1)e^{\iota 2\pi \frac{d}{\lambda}(N-1)\psi}\Big]^T, \tag{41}$$

and the variables of $\bar{\mathbf{w}}$ and $\mathbf{Q}$ are defined in (13).

Proof of Lemma 2: By differentiating $\ln p(\mathbf{y}|\boldsymbol{\theta})$ in (6) w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, we have

$$\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*} = -\frac{1}{\sigma_n^2}\Big(-\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^H\end{bmatrix}\widetilde{\mathbf{W}}\mathbf{S}^H\mathbf{y} + \rho\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^H\end{bmatrix}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix}\boldsymbol{\alpha}\Big) = \Big(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\Big)^*, \tag{42}$$

$$\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_\ell} = -\frac{1}{\sigma_n^2}\Big(-(\mathbf{y}-\mathbf{m})^H\frac{\partial \mathbf{m}}{\partial \psi_\ell} - \frac{\partial \mathbf{m}^H}{\partial \psi_\ell}(\mathbf{y}-\mathbf{m})\Big) \tag{43}$$

$$= -\frac{1}{\sigma_n^2}\Big(-\mathbf{y}^H\mathbf{S}\widetilde{\mathbf{W}}^H\widetilde{\mathbf{U}}_{\psi_\ell}\boldsymbol{\alpha} - \boldsymbol{\alpha}^H\widetilde{\mathbf{U}}_{\psi_\ell}^H\widetilde{\mathbf{W}}\mathbf{S}^H\mathbf{y} + \rho\boldsymbol{\alpha}^H\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^H\end{bmatrix}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H\widetilde{\mathbf{U}}_{\psi_\ell}\boldsymbol{\alpha} + \rho\boldsymbol{\alpha}^H\widetilde{\mathbf{U}}_{\psi_\ell}^H\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix}\boldsymbol{\alpha}\Big), \tag{44}$$

where we used $\mathbf{S}^H\mathbf{S} = \rho\mathbf{I}$ for constant-modulus training symbols in the quadratic terms in the RHSs of (42) and (44). By differentiating (42) and (43) w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ again, we have

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H = -\frac{\rho}{\sigma_n^2}\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^H\end{bmatrix}\begin{bmatrix}K & \bar{\mathbf{w}}^H\\ \bar{\mathbf{w}} & \mathbf{Q}\end{bmatrix}\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix} \tag{45}$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right)^H = -\frac{\rho}{\sigma_n^2}\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^T\end{bmatrix}\begin{bmatrix}K & \bar{\mathbf{w}}^T\\ \bar{\mathbf{w}}^* & \mathbf{Q}^*\end{bmatrix}\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^*\end{bmatrix} \tag{46}$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}\right)^H = \frac{\partial}{\partial \boldsymbol{\alpha}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H \mathbf{0}_{L+1}\mathbf{0}_{L+1}^T, \tag{47}$$

and the expression for $\frac{\partial}{\partial \psi_p}(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_q})^H$ is given in (36), which can further be detailed in a form like (44) based on $\mathbf{m} = \mathbf{S}\widetilde{\mathbf{W}}^H[\alpha_0, \mathbf{h}^T]^T$. (The detailed expression is omitted due to space limitation.) By using (42) and (44), the mixed second-order partial derivative is similarly derived as

$$\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_\ell}\right)^H = -\frac{1}{\sigma_n^2}\Big(-\widetilde{\mathbf{U}}_{\psi_\ell}^H\widetilde{\mathbf{W}}\mathbf{S}^H\Big(\mathbf{y} - \mathbf{S}\widetilde{\mathbf{W}}^H \times \begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}\end{bmatrix}\boldsymbol{\alpha}\Big) + \rho\begin{bmatrix}1 & \\ & \mathbf{U}_{\boldsymbol{\psi}}^H\end{bmatrix}\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H\widetilde{\mathbf{U}}_{\psi_\ell}\boldsymbol{\alpha}\Big), \tag{48}$$

and the expressions $\frac{\partial}{\partial \boldsymbol{\alpha}}(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \psi_\ell})^H$, $\frac{\partial}{\partial \psi_\ell}(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*})^H$, and $\frac{\partial}{\partial \psi_\ell}(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}})^H$ are given in (38). ■

Step 2: The FIM $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ is obtained by applying $-\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\{\cdot\}$ to the second-order derivatives of $\ln p(\mathbf{y},\boldsymbol{\theta}) = \ln p(\boldsymbol{\theta}) + \ln p(\mathbf{y}|\boldsymbol{\theta})$. First, consider the log-likelihood part $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ for which the second-order derivatives are given in Lemma 2. The corresponding submatrices are given by

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*}\right)^H\right\} = \frac{\rho}{\sigma_n^2}\begin{bmatrix}K & \nu_1^*\mathbf{1}_L^T\\ \nu_1\mathbf{1}_L & \zeta_1\mathbf{I}_L + \eta_1(\mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L)\end{bmatrix}, \tag{49}$$

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi},D} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\psi}}\right)^H\right\} = \frac{\rho}{\sigma_n^2}\mathbb{E}_{\boldsymbol{\theta}}\left\{\begin{bmatrix}K & \bar{\mathbf{w}}^H\\ \mathbf{U}_{\boldsymbol{\psi}}^H\bar{\mathbf{w}} & \mathbf{U}_{\boldsymbol{\psi}}^H\mathbf{Q}\end{bmatrix}\begin{bmatrix}\mathbf{0}_L^T\\ \acute{\mathbf{U}}_{\boldsymbol{\psi}}\end{bmatrix}\mathrm{diag}(\alpha_1,\ldots,\alpha_L)\right\} = \frac{\rho}{\sigma_n^2}\begin{bmatrix}\nu_2^*\mathbf{1}_L^T\\ \zeta_2\mathbf{I}_L + \eta_2(\mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L)\end{bmatrix}\mathbb{E}\{\mathrm{diag}(\alpha_1,\ldots,\alpha_L)\}, \tag{50}$$

$$\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial}{\partial \boldsymbol{\psi}}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\psi}}\right)^H\right\} = \frac{1}{\sigma_n^2}\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left\{\frac{\partial \mathbf{m}}{\partial \boldsymbol{\psi}}\frac{\partial \mathbf{m}^H}{\partial \boldsymbol{\psi}} + \left(\frac{\partial \mathbf{m}}{\partial \boldsymbol{\psi}}\frac{\partial \mathbf{m}^H}{\partial \boldsymbol{\psi}}\right)^*\right\} = \frac{2\rho}{\sigma_n^2}\mathrm{Re}\left\{\zeta_3\mathrm{diag}(\mathbb{E}\{|\alpha_1|^2\},\ldots,\mathbb{E}\{|\alpha_L|^2\}) + \eta_3\mathrm{diag}(\mathbb{E}\{\alpha_1^*\},\ldots,\mathbb{E}\{\alpha_L^*\})^H(\mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L)\ \mathrm{diag}(\mathbb{E}\{\alpha_1\},\ldots,\mathbb{E}\{\alpha_L\})\right\}, \tag{51}$$

where

$$\begin{aligned}
\nu_1 &= \boldsymbol{\varphi}_\psi^H\bar{\mathbf{w}},\\
\zeta_1 &= \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\varphi_\psi\Big(2\pi\frac{d}{\lambda}(n-m)\Big)\ [\mathbf{Q}]_{m+1,n+1},\\
\eta_1 &= \boldsymbol{\varphi}_\psi^H\mathbf{Q}\boldsymbol{\varphi}_\psi,\\
\nu_2 &= \acute{\boldsymbol{\varphi}}_\psi^H\bar{\mathbf{w}},\\
\zeta_2 &= \iota 2\pi\frac{d}{\lambda}\sum_{m=0}^{N-1}\sum_{n=1}^{N-1}n\varphi_\psi\Big(2\pi\frac{d}{\lambda}(n-m)\Big)\ [\mathbf{Q}]_{m+1,n+1},\\
\eta_2 &= \boldsymbol{\varphi}_\psi^H\mathbf{Q}\acute{\boldsymbol{\varphi}}_\psi,\\
\zeta_3 &= \Big(2\pi\frac{d}{\lambda}\Big)^2\sum_{m=1}^{N-1}\sum_{n=1}^{N-1}mn\varphi_\psi\Big(2\pi\frac{d}{\lambda}(n-m)\Big)\ [\mathbf{Q}]_{m+1,n+1},\\
\eta_3 &= \acute{\boldsymbol{\varphi}}_\psi^H\mathbf{Q}\acute{\boldsymbol{\varphi}}_\psi,\\
\acute{\boldsymbol{\varphi}}_\psi &\triangleq \iota 2\pi\frac{d}{\lambda}\big[\,0,\ \varphi_\psi(2\pi\tfrac{d}{\lambda}),\ \ldots,\ (N-1)\varphi_\psi\big(2\pi\tfrac{d}{\lambda}(N-1)\big)\,\big]^T\\
&\quad \text{with the characteristic function } \varphi_\psi(t) = \mathbb{E}\left\{e^{\iota t\psi}\right\},\\
\acute{\mathbf{U}}_{\boldsymbol{\psi}} &\triangleq \big[\,\tfrac{\partial \mathbf{u}_N(\psi_1)}{\partial \psi_1}\ \ldots\ \tfrac{\partial \mathbf{u}_N(\psi_L)}{\partial \psi_L}\,\big] \in \mathbb{C}^{N\times L}.
\end{aligned} \tag{52–55}$$

Under the parameter distribution assumptions of $\psi_\ell \sim \mathrm{Unif}[\Delta_1,\Delta_2]$, $\mathbb{E}\{\alpha_\ell\} = 0$, and $\mathbb{E}\{|\alpha_\ell|^2\} = \sigma^2$ for $\ell = 1,\ldots,L$, we have $\varphi_\psi\left(2\pi\frac{d}{\lambda}n\right) = \begin{cases}1, & \text{if } n=0\\ \frac{e^{\iota 2\pi(d/\lambda)n(\Delta_2+\Delta_1)/2}}{\pi(d/\lambda)(\Delta_2-\Delta_1)}\sin\big(2\pi\frac{d}{\lambda}n(\frac{\Delta_2-\Delta_1}{2})\big), & \text{if } n\neq 0\end{cases}$, $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi},D} = \mathbf{0}_{L+1}\mathbf{0}_L^T$, and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D} = \frac{2\rho\sigma^2}{\sigma_n^2}\zeta_3\mathbf{I}_L$. By Lemma 2,

$\mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*} = (\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}})^*$, and $\mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\psi},D}$, $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha},D}$, and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\alpha}^*,D}$ are given by the complex conjugate, Hermitian conjugate, and transpose of $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\psi},D}$, respectively, as seen in (38). The remaining off-diagonal matrices $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}^*,D}$ and $\mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha},D}$ become zero matrices by (39). Hence, we have the FIM $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$, as shown in Theorem 1 with (10) and (11).

Now consider the prior part $\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta},P}$. From $\ln p(\boldsymbol{\theta}) = \ln p(\boldsymbol{\alpha}) + \ln p(\boldsymbol{\psi})$ with $\psi_\ell \sim \text{Unif}[\Delta_1, \Delta_2]$, the elements of the FIM $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ are derived similarly to those of $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ as follows: $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} = -\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}(\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*})^H\}, \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*,P} = -\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\alpha}}(\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}})^H\}, \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = -\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\psi}}(\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}})^H\} = \mathbf{0}_L\mathbf{0}_L^T$, and the remaining off-diagonal matrices of $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ are all zero matrices. This completes the proof. ■

B. Proof of Corollary 1

By exploiting the distribution of the path angle $\psi_\ell \sim \text{Unif}[-1, 1]$ and the definitions of $\boldsymbol{\varphi}_\psi$ in (14) and $\acute{\boldsymbol{\varphi}}_\psi$ in (55), we have $\boldsymbol{\varphi}_\psi = [1, 0, \ldots, 0]^T$ and $\acute{\boldsymbol{\varphi}}_\psi = \mathbf{0}_N$, because $\varphi_\psi(n\pi) = \frac{\sin(\pi n)}{\pi n} = \begin{cases} 1, \text{ if } n=0 \\ 0, \text{ if } n\neq 0 \end{cases}$. In this case, the constituent parameters of $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D}$ and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ in (52) and (54) are given by $\nu_1 = \bar{w}_1$, $\zeta_1 = \text{tr}(\mathbf{Q})$, $\eta_1 = [\mathbf{Q}]_{1,1}$, $\zeta_3 = \pi^2 \sum_{m=1}^{N-1} m^2 [\mathbf{Q}]_{m+1,m+1}$, and $\eta_3 = 0$. By substituting these results into (49) - (51), we have the FIMs $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D}$ and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$, as shown in (15) and (16).

Based on the assumption of $\alpha_\ell \sim \mathcal{CN}(0, \sigma^2)$ for $\ell = 1, \ldots, L$, the diagonal submatrices of the FIM $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ are derived similarly to those of $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ as follows: $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P} = -\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}(\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*})^H\} = \frac{1}{\sigma^2}\mathbf{I}_{L+1} = \mathbf{J}_{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^*,P}$, and $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},P} = \mathbf{0}_L\mathbf{0}_L^T$. Hence, we have the claim. ■

C. Proof of Lemma 1

The FIM $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} + \mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha},P}$ with (10) and (12) can be rewritten as

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = (\rho/\sigma_n^2)\left(K'\mathbf{I}_{L+1} + \begin{bmatrix} 0 & \bar{w}_1^*\mathbf{1}_L^T \\ \bar{w}_1\mathbf{1}_L & \frac{\tau}{L}\mathbf{1}_L\mathbf{1}_L^T \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{0}_L^T \\ \mathbf{0}_L & (\frac{\tau}{L} - [\mathbf{Q}]_{1,1})((L-1)\mathbf{I}_L - \mathbf{I}_L^c) \end{bmatrix}\right), \quad (56)$$

where $K' = K + \frac{\sigma_n^2}{\rho\sigma^2}$, $\mathbf{I}_L^c \triangleq \mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L$ and τ is defined in (19). Then, by using *(F.1)* and *(F.2)* in Lemma 3 below, the second and the third terms in the RHS of (56) are jointly eigen-decomposed as in (57), shown at the bottom of the page, where $\kappa = \sqrt{\tau^2 + 4L|\bar{w}_1|^2} \geq 0$ and $\mathbf{E}$, $\mathbf{f}_1$ and $\mathbf{F}_2$ are given in Lemma 3. The equality (a) holds by $(L-1)\mathbf{I}_L - \mathbf{I}_L^c = \mathbf{F}_2\mathbf{F}_2^H$. Then, by (59) and (60), the matrix $\mathbf{B}$ defined in (57) is unitary due to the construction of $\mathbf{E}$ and $\mathbf{F}_2$ in Lemma 3 and $\mathbf{E}^H\begin{bmatrix} \mathbf{0}_{L-1}^T \\ \mathbf{F}_2 \end{bmatrix} = \begin{bmatrix} e_{21}^*\mathbf{1}_L^T\mathbf{F}_2 \\ e_{22}^*\mathbf{1}_L^T\mathbf{F}_2 \end{bmatrix} = \begin{bmatrix} e_{21}^*\sqrt{L}\mathbf{f}_1^T\mathbf{F}_2 \\ e_{22}^*\sqrt{L}\mathbf{f}_1^T\mathbf{F}_2 \end{bmatrix} = \mathbf{0}_2\mathbf{0}_{L-1}^T$, where e_{ij} are given in Lemma 3. Hence, (57) is the eigen-decomposition of the sum of the second and third terms of the RHS of (56). Then, by considering the first term $K'\mathbf{I}_{L+1}$ in the RHS of (56), the inverse of $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ is eigen-decomposed as in (18). ■

Lemma 3: For any $x \in \mathbb{C}$ and non-negative $\tau, L \in \mathbb{R}$, the following eigen-decompositions hold:

$$\textit{(F.1)} \quad \begin{bmatrix} 0 & x^*\mathbf{1}_L^T \\ x\mathbf{1} & \frac{\tau}{L}\mathbf{1}_L\mathbf{1}_L^T \end{bmatrix} = \mathbf{E}\begin{bmatrix} \frac{\tau-\kappa}{2} & 0 \\ 0 & \frac{\tau+\kappa}{2} \end{bmatrix}\mathbf{E}^H \quad (58)$$

$$\textit{(F.2)} \quad \mathbf{I}_L^c = \mathbf{1}_L\mathbf{1}_L^T - \mathbf{I}_L = \underbrace{\left[\mathbf{f}_1 \middle| \mathbf{F}_2\right]}_{=:\mathbf{F}} \times \left[\begin{array}{c|c} L-1 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{0}_{L-1} & -\mathbf{I}_{L-1} \end{array}\right]\left[\mathbf{f}_1 \middle| \mathbf{F}_2\right]^H, \quad (59)$$

where $\kappa = \sqrt{\tau^2 + 4L|x|^2}$, $\mathbf{f}_1 = \frac{1}{\sqrt{L}}\mathbf{1}_L$, $\mathbf{F}_2$ orthogonal to $\mathbf{f}_1$ is determined such that the matrix $\mathbf{F}$ in (59) is unitary, and the $L \times 2$ matrix $\mathbf{E}$ is given by

$$\mathbf{E} = \begin{cases} \begin{bmatrix} -\frac{\sqrt{2L}}{\kappa}\frac{x^*}{\sqrt{1-\frac{\tau}{\kappa}}} & \frac{\sqrt{2L}}{\kappa}\frac{x^*}{\sqrt{1+\frac{\tau}{\kappa}}} \\ \frac{1}{\sqrt{2L}}\sqrt{1-\frac{\tau}{\kappa}}\mathbf{1}_L & \frac{1}{\sqrt{2L}}\sqrt{1+\frac{\tau}{\kappa}}\mathbf{1}_L \end{bmatrix}, & \text{if } |x|^2 \neq 0 \\ \begin{bmatrix} 1 & 0 \\ \mathbf{0}_L & \frac{1}{\sqrt{L}}\mathbf{1}_L \end{bmatrix}, & \text{if } |x|^2 = 0 \end{cases} \triangleq \begin{bmatrix} e_{11} & e_{12} \\ e_{21}\mathbf{1}_L & e_{22}\mathbf{1}_L \end{bmatrix}. \quad (60)$$

Proof: Proof is by direct computation. ■

$$\begin{aligned} &\mathbf{E}\begin{bmatrix} \frac{\tau-\kappa}{2} & 0 \\ 0 & \frac{\tau+\kappa}{2} \end{bmatrix}\mathbf{E}^H + \left(\frac{\tau}{L} - [\mathbf{Q}]_{1,1}\right)\left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{f}_1 & \mathbf{F}_2 \end{array}\right]\left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{0}_{L-1} & L\mathbf{I}_{L-1} \end{array}\right]\left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{f}_1 & \mathbf{F}_2 \end{array}\right]^H \\ &\overset{(a)}{=} \mathbf{E}\begin{bmatrix} \frac{\tau-\kappa}{2} & 0 \\ 0 & \frac{\tau+\kappa}{2} \end{bmatrix}\mathbf{E}^H + \left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{f}_1 & \mathbf{F}_2 \end{array}\right]\left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{0}_{L-1} & (\tau - L\,[\mathbf{Q}]_{1,1})\,\mathbf{I}_{L-1} \end{array}\right]\left[\begin{array}{c|c} 0 & \mathbf{0}_{L-1}^T \\ \hline \mathbf{f}_1 & \mathbf{F}_2 \end{array}\right]^H \\ &= \mathbf{E}\begin{bmatrix} \frac{\tau-\kappa}{2} & 0 \\ 0 & \frac{\tau+\kappa}{2} \end{bmatrix}\mathbf{E}^H + \begin{bmatrix} \mathbf{0}_{L-1}^T \\ \mathbf{F}_2 \end{bmatrix}\left[(\tau - L\,[\mathbf{Q}]_{1,1})\,\mathbf{I}_{L-1}\right]\begin{bmatrix} \mathbf{0}_{L-1}^T \\ \mathbf{F}_2 \end{bmatrix}^H \\ &= \underbrace{\left[\mathbf{E}\middle|\begin{array}{c} \mathbf{0}_{L-1}^T \\ \mathbf{F}_2 \end{array}\right]}_{=:\mathbf{B}}\begin{bmatrix} \frac{\tau-\kappa}{2} & & \\ & \frac{\tau+\kappa}{2} & \\ & & (\tau - L\,[\mathbf{Q}]_{1,1})\,\mathbf{I}_{L-1} \end{bmatrix}\left[\mathbf{E}\middle|\begin{array}{c} \mathbf{0}_{L-1}^T \\ \mathbf{F}_2 \end{array}\right]^H, \end{aligned} \quad (57)$$

$$\breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\psi},D} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\alpha}|\boldsymbol{\psi}}\left\{\frac{\partial}{\partial \boldsymbol{\alpha}^*}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\alpha};\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right)^H\right\} = \mathbf{0}_{L+1}\mathbf{0}_L^T \tag{61}$$

$$\begin{aligned}[\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}]_{p,q} &= -\mathbb{E}_{\boldsymbol{\alpha}|\boldsymbol{\psi}}\left\{\mathbb{E}_{\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\psi}}\left\{\frac{\partial}{\partial \psi_p}\left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\alpha};\boldsymbol{\psi})}{\partial \psi_q}\right)^H\right\}\right\} = \frac{2}{\sigma_n^2}\mathbb{E}_{\boldsymbol{\alpha}|\boldsymbol{\psi}}\left\{\mathrm{Re}\left\{\frac{\partial \mathbf{m}^H}{\partial \psi_p}\frac{\partial \mathbf{m}}{\partial \psi_q}\right\}\right\} \\ &= \frac{2\rho}{\sigma_n^2}\mathbb{E}_{\boldsymbol{\alpha}|\boldsymbol{\psi}}\left\{\mathrm{Re}\left\{\alpha_p^*\alpha_q\frac{\partial \mathbf{u}_N^H(\psi_p)}{\partial \psi_p}\mathbf{Q}\frac{\partial \mathbf{u}_N(\psi_q)}{\partial \psi_q}\right\}\right\} \end{aligned} \tag{62}$$

$$= \begin{cases} \dfrac{2\rho}{\sigma_n^2}\left(2\pi\dfrac{d}{\lambda}\right)^2\mathbb{E}\left\{|\alpha_p|^2\right\}\displaystyle\sum_{m=1}^{N-1}\sum_{n=1}^{N-1} mne^{\iota 2\pi\frac{d}{\lambda}\psi_p(n-m)}\ [\mathbf{Q}]_{m+1,n+1}, & \text{if } p=q \\ \dfrac{2\rho}{\sigma_n^2}\left(2\pi\dfrac{d}{\lambda}\right)^2\mathrm{Re}\left\{\mathbb{E}\left\{\alpha_p^*\right\}\mathbb{E}\left\{\alpha_q\right\}\displaystyle\sum_{m=1}^{N-1}\sum_{n=1}^{N-1} mne^{\iota 2\pi\frac{d}{\lambda}(\psi_q n-\psi_p m)}\ [\mathbf{Q}]_{m+1,n+1}\right\}, & \text{if } p\neq q, \end{cases}$$

D. Proof of Theorem 2

Since the second-order derivative $\frac{\partial}{\partial \boldsymbol{\alpha}^*}(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^*})^H$ in (35) of Lemma 2 is not a function of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ is given, the FIM for $\boldsymbol{\alpha}$ is given by $\breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D} = \frac{\rho}{\sigma_n^2}\begin{bmatrix} K & \bar{\mathbf{w}}^H\mathbf{U}_{\boldsymbol{\psi}} \\ \mathbf{U}_{\boldsymbol{\psi}}^H\bar{\mathbf{w}} & \mathbf{U}_{\boldsymbol{\psi}}^H\mathbf{Q}\mathbf{U}_{\boldsymbol{\psi}} \end{bmatrix} = \breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},D}^*$. Applying $-\mathbb{E}_{\mathbf{y},\boldsymbol{\alpha}|\boldsymbol{\psi}}\{\cdot\}$ to (48) and (36), respectively, we obtain $\breve{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\psi},D}$ and $\breve{\mathbf{J}}_{\boldsymbol{\psi}\boldsymbol{\psi},D}$ as in (61) and (62), shown at the top of the page, where $\frac{\partial \mathbf{u}_N^H(\psi)}{\partial \psi}$ is defined in (41). Substituting the assumption of independent Rayleigh fading $\alpha_\ell \sim \mathcal{CN}(0,\sigma^2)$, $\ell = 1,\ldots,L$ into (62) yields $[\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi},D}]_{p,q} = 0$ for $p \neq q$. The remaining $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},D}$ and $\mathbf{J}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}},P}$ can be derived in a similar way to that in the proof of Theorem 1. ■

REFERENCES

[1] S. Noh, H. Yu, and Y. Sung, "Training signal design for sparse channel estimation in millimeter-wave communication with intelligent reflecting surfaces," in *Proc. IEEE Int. Conf. Commun. Workshops*, Montreal, QC, Canada, Jun. 2021, pp. 1–6.

[2] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[3] E. Basar *et al.*, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, Aug. 2019.

[4] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Apr. 2020.

[5] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[6] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[7] X. Ying, U. Demirhan, and A. Alkhateeb, "Relay aided intelligent reconfigurable surfaces: Achieving the potential without so many antennas," 2020, *arXiv:2006.06644*. [Online]. Available: http://arxiv.org/abs/2006.06644

[8] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[9] J. Zhang, C. Qi, P. Li, and P. Lu, "Channel estimation for reconfigurable intelligent surface aided massive MIMO system," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Atlanta, GA, USA, May 2020, pp. 1–5.

[10] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, early access, Apr. 12, 2021, doi: 10.1109/TCOMM.2021.3072729.

[11] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, U.K., May 2019, pp. 4659–4663.

[12] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[13] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Barcelona, Spain, May 2020, pp. 5000–5004.

[14] M. Xu, S. Zhang, C. Zhong, J. Ma, and O. A. Dobre, "Ordinary differential equation-based CNN for channel extrapolation over RIS-assisted communication," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1921–1925, Jun. 2021.

[15] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using the affine precoding framework and the CRB," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 940–954, Mar. 2006.

[16] D. P. Bertsekas, "On the Goldstein-Levitin-Polyak gradient projection method," *IEEE Trans. Autom. Control*, vol. AC-21, no. 2, pp. 174–184, Apr. 1976.

[17] S. Bubeck, *Convex Optimization: Algorithms Complexity*. Boston, MA, USA: Now, 2017.

[18] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multi path with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[19] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[20] W. Zhang, J. Xu, W. Xu, D. W. K. Ng, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, Feb. 2021.

[21] F.-L. Luo and C. J. Zhang, *Signal Processing for 5G: Algorithms and Implementations*. New York, NY, USA: Wiley, 2016.

[22] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.

[23] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

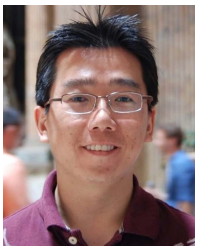
[25] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[26] A. Adhikary *et al.*, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, Jun. 2014.

[27] J. García-Morales, G. Femenias, and F. Riera-Palou, "Energy-efficient access-point sleep-mode techniques for cell-free mmWave massive MIMO networks with non-uniform spatial traffic density," *IEEE Access*, vol. 8, pp. 137587–137605, Jul. 2020.

[28] E. De Carvalho and D. T. M. Slock, "Cramér–Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Paris, France, Apr. 1997, pp. 129–132.
[29] S.-M. Omar, D. T. M. Slock, and O. Bazzi, "Bayesian and deterministic CRBs for semi-blind channel estimation in SIMO single carrier cyclic prefix systems," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Toronto, ON, Canada, Jan. 2011, pp. 1682–1686.
[30] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. H-Microw., Opt. Antennas*, vol. 130, no. 1, pp. 11–16, Feb. 1983.
[31] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020.
[32] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," in *Proc. IEEE Global Commun. Conf.*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
[33] T. Cuvelier and R. W. Heath, Jr., "MmWave MU-MIMO for aerial networks," in *Proc. 15th Int. Symp. Wireless Commun. Syst.*, Lisbon, Portugal, Aug. 2018, pp. 1–6.
[34] L. Wu, Z. Zhang, J. Dang, Y. Wu, H. Liu, and J. Wang, "Joint user identification and channel estimation over rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6803–6807, Jun. 2020.
[35] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[36] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.
[37] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[38] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.
[39] R. Borhani, J. Watt, and A. K. Katsaggelos, *Machine Learning Refined: Foundations, Algorithms, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[40] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Jun. 2020, pp. 1–6.
[41] B. Zheng, C. You, and R. Zhang, "Efficient channel estimation for double-IRS aided multi-user MIMO system," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3818–3832, Jun. 2021.
[42] B. Zheng *et al.*, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.
[43] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Jun. 2021.

Song Noh (Member, IEEE) received the B.S. degree in electrical engineering from Soongsil University, Seoul, South Korea, in 2008, the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015. From 2015 to 2018, he was a System Engineer with the Next Generation and Standards Group, Intel Corporation, Hillsboro, OR, USA, and participated in the design of PHY/MAC layer protocol for millimeter-wave mobile broadband systems. Since 2018, he has been a Faculty Member with the Department of Information Telecommunication Engineering, Incheon National University, Incheon, South Korea. His research interests include design and analysis of millimeter-wave communication systems and adaptive signal processing for wireless communications. Along with coauthors, he was a co-recipient of the Communication Theory Symposium Best Paper Award from the 2014 IEEE Global Communications Conference (GLOBECOM).



Heejung Yu (Senior Member, IEEE) received the B.S. degree in radio science and engineering from Korea University, Seoul, South Korea, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2001 and 2011, respectively. From 2001 to 2012, he was with the Electronics and Telecommunications Research Institute (ETRI), Daejeon. From 2012 to 2019, he was with Yeungnam University, South Korea. He is currently an Associate Professor with the Department of Electronics and Information Engineering, Korea University, Sejong, South Korea. His areas of research interests include statistical signal processing and communication theory.

Youngchul Sung (Senior Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2005. Before joining the Ph.D. program, he was with LG Electronics Ltd., Seoul, from 1995 to 2000. From 2005 to 2007, he was a Senior Engineer with the Corporate Research and Development Center, Qualcomm, Inc., San Diego, CA, USA, and participated in design of Qualcomm's 3GPP R6 WCDMA base station modem. Since 2007, he has been a Faculty Member with the School of Electrical Engineering, KAIST, Daejeon, South Korea. His research interests include signal processing for communications, statistical signal processing, and statistical inference and learning with applications to wireless communications and related areas. He is currently a Vice Director of the IEEE Communications Society Asia-Pacific Board and an Executive Editor of *ICT Express*. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2017 to 2021 and the IEEE SIGNAL PROCESSING LETTERS from 2012 to 2014.