

Angular-Domain Channel Estimation for One-Bit Massive MIMO Systems: Performance Bounds and Algorithms

Liu, Fangqing; Zhu, Heng; Li, Changheng; Li, Jian; Wang, Pu; Orlik, Philip V.

TR2020-009 January 17, 2020

Abstract

We consider angular-domain channel estimation in massive MIMO systems using one-bit analog-to-digital converters (ADCs) with various thresholding schemes at the receivers. We first derive the performance bounds for estimating angular domain channel parameters, including the angles-of-arrival (AoA), angles-of-departure (AoD) and the associated path gains. Specifically, we derive 1) the deterministic Cramer-Rao bound (CRB) when all of the angular-domain channel parameters are treated as deterministic unknowns; 2) the hybrid CRB when some parameters have known prior probability density functions (pdfs) while the rest are assumed to be deterministic unknowns; 3) the Bayesian CRB when all of them have known prior pdfs. We also consider using the maximum likelihood (ML) method for channel estimation and a computationally efficient relaxation based cyclic algorithm (referred to as 1bRELAX) to obtain the ML estimates. When the prior information is available, the maximum a posteriori (MAP) and joint ML-MAP (JML-MAP) estimators are derived. We also use the one-bit Bayesian information criterion (1bBIC) to determine the number of scattering paths. Numerical examples are provided to verify the derived performance bounds with different thresholding schemes and demonstrate the performance of the proposed channel estimation algorithms.

IEEE Transactions on Vehicular Technology

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Angular-Domain Channel Estimation for One-Bit Massive MIMO Systems: Performance Bounds and Algorithms

Fangqing Liu, Heng Zhu, Changheng Li, Jian Li, *Fellow, IEEE*, Pu Wang, *Senior Member, IEEE* and Philip V. Orlik, *Senior Member, IEEE*

Abstract—We consider angular-domain channel estimation in massive MIMO systems using one-bit analog-to-digital converters (ADCs) with various thresholding schemes at the receivers. We first derive the performance bounds for estimating angular-domain channel parameters, including the angles-of-arrival (AoA), angles-of-departure (AoD) and the associated path gains. Specifically, we derive 1) the deterministic Cramér-Rao bound (CRB) when all of the angular-domain channel parameters are treated as deterministic unknowns; 2) the hybrid CRB when some parameters have known prior probability density functions (pdfs) while the rest are assumed to be deterministic unknowns; 3) the Bayesian CRB when all of them have known prior pdfs. We also consider using the maximum likelihood (ML) method for channel estimation and a computationally efficient relaxation based cyclic algorithm (referred to as 1bRELAX) to obtain the ML estimates. When the prior information is available, the maximum *a posteriori* (MAP) and joint ML-MAP (JML-MAP) estimators are derived. We also use the one-bit Bayesian information criterion (1bBIC) to determine the number of scattering paths. Numerical examples are provided to verify the derived performance bounds with different thresholding schemes and demonstrate the performance of the proposed channel estimation algorithms.

Index Terms—Massive MIMO, angular-domain channel estimation, one-bit quantization, antenna-varying thresholds, time-varying thresholds, deterministic CRB, hybrid CRB, Bayesian CRB, ML, JML-MAP, MAP, 1bRELAX, 1bBIC.

I. INTRODUCTION

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems can offer significant throughput increase for wireless communications by taking advantage of the large available bandwidth at the mmWave frequency band [1]–[4]. Due to the small wavelengths and antenna sizes, hundreds of antennas can be fitted into a small space to provide a large array gain to compensate for the significant path losses at the mmWave band. However, the prohibitive cost and power consumption needed by a large number of high-resolution (e.g., 8-12 bits) analog-to-digital

converters (ADCs) at the antenna outputs make the system impractical.

One promising solution is to use low resolution ADCs (e.g., 1-4 bits) at the receivers [5]–[11], to drastically reduce the power consumption and cost of massive MIMO systems. Compared with the unquantized (i.e., infinite precision quantization) case, the capacity loss using one-bit ADCs is only 1.96 dB at low signal-to-noise ratios (SNRs) [12], [13].

Channel estimation has been studied for one-bit massive MIMO systems in [6], [14]–[18]. In [6], a least squares (LS) estimator was proposed by simply treating the quantization noise as additive Gaussian noise. A maximum likelihood (ML) estimator was developed in [14] by imposing a channel norm constraint on a convex optimization problem. In [16], a Bussgang decomposition based linear minimum mean-squared error (BLMMSE) estimator was proposed for one-bit uplink massive MIMO channel estimation. [7] proposed a Bayes-optimal joint channel-and-data (JCD) estimation algorithm, where the reliably detected payload data is in turn used to iteratively refine the channel estimate. In all of the aforementioned works, fixed zero-threshold is used for one-bit ADCs and the noise variance is assumed known. For non-zero thresholding schemes, [19] showed that an unknown non-zero threshold has a negative effect on the achievable channel estimation accuracy, especially at high SNRs. An adaptive quantization (AQ) scheme was proposed in [20], where the thresholds for the one-bit ADCs are assumed known and dynamically adjusted to converge to the optimal ones, to significantly improve the channel estimation performance. However, a large number of expensive high-resolution digital-to-analog converters (DACs) may be needed to generate the thresholds required by the AQ scheme.

All of the aforementioned methods focus on estimating each element of the channel matrix directly for one-bit massive MIMO systems. They can be regarded as non-parametric approaches. However, due to the sparse nature of the channel at mmWave bands (e.g., 28 GHz), the channel matrix can be represented in the angular domain in a parametric way with a much smaller set of parameters, i.e., angles-of-arrival (AoA), angles-of-departure (AoD) and path gains, which can be estimated with much shorter pilots. In this parametric approach, more accurate channel estimates can be expected since the number of unknown parameters is dramatically reduced. In addition, these angular states can be used for beamforming to compensate for the high path losses at mmWave bands [21].

F. Liu, H. Zhu and C. Li are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, Anhui, P.R. China (emails: fqliu@mail.ustc.edu.cn, zh2013@mail.ustc.edu.cn, lch8@mail.ustc.edu.cn).

J. Li is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, Anhui, P.R. China, and also with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611, USA (e-mail: li@dsp.ufl.edu).

P. Wang and P. V. Orlik are with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA (e-mail: pwang@merl.com, porlik@merl.com).

Hence, the angular-domain channel estimation is of significant interest. In [9], a generalized approximate message passing (GAMP) based algorithm was proposed by turning the one-bit angular-domain mmWave massive MIMO channel estimation problem into a noisy quantized compressed sensing problem, to provide a superior performance-complexity tradeoff.

In this paper, we focus on the angular-domain channel estimation for one-bit mmWave massive MIMO systems. Specifically, we consider zero-thresholding and *time-varying/antenna-varying* thresholding schemes for the one-bit ADCs, and propose computationally efficient one-bit angular-domain channel estimation algorithms. Our main contributions are summarized as follows:

- We derive the deterministic Cramér-Rao bounds (CRBs) for estimating the AoA, AoD and the associated path gains for various thresholding schemes. We pay particular attention to the practical case of unknown noise variance and highlight the ambiguity between the path gains and noise variance¹.
- For knowledge-aided channel estimation, we model the AoA, AoD and path gains as random parameters with known prior probability density functions (pdfs) and derive the performance bounds. In particular, we derive two hybrid CRBs when a subset of channel parameters have known prior pdfs while the remaining parameters are treated as deterministic unknowns. A Bayesian CRB is also derived by assuming all angular-domain channel parameters have known prior pdfs. We characterize the channel estimation performance of using different thresholding schemes.
- We consider the maximum likelihood (ML) estimator for channel estimation to deal with the deterministic unknowns. A computationally efficient relaxation based cyclic algorithm, referred to 1bRELAX, is used to obtain the ML estimates.
- For the knowledge-aided case, the maximum *a posteriori* probability (MAP) and joint ML and MAP (JML-MAP) channel estimation methods are derived.
- We use a simple and effective Bayesian information criterion (BIC), referred to as 1bBIC, to estimate the channel path number.

The rest of this paper is organized as follows. An angular-domain channel model for mmWave massive MIMO systems with one-bit ADCs is introduced in Section II. The deterministic CRB for angular-domain channel estimation is derived in Section III. The hybrid and Bayesian CRBs for knowledge-aided channel estimation are provided in Section IV. In Section V, we present the ML, JML-MAP and MAP algorithms for channel estimation and the 1bBIC rule for channel path number determination. The numerical examples are given in Section VI and we conclude the paper in Section VII.

Notations: $j \triangleq \sqrt{-1}$. Bold lower letters denote vectors. Bold uppercase letters denote matrices. \mathbf{I}_N denotes an $N \times N$

¹Part of these results were presented in [22] and [23]. Different from [20], the noise variance is considered as an unknown deterministic parameter in our angular-domain CRB analysis. To avoid the ambiguity problems, non-zero thresholds are necessary, especially low-cost practical thresholding schemes.

identity matrix. a_m denotes the m^{th} element of a vector \mathbf{a} , and $A_{p,q}$ denotes the $(p,q)^{\text{th}}$ element of a matrix \mathbf{A} . Superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ represent the complex conjugate, transpose, conjugate transpose and inverse operations, respectively. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the Euclidean and Frobenius norms, respectively. $\Re(\cdot)$ and $\Im(\cdot)$ denote taking the real and imaginary parts, respectively. \otimes denotes the Kronecker product. $\mathbf{A} \odot \mathbf{B}$ denotes the Khatri-Rao product between matrices \mathbf{A} and \mathbf{B} , which is given by $[\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_N \otimes \mathbf{b}_N]$, where \mathbf{a}_m and \mathbf{b}_m denote, respectively, the m^{th} columns of \mathbf{A} and \mathbf{B} . $\text{vec}(\mathbf{A})$ denotes the vectorization operation of stacking the columns of \mathbf{A} on top of each other. $\mathbb{E}_x[\cdot]$ denotes the expectation operation with respect to (w.r.t.) x .

II. SYSTEM MODEL

A. Angular-Domain mmWave Massive MIMO System Model

Consider a massive MIMO system with N_t transmit antennas at the mobile station (MS) and N_r receive antennas at the base station (BS) in Fig. 1. In the training phase, a pilot signal of length K is sent from the MS to the BS. Then the received signal $\mathbf{Y} \in \mathbb{C}^{N_r \times K}$ at the BS has the form:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}, \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K] \in \mathbb{C}^{N_t \times K}$ is the pilot signal with average transmit power $\mathbb{E}[\mathbf{x}_k^H \mathbf{x}_k] = \rho$, $1 \leq k \leq K$, and $\mathbf{V} \in \mathbb{C}^{N_r \times K}$ is the circularly symmetric complex-valued white Gaussian noise with i.i.d. $\mathcal{CN}(0, \sigma_v^2)$ entries.

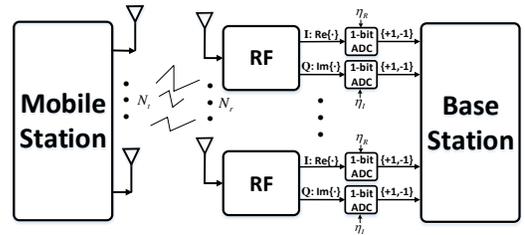


Fig. 1. One-bit massive MIMO system diagram.

Consider the channel model parameterized via spatial angles and gains associated with different propagation paths [24]. Assume that there are N_s active scatterers between the MS and the BS. Denote α_l as the complex gain of the l^{th} scattering path, θ_l and φ_l as the associated AoA and AoD, respectively. Then for a uniform linear array (ULA), the steering vectors for the l^{th} scatterer at the BS and MS are, respectively,

$$\mathbf{a}_{\text{BS}}(\theta_l) = \left[1, e^{j2\pi \sin(\theta_l) \frac{d_r}{\lambda}}, \dots, e^{j(N_r-1)2\pi \sin(\theta_l) \frac{d_r}{\lambda}} \right]^T, \quad (2a)$$

and

$$\mathbf{a}_{\text{MS}}(\varphi_l) = \left[1, e^{j2\pi \sin(\varphi_l) \frac{d_t}{\lambda}}, \dots, e^{j(N_t-1)2\pi \sin(\varphi_l) \frac{d_t}{\lambda}} \right]^T, \quad (2b)$$

where λ is the wavelength, and d_r and d_t are the inter-element spacings of the ULAs at the BS and MS, respectively. Then

the channel matrix \mathbf{H} is given by:

$$\mathbf{H} = \sum_{l=1}^{N_s} \alpha_l \mathbf{a}_{\text{BS}}(\theta_l) \mathbf{a}_{\text{MS}}^H(\varphi_l) \triangleq \mathbf{A}_{\text{BS}}(\boldsymbol{\theta}) \mathbf{H}_{\Lambda} \mathbf{A}_{\text{MS}}^H(\boldsymbol{\varphi}), \quad (3)$$

where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{N_s}]^T$, $\boldsymbol{\varphi} = [\varphi_1, \dots, \varphi_{N_s}]^T$ and $\mathbf{H}_{\Lambda} = \text{diag}\{\alpha_1, \dots, \alpha_{N_s}\}$. Also, the l^{th} columns of $\mathbf{A}_{\text{BS}}(\boldsymbol{\theta}) \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{A}_{\text{MS}}(\boldsymbol{\varphi}) \in \mathbb{C}^{N_t \times N_s}$ are given as $\mathbf{a}_{\text{BS}}(\theta_l)$ of (2a) and $\mathbf{a}_{\text{MS}}(\varphi_l)$ of (2b), respectively. It's worth noting that the angular-domain channel model of (3) is similar to that encountered in MIMO radar systems [25], [26]. Plugging (3) into (1), we can represent the received signal with the angular-domain parameters, $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, $\{\alpha_l\}_{l=1}^{N_s}$, as:

$$\mathbf{Y} = \mathbf{A}_{\text{BS}}(\boldsymbol{\theta}) \mathbf{H}_{\Lambda} \mathbf{A}_{\text{MS}}^H(\boldsymbol{\varphi}) \mathbf{X} + \mathbf{V}. \quad (4)$$

Let $\mathbf{y} = \text{vec}(\mathbf{Y})$. Then we have:

$$\begin{aligned} \mathbf{y} &= \text{vec}\left(\mathbf{A}_{\text{BS}}(\boldsymbol{\theta}) \mathbf{H}_{\Lambda} \mathbf{A}_{\text{MS}}^H(\boldsymbol{\varphi}) \mathbf{X}\right) + \mathbf{v} \\ &= \left[\left(\mathbf{X}^T \mathbf{A}_{\text{MS}}^*(\boldsymbol{\varphi})\right) \odot \mathbf{A}_{\text{BS}}(\boldsymbol{\theta})\right] \boldsymbol{\alpha} + \mathbf{v} \\ &\triangleq \boldsymbol{\Gamma}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \boldsymbol{\alpha} + \mathbf{v}, \end{aligned} \quad (5)$$

where $\mathbf{v} = \text{vec}(\mathbf{V})$ and $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{N_s}]^T$. It is seen from (5) that the angular parameters $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ are embedded in $\boldsymbol{\Gamma}$ in a nonlinear fashion, while the channel path gains $\boldsymbol{\alpha}$ appear as linear parameters. As shown in Fig. 1, the in-phase (I) and quadrature (Q) components of the received signal can be collected in a vector as:

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} = \bar{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \bar{\boldsymbol{\alpha}} + \bar{\mathbf{v}}, \quad (6)$$

where $\bar{\boldsymbol{\alpha}} = [\boldsymbol{\alpha}_R^T, \boldsymbol{\alpha}_I^T]^T \in \mathbb{R}^{2N_s \times 1}$, $\bar{\mathbf{v}} = [\mathbf{v}_R^T, \mathbf{v}_I^T]^T \in \mathbb{R}^{2KN_r \times 1}$ and

$$\bar{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \begin{bmatrix} \boldsymbol{\Gamma}_R & -\boldsymbol{\Gamma}_I \\ \boldsymbol{\Gamma}_I & \boldsymbol{\Gamma}_R \end{bmatrix} \in \mathbb{R}^{2KN_r \times 2N_s},$$

with $\mathbf{X}_R = \Re\{\mathbf{X}\}$ and $\mathbf{X}_I = \Im\{\mathbf{X}\}$. Note that the dependence of $\boldsymbol{\Gamma}$ on $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ may be neglected for notational brevity.

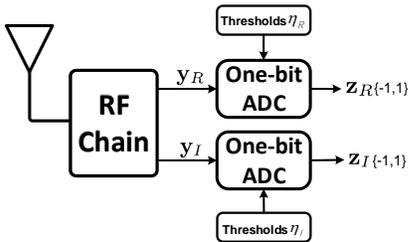


Fig. 2. The diagram for one-bit sampling.

B. One-Bit Quantization

In one-bit massive MIMO systems, two one-bit ADCs are used at the output of each receive antenna to quantize both I and Q components of the received signal. As shown in Fig. 2, the one-bit quantized data $\mathbf{z}_{R/I}$ (i.e., ± 1) is obtained by comparing the unquantized analog signal $\mathbf{y}_{R/I}$ to the thresholds $\boldsymbol{\eta}_{R/I}$:

$$\mathbf{z}_R = \text{sign}(\mathbf{y}_R - \boldsymbol{\eta}_R), \quad (7a)$$

$$\mathbf{z}_I = \text{sign}(\mathbf{y}_I - \boldsymbol{\eta}_I), \quad (7b)$$

where

$$\mathbf{y}_R = \boldsymbol{\Gamma}_R \boldsymbol{\alpha}_R - \boldsymbol{\Gamma}_I \boldsymbol{\alpha}_I + \mathbf{v}_R, \quad (8a)$$

$$\mathbf{y}_I = \boldsymbol{\Gamma}_I \boldsymbol{\alpha}_R + \boldsymbol{\Gamma}_R \boldsymbol{\alpha}_I + \mathbf{v}_I, \quad (8b)$$

and $\boldsymbol{\eta}_R \in \mathbb{R}^{KN_r \times 1}$ and $\boldsymbol{\eta}_I \in \mathbb{R}^{KN_r \times 1}$ are the given threshold vector used at the N_r pairs of the I and Q channels over K pilots, respectively. $\text{sign}(x)$ is the element-wise one-bit quantization function, which returns -1 if $x \leq 0$ and 1 otherwise.

All real-valued unknown parameters can be grouped as $\boldsymbol{\chi} = [\boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\alpha}_R^T, \boldsymbol{\alpha}_I^T]^T \in \mathbb{R}^{4N_s \times 1}$ when the noise variance σ_v^2 is known. When the noise variance is unknown, the unknown parameter vector can be expanded as $\boldsymbol{\psi} = [\boldsymbol{\chi}^T, \sigma_v]^T \in \mathbb{R}^{(4N_s+1) \times 1}$. The problem of interest herein is to estimate $\boldsymbol{\chi}$ with the quantized output $\bar{\mathbf{z}} = [\mathbf{z}_R^T, \mathbf{z}_I^T]^T$ and a given threshold vector $\bar{\boldsymbol{\eta}} = [\boldsymbol{\eta}_R^T, \boldsymbol{\eta}_I^T]^T$.

III. DETERMINISTIC CRB

A. Log-Likelihood Function

Note that $\mathbf{v}_R \sim \mathcal{N}(0, \sigma_v^2/2)$ and $\mathbf{v}_I \sim \mathcal{N}(0, \sigma_v^2/2)$. By invoking the independence of the signed measurements $\bar{\mathbf{z}}$ over K pilots, N_r receive antennas and between the I and Q channels, the log-likelihood function of $\bar{\mathbf{z}}$ is given by:

$$\ln p(\bar{\mathbf{z}}|\boldsymbol{\chi}) = \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}) + \ln \Phi(z_{I,m} \zeta_{I,m})], \quad (9)$$

where

$$\zeta_{R,m} = \frac{\kappa_{R,m} - \eta_{R,m}}{\sigma_v / \sqrt{2}}, \quad (10a)$$

$$\zeta_{I,m} = \frac{\kappa_{I,m} - \eta_{I,m}}{\sigma_v / \sqrt{2}}, \quad (10b)$$

with

$$\kappa_{R,m} \triangleq \boldsymbol{\Gamma}_R^T(m) \boldsymbol{\alpha}_R - \boldsymbol{\Gamma}_I^T(m) \boldsymbol{\alpha}_I, \quad (11a)$$

$$\kappa_{I,m} \triangleq \boldsymbol{\Gamma}_I^T(m) \boldsymbol{\alpha}_R + \boldsymbol{\Gamma}_R^T(m) \boldsymbol{\alpha}_I, \quad (11b)$$

denoting the real and imaginary parts of the noise-free receive signal, respectively. In addition, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ is the standard normal cumulative distribution function (cdf) and $\boldsymbol{\Gamma}_{R/I}^T(m)$ is the m^{th} row of $\boldsymbol{\Gamma}_{R/I}$.

B. Fisher Information Matrix

1) *Known Noise Variance*: The $(p, q)^{\text{th}}$ element of the Fisher information matrix (FIM) can be calculated as:

$$[\mathbf{J}_F(\boldsymbol{\chi})]_{p,q} = \mathbb{E}_{\bar{\mathbf{z}}} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}}|\boldsymbol{\chi})}{\partial \chi_p \partial \chi_q} \right], \quad (12)$$

where $\frac{\partial}{\partial \chi_p}$ represents the partial derivative w.r.t. the p^{th} element of $\boldsymbol{\chi}$. Given the grouped channel parameters, the FIM has the following block structure:

$$\mathbf{J}_F(\boldsymbol{\chi}) = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}, \boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\theta}, \boldsymbol{\varphi}} & \mathbf{F}_{\boldsymbol{\theta}, \boldsymbol{\alpha}_R} & \mathbf{F}_{\boldsymbol{\theta}, \boldsymbol{\alpha}_I} \\ \mathbf{F}_{\boldsymbol{\varphi}, \boldsymbol{\theta}}^T & \mathbf{F}_{\boldsymbol{\varphi}, \boldsymbol{\varphi}} & \mathbf{F}_{\boldsymbol{\varphi}, \boldsymbol{\alpha}_R} & \mathbf{F}_{\boldsymbol{\varphi}, \boldsymbol{\alpha}_I} \\ \mathbf{F}_{\boldsymbol{\alpha}_R, \boldsymbol{\theta}}^T & \mathbf{F}_{\boldsymbol{\alpha}_R, \boldsymbol{\varphi}}^T & \mathbf{F}_{\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_R} & \mathbf{F}_{\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I} \\ \mathbf{F}_{\boldsymbol{\alpha}_I, \boldsymbol{\theta}}^T & \mathbf{F}_{\boldsymbol{\alpha}_I, \boldsymbol{\varphi}}^T & \mathbf{F}_{\boldsymbol{\alpha}_I, \boldsymbol{\alpha}_R}^T & \mathbf{F}_{\boldsymbol{\alpha}_I, \boldsymbol{\alpha}_I} \end{bmatrix} \in \mathbb{R}^{4N_s \times 4N_s}. \quad (13)$$

The CRB matrix is given by the inverse of the FIM:

$$\mathbf{CRB}(\boldsymbol{\chi}) = [\mathbf{J}_F(\boldsymbol{\chi})]^{-1}. \quad (14)$$

The (p, p) th element of the CRB matrix establishes the best performance bound of an unbiased estimator.

To derive the FIM, we note that

$$\mathbb{E}_{\bar{\mathbf{z}}} \left\{ \frac{1}{\Phi^2(z_{R,m} \zeta_{R,m})} \right\} = \frac{1}{\Phi(\zeta_{R,m})} + \frac{1}{\Phi(-\zeta_{R,m})}, \quad (15a)$$

$$\mathbb{E}_{\bar{\mathbf{z}}} \left\{ \frac{1}{\Phi^2(z_{I,m} \zeta_{I,m})} \right\} = \frac{1}{\Phi(\zeta_{I,m})} + \frac{1}{\Phi(-\zeta_{I,m})}, \quad (15b)$$

where we have used the fact that the quantized outputs $z_{R,m}$ ($z_{I,m}$) are binary (± 1) random variables with probability $\Phi(\pm \zeta_{R,m})$ ($\Phi(\pm \zeta_{I,m})$), respectively.

Combining (9), (12) and (15), we have the exact expression of the FIM²:

$$[\mathbf{J}_F(\boldsymbol{\chi})]_{p,q} = \frac{1}{\pi \sigma_v^2} \sum_{m=1}^{KN_r} \left[f(\zeta_{R,m}) \frac{\partial \kappa_{R,m}}{\partial \chi_p} \frac{\partial \kappa_{R,m}}{\partial \chi_q} + f(\zeta_{I,m}) \frac{\partial \kappa_{I,m}}{\partial \chi_p} \frac{\partial \kappa_{I,m}}{\partial \chi_q} \right], \quad 1 \leq p, q \leq 4N_s, \quad (16)$$

where

$$f(\zeta_{R/I,m}) \triangleq \left(\frac{1}{\Phi(\zeta_{R/I,m})} + \frac{1}{\Phi(-\zeta_{R/I,m})} \right) e^{-\zeta_{R/I,m}^2}, \quad (17)$$

and the detailed expressions of the partial derivatives are given in Appendix A.

2) *Unknown Noise Variance*: When σ_v^2 is unknown, the FIM for estimating $\boldsymbol{\psi}$ is:

$$\mathbf{J}_F(\boldsymbol{\psi}) = \begin{bmatrix} \mathbf{J}_F(\boldsymbol{\chi}) & \mathbf{F}_{\boldsymbol{\chi}, \sigma_v} \\ \mathbf{F}_{\boldsymbol{\chi}, \sigma_v}^T & F_{\sigma_v, \sigma_v} \end{bmatrix} \in \mathbb{R}^{(4N_s+1) \times (4N_s+1)}, \quad (18)$$

where $\mathbf{J}_F(\boldsymbol{\chi})$ is given by (13), the p th element of the $4N_s \times 1$ vector $\mathbf{F}_{\boldsymbol{\chi}, \sigma_v}$ is calculated as:

$$[\mathbf{F}_{\boldsymbol{\chi}, \sigma_v}]_p = \mathbb{E}_{\bar{\mathbf{z}}} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}}|\boldsymbol{\chi})}{\partial \chi_p \partial \sigma_v} \right] = \frac{-1}{\pi \sigma_v^2} \sum_{m=1}^{KN_r} \left[f(\zeta_{R,m}) \frac{\partial \kappa_{R,m}}{\partial \chi_p} \frac{\kappa_{R,m} - \eta_{R,m}}{\sigma_v} + f(\zeta_{I,m}) \frac{\partial \kappa_{I,m}}{\partial \chi_p} \frac{\kappa_{I,m} - \eta_{I,m}}{\sigma_v} \right], \quad 1 \leq p \leq 4N_s, \quad (19)$$

and the scalar F_{σ_v, σ_v} is calculated as:

$$F_{\sigma_v, \sigma_v} = \mathbb{E}_{\bar{\mathbf{z}}} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}}|\boldsymbol{\chi})}{\partial \sigma_v \partial \sigma_v} \right] = \frac{1}{\pi \sigma_v^2} \sum_{m=1}^{KN_r} \left[f(\zeta_{R,m}) \left(\frac{\kappa_{R,m} - \eta_{R,m}}{\sigma_v} \right)^2 + f(\zeta_{I,m}) \left(\frac{\kappa_{I,m} - \eta_{I,m}}{\sigma_v} \right)^2 \right]. \quad (20)$$

Then, the CRB matrix for estimating the angular-domain channel parameter vector $\boldsymbol{\chi}$ is given by:

$$\mathbf{CRB}(\boldsymbol{\chi}) = \left[\mathbf{J}_F(\boldsymbol{\chi}) - \mathbf{F}_{\boldsymbol{\chi}, \sigma_v} F_{\sigma_v, \sigma_v}^{-1} \mathbf{F}_{\boldsymbol{\chi}, \sigma_v}^T \right]^{-1}. \quad (21)$$

²The more detailed derivations can be found in the supplementary material.

C. Remarks

1) *Singularity of the FIM*: When the noise variance is unknown, the FIM $\mathbf{J}_F(\boldsymbol{\psi})$ is singular when zero or optimal thresholds [20] are used, which means that there does not exist unbiased estimates with finite variances [27]. For the optimal thresholds, i.e., $\bar{\boldsymbol{\eta}} = \bar{\Gamma} \bar{\boldsymbol{\alpha}}$, the singularity comes from the fact that $\mathbf{F}_{\boldsymbol{\chi}, \sigma_v}$ and F_{σ_v, σ_v} in the FIM $\mathbf{J}_F(\boldsymbol{\psi})$ are zero vector and scalar, respectively, according to (19) and (20). When zero-thresholds are used, the FIM is singular due to the ambiguity between the path gains and noise variance. That is, according to (9), the parameter vector $[s\boldsymbol{\alpha}, s\sigma_v]$, $s \neq 0$, yields the same log-likelihood function as $[\boldsymbol{\alpha}, \sigma_v]$ for given one-bit measurements $\bar{\mathbf{z}}$. This ambiguity can be removed when known non-zero thresholds are used.

2) *Conversion to CRB for Channel Matrix*: The channel matrix \mathbf{H} can be recovered using the estimated angular-domain parameters $\boldsymbol{\chi}$. The channel matrix model in (3) can be vectorized as

$$\begin{aligned} \text{vec}(\mathbf{H}) &= \sum_{l=1}^{N_s} [\mathbf{a}_{\text{MS}}^*(\varphi_l) \otimes \mathbf{a}_{\text{BS}}(\theta_l)] \alpha_l \\ &= [\mathbf{A}_{\text{MS}}^*(\boldsymbol{\varphi}) \odot \mathbf{A}_{\text{BS}}(\boldsymbol{\theta})] \boldsymbol{\alpha} \\ &\triangleq \bar{\mathbf{h}}(\boldsymbol{\chi}). \end{aligned} \quad (22)$$

Let

$$\bar{\mathbf{h}}(\boldsymbol{\chi}) \triangleq \begin{bmatrix} \mathbf{h}_R(\boldsymbol{\chi}) \\ \mathbf{h}_I(\boldsymbol{\chi}) \end{bmatrix} \in \mathbb{R}^{2N_r N_t \times 1}. \quad (23)$$

We have [28]:

$$\mathbf{CRB}(\bar{\mathbf{h}}) = \frac{\partial \bar{\mathbf{h}}(\boldsymbol{\chi})}{\partial \boldsymbol{\chi}^T} \mathbf{CRB}(\boldsymbol{\chi}) \frac{\partial \bar{\mathbf{h}}(\boldsymbol{\chi})}{\partial \boldsymbol{\chi}}, \quad (24)$$

where $\mathbf{CRB}(\boldsymbol{\chi})$ is given in (14) for the known σ_v^2 case and in (21) for the unknown σ_v^2 case. The partial derivatives $\frac{\partial \bar{\mathbf{h}}(\boldsymbol{\chi})}{\partial \boldsymbol{\chi}}$ can be calculated similarly as those in Appendix A. The CRB for the channel matrix is trace $[\mathbf{CRB}(\bar{\mathbf{h}})]$.

IV. HYBRID AND BAYESIAN CRBS

A. Hybrid CRB with Prior Knowledge on Channel Path Gains

We then consider the case where the path gains have known prior pdfs while the angles are treated as deterministic unknowns. Specifically, we use a channel model for the 28-GHz outdoor mmWave channel [21], where the path gains have i.i.d. Gaussian distributions with known means and variances. The prior knowledge of channel path gains can be obtained from previous channel measurements for stationary channels.

The hybrid CRBs (HCRBs) establish the best performance bounds for both the deterministic and random parameters. When σ_v^2 is known, we rewrite the parameter vector as $\boldsymbol{\chi} = [\boldsymbol{\chi}_d, \boldsymbol{\chi}_r]^T$ with $\boldsymbol{\chi}_d = [\boldsymbol{\theta}^T, \boldsymbol{\varphi}^T]^T$ denoting the deterministic angle vector and $\boldsymbol{\chi}_r = [\boldsymbol{\alpha}_R^T, \boldsymbol{\alpha}_I^T]^T$ grouping the real and imaginary parts of the random path gains. Then we assume:

$$\boldsymbol{\chi}_r \sim \mathcal{N}(\mathbf{m}_\alpha, \mathbf{C}_\alpha), \quad (25)$$

where $\mathbf{m}_\alpha = [m_{R,1}, \dots, m_{R,N_s}, m_{I,1}, \dots, m_{I,N_s}]^T \in \mathbb{R}^{2N_s \times 1}$ and $\mathbf{C}_\alpha = \text{diag}\{\sigma_{R,1}^2, \dots, \sigma_{R,N_s}^2, \sigma_{I,1}^2, \dots, \sigma_{I,N_s}^2\} \in \mathbb{R}^{2N_s \times 2N_s}$ with $m_{R/I,l}$ and $\sigma_{R/I,l}^2$ denoting the known mean and variance of $\alpha_{R/I,l}$, respectively. Therefore, the (p, q) th

element of the hybrid information matrix (HIM) $\mathbf{J}_H(\boldsymbol{\chi})$ is given by:

$$\begin{aligned} [\mathbf{J}_H(\boldsymbol{\chi})]_{p,q} &= \mathbb{E}_{\bar{\mathbf{z}}, \boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}}, \boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \chi_p \partial \chi_q} \right] \\ &= \mathbb{E}_{\bar{\mathbf{z}}, \boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}} | \boldsymbol{\chi}_r, \boldsymbol{\chi}_d)}{\partial \chi_p \partial \chi_q} \right] \\ &\quad + \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \chi_p \partial \chi_q} \right] \\ &= [\mathbf{J}_D(\boldsymbol{\chi})]_{p,q} + [\mathbf{J}_P(\boldsymbol{\chi})]_{p,q}, \end{aligned} \quad (26)$$

where $\mathbb{E}_{\bar{\mathbf{z}}, \boldsymbol{\chi}_r | \boldsymbol{\chi}_d}[\cdot]$ denotes the conditional expectation operator w.r.t. both $\bar{\mathbf{z}}$ and $\boldsymbol{\chi}_r$ for given $\boldsymbol{\chi}_d$, $p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)$ is the known prior pdf of $\boldsymbol{\chi}_r$ for given $\boldsymbol{\chi}_d$ and $\ln p(\bar{\mathbf{z}} | \boldsymbol{\chi}_r, \boldsymbol{\chi}_d)$ is given in (9). Recall from the FIM expression of (12), we have:

$$\mathbf{J}_D(\boldsymbol{\chi}) = \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} [\mathbf{J}_F(\boldsymbol{\chi})] = \int \mathbf{J}_F(\boldsymbol{\chi}) \cdot p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d) d\boldsymbol{\chi}_r, \quad (27)$$

and

$$\mathbf{J}_P(\boldsymbol{\chi}) = \begin{bmatrix} \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \chi_d \partial \chi_d^T} \right] & \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \boldsymbol{\chi}_r \partial \chi_d^T} \right] \\ \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \chi_d \partial \boldsymbol{\chi}_r^T} \right] & \mathbb{E}_{\boldsymbol{\chi}_r | \boldsymbol{\chi}_d} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r | \boldsymbol{\chi}_d)}{\partial \boldsymbol{\chi}_r \partial \boldsymbol{\chi}_r^T} \right] \end{bmatrix}. \quad (28)$$

When the random vector $\boldsymbol{\chi}_r$ does not depend on the deterministic unknown vector $\boldsymbol{\chi}_d$, $\mathbf{J}_P(\boldsymbol{\chi})$ reduces to:

$$\mathbf{J}_P(\boldsymbol{\chi}) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_\alpha^{-1} \end{bmatrix} \in \mathbb{R}^{4N_s \times 4N_s}. \quad (29)$$

Note that in (26), $\mathbf{J}_D(\boldsymbol{\chi})$ and $\mathbf{J}_P(\boldsymbol{\chi})$ represent the information from the data and the prior distribution, respectively. Then, the HCRB matrix is given by:

$$\mathbf{HCRB}(\boldsymbol{\chi}) = [\mathbf{J}_H(\boldsymbol{\chi})]^{-1}. \quad (30)$$

Unfortunately, there is no analytical expression for $\mathbf{J}_D(\boldsymbol{\chi})$ due to the nonlinear function $f(\zeta_{R/I,m})$ of (17) involved in $\mathbf{J}_F(\boldsymbol{\chi})$ of (16). In the following, approximate expressions of $\mathbf{J}_D(\boldsymbol{\chi})$ are derived to avoid numerical integrations.

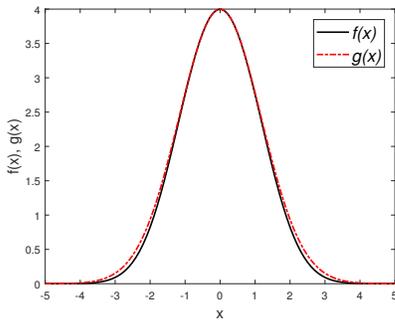


Fig. 3. $f(x)$ of (17) compared with $g(x) = 4e^{-(1-\frac{2}{\pi})x^2}$.

1) A Loose Lower Bound with Low SNR Approximations:

For low SNR (i.e., $\zeta_{R/I,m} \rightarrow 0$), $f(\zeta_{R/I,m})$ in $[\mathbf{J}_F(\boldsymbol{\chi})]_{p,q}$ is close to 4 (see Fig. 3), and we can obtain an approximate

$[\mathbf{J}_F(\boldsymbol{\chi})]_{p,q}$ by simply replacing $f(\zeta_{R/I,m})$ with 4:

$$\begin{aligned} [\hat{\mathbf{J}}_F(\boldsymbol{\chi})]_{p,q} &= \frac{4}{\pi\sigma_v^2} \sum_{m=1}^{KN_r} \frac{\partial \kappa_{R,m}}{\partial \chi_p} \frac{\partial \kappa_{R,m}}{\partial \chi_q} + \frac{4}{\pi\sigma_v^2} \sum_{m=1}^{KN_r} \frac{\partial \kappa_{I,m}}{\partial \chi_p} \frac{\partial \kappa_{I,m}}{\partial \chi_q} \\ &\triangleq [\hat{\mathbf{J}}_{F,R}(\boldsymbol{\chi})]_{p,q} + [\hat{\mathbf{J}}_{F,I}(\boldsymbol{\chi})]_{p,q}, \end{aligned} \quad (31)$$

and hence, an approximate expression of $[\mathbf{J}_D(\boldsymbol{\chi})]_{p,q}$ of (27) can be obtained accordingly:

$$[\hat{\mathbf{J}}_D(\boldsymbol{\chi})]_{p,q} = [\hat{\mathbf{J}}_{D,R}(\boldsymbol{\chi})]_{p,q} + [\hat{\mathbf{J}}_{D,I}(\boldsymbol{\chi})]_{p,q}, \quad (32)$$

where

$$[\hat{\mathbf{J}}_{D,R}(\boldsymbol{\chi})]_{p,q} = \frac{4}{\pi\sigma_v^2} \sum_{m=1}^{KN_r} \int \frac{\partial \kappa_{R,m}}{\partial \chi_p} \frac{\partial \kappa_{R,m}}{\partial \chi_q} p(\boldsymbol{\chi}_r) d\boldsymbol{\chi}_r, \quad (33)$$

and $[\hat{\mathbf{J}}_{D,I}(\boldsymbol{\chi})]_{p,q}$ can be obtained similarly. According to (64) in Appendix A, the partial derivatives $\frac{\partial \kappa_{R,m}}{\partial \chi_p}$ are linear functions of the path gain vector $\boldsymbol{\chi}_r$. Also, since $p(\boldsymbol{\chi}_r)$ is the Gaussian pdf of the random path gain vector, the analytical expressions of the integral in (33) can be obtained easily by plugging (64) into (33) (omitted here).

Note that in the unquantized case, $\mathbf{J}_P(\boldsymbol{\chi})$ remains unchanged and $\mathbf{J}_D(\boldsymbol{\chi})$ is calculated using the FIM for unquantized systems³, where $f(\zeta_{R/I,m})$ is replaced by a constant of 2π . Hence, an exact analytical expression of $\mathbf{J}_H(\boldsymbol{\chi})$ can be obtained for the said case using the aforementioned derivations.

2) *Tighter Lower Bounds:* In [29], $g(x) = 4e^{-(1-\frac{2}{\pi})x^2}$ is proved to be an upper and close bound of $f(x)$ (see Fig. 3). Then, with $f(\zeta_{R/I,m})$ in (16) replaced by $g(\zeta_{R/I,m})$, we can obtain a much closer approximation of $[\mathbf{J}_D(\boldsymbol{\chi})]_{p,q}$:

$$[\hat{\mathbf{J}}_{D,R}(\boldsymbol{\chi})]_{p,q} = \frac{1}{\pi\sigma_v^2} \sum_{m=1}^{KN_r} \int \frac{\partial \kappa_{R,m}}{\partial \chi_p} \frac{\partial \kappa_{R,m}}{\partial \chi_q} g(\zeta_{R,m}) p(\boldsymbol{\chi}_r) d\boldsymbol{\chi}_r. \quad (34)$$

Fortunately, it turns out that $g(\zeta_{R,m}) p(\boldsymbol{\chi}_r)$ can be reformulated into the Gaussian pdf of the random path gain vector $\boldsymbol{\chi}_r$ but with a different mean vector and covariance matrix, dependent on m (see Appendix B). Hence, a tighter approximate expression of $\mathbf{J}_D(\boldsymbol{\chi})$ in (32) can be derived similarly, and the resulting HCRB provides a lower bound to the true HCRB of (30).

3) *Unknown Noise Variance:* Note that when σ_v^2 is deterministic unknown, we can rewrite the expanded hybrid parameter vector $\boldsymbol{\psi}$ as: $\boldsymbol{\psi} = [\boldsymbol{\psi}_r, \boldsymbol{\psi}_d]^T$ with $\boldsymbol{\psi}_r = \boldsymbol{\chi}_r$ and $\boldsymbol{\psi}_d = [\boldsymbol{\chi}_d^T, \sigma_v^2]^T$. In this case, the HIM component $\mathbf{J}_P(\boldsymbol{\chi})$ in (28) is expanded to $\mathbf{J}_P(\boldsymbol{\psi})$ as:

$$\mathbf{J}_P(\boldsymbol{\psi}) = \begin{bmatrix} \mathbf{J}_P(\boldsymbol{\chi}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(4N_s+1) \times (4N_s+1)}. \quad (35)$$

And $\mathbf{J}_D(\boldsymbol{\psi})$, $\mathbf{J}_H(\boldsymbol{\psi})$ can be obtained straightforwardly and $\mathbf{HCRB}(\boldsymbol{\chi})$ can be calculated similarly to (21). The approximate HCRBs for the unknown σ_v^2 case can be derived similarly to its counterpart of known σ_v^2 .

³The derivation of the unquantized FIM is given in the supplementary material.

B. Hybrid CRB with Prior Knowledge on Angular Parameters

It is shown in [30] that the out-of-band information (mainly the angular information) extracted by the sub 6-GHz communication systems can provide useful channel characteristics for mmWave band channels. Here we study the case where the path gains are deterministic unknowns and the angles form a Gaussian random vector with known mean vector \mathbf{m}_ϕ and covariance matrix \mathbf{C}_ϕ . When σ_v^2 is known, the angular-domain channel parameter vector can be rearranged as $\boldsymbol{\chi} = [\boldsymbol{\chi}_r, \boldsymbol{\chi}_d]^T$ with $\boldsymbol{\chi}_r = [\boldsymbol{\theta}^T, \boldsymbol{\varphi}^T]^T$ denoting the random angle vector and $\boldsymbol{\chi}_d = [\boldsymbol{\alpha}_R^T, \boldsymbol{\alpha}_I^T]^T$. Then, we assume:

$$\boldsymbol{\chi}_r \sim \mathcal{N}(\mathbf{m}_\phi, \mathbf{C}_\phi). \quad (36)$$

The corresponding HIM can be calculated similarly with (26), where

$$\mathbf{J}_P(\boldsymbol{\chi}) = \begin{bmatrix} \mathbb{E}_{\boldsymbol{\chi}_r} \left[-\frac{\partial^2 \ln p(\boldsymbol{\chi}_r)}{\partial \boldsymbol{\chi}_r \partial \boldsymbol{\chi}_r^T} \right] \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_\phi^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (37)$$

In this scenario, there are no analytical or approximate expressions for $\mathbf{J}_D(\boldsymbol{\chi})$, and we calculate $\mathbf{J}_D(\boldsymbol{\chi})$ using numerical integrations. When σ_v^2 is unknown, the HIM and HCRB can be similarly derived (omitted here).

C. Bayesian CRB

1) *Bayesian CRB for Channel Parameters:* Here we assume that σ_v^2 is known, and the path gains as well as the angles are independent Gaussian random vectors with known mean vectors \mathbf{m}_α , \mathbf{m}_ϕ and covariance matrices \mathbf{C}_α , \mathbf{C}_ϕ , respectively. Then the (p, q) th element of the Bayesian information matrix (BIM) $\mathbf{J}_B(\boldsymbol{\chi})$ is given by [28]:

$$[\mathbf{J}_B(\boldsymbol{\chi})]_{p,q} = \mathbb{E}_{\bar{\mathbf{z}}, \boldsymbol{\chi}} \left[-\frac{\partial^2 \ln p(\bar{\mathbf{z}}, \boldsymbol{\chi})}{\partial \chi_p \partial \chi_q} \right]. \quad (38)$$

Since $\ln p(\bar{\mathbf{z}}, \boldsymbol{\chi}) = \ln p(\bar{\mathbf{z}}|\boldsymbol{\chi}) + \ln p(\boldsymbol{\chi})$, where $p(\boldsymbol{\chi})$ is the known prior pdf of $\boldsymbol{\chi}$, we have:

$$\mathbf{J}_B(\boldsymbol{\chi}) = \mathbf{J}_D(\boldsymbol{\chi}) + \mathbf{J}_P(\boldsymbol{\chi}), \quad (39)$$

where $\mathbf{J}_D(\boldsymbol{\chi}) = \mathbb{E}_{\boldsymbol{\chi}} [\mathbf{J}_F(\boldsymbol{\chi})]$ and

$$\mathbf{J}_P(\boldsymbol{\chi}) = \mathbb{E}_{\boldsymbol{\chi}} \left[\frac{\partial^2 \ln p(\boldsymbol{\chi})}{\partial \boldsymbol{\chi} \partial \boldsymbol{\chi}^T} \right] = \begin{bmatrix} \mathbf{C}_\phi^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_\alpha^{-1} \end{bmatrix} \in \mathbb{R}^{4N_s \times 4N_s}. \quad (40)$$

$\mathbf{J}_D(\boldsymbol{\chi})$ is obtained by numerical integrations (note that the involved numerical integrations w.r.t. the random path gains can be avoided using the approximations in Section IV-A). The Bayesian CRB (BCRB) matrix is computed as:

$$\text{BCRB}(\boldsymbol{\chi}) = [\mathbf{J}_B(\boldsymbol{\chi})]^{-1}. \quad (41)$$

For unknown σ_v^2 , the BCRB reduces to the HCRB with the only deterministic unknown given by σ_v . The derivation of this HCRB is similar to that in Section IV-A3.

2) *Conversion to Bayesian CRB for Channel Matrix:* Similar to Section III-C2, the BCRB for the channel matrix estimation is obtained as [28]:

$$\text{BCRB}(\bar{\mathbf{h}}) = \mathbb{E}_{\boldsymbol{\chi}} \left[\frac{\partial \bar{\mathbf{h}}(\boldsymbol{\chi})}{\partial \boldsymbol{\chi}^T} \right] \text{BCRB}(\boldsymbol{\chi}) \mathbb{E}_{\boldsymbol{\chi}} \left[\frac{\partial \bar{\mathbf{h}}(\boldsymbol{\chi})}{\partial \boldsymbol{\chi}} \right], \quad (42)$$

where $\bar{\mathbf{h}}$ is given in (23), and the expectations of the partial derivatives w.r.t. the random path gains have closed

form expressions while the expectations w.r.t. the random angles can only be obtained via numerical integrations. For any Bayesian estimate $\hat{\mathbf{H}}$, its performance lower bound is trace $[\text{BCRB}(\bar{\mathbf{h}})]$.

V. ANGULAR-DOMAIN CHANNEL ESTIMATION

A. ML Estimator

For the case of known σ_v^2 , the ML estimates are obtained by minimizing the negative log-likelihood function of (9):

$$\hat{\boldsymbol{\chi}}_{\text{ML}} = \arg \min_{\boldsymbol{\chi}} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}) + \ln \Phi(z_{I,m} \zeta_{I,m})], \quad (43)$$

where $\zeta_{R/I,m}$ of (10) are nonlinear w.r.t. the angles AoA and AoD and linear w.r.t. the real-valued path gains $\boldsymbol{\alpha}_R$ and $\boldsymbol{\alpha}_I$. Given $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, $\boldsymbol{\alpha}_R$ and $\boldsymbol{\alpha}_I$ can be estimated by minimizing the cost function of (43) as an unconstrained convex optimization problem [31].

When σ_v^2 is unknown, we can reparameterize (43) by defining $\varepsilon = \frac{1}{\sigma_v}$, $\boldsymbol{\beta}_R = \frac{\boldsymbol{\alpha}_R}{\sigma_v}$, and $\boldsymbol{\beta}_I = \frac{\boldsymbol{\alpha}_I}{\sigma_v}$. With $\bar{\boldsymbol{\psi}} = [\boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\beta}_R^T, \boldsymbol{\beta}_I^T, \varepsilon]^T$, the ML estimation problem can be cast as:

$$\hat{\bar{\boldsymbol{\psi}}}_{\text{ML}} = \arg \min_{\bar{\boldsymbol{\psi}}} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \bar{\zeta}_{R,m}) + \ln \Phi(z_{I,m} \bar{\zeta}_{I,m})], \quad (44)$$

where

$$\bar{\zeta}_{R,m} = \sqrt{2} \left[\boldsymbol{\Gamma}_R^T(m) \boldsymbol{\beta}_R - \boldsymbol{\Gamma}_I^T(m) \boldsymbol{\beta}_I - \varepsilon \boldsymbol{\eta}_{R/I,m} \right], \quad (45a)$$

$$\bar{\zeta}_{I,m} = \sqrt{2} \left[\boldsymbol{\Gamma}_I^T(m) \boldsymbol{\beta}_R + \boldsymbol{\Gamma}_R^T(m) \boldsymbol{\beta}_I - \varepsilon \boldsymbol{\eta}_{I,m} \right]. \quad (45b)$$

For given $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, the cost function in (44) is, again, convex w.r.t. $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_I$ and ε . Note that when $\boldsymbol{\eta}_{R/I} = \mathbf{0}$, i.e., when the zero-thresholds are used, ε in (45) vanishes and one can only estimate $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_I$, i.e., the ratio between $\boldsymbol{\alpha}_{R/I}$ and σ_v .

For the case of unknown noise variance, we could discretize the AoA and AoD in the set of $(-\pi/2, \pi/2]$. Then, for a given pair of $(\boldsymbol{\theta}, \boldsymbol{\varphi})$, we can estimate the corresponding $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_I$, and ε by solving (44) using the Newton's method. We can then repeat the previous step over all possible combinations of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, and then select the ML estimates giving the minimum cost in (44). The path gains and noise variance can be recovered as $\hat{\boldsymbol{\alpha}}_R = \frac{\hat{\boldsymbol{\beta}}_R}{\hat{\varepsilon}}$, $\hat{\boldsymbol{\alpha}}_I = \frac{\hat{\boldsymbol{\beta}}_I}{\hat{\varepsilon}}$ and $\hat{\sigma}_v^2 = \frac{1}{\hat{\varepsilon}^2}$.

This direct ML approach, however, requires a $2N_s$ -dimensional *coarse search* over the angular parameter space. With L_θ discretized grid points for the AoA and L_φ discretized grid points for the AoD, we need to solve the convex optimization problem $L_\theta^{N_s} L_\varphi^{N_s}$ times. As a result, the resulting ML estimator is computationally prohibitive even for small N_s .

B. 1bRELAX Algorithm for ML Estimation

We introduce below an extension of a computationally efficient relaxation based cyclic algorithm (RELAX), which was originally proposed for sinusoidal parameter estimation [32], to obtain the ML channel estimates. The extension is referred to as the 1bRELAX channel estimation algorithm.

The 1bRELAX algorithm breaks down the joint $2N_s$ -dimensional searches with iterative 2-dimensional searches.

Denote $\chi_l = [\theta_l, \varphi_l, \alpha_{R,l}, \alpha_{I,l}]^T$ as the angular-domain parameter vector corresponding to the l^{th} path. 1bRELAX begins, in Step 1, by assuming there is only one dominating scattering path between the BS and MS and solving the following problem:

$$[\hat{\chi}_1, \hat{\sigma}_v] = \arg \min_{\chi_1, \sigma_v} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}^{1,1}) + \ln \Phi(z_{I,m} \zeta_{I,m}^{1,1})], \quad (46)$$

where $\zeta_{R,m}^{p,q}$ and $\zeta_{I,m}^{p,q}$ are defined, respectively, as:

$$\zeta_{R,m}^{p,q} = \frac{\sum_{l=1, l \neq q}^p \hat{\kappa}_{R,m}^l + \kappa_{R,m}^q - \eta_{R,m}}{\sigma_v / \sqrt{2}} \quad (47a)$$

and

$$\zeta_{I,m}^{p,q} = \frac{\sum_{l=1, l \neq q}^p \hat{\kappa}_{I,m}^l + \kappa_{I,m}^q - \eta_{I,m}}{\sigma_v / \sqrt{2}} \quad (47b)$$

with

$$\kappa_{R,m}^l = \gamma_{R,m}^l \alpha_{R,l} - \gamma_{I,m}^l \alpha_{I,l}, \quad (48a)$$

and

$$\kappa_{I,m}^l = \gamma_{I,m}^l \alpha_{R,l} + \gamma_{R,m}^l \alpha_{I,l}, \quad (48b)$$

denoting the real and imaginary parts of the received signal from the l^{th} path, respectively. The $\gamma^l(\theta_l, \varphi_l)$ is given in (63) and the $\hat{\kappa}_{R/I,m}^l$ is the reconstructed received signal using the estimated angular-domain channel parameter vector $\hat{\chi}_l$ in the previous iteration step. Equation (46) can be solved by a two-dimensional *coarse search* over (θ_1, φ_1) and followed by using the gradient method for $(\alpha_{R,1}, \alpha_{I,1}, \sigma_v)$. Let $\{\theta'_1, \varphi'_1, \alpha'_{R,1}, \alpha'_{I,1}, \sigma'_v\}$ denote the estimated parameters that minimize the cost function of (46). Then we perform a *grid-less fine search* over the AoA interval $[\theta'_1 - \frac{\pi}{L_\theta}, \theta'_1 + \frac{\pi}{L_\theta}]$ and AoD interval $[\varphi'_1 - \frac{\pi}{L_\varphi}, \varphi'_1 + \frac{\pi}{L_\varphi}]$ by using the interior point based bounded optimization method (e.g., “fmincon” of MATLAB) to find the estimate $\{\hat{\chi}_1, \hat{\sigma}_v\}$ of $\{\chi_1, \sigma_v\}$.

In Step 2, we assume that there are two scattering paths, i.e., $N_s = 2$. The 1bRELAX algorithm finds the parameters of the second strongest path by minimizing the cost function below:

$$\hat{\chi}_2 = \arg \min_{\chi_2} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}^{2,2}) + \ln \Phi(z_{I,m} \zeta_{I,m}^{2,2})], \quad (49)$$

where $\zeta_{R,m}^{2,2}$ and $\zeta_{I,m}^{2,2}$ can be computed from (47) with $\hat{\kappa}_{R,m}^1$ and $\hat{\kappa}_{I,m}^1$ reconstructed with the estimated angular-domain channel parameters corresponding to the strongest path from Step 1:

$$\hat{\kappa}_{R,m}^1 = \hat{\gamma}_{R,m}^1(\hat{\theta}_1, \hat{\varphi}_1) \hat{\alpha}_{R,1} - \hat{\gamma}_{I,m}^1(\hat{\theta}_1, \hat{\varphi}_1) \hat{\alpha}_{I,1}, \quad (50a)$$

$$\hat{\kappa}_{I,m}^1 = \hat{\gamma}_{I,m}^1(\hat{\theta}_1, \hat{\varphi}_1) \hat{\alpha}_{R,1} + \hat{\gamma}_{R,m}^1(\hat{\theta}_1, \hat{\varphi}_1) \hat{\alpha}_{I,1}. \quad (50b)$$

In a similar fashion, $(\hat{\chi}_1, \hat{\sigma}_v)$ is refined using the $\hat{\chi}_2$ to solve the problem below via a *fine search*:

$$[\hat{\chi}_1, \hat{\sigma}_v] = \arg \min_{\chi_1, \sigma_v} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}^{2,1}) + \ln \Phi(z_{I,m} \zeta_{I,m}^{2,1})]. \quad (51)$$

This procedure of iteratively refining parameter vectors $\{\hat{\chi}_1, \hat{\sigma}_v\}$ and $\hat{\chi}_2$ of the two strongest paths with *fine searches*

continues until convergence.

Step 3 of 1bRELAX assumes $N_s = 3$, and the algorithm continues by solving the problem below to estimate the parameter vector of the third strongest channel path using $\hat{\chi}_2$ and $\{\hat{\chi}_1, \hat{\sigma}_v\}$ from Step 2:

$$\hat{\chi}_3 = \arg \min_{\chi_3} - \sum_{m=1}^{KN_r} [\ln \Phi(z_{R,m} \zeta_{R,m}^{3,3}) + \ln \Phi(z_{I,m} \zeta_{I,m}^{3,3})]. \quad (52)$$

Then $\{\hat{\chi}_1, \hat{\sigma}_v\}$, $\hat{\chi}_2$ and $\hat{\chi}_3$ are iteratively refined with *fine searches* until convergence. The algorithm continues until the parameter vectors of all paths are estimated.

1) *Computational Complexity*: Compared to the direct ML method, 1bRELAX requires only one two-dimensional search over the parameter space of AoA and AoD for each path in each step, resulting in significantly reduced computational complexities. To obtain the *coarse estimates*, the 1bRELAX algorithm needs to calculate the corresponding path gains for $L_\theta L_\varphi$ angular grid points via, e.g., the Newton’s method for each of the N_s paths. In each iteration of the Newton’s method, the computational load is dominated by the calculation of a 2×2 Hessian matrix and a 2×1 gradient vector, which needs to construct the unquantized received signal $\kappa_{R/I}$ and has a computational cost of $\mathcal{O}(KN_r N_t)$. Our numerical results show that typically the Newton’s method can converge in less than 5 iterations. For the refinement steps in 1bRELAX, the computational cost of the interior point based method can be neglected compared with the *coarse searches*. As a result, the total computational complexity for the 1bRELAX algorithm is on the order of $\mathcal{O}(N_s K L_\theta L_\varphi N_r N_t)$, where L_θ and L_φ are several times (e.g., 2) larger than N_r and N_t , respectively. In comparison, the computational costs of the recently proposed algorithms BLMMSE and GAMP are on the order of $\mathcal{O}(K^3 N_r^3)$ and $\mathcal{O}(K \max(N_r N_t, N_r \log(N_r), N_t \log(N_t)))$, respectively.

C. 1bBIC for Path Number Determination

When N_s is unknown, the Bayesian information criterion (BIC) can be used with 1bRELAX to provide a consistent estimate of the channel path number [33], [34]. Below we derive the customized BIC rule, referred to as the 1bBIC, to determine the number of channel paths.

By assuming that the prior pdf of ψ , $p(\psi)$, is flat around the ML estimate $\hat{\psi}_{\text{ML}}$ and does not depend on the K , N_r and N_t , the channel path number estimate according to the BIC is obtained by minimizing the criterion below [33], [34]:

$$-2 \ln p_{\hat{N}_s}(\bar{\mathbf{z}} | \hat{\psi}_{\text{ML}}) + \ln |\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\psi}_{\text{ML}})|, \quad (53)$$

where $p_{\hat{N}_s}(\bar{\mathbf{z}} | \hat{\psi}_{\text{ML}})$ is the likelihood function under the hypothesis that the path number is \hat{N}_s , and $|\cdot|$ denotes the determinant of a matrix. The second term is a penalty term that penalizes an overestimated path number with $\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\psi}_{\text{ML}})$ defined as:

$$\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\psi}_{\text{ML}}) = - \frac{\partial^2 \ln p_{\hat{N}_s}(\bar{\mathbf{z}} | \psi)}{\partial \psi \partial \psi^T} \Big|_{\psi = \hat{\psi}_{\text{ML}}}. \quad (54)$$

Algorithm 1 1bRELAX with 1bBIC

Input: Measurements $\bar{\mathbf{z}}$, threshold vector $\bar{\boldsymbol{\eta}}$, pilot signal \mathbf{X} , maximum path number N_{\max}

Output: Channel path number \hat{N}_s , angular parameters and noise variance $\hat{\boldsymbol{\psi}}$

Set $L = 1$, obtain $\Xi_1 = [\hat{\boldsymbol{\chi}}_1, \hat{\sigma}_v]$ via a *coarse search* followed by a *fine search*

Compute the 1bBIC cost function value $C(1)$ from (59) with $\hat{\zeta}_{R/I,m}$ obtained from $[\hat{\boldsymbol{\chi}}_1, \hat{\sigma}_v]$

for $L = 2$ to N_{\max} **do**

 Obtain $\hat{\boldsymbol{\chi}}_L$ via a *coarse search* followed by a *fine search* and set $k = 1$

repeat

if $k = 1$ **then**

 Update $[\hat{\boldsymbol{\chi}}_1, \hat{\sigma}_v]$ via a *fine search*

else

 Update $\hat{\boldsymbol{\chi}}_k$ via a *fine search*

end if

$k = (k \bmod L) + 1$

until the ML cost function of (43) converges

$\Xi_L = [\{\hat{\boldsymbol{\chi}}_l\}_{l=1}^L, \hat{\sigma}_v]^T$

 Compute the 1bBIC cost function value $C(L)$ with $\hat{\zeta}_{R/I,m}$ obtained from $\{\hat{\boldsymbol{\chi}}_l\}_{l=1}^L$ and $\hat{\sigma}_v$

end for

$\hat{\boldsymbol{\psi}} = \Xi_{\hat{N}_s}$ where $\hat{N}_s = \arg \min_l C(l)$

Under mild conditions [33], the matrix $\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}})$ is shown to have the following asymptotic relationship with the FIM $\mathbf{J}_F(\boldsymbol{\psi})$ of (18):

$$\left[\mathbf{P}^{-1} \mathbf{J}_F(\boldsymbol{\psi}) \mathbf{P}^{-1} - \mathbf{P}^{-1} \hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}}) \mathbf{P}^{-1} \right] \rightarrow \mathbf{0} \quad \text{as } N' \rightarrow \infty, \quad (55)$$

where $N' \triangleq \{K, N_r, N_t\}$ and \mathbf{P} is a normalization matrix dependent on N' . Equation (55) implies that, after a proper normalization, $\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}})$ can be substituted by $\mathbf{J}_F(\boldsymbol{\psi})$ asymptotically. This can be used to obtain a simpler expression for the penalty term in (53).

We now find a normalization matrix \mathbf{P} to satisfy:

$$\mathbf{P}^{-1} \mathbf{J}_F(\boldsymbol{\psi}) \mathbf{P}^{-1} \stackrel{N' \rightarrow \infty}{\approx} \mathcal{O}(\mathbf{1}). \quad (56)$$

As shown in the supplementary material, when the pilot sequences are orthogonal to one another (i.e., $\mathbf{X}\mathbf{X}^H = \frac{K\rho}{N_t} \mathbf{I}_{N_t}$), a proper normalization matrix \mathbf{P} is found to be:

$$\mathbf{P} = \begin{bmatrix} N_t N_r^{\frac{1}{2}} K^{\frac{1}{2}} \mathbf{I}_{N_s} & & \\ & N_r^{\frac{3}{2}} K^{\frac{1}{2}} \mathbf{I}_{N_s} & \\ & & N_r^{\frac{1}{2}} K^{\frac{1}{2}} \mathbf{I}_{2N_s+1} \end{bmatrix}. \quad (57)$$

Given the normalization matrix of (57) and invoking (55) with $\hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}})$ replaced by $\mathbf{J}_F(\boldsymbol{\psi})$, the penalty term of (53) is given by

$$\begin{aligned} \ln \left| \hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}}) \right| &= \ln |\mathbf{P}^2| + \ln \left| \mathbf{P}^{-1} \hat{\mathbf{J}}(\bar{\mathbf{z}}, \hat{\boldsymbol{\psi}}_{\text{ML}}) \mathbf{P}^{-1} \right| \quad (58) \\ &= (6N_s + 1) \ln(N_r) + (4N_s + 1) \ln(K) \\ &\quad + 2N_s \ln(N_t) + \mathcal{O}(1). \end{aligned}$$

As a result, the 1bBIC rule of (53) reduces to:

$$\begin{aligned} \hat{N}_s = \arg \min_{N_s} \{ &-2 \sum_{m=1}^{KN_r} \left[\ln \Phi(z_{R,m} \hat{\zeta}_{R,m}) + \ln \Phi(z_{I,m} \hat{\zeta}_{I,m}) \right] \\ &+ 6N_s \ln(N_r) + 4N_s \ln(K) + 2N_s \ln(N_t) \}. \quad (59) \end{aligned}$$

To use 1bBIC with 1bRELAX, we can compute the cost function in (59) at the end of each step of 1bRELAX for up to a prescribed maximum possible path number N_{\max} . We give a flow chart in Algorithm 1 to show the detailed steps of 1bRELAX and how the 1bBIC is integrated into 1bRELAX.

D. JML-MAP and MAP Estimators

For the hybrid case, we utilize the prior pdf of random parameters and introduce the JML-MAP estimator for both the random and deterministic parameters:

$$\hat{\boldsymbol{\chi}}_{\text{JML/MAP}} = \arg \min_{\boldsymbol{\chi}} - [\ln p(\bar{\mathbf{z}}|\boldsymbol{\chi}_r, \boldsymbol{\chi}_d) + \ln p(\boldsymbol{\chi}_r|\boldsymbol{\chi}_d)]. \quad (60)$$

For the case of all random unknowns, on the other hand, we introduce the MAP method to estimate the random parameters:

$$\hat{\boldsymbol{\chi}}_{\text{MAP}} = \arg \min_{\boldsymbol{\chi}} - [\ln p(\bar{\mathbf{z}}|\boldsymbol{\chi}) + \ln p(\boldsymbol{\chi})]. \quad (61)$$

It can be shown that the above two problems are also unconstrained convex optimization problems to find the optimal path gains for given AoA and AoD, and hence the cost functions in (60) and (61) can be efficiently minimized via 1bRELAX procedures with small modifications.

When the noise variance σ_v^2 is unknown, the formulation of (44) cannot be utilized if the prior knowledge is on the channel path gains due to the definition of β_R and β_I . In this case, one can estimate σ_v^2 first during an idle period [20], [35], i.e., when there is no transmission between the BS and the MS. The relevant ML problem is:

$$\begin{aligned} \hat{\varepsilon} = \arg \min_{\varepsilon} - &\sum_{m=1}^{KN_r} \left[\ln \Phi(-\sqrt{2\varepsilon} z_{R,m} \eta_{R,m}) \right. \\ &\left. + \ln \Phi(-\sqrt{2\varepsilon} z_{I,m} \eta_{I,m}) \right]. \quad (62) \end{aligned}$$

Then $\hat{\sigma}_v^2$ can be calculated as $\hat{\sigma}_v^2 = \frac{1}{\hat{\varepsilon}}$. Note that (62) is a one-dimensional unconstrained convex optimization problem for $\bar{\boldsymbol{\eta}} \neq \mathbf{0}$, which can be solved efficiently via, e.g., the Newton's method. After that, we can solve (60) and (61) by replacing σ_v^2 with $\hat{\sigma}_v^2$.

VI. NUMERICAL EXAMPLES

Numerical examples are provided below to compare the CRBs using different thresholding schemes and verify the effectiveness of the proposed 1bRELAX algorithm for channel estimation and 1bBIC for path number determination. We consider a massive MIMO system model with $N_r = 64$, $N_t = 32$. Unless otherwise specified, the path number N_s is set to 4 and random QPSK symbols are used as the pilot signal. The channel matrix is normalized as $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_r N_t$ and the transmit signal power is normalized as $\rho = 1$. The mean-squared error (MSE) of the channel matrix estimate is defined as $\mathbb{E}[\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2] / (N_r N_t)$. The MSEs are compared with the

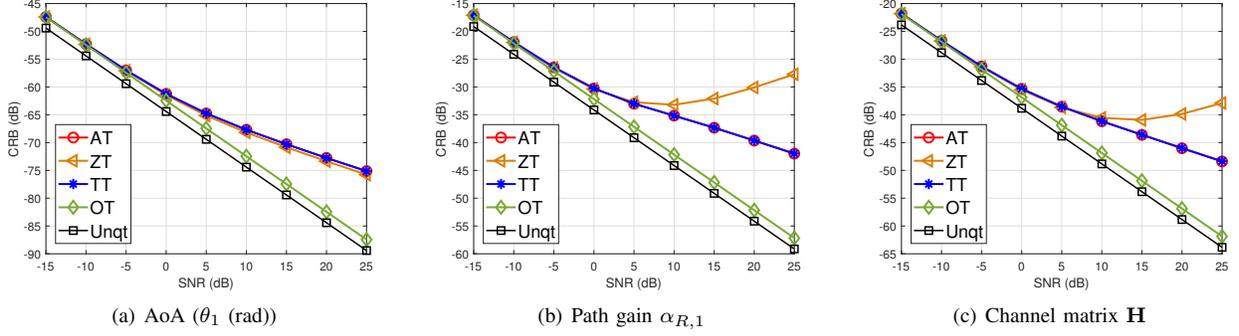


Fig. 4. Deterministic CRB comparison as a function of SNR with $K = 32$ and known σ_v^2 .

CRBs (i.e., $\text{trace}[\mathbf{CRB}(\bar{\mathbf{h}})]$ and $\text{trace}[\mathbf{BCRB}(\bar{\mathbf{h}})]$) normalized by $N_r N_t$. The average SNR at each receive antenna output is defined as $\text{SNR} = \mathbb{E}[\|\Gamma\alpha\|_2^2]/(KN_r\sigma_v^2) = 1/\sigma_v^2$.

We consider the following thresholding schemes: 1) zero-thresholding (ZT) (i.e., $\bar{\eta} = \mathbf{0}$); 2) optimal thresholding (OT), which requires the knowledge of noise-free signals (i.e., $\bar{\eta} = \bar{\Gamma}\bar{\alpha}$); 3) *time-varying* thresholding (TT) [36], which selects the thresholds randomly from a predefined discrete set, e.g., $[-h_{\max}, -h_{\max} + \Delta, \dots, h_{\max} - \Delta, h_{\max}]$ at each sampling instant with h_{\max} being the dynamic range and Δ being the stepsize; 4) *antenna-varying* thresholding (AT), which selects the thresholds from the discrete set randomly once and fixed at all times for each antenna output. For the AT and TT schemes, the thresholds are selected randomly from 8 discrete values uniformly distributed in $[-h_{\max}, h_{\max}]$ with $h_{\max} = 0.5$. Due to the randomly selected thresholds, the CRBs for the AT and TT schemes are obtained by averaging over 100 realizations of the randomly selected thresholds. The CRBs for unquantized systems are also considered here for comparison purposes. To compute the HCRBs and BCRBs, the first terms of HIM and BIM (i.e., $\mathbf{J}_D(\chi)$ and $\mathbf{J}_D(\psi)$) are obtained by averaging the corresponding FIM (i.e., $\mathbf{J}_F(\chi)$ and $\mathbf{J}_F(\psi)$) over 10^4 realizations of the random parameters according to the prior distributions.

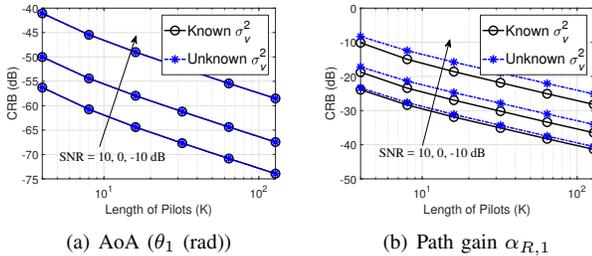


Fig. 5. Deterministic CRB for the AT scheme as a function of pilot length K when $\text{SNR} = -10, 0, 10$ dB.

A. Channel Estimation Performance Bounds

1) *Comparison of Deterministic CRBs*: Consider the following deterministic parameters, which are generated randomly once and fixed at all times.

- AoA vector $\theta = [40^\circ, 41^\circ, -70^\circ, -65^\circ]^T$,

- AoD vector $\varphi = [-25^\circ, -42.5^\circ, 55^\circ, 52^\circ]^T$,
- Path gain vector $\alpha = [0.5775 - 0.1733j, -0.3119 + 0.2599j, 0.2714 - 0.4505j, 0.4331 + 0.0866j]^T$.

Fig. 4 shows the CRBs for estimating the angular-domain parameters (e.g., AoA (θ_1) and $\alpha_{R,1}$) and channel matrix (i.e., \mathbf{H}) as functions of SNR for σ_v^2 known. Fig. 4 shows that the CRB curves for the OT scheme and the unquantized case (Unqt) are log-linear w.r.t. SNR. With a loss of 1.96 dB compared with the unquantized CRB, the OT scheme provides the lowest bounds among all one-bit thresholding schemes, as expected. At low SNRs, the one-bit CRBs converge due to the dominating noise effect. Compared with the fixed ZT scheme, the random thresholding schemes (i.e., the AT and TT schemes) provide slightly worse results for the AoA estimation in Fig. 4(a) but much better results for the path gain estimation in Fig. 4(b) as well as for the overall channel matrix estimation in Fig. 4(c). AT gives almost identical performance to its TT counterpart, but with a lower system complexity and cost. AT and ZT have similar hardware cost, but ZT suffers for ambiguity problems when σ_v^2 is unknown. Therefore, we focus on AT hereafter.

Fig. 5 shows the CRBs as functions of pilot length K using the AT scheme. First, it is observed that the CRB curves are log-linear w.r.t. the pilot length. Secondly, from Fig. 5, the unknown σ_v^2 has almost no impact on the angle CRB, but leads to higher CRBs for the path gain.

2) *Hybrid CRB with Prior Knowledge on Channel Path Gains*: Assume that the path gains have known Gaussian pdf with mean $\mathbf{m}_\alpha = \sqrt{\frac{1-\tau}{2N_s}}[1, \dots, 1]^T \in \mathbb{R}^{2N_s \times 1}$ and covariance matrix $\mathbf{C}_\alpha = \frac{\tau}{2N_s}\mathbf{I}_{2N_s}$ with $\tau = 0.1$. The angles are deterministic unknowns, given as follows:

- AoA vector $\theta = [62^\circ, -36^\circ, -45^\circ, 15^\circ]^T$,
- AoD vector $\varphi = [-45^\circ, 34^\circ, 25^\circ, -70^\circ]^T$.

Fig. 6 compares several versions of the path gain HCRB with the AT scheme as a function of SNR when σ_v^2 is unknown. Specifically, we consider the numerically implemented HCRB via Monte-Carlo trials (denoted as MC HCRB), the unquantized HCRB computed using the exact analytical expressions (denoted as Unqt HCRB), the low-SNR approximate HCRB in Section IV-A1 (denoted as LowHCRB), the tighter approximate HCRB in Section IV-A2 (denoted as ApproxHCRB) and the expectation of the deterministic CRB (ECRB) which serves as a tight bound for the prior-aided estimate in the asymptotic

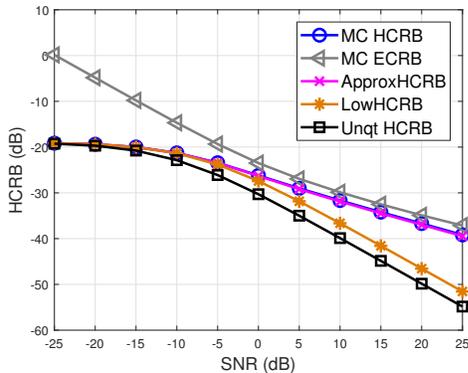


Fig. 6. Comparison of HCRB for $\alpha_{R,1}$ using the AT scheme as a function of SNR with prior knowledge on path gain for $K = 15$ and unknown σ_v^2 .

region (i.e., $\text{SNR} \rightarrow \infty$ or $K \rightarrow \infty$) [28]. The ECRB is obtained by averaging the deterministic CRB over its random realizations. Fig. 6 shows that LowHCRB is looser than MC HCRB when the $\text{SNR} > -5$ dB. ApproxHCRB is almost the same as MC HCRB in all SNRs we consider. At low SNRs, the ECRB is much higher than its HCRB counterparts due to the lack of prior knowledge, and the one-bit HCRBs coincide with the unquantized HCRB since the prior knowledge dominates over the information from data.

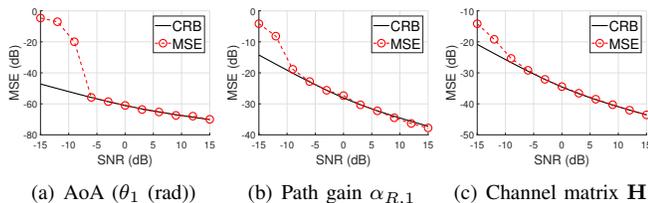


Fig. 7. MSE of ML estimate for the AT scheme as a function of SNR with $K = 32$ and unknown σ_v^2 .

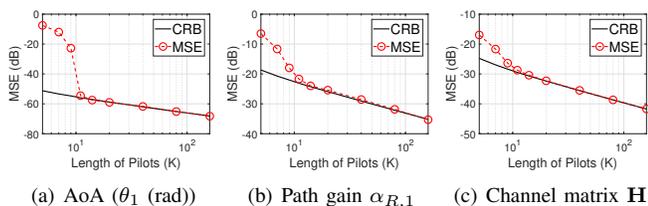


Fig. 8. MSE of ML estimate for the AT scheme as a function of pilot length K with $\text{SNR} = 0$ dB and unknown σ_v^2 .

B. ML, JML-MAP and MAP Parameter Estimation

We now evaluate the performance of the ML, JML-MAP and MAP channel estimates, all obtained via the variations of the 1bRELAX algorithm, where the number of the discretized grid points over the angle range $(-\pi/2, \pi/2]$ is $L_\theta = 192$ for the AoA and $L_\varphi = 96$ for the AoD. The MSEs are obtained by averaging over 300 Monte-Carlo trials.

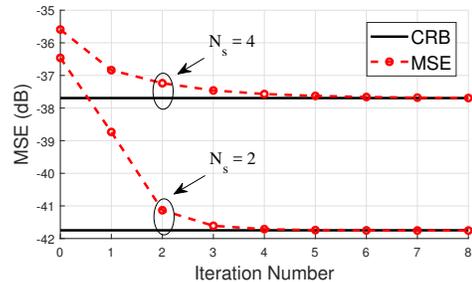


Fig. 9. MSE of ML channel matrix estimate for the AT scheme as a function of the iteration number in the *fine search* step of 1bRELAX when $K = 32$ and $\text{SNR} = 0$ dB.

TABLE I
AVERAGE RUNNING TIME OF 1bRELAX

Aver. Run. Time (s)	$K = 10$	$K = 20$	$K = 40$
$N_s = 2$	8.69	13.67	24.59
$N_s = 4$	18.89	28.84	49.76

1) *ML Estimation for Deterministic Unknown Parameters:* We assume that the path number N_s is known and use the same deterministic channel parameters as those in Section VI-A1. Note that the angle separations between the AoA of the first and second paths as well as the AoD of the third and fourth paths are less than the corresponding Rayleigh resolutions. The MSEs of the ML estimates are compared to the corresponding CRBs in Fig. 7 as functions of SNR and in Fig. 8 as functions of the pilot length K , when σ_v^2 is unknown. It is seen that the MSEs of the ML estimates obtained via 1bRELAX can approach the corresponding CRBs at high SNRs or with long pilot sequences. It is also observed that there is a *threshold effect* (e.g., $\text{SNR} = -6$ dB in Fig. 7 and $K = 11$ in Fig. 8) below which the MSEs deviate abruptly from the corresponding CRBs.

Fig. 9 shows the MSE of ML channel matrix estimate as a function of the iteration number in the *fine search* step of 1bRELAX when the path number is $N_s = 2$ and $N_s = 4$. The iteration number here refers to the iterations needed in the final step of refining the parameters of N_s paths. It is seen from Fig. 9 that 1bRELAX algorithm can converge with 4 iterations for $N_s = 2$ and 6 iterations for $N_s = 4$, respectively. Fig. 9 also reveals that the *grid-less* refinement step is necessary to enhance the channel estimation performance of 1bRELAX. We give the average running time of 1bRELAX for different pilot lengths in Table I. The simulations are conducted on a PC with Intel(R) Core(TM) i7-6700 CPU (3.40 GHz) and 64.0 GB RAM.

Then we compare the channel estimation performance of 1bRELAX with those of BLMMSE [16] and GAMP [9] in Fig. 10. Note that both the BLMMSE and GAMP algorithms need a prior knowledge on the noise variance, and hence σ_v^2 is assumed to be known for all methods. It is shown in Fig. 10 that GAMP provides a performance gain of at least 5 dB over BLMMSE. 1bRELAX significantly outperforms the other two algorithms, especially at high SNRs. For example, 1bRELAX

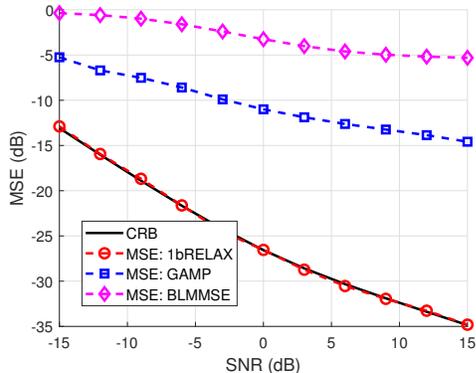


Fig. 10. Comparison of channel matrix estimation performance using the AT scheme as a function of SNR with $K = 128$ and known σ_v^2 .

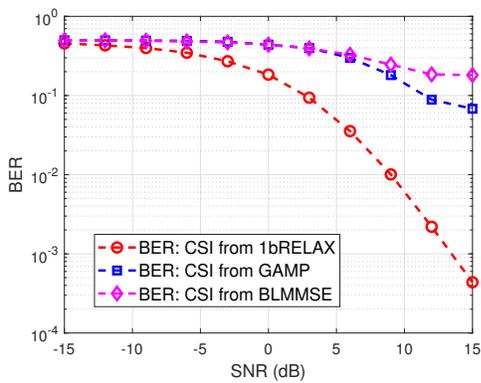


Fig. 11. Comparison of BER performance using the AT scheme with imperfect CSI as a function of SNR when σ_v^2 is known.

outperforms GAMP and BLMMSE by more than 20 dB when $\text{SNR} = 15$ dB. Fig. 11 plots the average bit-error rate (BER) curves obtained via the one-bit ML detector in [14] using imperfect channel state information (CSI) obtained from the aforementioned three algorithms, respectively. For simplicity, we consider a typical case where there is only one independent QPSK data stream between the transmitter and the receiver. The transmitted data signal has the same power as the pilot signal. It is shown in Fig. 11 that, as expected, 1bRELAX can provide a much better BER performance than GAMP and BLMMSE at high SNRs.

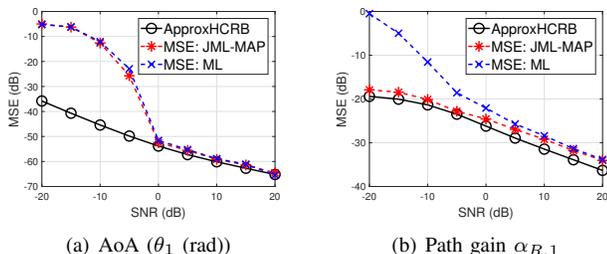


Fig. 12. MSE of JML-MAP estimate for the AT scheme with prior knowledge on path gains as a function of SNR for $K = 15$ and known σ_v^2 .

2) *JML-MAP Estimation for Hybrid Parameters*: For the knowledge-aided JML-MAP and MAP estimators, we assume

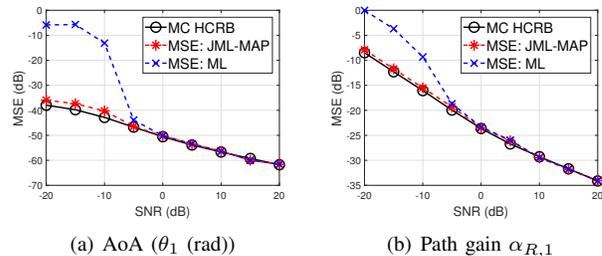


Fig. 13. MSE of JML-MAP estimate for the AT scheme with prior knowledge on angular parameters as a function of SNR when $K = 15$ and known σ_v^2 .

that the noise variance σ_v^2 is known. We first consider the case where the path gains have known prior pdfs and the angular parameters are deterministic unknowns. The simulation parameters used here are the same as those in Section VI-A2. Fig. 12 shows the MSEs of the JML-MAP estimates and the deterministic ML estimates for the AT scheme. It is shown in Fig. 12(b) that the JML-MAP path gain estimates are better than the ML estimates, especially at low SNRs. The ApproxHCRBs can provide reasonable lower bounds for the JML-MAP estimates for the SNRs we consider.

We then assume the angles have known truncated Gaussian distributions with angle intervals constructed by the *three-sigma rule*, i.e., all the random realizations of $\theta_l \in \mathcal{N}(m_{\theta_l}, \sigma_{\theta_l}^2)$ are in the interval $[m_{\theta_l} - 3\sigma_{\theta_l}, m_{\theta_l} + 3\sigma_{\theta_l}]$. Specifically, we consider the following parameters:

- AoA intervals: $\theta_1 \in [-74.5^\circ, -69.5^\circ]$, $\theta_2 \in [31^\circ, 36^\circ]$, $\theta_3 \in [-52.5^\circ, -47.5^\circ]$, $\theta_4 \in [59.5^\circ, 64.5^\circ]$,
- AoD intervals: $\varphi_1 \in [63.5^\circ, 68.5^\circ]$, $\varphi_2 \in [-45^\circ, -40^\circ]$, $\varphi_3 \in [17.5^\circ, 22.5^\circ]$, $\varphi_4 \in [-27.5^\circ, -22.5^\circ]$,
- Path gain vector $\alpha = [0.4058 + 0.3170j, -0.2536 + 0.4438j, 0.5706 - 0.1268j, -0.1014 - 0.3487j]^T$.

We plot the MSEs of JML-MAP estimates for the AT scheme in Fig. 13. Since no analytical expression for the HCRB exists in this case, we compute the HCRB via Monte-Carlo trails. It is seen that the *threshold effect* for angle estimation is suppressed due to the prior knowledge for the JML-MAP, which has a much better performance than the ML for low SNR. Moreover, since accurate angle estimates are available for all SNRs, the *threshold effect* of JML-MAP path gain estimate also vanishes and the corresponding MSE can approach the MC HCRB.

3) *MAP Estimation for Bayesian Parameters*: Finally, consider the case where the channel path gains have the same Gaussian distribution as in Section VI-A2 and the AoA and AoD angles have the same truncated Gaussian distributions as in Section VI-B2. We plot the MSE of the MAP channel matrix estimate for the AT scheme in Fig. 14. It is observed that the MSE of MAP channel estimates cannot approach the corresponding MC BCRB. The gap between them is about 5 dB.

C. 1bBIC for Path Number Determination

We now use the 1bBIC for channel path number determination for the same deterministic channel parameters as in Section VI-A and unknown σ_v^2 . Two different pilot sequences

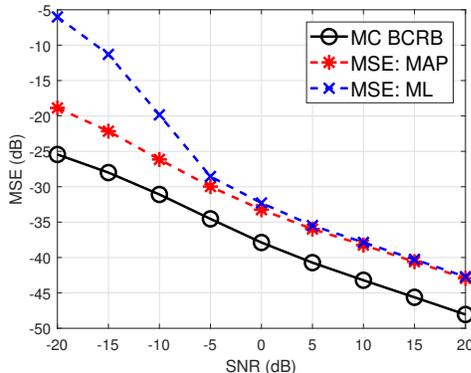


Fig. 14. MSE of MAP channel matrix estimate for the AT scheme as a function of SNR for $K = 15$ and known σ_v^2 .

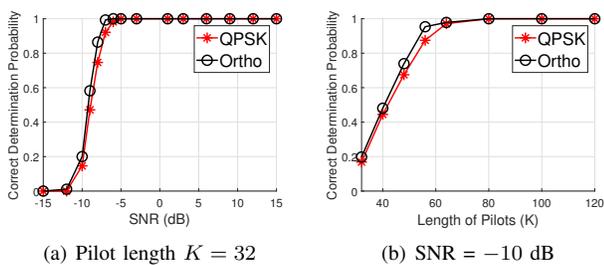


Fig. 15. Correct determination probability of 1bBIC when σ_v^2 is unknown, (a) as a function of SNR and (b) as a function of pilot length.

are considered here: random QPSK pilots (QPSK) and random orthogonal pilots (Ortho). The maximum possible path number N_{\max} is set to 8. We compare the correct determination probability of 1bBIC with the two different pilot sequences as functions of the SNR in Fig. 15(a) when the pilot length $K = 32$ and as functions of the pilot length in Fig. 15(b) when SNR = -10 dB. It can be seen from Fig. 15 that for both pilot sequences, the probability of correct path number determination reaches 1 for SNR > -5 dB in Fig. 15(a) or $K > 80$ in Fig. 15(b).

VII. CONCLUSIONS

We have focused on the angular-domain channel estimation for one-bit mmWave massive MIMO systems. In particular, we have derived the deterministic CRB and the knowledge-aided hybrid and Bayesian CRBs, for estimating the AoA, AoD and the associated path gains. We have also presented the ML estimator for deterministic channel parameter estimation. When the prior information is available, the JML-MAP and MAP estimators have been derived for the hybrid and Bayesian parameter estimation, respectively. Variations of the computationally efficient 1bRELAX algorithm have been used to obtain the ML, JML-MAP and MAP channel estimates. A simple and effective rule, referred to as 1bBIC, has been derived for path number determination. Numerical examples have been given to compare the performance bounds using different one-bit thresholding schemes and to verify the effectiveness of the parameter estimators and the 1bBIC rule.

We have demonstrated that accurate channel parameter estimates can be obtained using our estimators. The correct determination probability of 1bBIC approaches 1 even for a low SNR or a moderate pilot length. In addition, we have found that the simple and inexpensive *antenna-varying* thresholding scheme allows for the noise variance to be unknown and outperforms its zero-thresholding counterpart, especially at high SNRs. We have also shown that the prior knowledge of the angular-domain parameters can lead to significantly improved channel estimation performance, especially at low SNRs.

APPENDIX A

DERIVATION OF THE PARTIAL DERIVATIVES: $\frac{\kappa_{R/I,m}}{\partial \chi_p}$

Recall from the unquantized received signal model in (5), the complex-valued signal $\mathbf{\Gamma}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \boldsymbol{\alpha}$ is the superposition of the signals from N_s paths:

$$\begin{aligned} \mathbf{\Gamma}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \boldsymbol{\alpha} &= \sum_{l=1}^{N_s} \left[\left(\mathbf{X}^T \mathbf{a}_{\text{MS}}^*(\varphi_l) \right) \otimes \mathbf{a}_{\text{BS}}(\theta_l) \right] \alpha_l \\ &\triangleq \sum_{l=1}^{N_s} \gamma^l(\theta_l, \varphi_l) \alpha_l, \end{aligned} \quad (63)$$

where $\gamma^l(\theta_l, \varphi_l) \in \mathbb{C}^{KN_r \times 1}$. We ignore the dependence of γ^l on θ_l and φ_l for notation brevity in the following. The partial derivatives of $\kappa_{R,m}$ w.r.t. the l^{th} elements of $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\alpha}_R$ and $\boldsymbol{\alpha}_I$ in (16) can be computed, respectively, as:

$$\frac{\partial \kappa_{R,m}}{\partial \theta_l} = \frac{\partial \gamma_{R,m}^l}{\partial \theta_l} \alpha_{R,l} - \frac{\partial \gamma_{I,m}^l}{\partial \theta_l} \alpha_{I,l}, \quad (64a)$$

$$\frac{\partial \kappa_{R,m}}{\partial \varphi_l} = \frac{\partial \gamma_{R,m}^l}{\partial \varphi_l} \alpha_{R,l} - \frac{\partial \gamma_{I,m}^l}{\partial \varphi_l} \alpha_{I,l}, \quad (64b)$$

and

$$\frac{\partial \kappa_{R,m}}{\partial \alpha_{R,l}} = \gamma_{R,m}^l, \quad (64c)$$

$$\frac{\partial \kappa_{R,m}}{\partial \alpha_{I,l}} = -\gamma_{I,m}^l, \quad (64d)$$

with

$$\frac{\partial \gamma^l}{\partial \theta_l} = \left(\mathbf{X}^T \mathbf{a}_{\text{MS}}^*(\varphi_l) \right) \otimes \frac{\partial \mathbf{a}_{\text{BS}}(\theta_l)}{\partial \theta_l},$$

and

$$\frac{\partial \gamma^l}{\partial \varphi_l} = \left(\mathbf{X}^T \frac{\partial \mathbf{a}_{\text{MS}}^*(\varphi_l)}{\partial \varphi_l} \right) \otimes \mathbf{a}_{\text{BS}}(\theta_l),$$

where $\frac{\partial \mathbf{a}_{\text{BS}}(\theta_l)}{\partial \theta_l}$ and $\frac{\partial \mathbf{a}_{\text{MS}}^*(\varphi_l)}{\partial \varphi_l}$ are given by:

$$j2\pi \cos \theta_l \frac{d_r}{\lambda} \left[0, e^{j2\pi \sin \theta_l \frac{d_r}{\lambda}}, \dots, (N_r - 1) e^{j(N_r - 1)2\pi \sin \theta_l \frac{d_r}{\lambda}} \right],$$

and

$$j2\pi \cos \varphi_l \frac{d_t}{\lambda} \left[0, e^{j2\pi \sin \varphi_l \frac{d_t}{\lambda}}, \dots, (N_t - 1) e^{j(N_t - 1)2\pi \sin \varphi_l \frac{d_t}{\lambda}} \right],$$

respectively.

The partial derivatives of $\kappa_{I,m}$, i.e., $\frac{\partial \kappa_{I,m}}{\partial \theta_l}$, $\frac{\partial \kappa_{I,m}}{\partial \varphi_l}$, $\frac{\partial \kappa_{I,m}}{\partial \alpha_{R,l}}$ and $\frac{\partial \kappa_{I,m}}{\partial \alpha_{I,l}}$, can be obtained similarly.

APPENDIX B
REFORMULATION OF $g(\zeta_{R,m})p(\chi_r)$

Recall from the definition of $\kappa_{R,m}$ in (11a), and according to (63), we can reformulate $\kappa_{R,m}$ as:

$$\kappa_{R,m} = \sum_{l=1}^{N_s} \left[\gamma_{R,m}^l \alpha_{R,l} - \gamma_{I,m}^l \alpha_{I,l} \right] \triangleq \bar{\gamma}_m \chi_r, \quad (65)$$

where $\bar{\gamma}_m = [\gamma_{R,m}^1, \dots, \gamma_{R,m}^{N_s}, -\gamma_{I,m}^1, \dots, -\gamma_{I,m}^{N_s}]^T$. Plugging (65) into (10a), we have:

$$g(\zeta_{R,m}) = 4e^{-(1-\frac{2}{\pi})\zeta_{R,m}^2} = 4e^{-t\eta_{R,m}^2} \cdot e^{-\frac{1}{2}(\chi_r^T (2t\mathbf{A}_m)\chi_r - \mathbf{p}_m^T \chi_r)}, \quad (66)$$

where $t = (1 - \frac{2}{\pi}) \frac{2}{\sigma_v^2}$, $\mathbf{A}_m = \bar{\gamma}_m \bar{\gamma}_m^T$ and $\mathbf{p}_m = 4t\eta_{R,m} \bar{\gamma}_m$. Then,

$$\begin{aligned} g(\zeta_{R,m})p(\chi_r) &= \frac{4e^{-t\eta_{R,m}}}{\sqrt{(2\pi)^{2N_s} |\mathbf{C}_\alpha|}} \\ &\cdot e^{-\frac{1}{2}(\chi_r^T (2t\mathbf{A}_m)\chi_r - \mathbf{p}_m^T \chi_r + (\chi_r - \mathbf{m}_\alpha)^T \mathbf{C}_\alpha^{-1} (\chi_r - \mathbf{m}_\alpha))} \\ &= \delta_m \cdot \frac{1}{\sqrt{(2\pi)^{2N_s} |\mathbf{Q}_m|}} e^{-\frac{1}{2}(\chi_r - \mathbf{u}_m)^T \mathbf{Q}_m^{-1} (\chi_r - \mathbf{u}_m)}, \end{aligned} \quad (67)$$

where

$$\begin{aligned} \delta_m &= 4 \sqrt{\frac{|\mathbf{Q}_m|}{|\mathbf{C}_\alpha|}} e^{-t\eta_{R,m}^2 - \frac{1}{2} \mathbf{m}_\alpha^T \mathbf{C}_\alpha^{-1} \mathbf{m}_\alpha + \epsilon_m}, \\ \epsilon_m &= \frac{1}{8} (2\mathbf{C}_\alpha^{-1} \mathbf{m}_\alpha + \mathbf{p}_m)^T \mathbf{Q}_m (2\mathbf{C}_\alpha^{-1} \mathbf{m}_\alpha + \mathbf{p}_m), \\ \mathbf{u}_m &= \frac{1}{2} \mathbf{Q}_m (2\mathbf{C}_\alpha^{-1} \mathbf{m}_\alpha + \mathbf{p}_m), \\ \mathbf{Q}_m &= (2t\mathbf{A}_m + \mathbf{C}_\alpha^{-1})^{-1}. \end{aligned}$$

According to (67), $g(\zeta_{R,m})p(\chi_r)$ has been reformulated to a constant of δ_m times a Gaussian pdf of χ_r with mean vector \mathbf{u}_m and covariance matrix \mathbf{Q}_m .

REFERENCES

- [1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [2] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.
- [3] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
- [4] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [5] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *International ITG Workshop on Smart Antennas (WSA)*, Bremen, Germany, Feb. 2010, pp. 143–150.
- [6] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv preprint arXiv:1404.7736*, 2014.
- [7] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
- [8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.
- [9] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
- [10] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.
- [11] M. Sarajlic, L. Liu, and O. Edfors, "When are low resolution ADCs energy efficient in massive MIMO?" *IEEE Access*, vol. 5, pp. 14837–14853, Jul. 2017.
- [12] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007, pp. 1286–1289.
- [13] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
- [14] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
- [15] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *IEEE International Conference on Communication Workshop (ICCW)*, London, United Kingdom, Jun. 2015, pp. 1304–1309.
- [16] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.
- [17] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.
- [18] J. Zhu, H. Cao, and Z. Xu, "A two-stage approach to estimate CFO and channel with one-bit ADCs," *Signal Process.*, vol. 168, Mar. 2020.
- [19] M. S. Stein, S. Bar, J. A. Nossek, and J. Tabrikian, "Performance analysis for channel estimation with 1-bit ADC and unknown quantization threshold," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2557–2571, May 2018.
- [20] F. Wang, J. Fang, H. Li, Z. Chen, and S. Li, "One-bit quantization design and channel estimation for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10921–10934, Nov. 2018.
- [21] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [22] P. Wang, J. Li, M. Pajovic, P. T. Boufounos, and P. V. Orlik, "On angular-domain channel estimation for one-bit massive MIMO system with fixed and time-varying thresholds," in *51th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2017.
- [23] F. Liu, H. Zhu, J. Li, P. Wang, and P. V. Orlik, "Massive MIMO channel estimation using signed measurements with antenna-varying thresholds," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*, Freiburg, Germany, Jun. 2018, pp. 188–192.
- [24] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [25] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, Sep. 2007.
- [26] J. Li and P. Stoica, Eds., *MIMO radar signal processing*. NJ: John Wiley & Sons, 2009.
- [27] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.
- [28] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.
- [29] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.
- [30] N. González-Prelcic, A. Ali, V. Va, and R. W. Heath, "Millimeter-wave communication with out-of-band information," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 140–146, Dec. 2017.
- [31] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, United Kingdom: Cambridge University Press, 2004.
- [32] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

- [33] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [34] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, May 1978.
- [35] J. Fang and H. Li, "Adaptive distributed estimation of signal power from one-bit quantized data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 4, pp. 1893–1905, Oct. 2010.
- [36] C. Gianelli, L. Xu, J. Li, and P. Stoica, "One-bit compressive sampling with time-varying thresholds: Maximum likelihood and the Cramér-Rao bound," in *50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2016, pp. 399–403.



Fangqing Liu received the B.E. degree in communication engineering from the Xidian University, Xi'an, China in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His current research interests include spectral estimation, array signal processing and their applications to communication and radar systems.



Heng Zhu received the B.S. degree in physics from the University of Science and Technology of China in 2017. Since September 2017, he has been working towards his M.S. degree in electrical engineering with the University of Science and Technology of China. His current research interests include statistical signal processing and compressed sensing.



Changheng Li received the B.S. degree in applied mathematics from the University of Science and Technology of China, Hefei, China, in 2017, where he is currently working toward the M.S. degree at the Department of Electrical Engineering and Information Science. His research interests include spectral estimation, statistical signal processing, and their applications.



Jian Li (F'05) received the M.Sc. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, OH, USA, in 1987 and 1991, respectively. She is currently a Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. Her current research interests include spectral estimation, statistical and array signal processing, and their applications to radar, sonar, and biomedical engineering. Her publications include *Robust Adaptive Beamforming* (2005, Wiley), *Spectral Analysis: the Missing Data Case* (2005, Morgan and Claypool), *MIMO Radar Signal Processing* (2009, Wiley), and *Waveform Design for Active Sensing Systems A Computational Approach* (2011, Cambridge University Press).

Dr. Li is a Fellow of IET and also a Fellow of the European Academy of Sciences (Brussels). She was the recipient of the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, May 2002. She was an Associate Editor for the *IEEE TRANSACTIONS ON SIGNAL PROCESSING* from 1999 to 2005, an Associate Editor for the *IEEE SIGNAL PROCESSING MAGAZINE* from 2003 to 2005, and a member of the Editorial Board of *Signal Processing*, a publication of the European Association for Signal Processing, from 2005 to 2007. She was a member of the Editorial Board of the *IEEE SIGNAL PROCESSING MAGAZINE* from 2010 to 2012. She is currently a member of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society. She is a co-author of the paper that has received the M. Barry Carlton Award for the best paper published in the *IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS* in 2005 and also a co-author of a paper published in the *IEEE TRANSACTIONS ON SIGNAL PROCESSING* that has received the Best Paper Award in 2013 from the IEEE Signal Processing Society.



Pu Wang (S'05-M'12-SM'18) received the Ph.D. degree in Electrical Engineering from the Stevens Institute of Technology, Hoboken, NJ, USA, in 2011.

He is now a Principal Research Scientist at the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, where he was an intern in the summer of 2010. Before returning to MERL, he was a Research Scientist at the Schlumberger-Doll Research, Cambridge, MA, contributing to developments of next-generation logging-while-drilling Acoustics/NMR products. His current research interests include signal processing, Bayesian inference, statistical learning, and their applications to (mmWave and THz) sensing, wireless communications, networks and automotive applications.

Dr. Wang was selected as a Distinguished Speaker of the Society of Petrophysicists and Well Log Analysts (SPWLA) in 2017 for the work "Dipole Shear Anisotropy Using Logging-While-Drilling Sonic Tools". He was a recipient of the Conrad Schlumberger Award for Technical Depth twice in 2017 and 2015 from the Schlumberger (North America) Reservoir Symposium, the IEEE Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society in 2013, and the Outstanding Paper Award from the IEEE AFICON Conference in 2011. He also received the Outstanding Doctoral Thesis in Electrical Engineering Award in 2011, the Edward Peskin Award in 2011, the Francis T. Boesch Award in 2008, and the Outstanding Research Assistant Award in 2007 from the Stevens Institute of Technology. He is an Associate Editor for *IEEE Signal Processing Letters*.

Dr. Wang was selected as a Distinguished Speaker of the Society of Petrophysicists and Well Log Analysts (SPWLA) in 2017 for the work "Dipole Shear Anisotropy Using Logging-While-Drilling Sonic Tools". He was a recipient of the Conrad Schlumberger Award for Technical Depth twice in 2017 and 2015 from the Schlumberger (North America) Reservoir Symposium, the IEEE Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society in 2013, and the Outstanding Paper Award from the IEEE AFICON Conference in 2011. He also received the Outstanding Doctoral Thesis in Electrical Engineering Award in 2011, the Edward Peskin Award in 2011, the Francis T. Boesch Award in 2008, and the Outstanding Research Assistant Award in 2007 from the Stevens Institute of Technology. He is an Associate Editor for *IEEE Signal Processing Letters*.



Philip V. Orlik (M'97-11, SM'12) was born in New York, NY in 1972. He received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York (SUNY) at Stony Brook. In 1999 he earned his Ph.D. in electrical engineering also from SUNY Stony Brook.

Since 2000 he has been with Mitsubishi Electric Research Laboratories Inc. located in Cambridge, MA where he is currently the Manager of the Signal Processing Group. His primary research focus is on advanced wireless and wired communications, sensor/IoT networks. Other research interests include vehicular/car-to-car communications, mobility modeling, performance analysis, and queuing theory.