Channel Estimation in One-Bit Massive MIMO Systems: Angular Versus Unstructured Models

Shilpa Rao, *Student Member, IEEE*, Amine Mezghani, *Member, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*—Millimeter wave (mmWave) massive MIMO cellular systems will be characterized by increased bandwidths and the ability to handle a large number of users. To cope with the power consumption problem due to an increased number of receive antennas, the idea of equipping one-bit ADCs at the base station has been proposed. The goal of this paper is to establish performance bounds on the channel estimation of one-bit mmWave massive MIMO receivers for different types of channel models. The Cramér-Rao bound (CRB) is a lower bound on the performance of unbiased estimators and sets a benchmark for the design of channel estimators. We consider both a structured channel model for a single user where the channel is composed of a superposition of multipaths characterized by path delays and directions-of-arrival (DOAs), and an unstructured channel model where the channel is a generic FIR filter. The Fisher information matrix (FIM) for these channel models are derived in closed form. The CRB is also extended to a dictionary-based channel model, where the path delays and DOAs are selected from small perturbations on a discrete grid, and a sparsity constraint applies to the vector of path loss components. We also derive the Bayesian CRB when the array response is imperfectly known and is affected by perturbations in the sensor pattern or position. The CRBs are evaluated numerically and the effects of various system parameters on the CRB are studied. The dependencies between channel parameters and the effect of array perturbations are also investigated.**

***Index Terms*—Massive multiple input multiple-output, wideband channel estimation, one-bit analog-to-digital converters, Cramér-Rao bounds, array calibration.**

I. Introduction

MILLIMETER wave (mmWave) massive multiple-input multiple-output (MIMO) is a promising area for next-generation wireless communication systems. These systems employ arrays with many antennas, of the order of a hundred or more, at the base station (BS) and operate in the 30–300 GHz frequency range. Furthermore, the rising user demands for capacity can be met through higher bandwidths and spatial multiplexing. Massive MIMO is also capable of concentrating energy in very selective directions, thus significantly increasing the energy efficiency. However, at mmWave, the benefits of massive MIMO are limited by the low SNR per antenna a result of increased propagation losses, diminished scattering, atmospheric absorption and higher noise bandwidths.

Channel estimation, which is key for exploiting the potential gains offered by mmWave massive MIMO, is challenging due to the aforementioned limitations. Traditional channel models that assume Rayleigh fading are not suitable for mmWave systems because the scattering environment of mmWave channels is not dense, but rather sparse with line-of-sight (LOS) and a few reflected propagation paths contributing to the effective channel [1], [2]. Hence, much of the work in mmWave has focused on direction-of-arrival (DOA)-based channel estimation. Previous works in mmWave channel estimation exploit the sparsity [3] of these channels in the delay and angle domains [4]–[11]. The authors of [1] focus on pilot-aided mmWave DOA-based channel estimation, channel subspace estimation is studied in [9] and mmWave channel estimation with hybrid architectures is considered in [12] and [13].

The large bandwidths at mmWave limit the performance of analog-to-digital converters (ADCs) at the receiver. It has been shown that the power consumption of ADCs increases drastically with increasing quantization resolution and sampling rate requirements [14], [15]. One-bit ADCs are cost-effective, consume less power and do not require automatic gain control [16], [17]. Moreover, it has been shown that one-bit ADCs suffer from a low power penalty (approximately $\pi/2$) at low SNRs [18]. Therefore, one-bit quantization is pertinent to mmWave massive MIMO systems since they suffer from low SNR per antenna. Prior work focuses on channel estimation using the Bussgang decomposition with a Rayleigh fading channel model [19], quantization design with non-zero threshold and channel estimation [20], throughput analysis [21], blind and semi-blind channel estimation with time division duplexing [21] in one-bit massive MIMO systems. The authors of [22] provide a blind sparse channel estimation algorithm based on a maximum likelihood (ML) formulation. In [5], approximate message passing algorithms to exploit the joint sparsity of the broadband channel in the angle and delay domains with few-bit ADCs are proposed. However, in the above-mentioned DOA-based channel models, it is assumed that the inter-element time delay between antennas in the array is small compared to the inverse signal bandwidth. For mmWave massive MIMO systems, this assumption will typically not be true. This effect, sometimes referred to as (beam) "squint", has been observed to cause a serious mismatch in the array

Manuscript received February 15, 2019; revised June 19, 2019; accepted July 13, 2019. Date of publication August 5, 2019; date of current version September 20, 2019. This work was supported by National Science Foundation under Grant ECCS-1547155 and Grant CCF-1703635. The guest editor coordinating the review of this article and approving it for publication was Dr. Shi Jin. *(Corresponding author: Shilpa Rao.)*

S. Rao and A. L. Swindlehurst are with the Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA 92697 USA (e-mail: shilpar1@uci.edu; swindle@uci.edu).

A. Mezghani is with the Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA 92697 USA. He is now with Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: amine.mezghani@utexas.edu).

Digital Object Identifier 10.1109/JSTSP.2019.2933163

response, and if ignored can significantly degrade performance [23]–[26].

In this paper, we focus on Cramér-Rao performance bounds for channel estimation in one-bit mmWave massive MIMO systems. We will explore both deterministic and Bayesian Cramér-Rao bounds (CRBs) depending on whether or not prior information about the parameters is available. In the Bayesian setting, the a priori information is taken into account in the computation of the joint probability density function (pdf). Misspecified Cramér-Rao bounds (MCRBs), when the assumed data model is different from the true model, can also be derived [27]. In [28], a lower bound on the Fisher information matrix (FIM) for the exponential family of distributions is derived and the ML estimator based on the "pessimistic" CRB for the DOA parameter is derived in [29]. In [30] and [31], the CRB for the channel parameterized by DOAs and path gains is derived when the channel and the array responses are frequency flat. In our work, we will take into account the channel estimation error for DOA-based models when the array response does not exactly match the assumed array model, similar to the array perturbation studies of [32]–[34]. In particular, we are interested in the level of array calibration accuracy needed for DOA-based methods to maintain their advantage compared with less parsimonious unstructured models.

Our analysis focuses on a one-bit mmWave single-input-multiple-output (SIMO) pilot-based single-carrier transmission system where a single transmitter is equipped with one antenna and the base station employs an antenna array. We consider channel models that are either "structured" (DOA-based, arbitrary delays) or "unstructured" (FIR, uniformly-spaced delays), and compare the resulting CRBs for the channel estimates. Under the structured channel model, we assume that the channel is parameterized by the multipath fading coefficient, the DOA, and the delay associated with each of the paths. We also take into account perturbations to the array response, as mentioned above. The channel is modeled as having a finite duration impulse response composed of a discrete number of arbitrary delays. To simplify the analysis, we do not assume any Doppler spread is present, so the angle and delay parameters are assumed to be time-invariant over the channel estimation period. We also take into account the squint effect mentioned above, where for wideband signals the time delay from one end of the array to the other cannot simply be represented as a phase shift. Incorporating these temporal shifts is important because the end-to-end delay for the antenna array is of the same order as the symbol duration.

For the unstructured model, the channel is represented as an FIR filter. While this model has a much larger number of parameters than the structured case when the number of multipaths arrivals at each delay is not large, the model depends linearly on the parameters, and thus it is easier to estimate. The interesting issue to be addressed is whether or not the increased difficulty in estimating the structured channel is worth the potential gain in estimation performance.

We also consider a dictionary-based channel model used in the compressive sensing literature [6], [10], [22]. The dictionary is based on a discrete set of DOAs and path delays obtained from a grid, where the grid size is greater than the number of antennas. This formulation is commonly used in mmWave channel estimation since compressive sensing based algorithms which exploit the underlying sparse multipath structure can be used. We consider a "dictionary mismatched" channel model, where the multipath DOAs and delays are matched to the nearest grid point and the difference between the dictionary and the true source parameters, or the grid mismatch, are parameters to be estimated.

We derive the FIM for the parameters of interest assuming that the "length" of the channel (i.e., the number of discrete multipaths in the structured case, or the maximum length of the FIR filter in the unstructured case) and the noise variance are known. For the structured model, these parameters are the path loss components, DOAs, path delays and the array perturbation parameters. The perturbations are modeled as complex Gaussian with a known covariance matrix. The FIM is derived for the specific case of uniform linear arrays (ULAs), a root-raised-cosine shaping function, and sensor position and pattern perturbations. For the unstructured channel model, the parameters are simply the path loss components, whereas for the dictionary-based channel, the parameters are the path loss components, the grid mismatch error parameters and the array perturbation parameters.

A number of numerical experiments are performed to evaluate the CRBs, and a comparison of the structured, unstructured and dictionary-based channels as a function of the SNR is performed. The effects of perturbation, bandwidth, the channel delay-tap length, and the number of receive antennas on the CRBs of the one-bit quantized system are also studied. We also include comparisons with the CRB obtained when there is no quantization error. The numerical results provide insight into the relative impact of the various factors that influence of the channel estimate, including the precision of the array calibration, the model parsimony, the one-bit quantization, size of the array, SNR, bandwidth, etc.

Notation: Boldface lowercase $\mathbf{a}$ denotes a vector and boldface uppercase $\mathbf{A}$ denotes a matrix. $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. The ith element of $\mathbf{a}$ and the (i,j)th entry of $\mathbf{A}$ are represented by a_i and $[\mathbf{A}]_{ij}$ respectively. The Hadamard (element-wise) product is represented by $\odot$, the Kronecker product by $\otimes$ and the convolution operation by $\circledast$. The operation $\mathrm{vec}(\cdot)$ denotes the vectorization operation, i.e. the stacking of the columns of a matrix one below the other. Real and imaginary parts are given by $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ respectively. $\mathbb{E}[\cdot]$ is the expectation operator. The ith row and jth column of the matrix $\mathbf{A}$ are given by $\mathbf{A}^{(i,)}$ and $\mathbf{A}^{(,j)}$ respectively. $\mathbf{A} \succeq \mathbf{B}$ and $\mathbf{A} \succ \mathbf{B}$ mean that $\mathbf{A} - \mathbf{B}$ is positive semidefinite and positive definite respectively. The cumulative distribution function (cdf) of the standard normal distribution is given by $\Phi(x)$, and its derivate $\Phi'(x)$ is the pdf of the standard normal distribution.

II. MmWave Channel Model

We consider an uplink mmWave MIMO system with a single-antenna user terminal and M receive antennas at the base station. We assume that the wireless communication channel is linear and its properties change slowly with respect to the

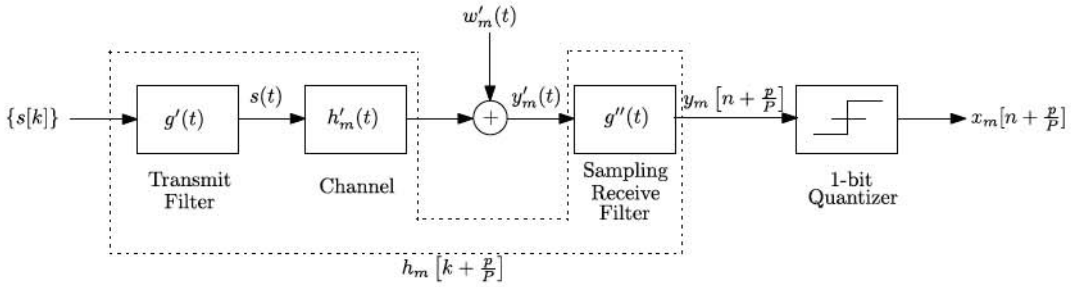


Fig. 1. System block diagram.

signal duration. The communication system block diagram is illustrated in Fig. 1. The received baseband signal at the mth antenna is given by

$$y'_m(t) = \sqrt{P_t}\int_{-\infty}^{\infty} h'_m(t')s(t-t')dt' + w'_m(t), \qquad (1)$$

where P_t is the transmit signal power, $h'_m(t)$ is the impulse response of the channel from the transmitter to antenna m at time t and $w'_m(t)$ is the corresponding noise.

The source signal is assumed to be band-limited to $[-B/2, B/2]$, and $w'_m(t)$ is a complex circularly symmetric and temporally white Gaussian process with power spectral density N_0. The source signal $s(t)$ is encoded as a digital signal with a common pulse period T_s, where the complex valued symbols $\{s[k]\}$ are modulated by a pulse shape function $g'(t)$ as

$$s(t) = \sum_{k=-\infty}^{\infty} s[k]g'(t-kT_s).$$

The received signal in (1) is assumed to be fractionally sampled by the filter $g''(t)$ by a factor of P, and can be equivalently represented in sampled time as

$$y_m\left[n+\frac{p}{P}\right] = \sqrt{P_t}\sum_{l=-\infty}^{\infty} h_m\left[l+\frac{p}{P}\right]s[n-l] + w_m\left[n+\frac{p}{P}\right], \qquad (2)$$

where $p = 0, 1, \ldots, P-1$, $h_m\left[l+\frac{p}{P}\right]$ is the equivalent discrete-time channel and $w_m\left[n+\frac{p}{P}\right]$ is the discrete-time noise, given by

$$h_m\left[l+\frac{p}{P}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g''\left(\left(l+\frac{p}{P}\right)T_s - t''\right) h'_m(t''-t')g'(t')dt'dt''$$

$$w_m\left[n+\frac{p}{P}\right] = \int_{-\infty}^{\infty} g''(t'')w'_m\left(\left(n+\frac{p}{P}\right)T_s - t''\right)dt''.$$

We will absorb the effects of the transmit and receive filters, $g'(t)$ and $g''(t)$, into $g(t) = g''(t) \circledast g'(t)$. We assume that the convolutive channel is frequency selective with maximum delay length of L symbol periods, so that the FIR assumption means that at most L consecutive symbols play a role in the received signal, i.e. $h_m\left[l+\frac{p}{P}\right]$ is zero outside the interval $[0, L-1]$. Generally, oversampling w.r.t. the Nyquist rate results in noise correlation [35]. However, if the receive filter $g''(t)$ is chosen to be a root-Nyquist pulse, the discrete-time noise $w_m[n]$ is white [36]. Consequently, the root-raised-cosine filter is chosen for the transmit and receive filters in our analysis, so that

$$g(t) = \frac{\sin \pi t/T_s}{\pi t/T_s}\frac{\cos \pi\alpha t/T_s}{1-4\alpha^2t^2/T_s^2},$$

where α is the roll-off factor and the noise $w_m[n] \sim \mathcal{CN}(0, \sigma^2)$, $\sigma^2 = N_0B$. In our analysis, we assume that the noise variance σ^2 is known, since the channel gains and σ^2 are not separately identifiable when one-bit quantization is used, which leads to a singular FIM [30], [31].

A. Unstructured Channel Model

In the unstructured case, the channel is modeled as a uniformly sampled FIR filter characterized by the complex gains of each path to the receiver. Let $\beta_{r,m}$ be the complex path gain of the rth path to the mth antenna so that the channel between the source and antenna m is

$$h'_m(t) = \sum_{r=1}^{R}\beta_{r,m}\delta(t-(\tau_0+(r-1)\Delta')),$$

where τ_0 corresponds to the delay of the first multipath arrival, and the value of Δ' is determined by the minimum resolvable time difference between different paths. Under this FIR model, the discrete channel can be expressed as $\mathbf{h}[k] = [h_1[k], h_2[k], \ldots, h_M[k]]^T$, where $k = l + \frac{p}{P}$, and

$$\mathbf{h}[k] = \underbrace{\begin{bmatrix}\beta_{1,1} & \ldots & \beta_{R,1}\\ \beta_{1,2} & \ldots & \beta_{R,2}\\ \vdots & \ldots & \vdots\\ \beta_{1,M} & \ldots & \beta_{R,M}\end{bmatrix}}_{\tilde{\boldsymbol{\beta}}\in\mathbb{C}^{M\times R}}\begin{bmatrix} g(kT_s-\tau_0)\\ g(kT_s-(\tau_0+\Delta'))\\ \vdots\\ g(kT_s-(\tau_0+(R-1)\Delta'))\end{bmatrix}. \qquad (3)$$

The parameters of the unstructured model are comprised by $\boldsymbol{\beta}$, the vector of all complex path gains, i.e. $\boldsymbol{\beta} = \mathrm{vec}(\tilde{\boldsymbol{\beta}})$, and we also define $\boldsymbol{\beta}^{\mathrm{Re}} = \mathrm{Re}(\boldsymbol{\beta}) \in \mathbb{R}^{MR\times 1}$ and $\boldsymbol{\beta}^{\mathrm{Im}} = \mathrm{Im}(\boldsymbol{\beta}) \in \mathbb{R}^{MR\times 1}$.

B. Structured Channel Model

The structured channel model is a geometric channel parameterized by path loss components, path delays and DOAs. The response of the antenna array to a waveform arriving from direction θ is denoted by $\mathbf{a}(\theta, \rho) \in \mathbb{C}^{M\times 1}$. The vector ρ represents

parameters on which the array response depends (e.g., antenna positions, gain and phase response, etc.). We will assume these parameters are a priori known to be Gaussian with some nominal mean value $\boldsymbol{\rho}_0$ and covariance $\boldsymbol{\Omega}$.

Let the DOA of the rth multipath, measured clockwise with respect to the y-axis, be $\theta_r, r = 1, 2, \ldots, R$, and assume that the antenna elements are close enough together so that they share a common complex path gain γ_r for the rth path. The path gain γ_r is an aggregate of the large-scale fading, namely the path loss and shadowing, as well as the small scale fading. More specifically,

$$\beta_{r,m} = \gamma_r q_m(\theta_r, \boldsymbol{\rho}) \exp(-j\omega_c \tau_{r,m}), \tag{4}$$

where ω_c is the carrier frequency and $q_m(\theta_r, \boldsymbol{\rho})$ is the sensor pattern of the mth sensor in the direction θ_r. The channel between the source and the mth antenna is

$$h'_m(t) = \sum_{r=1}^{R} \gamma_r q_m(\theta_r, \boldsymbol{\rho}) \exp(-j\omega_c \tau_{r,m}) \delta(t - \tau_{r,m}).$$

Incorporating the transmit and receive pulse shaping, the discrete-time equivalent channel becomes $h_m[k] = \sum_{r=1}^{R} \gamma_r q_m(\theta_r, \boldsymbol{\rho}) g(kT_s - \tau_{r,m}) \exp(-j\omega_c \tau_{r,m})$ and stacking the discrete-time channels from all M antennas, we get

$$\mathbf{h}[k] = \sum_{r=1}^{R} \gamma_r \begin{bmatrix} g(kT_s - \tau_{r,1}) \\ g(kT_s - \tau_{r,2}) \\ \vdots \\ g(kT_s - \tau_{r,M}) \end{bmatrix} \odot \mathbf{a}(\theta_r, \boldsymbol{\rho}), \tag{5}$$

where $\odot$ is the Hadamard (element-wise) product and the term $\exp(-j\omega_c \tau_{r,m})$ in (4) was absorbed into the expression for the array response $\mathbf{a}(\theta_r, \boldsymbol{\rho})$ by writing it in terms of the array sensor coordinates. Note that (5) uses time delays rather than phase shifts in the pulse shaping functions because the inverse bandwidth of the signals at mmWave frequencies may approach the inter-element delay between antennas in the array.

As an example of the array perturbation parameter $\boldsymbol{\rho}$, consider the case of sensor position and pattern perturbation. The array response vector $\mathbf{a}(\theta_r, \boldsymbol{\rho})$ can be written as

$$\mathbf{a}(\theta_r, \boldsymbol{\rho}) = \begin{bmatrix} q_1(\theta_r, \boldsymbol{\rho}) \exp\left(-j2\pi(x_1(\boldsymbol{\rho}) \sin\theta_r + y_1(\boldsymbol{\rho}) \cos\theta_r)/\lambda\right) \\ \vdots \\ q_M(\theta_r, \boldsymbol{\rho}) \exp\left(-j2\pi(x_M(\boldsymbol{\rho}) \sin\theta_r + y_M(\boldsymbol{\rho}) \cos\theta_r)/\lambda\right) \end{bmatrix},$$

where $(x_m(\boldsymbol{\rho}), y_m(\boldsymbol{\rho}))$ are the coordinates of the mth sensor. In the presence of an antenna pattern perturbation only [34], we have $q_m(\theta_r, \boldsymbol{\rho}) = q_{m,0}(\theta_r) + \rho_m$, where $q_{m,0}(\theta_r)$ is the nominal pattern, ρ_m is the complex perturbation and $\boldsymbol{\rho}$ stacks both the real and imaginary parts of the complex perturbation for all antennas, so that $\boldsymbol{\rho} \in \mathbb{R}^{2M \times 1}$. For sensor position perturbations only, the sensor coordinates can be modeled in a recursive manner as in [32]. The following piecewise linear model applies to flexible array structures: $(x_m(\boldsymbol{\rho}), y_m(\boldsymbol{\rho})) = (x_{m-1} + \delta \sin\phi_m(\boldsymbol{\rho}),\ y_{m-1} + \delta \cos\phi_m(\boldsymbol{\rho}))$, where $\phi_m(\boldsymbol{\rho}) = \phi_{m-1} + \rho_{m-1}$, with initial conditions $\phi_1 = \pi/2$, $x_1 = y_1 = 0$, and δ is the spacing between antenna elements. Thus, ρ_m is the incremental angular perturbation of the mth sensor and $\boldsymbol{\rho} \in \mathbb{R}^{M-1 \times 1}$.

A special case of the above modeling is a uniform linear array operating with a nominal omnidirectional sensor pattern, $q_m(\theta_r) = 1$ and $\boldsymbol{\rho} = \mathbf{0}$. The path delay to the mth sensor is given by $\tau_{r,m} = \tau_r + (m-1)\frac{\delta}{c}\sin\theta_r$, where c is the speed of light and $\tau_r = \tau_{r,1}$ is the time delay of the rth propagation path to the first antenna element. Let $\mathbf{g}_k(\tau_r, \theta_r) = [g(kT_s - \tau_r), \ldots, g(kT_s - \tau_r - \frac{(M-1)\delta\sin\theta_r}{c})]^T$, $\mathbf{a}(\theta_r) = [1, e^{-j\frac{2\pi\delta}{\lambda}\sin\theta_r}, \ldots, e^{-j\frac{2\pi\delta}{\lambda}(M-1)\sin\theta_r}]^T$ where λ is the wavelength, and $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_R]^T$. The lth delay-tap of the channel can then be written as

$$\mathbf{h}[k] = \underbrace{[\mathbf{g}_k(\tau_1, \theta_1), \mathbf{g}_k(\tau_2, \theta_2), \ldots, \mathbf{g}_k(\tau_R, \theta_R)]}_{\mathbf{G}_k \in \mathbb{R}^{M \times R}} \odot \underbrace{[\mathbf{a}(\theta_1)\ \mathbf{a}(\theta_2)\ \ldots\ \mathbf{a}(\theta_R)]}_{\mathbf{A} \in \mathbb{C}^{M \times R}} \boldsymbol{\gamma}. \tag{6}$$

In the derivation of the CRB, we use the following notation: $\boldsymbol{\gamma}^{\mathrm{Re}} = \mathrm{Re}(\boldsymbol{\gamma})$ and $\boldsymbol{\gamma}^{\mathrm{Im}} = \mathrm{Im}(\boldsymbol{\gamma})$.

Note: Although the path delays, $\tau_{r,m}$, are functions of $\boldsymbol{\rho}$ (in the case of position perturbation, for example), we do not consider this effect in the derivation of the FIM since the derivative of $\boldsymbol{\tau}$ with respect to $\boldsymbol{\rho}$ is negligible.

C. Dictionary Based Channel Model

In mmWave transmission, the propagation channel is often described using a sparse scattering model. The underlying channel is still parameterized by DOAs, path delays and complex path gains as in the structured model, but the DOAs and delays are assumed to lie on a fixed grid, and the channel estimation is formulated as a sparse recovery problem. This approach leverages tools available in compressive sensing to design efficient algorithms for determining the channel. A grid mismatch occurs if a particular DOA is not present in the possible DOA set. In our analysis, we model the true DOA as a perturbation to the nearest DOA in the grid. Let the uniform grid of DOAs consist of N_a points with $N_a \geq M$, so that the DOA dictionary is the set $\theta' \in \{0, 2\pi/N_a, \ldots, 2\pi(N_a - 1)/N_a\}$. Then, a Taylor interpolation of $\mathbf{a}(\theta, \boldsymbol{\rho})$ around the nearest DOA in the grid, θ', yields

$$\mathbf{a}(\theta' + \theta, \boldsymbol{\rho}) = \mathbf{a}(\theta', \boldsymbol{\rho}) + \theta \frac{\partial \mathbf{a}(\theta, \boldsymbol{\rho})}{\partial \theta}\bigg|_{\theta=\theta'}. \tag{7}$$

Here, we have used θ to denote the grid mismatch between the nearest DOA grid point and the DOA of the corresponding multipath, rather than the multipath DOA as in the structured channel.

Similarly, the uniform delay grid is the set $\tau' \in \{0, \frac{(L-1)T_s}{N_d}, \ldots, \frac{(L-1)(N_d-1)T_s}{N_d}\}$ and a Taylor interpolation similar to (7) can be performed to obtain

$$g(kT_s - \tau' - \tau) = g(kT_s - \tau') + \tau \frac{\partial g(kT_s - \tau)}{\partial \tau}\bigg|_{\tau=\tau'}. \tag{8}$$

Here, we have used τ to denote the grid mismatch between the nearest delay grid point and the delay of the corresponding multipath. Having taken into account the dictionary errors, the

discrete-time channel $\mathbf{h}[k]$ for the dictionary-based channel model is given by

$$\mathbf{h}[k] = (\mathbf{G}_D(kT_s, \boldsymbol{\tau}) \otimes \mathbf{A}_D(\boldsymbol{\theta}, \boldsymbol{\rho}))\,\boldsymbol{\gamma}, \tag{9}$$

where $\mathbf{A}_D(\boldsymbol{\theta}, \boldsymbol{\rho}) \in \mathbb{C}^{M \times N_a}$ and $\mathbf{G}_D(kT_s, \boldsymbol{\tau}) \in \mathbb{R}^{P \times N_d}$ are the angular and delay domain perturbed dictionary matrices, respectively, and $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are the vectors of grid mismatch errors to be estimated. The columns of $\mathbf{G}_D(kT_s, \boldsymbol{\tau})$ are of the form $[g(kT_s - \tau' - \tau), \ldots, g((k + \frac{P-1}{P})T_s - \tau' - \tau)]^T$, and the columns of $\mathbf{A}_D(\boldsymbol{\theta}, \boldsymbol{\rho})$ are of the form $\mathbf{a}(\theta' + \theta, \boldsymbol{\rho})$. The complex unknown vector $\boldsymbol{\gamma}$ is a sparse $N_a N_d \times 1$ vector that carries the path gains from the corresponding DOAs and delays in the dictionary. The sparse formulation implies that $\boldsymbol{\gamma}$ only has $R \ll N_a N_d$ non-zero elements. In our analysis of the CRB for this model, and unlike the structured and unstructured models described above, we assume that the locations of the non-zero elements in $\boldsymbol{\gamma}$, and therefore the nearest angle and delay grid points are known a priori. The benefit of this a priori information will depend on the resolution of the grid, and the ability of dictionary-based methods to correctly identify the correct grid points. The assumption should be a reasonable one for dictionaries whose grids are not too finely spaced.

D. System Model

Gathering the received signals at the M antennas from (2) in $\mathbf{y}$ and the noise in $\mathbf{w}$,

$$\mathbf{y}\left[n + \frac{p}{P}\right] = \sqrt{P_t} \sum_{l=0}^{L-1} \mathbf{h}\left[l + \frac{p}{P}\right] s[n-l] + \mathbf{w}\left[n + \frac{p}{P}\right],$$

where $\mathbf{h}\left[l + \frac{p}{P}\right]$ corresponds to (6) for the structured channel, (3) for the unstructured channel and (9) for the dictionary based channel model. We collect samples from N source symbol periods, where the coherence time of the channel is greater than NT_s, and sample the received signal at each antenna at P times the symbol rate. We collect the $MP \times N$ spatial and temporal samples of the received signal in the matrix $\mathbf{Y}$ to get

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}[0] & \mathbf{y}[1] & \ldots & \mathbf{y}[N-1] \\ \mathbf{y}[\frac{1}{P}] & \mathbf{y}[1+\frac{1}{P}] & \ldots & \mathbf{y}[N-1+\frac{1}{P}] \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}[\frac{P-1}{P}] & \mathbf{y}[1+\frac{P-1}{P}] & \ldots & \mathbf{y}[N-1+\frac{P-1}{P}] \end{bmatrix} \tag{10}$$

$$= \sqrt{P_t} \underbrace{\begin{bmatrix} \mathbf{h}[0] & \mathbf{h}[1] & \ldots & \mathbf{h}[L-1] \\ \mathbf{h}[\frac{1}{P}] & \mathbf{h}[1+\frac{1}{P}] & \ldots & \mathbf{h}[L-1+\frac{1}{P}] \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{h}[\frac{P-1}{P}] & \mathbf{h}[1+\frac{P-1}{P}] & \ldots & \mathbf{h}[L-1+\frac{P-1}{P}] \end{bmatrix}}_{\mathbf{H} \in \mathbb{C}^{MP \times L}}$$

$$\underbrace{\begin{bmatrix} s[0] & s[1] & \ldots & s[N-1] \\ s[-1] & s[0] & \ldots & s[N-2] \\ \vdots & \vdots & \vdots & \vdots \\ s[-L+1] & s[-L+2] & \ldots & s[N-L] \end{bmatrix}}_{\tilde{\mathbf{S}} \in \mathbb{C}^{L \times N}} + \mathbf{W}.$$

Vectorizing (10) and taking the real and imaginary parts separately, we have

$$\mathbf{y} = \begin{bmatrix} \mathrm{Re}\,(\mathrm{vec}(\mathbf{Y})) \\ \mathrm{Im}\,(\mathrm{vec}(\mathbf{Y})) \end{bmatrix} = \mathbf{S}\mathbf{h} + \mathbf{w}, \tag{11}$$

where,

$$\mathbf{S} = \begin{bmatrix} \mathrm{Re}\left(\tilde{\mathbf{S}}^{\mathrm{T}} \otimes \sqrt{\mathrm{P_t}}\mathbf{I}_{\mathrm{MP}}\right) & -\mathrm{Im}\left(\tilde{\mathbf{S}}^{T} \otimes \sqrt{P_t}\mathbf{I}_{MP}\right) \\ \mathrm{Im}\left(\tilde{\mathbf{S}}^{T} \otimes \sqrt{P_t}\mathbf{I}_{MP}\right) & \mathrm{Re}\left(\tilde{\mathbf{S}}^{T} \otimes \sqrt{P_t}\mathbf{I}_{MP}\right) \end{bmatrix} \in \mathbb{R}^{2MNP \times 2LMP},$$

$$\mathbf{h} = \begin{bmatrix} \mathrm{Re}\,(\mathrm{vec}(\mathbf{H})) \\ \mathrm{Im}\,(\mathrm{vec}(\mathbf{H})) \end{bmatrix} \in \mathbb{R}^{2LMP \times 1},$$

$$\mathbf{w} = \begin{bmatrix} \mathrm{Re}\,(\mathrm{vec}(\mathbf{W})) \\ \mathrm{Im}\,(\mathrm{vec}(\mathbf{W})) \end{bmatrix} \in \mathbb{R}^{2MNP \times 1},$$

We note that $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I})$. The per-antenna SNR at the receiver is defined as

$$\mathrm{SNR} = \frac{P_t}{\sigma^2} \sum_{l=0}^{L-1} \sum_{p=0}^{P-1} \mathbb{E}\left[|h_m[l + p/P]|^2\right].$$

As in [19], we define the quantization operation as $\mathcal{Q}(\cdot) = \frac{1}{\sqrt{2}}(\mathrm{sign}(\cdot)))$, where the sign operation is performed separately for the real and imaginary parts, so the quantized output $\mathbf{x}$ is

$$\mathbf{x} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{S}\mathbf{h} + \mathbf{w}).$$

A distinction between the unstructured, structured and dictionary based models is that in the structured model and the dictionary based model, the channel parameters are estimated, whereas in the unstructured model, the channel $\mathbf{h}$ is estimated.

III. Cramér- Rao Bound

In this section, we derive the CRB for parameters of the spatially structured and unstructured channel models. For the spatially structured channel model, the parameter vector consists of both deterministic and stochastic components. The deterministic components are, namely, the DOAs, path delays and complex path gains. When considering sensor position perturbations only, we have $\boldsymbol{\rho} \in \mathbb{R}^{M-1 \times 1}$, and for pattern perturbations only, we have $\boldsymbol{\rho} = \left[\mathrm{Re}(\rho_1), \mathrm{Re}(\rho_2), \ldots, \mathrm{Re}(\rho_{\mathrm{M}}), \mathrm{Im}(\rho_1), \mathrm{Im}(\rho_2), \ldots, \mathrm{Im}(\rho_{\mathrm{M}})\right] \in \mathbb{R}^{2M \times 1}$. The full list of parameters is $\boldsymbol{\Theta} = \left[\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}^{\mathrm{Re}}, \boldsymbol{\gamma}^{\mathrm{Im}}, \boldsymbol{\rho}\right]$ under the structured and dictionary-based channel models, and $\boldsymbol{\Theta} = \left[\boldsymbol{\beta}^{\mathrm{Re}}, \boldsymbol{\beta}^{\mathrm{Im}}\right]$ for the unstructured model. Assuming independent observations, the log-likelihood

for the spatially structured and unstructured models, $l(\mathrm{x};\boldsymbol{\Theta})$, can be derived in a manner similar to [19], to obtain

$$l(\mathrm{x};\boldsymbol{\Theta}) = \sum_{k=1}^{2MNP} \ln \Phi\left(\frac{2}{\sigma} x_k u_k\right),$$

where x_k is the kth element of x, $u_k = \left(\mathrm{s}^{(k,)}\right)^T \mathrm{h}$, $\mathrm{s}^{(k,)}$ is the kth row of S, and $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

Assuming that the regularity condition of the log-likelihood holds, the FIM has the following form [37]:

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P,$$

where $\mathbf{J}_D$ and $\mathbf{J}_P$ are the information matrices obtained from the data and the a priori information, respectively. The (i,j)th elements of $\mathbf{J}_D$ and $\mathbf{J}_P$ are given by

$$[\mathbf{J}_D]_{i,j} = -\mathbb{E}_{\mathrm{x},\boldsymbol{\rho}}\left[\frac{\partial^2}{\partial\Theta_i\partial\Theta_j} l(\mathrm{x};\boldsymbol{\Theta})\right]$$

$$[\mathbf{J}_P]_{i,j} = \mathbb{E}_{\boldsymbol{\rho}}\left[\frac{\partial^2}{\partial\Theta_i\partial\Theta_j}(\boldsymbol{\rho}-\boldsymbol{\rho}_0)^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\rho}-\boldsymbol{\rho}_0)\right],$$

where $\boldsymbol{\Omega}$ is the covariance matrix of $\boldsymbol{\rho}$. Since only $\boldsymbol{\rho}$ is random, the matrix $\mathbf{J}_P$ is given by

$$\mathbf{J}_P = \mathrm{blkdiag}\{\mathrm{diag}(\mathbf{0}_D), \boldsymbol{\Omega}^{-1}\},$$

where $\mathrm{blkdiag}\{\cdot\}$ is a block-diagonal matrix where the arguments form the diagonal blocks, and $D = 4R$, $D = 2MR$ and $D = 4R$, for the structured, unstructured and dictionary based channel models respectively.

The expectation with respect to the joint distribution of x and $\boldsymbol{\rho}$ in $\mathbf{J}_D$ is difficult to compute. Instead, we follow the approach in [32], [34]. If the perturbations are small, $\mathbf{J}_D$ can be approximated to order $O(1)$ around $\boldsymbol{\rho}_0$, in which case

$$[\mathbf{J}_D]_{i,j} \approx -\mathbb{E}_{\mathrm{x}}\left[\frac{\partial^2}{\partial\Theta_i\partial\Theta_j} l(\mathrm{x};\boldsymbol{\Theta})\right]\Bigg|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}. \tag{12}$$

Then, the CRB for the ith parameter of an unbiased estimator with $\mathbb{E}\left[\hat{\boldsymbol{\Theta}}\right] = \boldsymbol{\Theta}$ is given by the (i,i) element of the inverse of the FIM, where the FIM is computed at the "true" values of $\boldsymbol{\Theta}$ and $\boldsymbol{\rho}$. That is,

$$\mathrm{var}(\hat{\Theta}_i) \geq \left[\mathbf{J}^{-1}(\boldsymbol{\Theta})\right]_{i,i}.$$

The approximation in (12) is sufficiently accurate for values of $\boldsymbol{\rho}$ commonly encountered in real calibrated systems (see [34] for more details).

A. Unstructured Channel

For the unstructured channel model, the FIM is given by

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Re}}} & \mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Re}}\boldsymbol{\beta}^{\mathrm{Im}}} \\ \mathbf{J}^T_{\boldsymbol{\beta}^{\mathrm{Re}}\boldsymbol{\beta}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Im}}} \end{bmatrix} \in \mathbb{R}^{2MR\times 2MR}, \tag{13}$$

with

$$\mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Re}}} = \mathbb{E}\left[\nabla_{\boldsymbol{\beta}^{\mathrm{Re}}} l(\mathrm{x};\boldsymbol{\Theta})(\nabla_{\boldsymbol{\beta}^{\mathrm{Re}}} l(\mathrm{x};\boldsymbol{\Theta}))^T\right] \in \mathbb{R}^{MR\times MR},$$

$$\mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Im}}} = \mathbb{E}\left[\nabla_{\boldsymbol{\beta}^{\mathrm{Im}}} l(\mathrm{x};\boldsymbol{\Theta})(\nabla_{\boldsymbol{\beta}^{\mathrm{Im}}} l(\mathrm{x};\boldsymbol{\Theta}))^T\right] \in \mathbb{R}^{MR\times MR},$$

$$\mathbf{J}_{\boldsymbol{\beta}^{\mathrm{Re}}\boldsymbol{\beta}^{\mathrm{Im}}} = \mathbb{E}\left[\nabla_{\boldsymbol{\beta}^{\mathrm{Re}}} l(\mathrm{x};\boldsymbol{\Theta})(\nabla_{\boldsymbol{\beta}^{\mathrm{Im}}} l(\mathrm{x};\boldsymbol{\Theta}))^T\right] \in \mathbb{R}^{MR\times MR}. \tag{14}$$

The expressions for the Jacobians as follows:

$$\mathbf{D}_{\boldsymbol{\beta}^{\mathrm{Re}}} = \left[\frac{\partial \mathrm{h}}{\partial\beta^{\mathrm{Re}}_{1,1}}, \frac{\partial \mathrm{h}}{\partial\beta^{\mathrm{Re}}_{1,2}}, \ldots, \frac{\partial \mathrm{h}}{\partial\beta^{\mathrm{Re}}_{R,M}}\right],$$

$$\frac{\partial \mathrm{h}[k]}{\partial\beta^{\mathrm{Re}}_{r,m}} = g(kT_s - (\tau_0 + (r-1)\Delta')),$$

$$\mathbf{D}_{\boldsymbol{\beta}^{\mathrm{Im}}} = j\mathbf{D}_{\boldsymbol{\beta}^{\mathrm{Re}}},$$

where $\mathbf{e}_m$ is the unit vector with a 1 at the mth index.

B. Structured and Dictionary Based Channels

The regularity condition for the pdf of x can be easily verified. For the structured and the grid mismatched dictionary based channel models, $\mathbf{J}_D$ is block-partitioned and symmetric and is given by

$$\mathbf{J}_D = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta}} & \mathbf{J}_{\boldsymbol{\theta\tau}} & \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\gamma}^{\mathrm{Re}}} & \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\theta\rho}} \\ \mathbf{J}^T_{\boldsymbol{\theta\tau}} & \mathbf{J}_{\boldsymbol{\tau}} & \mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\gamma}^{\mathrm{Re}}} & \mathbf{J}_{\boldsymbol{\tau}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\tau\rho}} \\ \mathbf{J}^T_{\boldsymbol{\theta}\boldsymbol{\gamma}^{\mathrm{Re}}} & \mathbf{J}^T_{\boldsymbol{\tau}\boldsymbol{\gamma}^{\mathrm{Re}}} & \mathbf{J}_{\boldsymbol{\gamma}^{\mathrm{Re}}} & \mathbf{J}_{\boldsymbol{\gamma}^{\mathrm{Re}}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\gamma}^{\mathrm{Re}}\boldsymbol{\rho}} \\ \mathbf{J}^T_{\boldsymbol{\theta}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}^T_{\boldsymbol{\tau}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}^T_{\boldsymbol{\gamma}^{\mathrm{Re}}\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\gamma}^{\mathrm{Im}}} & \mathbf{J}_{\boldsymbol{\gamma}^{\mathrm{Im}}\boldsymbol{\rho}} \\ \mathbf{J}^T_{\boldsymbol{\theta\rho}} & \mathbf{J}^T_{\boldsymbol{\tau\rho}} & \mathbf{J}^T_{\boldsymbol{\gamma}^{\mathrm{Re}}\boldsymbol{\rho}} & \mathbf{J}^T_{\boldsymbol{\gamma}^{\mathrm{Im}}\boldsymbol{\rho}} & \mathbf{J}_{\boldsymbol{\rho}} \end{bmatrix}, \tag{15}$$

where the expression for each matrix block is provided in Appendix B. Let $u_k = (\mathrm{s}^{(k)})^T\mathrm{h}$, $(\cdot)^{(i,)}$ be the ith row of the argument and $(\cdot)^{(,i)}$ be the ith column of the argument. Denoting the Jacobian of h with respect to θ by $\mathbf{D}_{\boldsymbol{\theta}}$, we show in Appendix A that $\mathbf{J}_{\boldsymbol{\theta}}$ can be written as

$$[\mathbf{J}_{\boldsymbol{\theta}}]_{i,j} = \mathbb{E}\left[\frac{\partial l(\mathrm{x};\boldsymbol{\Theta})}{\partial\theta_i}\frac{\partial l(\mathrm{x};\boldsymbol{\Theta})}{\partial\theta_j}\right] = \left(\mathbf{D}^{(,i)}_{\boldsymbol{\theta}}\right)^T \mathbf{S}^T\ \boldsymbol{\Phi}_D\ \mathbf{S}\ \mathbf{D}^{(,j)}_{\boldsymbol{\theta}}, \tag{16}$$

where $\boldsymbol{\Phi}_D$ is a diagonal matrix with $[\boldsymbol{\Phi}_D]_{k,k} = \frac{2}{\sigma^2}\frac{(\Phi'(\frac{\sqrt{2}}{\sigma}u_k))^2}{\Phi(\frac{\sqrt{2}}{\sigma}u_k)(1-\Phi(\frac{\sqrt{2}}{\sigma}u_k))}$. In a similar manner, we can derive expressions for the other sub-matrices of (15) in terms of $\mathbf{D}_{\boldsymbol{\theta}}$, $\mathbf{D}_{\boldsymbol{\tau}}$, $\mathbf{D}_{\boldsymbol{\gamma}^{\mathrm{Re}}}$ and $\mathbf{D}_{\boldsymbol{\gamma}^{\mathrm{Im}}}$. For instance, the sub-matrix $\mathbf{J}_{\boldsymbol{\theta\tau}}$ is given by

$$\mathbf{J}_{\boldsymbol{\theta\tau}} = (\mathbf{D}_{\boldsymbol{\theta}})^T\mathbf{S}^T\ \boldsymbol{\Phi}_D\ \mathbf{S}\ \mathbf{D}_{\boldsymbol{\tau}}.$$

The expressions for the Jacobian matrices are derived in Appendix B, and are evaluated at the nominal perturbation value $\boldsymbol{\rho}_0$.

IV. Simulation Results

We consider a ULA with half-wavelength spacing between antenna elements, i.e. $\delta = 0.5\lambda$, $\lambda = c/f_c$ with the carrier frequency $f_c = 60$ GHz. The source signal $s[n], n = -L+1, \ldots, N-1$, is taken to be a randomly generated quadrature phase shift keying (QPSK) sequence. We choose r to be the

raised cosine filter with roll-off factor $\alpha = 0.8$. For wideband systems, the channel bandwidths can be as high as 1 GHz [38]. For this reason, in our simulations the null-to-null bandwidth is set to 1 GHz, the symbol duration $T_s = 1$ ns, and the received signal is oversampled at the sampling frequency $f_s = P \times \text{null} - \text{to} - \text{null bandwidth}$. The oversampling factor is $P = 3$ in all plots except Fig. 7.

A comparison of the CRBs of the different channel models is accomplished by translating the CRB of the parameters $\boldsymbol{\Theta}$ into the CRB of the channel by the following transformation [39]

$$\text{CRB}(\mathbf{h}) \succeq \nabla_{\boldsymbol{\Theta}}\mathbf{h} \;\; \mathbf{J}^{-1} \;\; \nabla_{\boldsymbol{\Theta}}^{T}\mathbf{h},$$

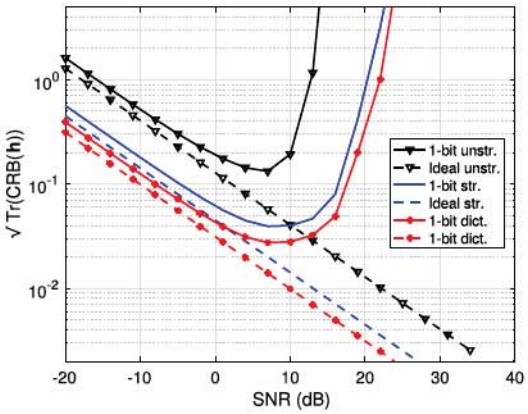
where $\nabla_{\boldsymbol{\Theta}}(\cdot)$ is the Jacobian with respect to $\boldsymbol{\Theta}$. The parameters for the structured channel model are selected as follows. The angles of arrival of the multipaths are assumed to be distributed independently and uniformly in $[0, 2\pi)$. The R complex gains of the multipaths γ_r are a. The path delays $\tau_r, r = 1, 2, \ldots, R$, are chosen to be integer multiples of the sampling interval $1/f_s$. For the first few simulations, we ignore the effect of array perturbations. Similarly, for the unstructured channel, the elements of the spatial signatures $\beta_{m,r}, m = 1, 2, \ldots, M,\ r = 1, 2, \ldots, R$, are also assumed to be generated independently and identically from a complex normal distribution with unit variance, and $\Delta' = 1/f_s$.

Finally, for the dictionary matrix in (9), we have selected $N_a = 2M$ and $N_d = 2L - 1$. The angular and delay domain mismatch errors, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$, are generated independently and uniformly in $[-\frac{\pi}{2N_a}, \frac{\pi}{2N_a}]$ and $[-\frac{(L-1)T_s}{2N_d}, \frac{(L-1)T_s}{2N_d}]$, respectively. The square root of the trace of the CRB matrix for each of the parameters is computed and averaged over 50 realizations of the channel.

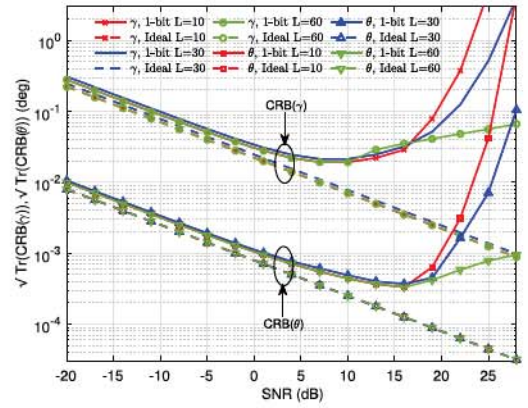
A. *Performance vs. SNR*

Fig. 2(a) shows the square root of the trace of the channel estimate CRB as a function of the SNR for a single line-of-sight path ($R = 1$) to the receiver and a single tap channel ($L = 1$) with $M = 32$ receive antennas and pilot length $N = 20$. This is the frequency-flat fading case with the unstructured model corresponding to the Rayleigh fading case.

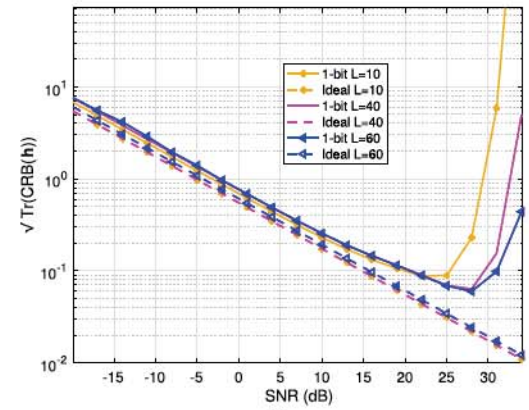
The increase in the bound at high SNR in Fig. 2(a) is commonly observed in one-bit sampled systems, and is due to the loss of information in the channel gain as the amplitude of the received signal grows, and the FIM becomes rank deficient. This illustrates the benefit behind dithering (stochastic resonance), where adding noise (lowering the SNR) can improve estimation performance with coarsely sampled data. We also observe that at low-moderate SNRs, which is common in mmWave, the gap due to the quantization error between the unquantized and one-bit bounds is about 1.96 dB as expected [40]. At higher SNRs, however, the gap is much higher since the FIM becomes increasingly ill-conditioned due to reduced identifiability of the channel gains. The CRBs for the unstructured one-bit and unquantized channels are higher than the other CRBs since they reflect the estimation error for a total of 64 real-valued parameters compared to only 3 structured and dictionary based channel parameters



(a)



(b)



(c)

Fig. 2. (a) $\text{CRB}(\mathbf{h})$ with 32 antennas and a single LOS path to the receiver. The dashed lines correspond to the CRBs of unquantized (ideal) systems whereas solid lines are used to plot the CRBs of one-bit cases. (b) $\text{CRB}(\boldsymbol{\theta})$ and $\text{CRB}(\boldsymbol{\gamma})$ of the structured channel with 32 antennas and a single LOS path to the receiver. The CRB is plotted for different values of the channel delay-tap length. (c) $\text{CRB}(\mathbf{h})$ of unstructured channels with 32 antennas for $L = 10$, $L = 40$ and $L = 60$.

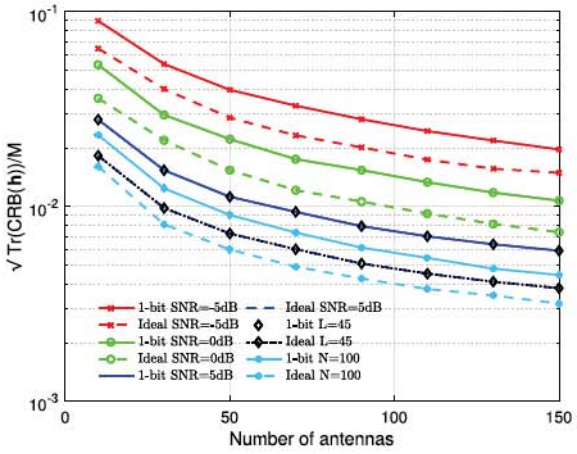
(2 real-valued path gains and one DOA). Furthermore, the dictionary based CRB is lower than the unstructured and structured counterparts. This is because approximate knowledge of the DOAs and delays is available and the CRB reflects the estimation error for the grid mismatch. Fig. 2(a) provides a very interesting observation concerning the structured vs. unstructured models

and one-bit quantization. Note that the channel estimation lower bound for the structured model under one-bit quantization is significantly lower than the bound for the unstructured model without quantization (perfect resolution), provided that the SNR is below 10 dB, which would be the typical case for mmWave systems. Thus, the gain in parsimony provided by the structured model more than compensates for the loss due to the coarse quantization, and this provides a strong argument for the use of the structured model when the propagation environment is relatively simple.

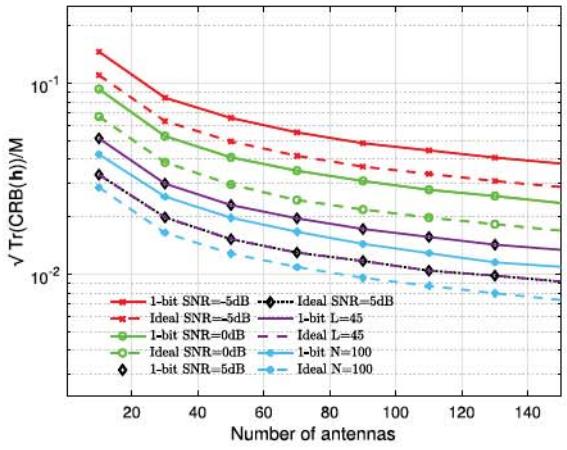
When the channel is strongly frequency selective (large delay spread L), the received signal power is spread evenly over time, the quantization noise becomes circularly symmetric, and the resulting amplitude distortion caused by the quantization is reduced (see [41] for details). This is evident in the CRB results shown in Figs. 2(b) and (c). Fig. 2(b) illustrates the CRBs of γ and θ for different channel lengths $L = 10$, $L = 30$ and $L = 60$. The number of pilots is set at $N = 80$ in all cases. It is seen that as the number of channel taps increases, the degradation in performance due to quantization at higher SNRs also decreases. For the same number of pilots, no effect on the CRB is observed at low-to-moderate SNRs. The performance of the ideal system is insensitive to L, and therefore, the dashed curves corresponding to the ideal system overlap. A final observation regarding Fig. 2(b) is that the CRB result for θ is not inconsistent with the CRB results in [30] and [31], but rather our definition of SNR is different from theirs and, as a result, we are plotting the CRB over a wider SNR range. A similar effect is seen Fig. 2(c) where the CRB of the unstructured channel is plotted as a function of the SNR for different values of L. Here, a single multipath is considered with the number of pilots fixed to $N = 80$. Since the size of the β grows linearly with R and L, the CRB is normalized by MR. Thus, the effect of quantization error from one-bit ADCs in longer frequency-selective channels is less severe at high SNRs.

B. Performance vs Number of Antennas

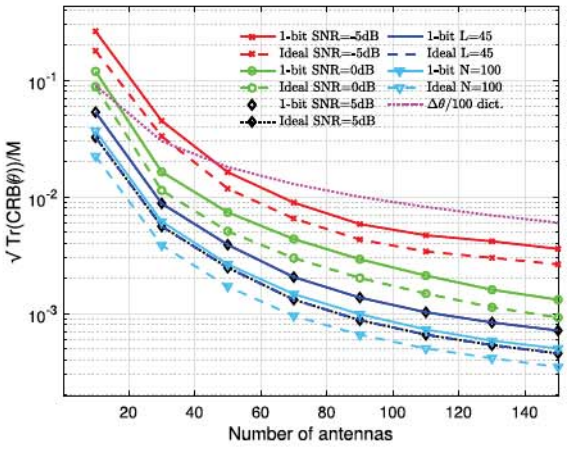
The effect of increasing the number of antennas on the CRB is studied next. In Fig. 3(a), we plot the CRB normalized by the number of receive antennas ($\sqrt{\mathrm{Tr}(\mathrm{CRB}(\mathbf{h}))/M}$) for the structured channel case, for varying values of M at different SNRs and delay tap lengths. The number of pilots is fixed to $N = 60$ and the number of multipaths is $R = 4$. As expected, increasing the number of antennas reduces the average estimation error in each channel coefficient. In [41], it was found that, for a given SNR, the number of antennas in one-bit systems with an unstructured channel model should increase by approximately 2.5 times to meet the achievable rate of an otherwise equivalent ideal unquantized system. At $\mathrm{SNR} = -5$ dB, fewer than twice the number of antennas are required for the one-bit system to achieve the same channel estimation performance as the unquantized system. For example, the one-bit CRB at $M = 80$ is equal to the unquantized CRB at $M = 50$. At 5 dB SNR, it is seen that the number of antennas should be increased by slightly more than a factor of two; the one-bit CRB at $M = 110$ is equal to the unquantized CRB at $M = 50$. At 5 dB SNR,



(a)



(b)



(c)

Fig. 3. (a) $\mathrm{CRB}(\mathbf{h})$ of the structured channels as a function of the number of receive antennas. For the case with $L = 45$, the SNR is 5 dB, and for the $N = 100$ case- $L = 15$ and $\mathrm{SNR} = 5$ dB. For the other plots, $L = 15$. (b). $\mathrm{CRB}(\mathbf{h})$ of the unstructured channels as a function of the number of receive antennas for different values of L and SNRs. The constrained CRB is also illustrated for $\mathrm{SNR} = 5$ dB and $L = 15$. (c) $\mathrm{CRB}(\theta)$ of the structured channels as a function of the number of receive antennas. As in (a), for the case of $L = 45$, the SNR is fixed to 5 dB. The final plot with $N = 100$ corresponds to $L = 15$ and $\mathrm{SNR} = 5$ dB. The dotted line indicates the scaled angular resolution of the dictionary based channel model.

even the highly frequency-selective channel ($L = 45$) has the same CRB as a channel with $L = 15$ as seen by the overlapping blue and black curves. At higher SNRs, it can be expected that one-bit systems are advantageous for highly frequency-selective channels.

We also plot the CRB as a function of M for unstructured channels in Fig. 3(b). Since the number of parameters for the unstructured channel scales with M (more specifically, the number of parameters is $2MR$), the CRB is normalized by MR, and $N = 80$. As expected, increasing the number of antennas reduces the average estimation error in each channel coefficient. At low SNRs, the number of antennas needed by one-bit systems has to again increase by almost two times to cope with the loss due to quantization distortion. For instance, at $\mathrm{SNR} = 0$ dB, the one-bit system with $M = 100$ achieves the same CRB as the unquantized system with $M = 50$. As the SNR is increased, it is found that the number of antennas should be increased even more to meet the CRB of the unquantized system.

Fig. 3(c) shows the CRB for θ as a function of the number of antennas. The figure also shows the grid spacing of the dictionary based channel model scaled appropriately. At $\mathrm{SNR} = -5$ dB, the number of antennas should be increased by less than 1.5 times to achieve the unquantized CRB, significantly less than for the channel itself in Fig. 3(a). The one-bit CRB at $M = 110$ is equal to the unquantized CRB at $M = 90$. At 5 dB, the factor increases to 1.5, still fewer than for the channel. Upon examining equation (V), it can be seen that the Jacobian scales linearly with M, and thus, the CRB reduces with a factor of M^2. Comparing Figs. 3(a) and (c) with (b), it is seen that, to match the CRB of the unquantized systems, the unstructured models require almost $2.5 - 3$ times the number of antennas, whereas the structured models typically require much less than twice the number of antennas. Thus, this example illustrates another advantage of using the structured model together with one-bit quantization: fewer additional antennas are needed to achieve the same performance as an ideal unquantized system than in the case of an unstructured channel model. Note that in this example, the structured model provides DOA estimates that are well beyond the resolution of the grid used in the dictionary-based model, especially for larger array sizes.

C. Effect of Array Calibration Errors

We now consider the effect of array perturbations on the CRB. First, we assume a gain pattern perturbation only with $\mathbf{\Omega} = \sigma_{\boldsymbol{\rho}}^2\, \mathbf{I}_M$. Here, $M = 64$, the number of pilots is $N = 50$, the channel length $L = 5$ and the number of multipaths $R = 5$. Fig. 4 shows the CRB of the structured channel for a standard deviation of $\boldsymbol{\rho}$, $\sigma_{\boldsymbol{\rho}} = 0.1$. At $\sigma_{\boldsymbol{\rho}} = 0.1$, the structured model is an appropriate model to assess the channel performance at lower SNRs but the CRB degrades faster. On the other hand, the unstructured channel is a more suitable model when the perturbations are large and unknown. Fig. 4 amplifies the result of Fig. 2(a), showing that the structured model with one-bit quantization achieves better channel estimation performance than the ideal unstructured model up to about 10 dB SNR even when there are array perturbations at the level of $\sigma_{\boldsymbol{\rho}} = 0.1$. Thus,

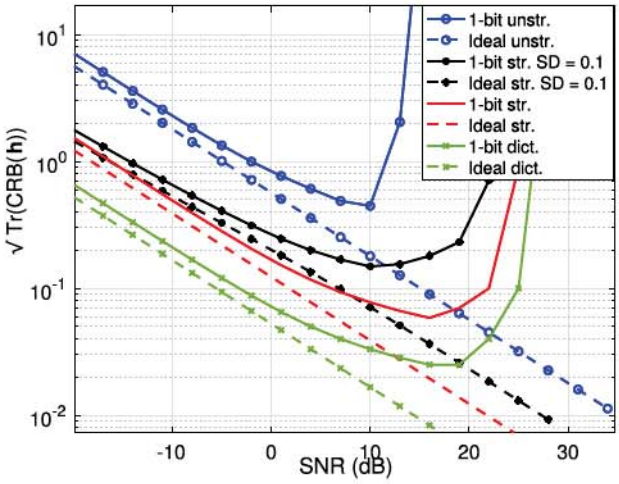


Fig. 4. $\mathrm{CRB}(\mathbf{h})$ as a function of the SNR for varying standard deviations (SDs) of the pattern perturbation.

for low-to-moderate SNRs where mmWave systems operate, structured models that have even imprecise calibration and use only one-bit quantization perform better than using unstructured models with perfect quantization.

We also study the effect of position perturbation on the CRB. The CRBs of γ and θ as a function of the standard deviation, $\sigma_{\boldsymbol{\rho}}$, are plotted in Fig. 5(a) and (b), respectively, for two different values of the SNR, 5 dB and 10 dB. It is seen in Fig. 5(a) that for small values of $\sigma_{\boldsymbol{\rho}}$, increasing the SNR is advantageous since the CRB is lower at $\mathrm{SNR} = 10$ dB than at $\mathrm{SNR} = 5$ dB. However, more interestingly, increasing the perturbation causes the CRB to degrade rapidly at $\mathrm{SNR} = 10$ dB due to the near-singularity of the FIM. The same effect was observed in Fig. 4. Similarly, the CRB of θ also degrades for $\mathrm{SNR} = 10$ dB with an increasing perturbation in Fig. 5(b). The gap between the one-bit and unquantized CRBs is greater for $\mathrm{SNR} = 10$ dB and it becomes more pronounced upon increasing the perturbation.

D. Effect of Bandwidth and Oversampling

Fig. 6(a) and (b) show the CRBs of γ and θ, and τ respectively, as a function of the signal bandwidth for $M = 32$ and $M = 64$ in the structured channel model. The delay spread is fixed to $0.2\ \mu$s and the SNR is 0 dB. The number of pilots is $N = 60$ and it is assumed that $R = 4$ multipaths are present. The null-to-null bandwidth is varied from 1 MHz to 1 GHz and the oversampling factor is kept at $P = 3$. Since the delay spread is kept constant, the channel length increases with the bandwidth, making the channel more frequency-selective. For both γ and θ, increasing the BW from 1 MHz to 1 GHz decreases the CRB by almost an order of magnitude. Therefore, for a given number of estimation parameters, a broadband system exhibits a lower estimation error for structured channel parameters than a narrowband system. Similarly, increasing the bandwidth also provides better resolution for the estimation of the delay parameters τ as seen in Fig. 6(b).

The effect of oversampling on the CRB is shown in Fig. 7. Here, the gap between the one-bit CRB and the unquantized

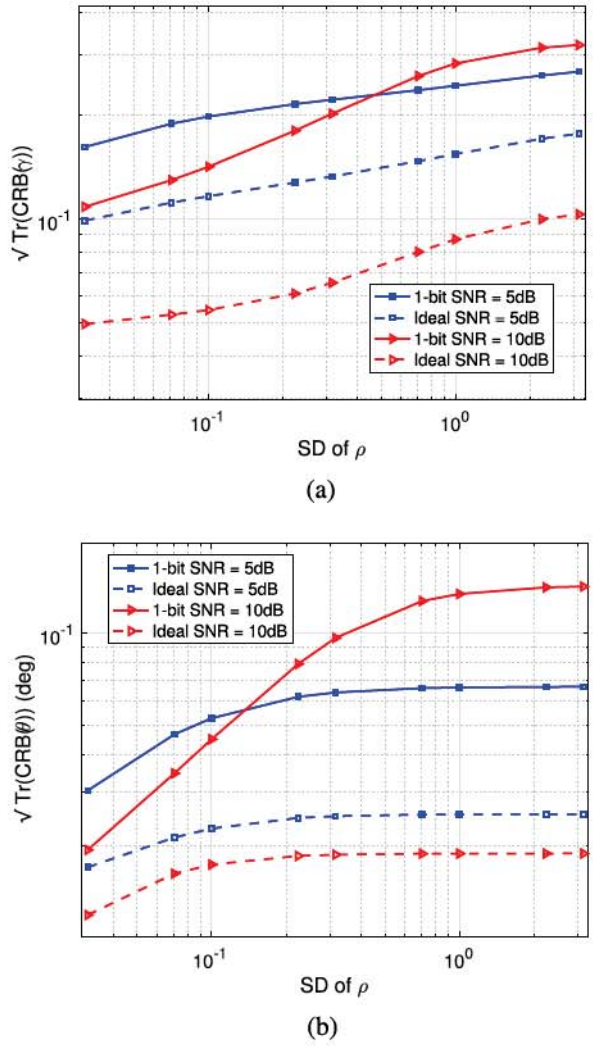


Fig. 5. (a) $\mathrm{CRB}(\gamma)$ as a function of the SD of ρ. (b) $\mathrm{CRB}(\theta)$ antennas as a function of the SD of ρ.

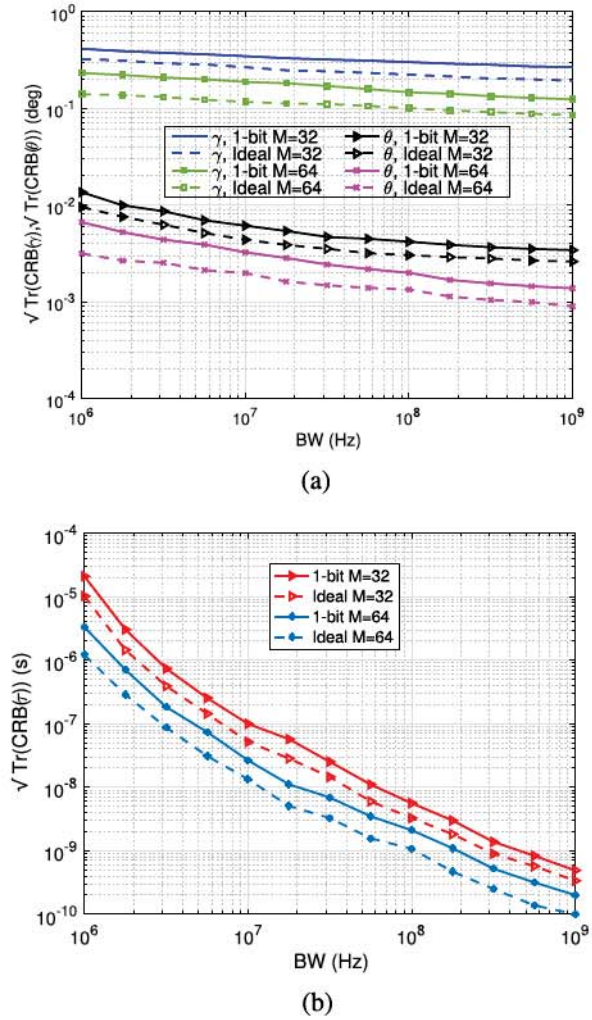


Fig. 6. (a) $\mathrm{CRB}(\gamma)$ and $\mathrm{CRB}(\theta)$ of the structured channels as a function of the system bandwidth for a fixed delay spread. (b) $\mathrm{CRB}(\tau)$ as a function of the system bandwidth.

CRB is plotted as a function of the oversampling factor. More specifically, we plot $\mathrm{Tr}[\mathrm{CRB}_{1-\mathrm{bit}}(\cdot)]/\mathrm{Tr}[\mathrm{CRB}_{\infty}(\cdot)]$ for τ, γ and θ. The number of antennas is kept fixed at 32 and the oversampling factor above the Nyquist rate is varied from 1 to 5. It is seen that performance loss from quantization is the least at $\mathrm{SNR} = -20$ dB and increases upon increasing the SNR. However, the effect of oversampling is to reduce the loss beyond the $2/\pi$ limit. At low SNRs, for example at -20 dB, increasing the oversampling factor does not have any effect and the plots for all parameters coincide. However, improvements would still be possible at low SNRs if the analog filter prior to sampling is optimized [42].

E. Effect of Path Separation and Number of Multipaths

A disadvantage of the structured channel model is that the FIM becomes ill-conditioned when two multipaths arrive with similar DOAs and path delays. Fig. 8 illustrates the CRB as a function of the fractional path delay difference. In this setup, $R = 2$ and

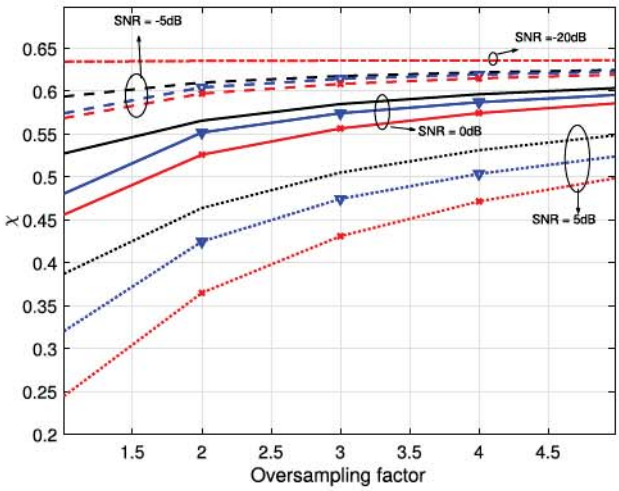


Fig. 7. $\chi_{\mathbf{p}} = \frac{\mathrm{Tr}[\mathrm{CRB}_{1-\mathrm{bit}}(\mathbf{p})]}{\mathrm{Tr}[\mathrm{CRB}_{\infty}(\mathbf{p})]}$ for $\mathbf{p} \in \{\gamma, \theta, \tau\}$ of structured channels as a function of the oversampling factor, P. The plots with the $\triangledown$ and $*$ markers represent parameters τ and γ, respectively. The plots for θ are unmarked. For $\mathrm{SNR} = -20$ dB, χ for all parameters are equal.

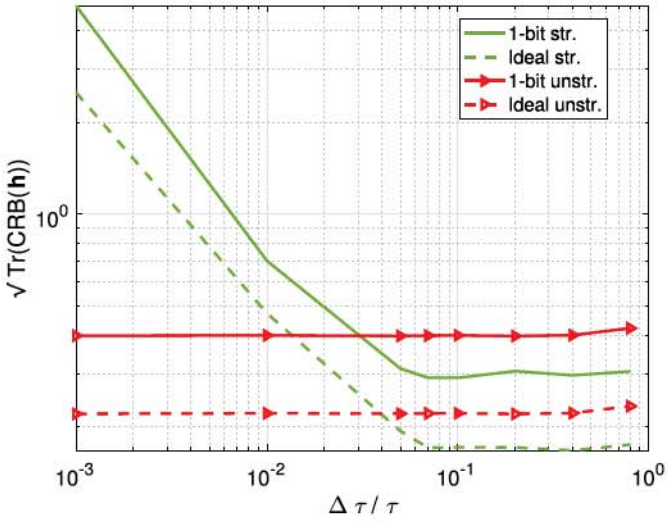
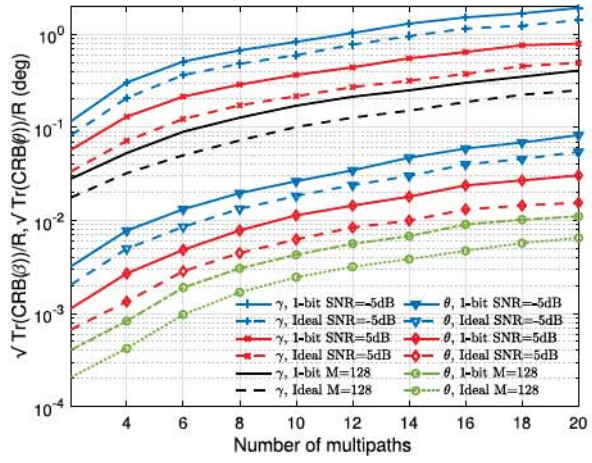
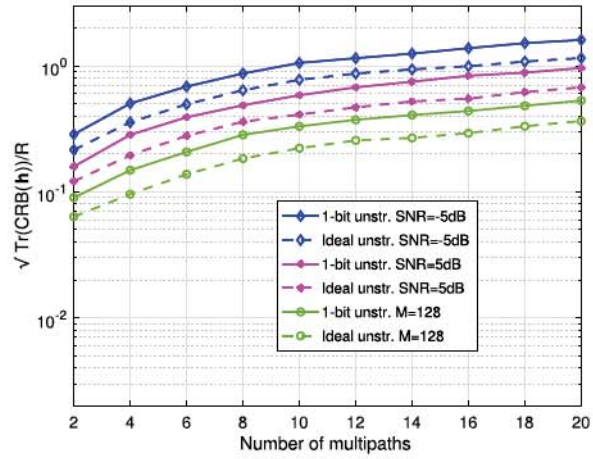


Fig. 8. The CRB of the channel as a function of the fractional separation between path delays. The SNR is 0 dB, $M = 64$, $N = 40$ and $L = 10$.



(a)



(b)

Fig. 9. (a) $\mathrm{CRB}(\gamma)$ and $\mathrm{CRB}(\theta)$ as a function of the number of multipaths. For the $M = 128$ plots, the SNR $= 5$ dB. (b) $\mathrm{CRB}(\mathbf{h})$ of unstructured channels as a function of the number of multipaths.

the two paths have different path gains but are configured to have the same DOA under the structured model. It is seen that as the path delays become closer, the structured model is not accurate and the unstructured model is better suited for resolving the two paths. When the paths are well separated in time, the structured channel model yields better results. For the given example, it is seen that the structured model only begins to break down when the difference in path delays is on the order of $3 - 4\%$ of the symbol period, but this illustrates the necessity of choosing the proper model order for the structured case, which is the key drawback of this method.

Fig. 9(a) and (b) show the normalized CRB $(\sqrt{\mathrm{Tr}(\mathrm{CRB}(\mathrm{h}))/R})$ as a function of the number of multipaths for $M = 64$, $N = 80$ and $L = 25$. It is seen that the CRB increases almost linearly on the log scale as the number of parameters also increases linearly with R.

V. Conclusion

In this work, we considered performance bounds for channel estimation in one-bit mmWave massive MIMO systems. Three channel models were considered, namely, (1) the structured channel where the scattering environment consists of a small number of scatterers characterized by distinct path delays, directions of arrival and path gains to the receiver, (2) the unstructured channel which is appropriate when the number of paths to the receiver is large and estimation of the path parameters is too difficult, and (3) the dictionary-based formulation with grid mismatch which is useful for sparse mmWave multipath channels. Closed-form expressions for the FIMs were derived when the receiver uses a uniform linear antenna array and a root-raised cosine filter. The derivation for the FIM for the structured and dictionary based channels also included the effect of array calibration errors, and the cases of gain pattern and antenna position perturbations were considered as specific examples. A number of numerical experiments were performed to evaluate the CRBs. A comparison in the CRBs of the structured, unstructured and dictionary-based channels as a function of the SNR indicated that the structured CRB is lower than that of the unstructured channel since fewer parameters contribute to the expression of the CRB. The effects of perturbation, bandwidth, the channel length, and the number of receive antennas on the CRBs of the one-bit quantized system were also considered. It was found that pertubation caused the CRB to degrade and approach that of an unstructured model at high SNRs. However, at low-to-moderate per-antenna SNRs, which is common in mmWave, the structured one-bit channel models have better channel estimation performance than the unquantized unstructured models. Furthermore, increasing the bandwidth and the oversampling factor caused the estimation error variance to decrease due to improved temporal resolution. It is also seen that to achieve the same error variance as an unquantized system, the one-bit structured system required significantly less than twice the number of antennas.

One of the principal observations of our results is that a significantly lower channel estimation error can be achieved by using a structured rather than an unstructured channel model, even

when the underlying array calibration is not precisely known. Thus, the extra computational cost required for the resulting non-linear optimization is often well worth the effort. The study of dependencies between the various system parameters can be useful in the design of channel estimation for mmWave massive MIMO with one-bit quantizers at the receiver.

Appendix A

For the structured channel,

$$
\begin{aligned}
[\mathbf{J}_{\boldsymbol{\theta}}]_{i,j} &= \mathbb{E}\left[\frac{\partial l(\mathbf{x};\boldsymbol{\Theta})}{\partial \theta_i}\frac{\partial l(\mathbf{x};\boldsymbol{\Theta})}{\partial \theta_j}\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{2MN}\sum_{l=1}^{2MN}\frac{4}{\sigma^2}\frac{\Phi'\left(\frac{2}{\sigma}x_k u_k\right)\Phi'\left(\frac{2}{\sigma}x_l u_l\right)}{\Phi\left(\frac{2}{\sigma}x_k u_k\right)\Phi\left(\frac{2}{\sigma}x_l u_l\right)}x_k x_l \right. \\
&\qquad \left. \times \left(\mathbf{D}_{\boldsymbol{\theta}}^{(,i)}\right)^T \mathbf{s}^{(k,)}\left(\mathbf{s}^{(l,)}\right)^T \mathbf{D}_{\boldsymbol{\theta}}^{(,j)}\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{2MN}\left(\frac{2x_k}{\sigma}\frac{\Phi'\left(\frac{2}{\sigma}x_k u_k\right)}{\Phi\left(\frac{2}{\sigma}x_k u_k\right)}\right)^2 \left(\mathbf{D}_{\boldsymbol{\theta}}^{(,i)}\right)^T \mathbf{s}^{(k,)}\left(\mathbf{s}^{(k,)}\right)^T \mathbf{D}_{\boldsymbol{\theta}}^{(,j)}\right. \\
&\qquad + \sum_{k=1}^{2MN}\sum_{l\neq k}\frac{4}{\sigma^2}\frac{\Phi'\left(\frac{2}{\sigma}x_k u_k\right)\Phi'\left(\frac{2}{\sigma}x_l u_l\right)}{\Phi\left(\frac{2}{\sigma}x_k u_k\right)\Phi\left(\frac{2}{\sigma}x_l u_l\right)}x_k x_l \\
&\qquad \left. \times \left(\mathbf{D}_{\boldsymbol{\theta}}^{(,i)}\right)^T \mathbf{s}^{(k,)}\left(\mathbf{s}^{(l,)}\right)^T \mathbf{D}_{\boldsymbol{\theta}}^{(,j)}\right] \\
&= \sum_{k=1}^{2MN}\frac{2}{\sigma^2}\frac{\left(\Phi'\left(\frac{\sqrt{2}}{\sigma}u_k\right)\right)^2}{\Phi\left(\frac{\sqrt{2}}{\sigma}u_k\right)\left(1-\Phi\left(\frac{\sqrt{2}}{\sigma}u_k\right)\right)} \\
&\quad \times \left(\mathbf{D}_{\boldsymbol{\theta}}^{(,i)}\right)^T \mathbf{s}^{(k,)}\left(\mathbf{s}^{(k,)}\right)^T \mathbf{D}_{\boldsymbol{\theta}}^{(,j)} \\
&= \left(\mathbf{D}_{\boldsymbol{\theta}}^{(,i)}\right)^T \mathbf{S}^T\, \boldsymbol{\Phi}_D\, \mathbf{S}\, \mathbf{D}_{\boldsymbol{\theta}}^{(,j)}.
\end{aligned}
$$

Other sub-matrices can be derived in a similar manner.

Appendix B

The Jacobian matrices are computed below.

$$
\mathbf{D}_{\boldsymbol{\theta}} = \left[\frac{\partial \mathbf{h}}{\partial \theta_1}, \frac{\partial \mathbf{h}}{\partial \theta_2}, \ldots, \frac{\partial \mathbf{h}}{\partial \theta_R}\right]
$$

$$
\frac{\partial \mathbf{h}}{\partial \theta_r} = \begin{bmatrix} \mathrm{Re}\left(\mathrm{vec}\left(\frac{\partial \mathbf{H}}{\partial \theta_r}\right)\right) \\ \mathrm{Im}\left(\mathrm{vec}\left(\frac{\partial \mathbf{H}}{\partial \theta_r}\right)\right) \end{bmatrix}
$$

$$
\frac{\partial \mathbf{H}}{\partial \theta_r} = \begin{bmatrix} \frac{\partial \mathbf{h}[0]}{\partial \theta_r} & \frac{\partial \mathbf{h}[1]}{\partial \theta_r} & \cdots & \frac{\partial \mathbf{h}[L-1]}{\partial \theta_r} \\ \frac{\partial \mathbf{h}[\frac{1}{P}]}{\partial \theta_r} & \frac{\partial \mathbf{h}[1+\frac{1}{P}]}{\partial \theta_r} & \cdots & \frac{\partial \mathbf{h}[L-1+\frac{1}{P}]}{\partial \theta_r} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \mathbf{h}[\frac{P-1}{P}]}{\partial \theta_r} & \frac{\partial \mathbf{h}[1+\frac{P-1}{P}]}{\partial \theta_r} & \cdots & \frac{\partial \mathbf{h}[L-1+\frac{P-1}{P}]}{\partial \theta_r} \end{bmatrix}.
$$

It is sufficient to compute one of the blocks since the other blocks can be computed in a recursive manner:

$$
\begin{aligned}
\frac{\partial \mathbf{h}[k]}{\partial \theta_r} &= \gamma_r \begin{bmatrix} \frac{\partial g(kTs-\tau_{r,1})}{\partial \theta_r} \\ \frac{\partial g\left(kT_s-\tau_{r,2}\right)}{\partial \theta_r} \\ \vdots \\ \frac{\partial g\left(kT_s-\tau_{r,M}\right)}{\partial \theta_r} \end{bmatrix} \odot \mathbf{a}(\theta_r,\boldsymbol{\rho}) \\
&\quad + \gamma_r \begin{bmatrix} g(kTs-\tau_{r,1}) \\ g\left(kT_s-\tau_{r,2}\right) \\ \vdots \\ g\left(kT_s-\tau_{r,M}\right) \end{bmatrix} \odot \frac{\partial \mathbf{a}(\theta_r,\boldsymbol{\rho})}{\partial \theta_r}.
\end{aligned}
$$

If we further assume a uniform linear array, the algebra is straightforward and, at the nominal array perturbation, we get

$$
\begin{aligned}
\frac{\partial \mathbf{h}[k]}{\partial \theta_r} &= \gamma_r \begin{bmatrix} 0 \\ \frac{\partial g\left(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c}\right)}{\partial \theta_r} \\ \vdots \\ \frac{\partial g\left(kT_s-\tau_r-\frac{\delta(M-1)\sin\theta_r}{c}\right)}{\partial \theta_r} \end{bmatrix} \odot \mathbf{a}(\theta_r,\boldsymbol{\rho}_0) \\
&\quad + \gamma_r \begin{bmatrix} g(kTs-\tau_r) \\ g\left(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c}\right) \\ \vdots \\ g\left(kT_s-\tau_r-\frac{\delta(M-1)\sin\theta_r}{c}\right) \end{bmatrix} \odot \frac{\partial \mathbf{a}(\theta_r,\boldsymbol{\rho}_0)}{\partial \theta_r}.
\end{aligned}
$$

The derivative of $g(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})$ with respect to θ_r is shown in (17) at the bottom of the next page. The derivatives of the array steering vector can also be computed in a similar manner. For the above uniform linear array, under nominal perturbation values ($q_{m,0}=1, \rho_m=0$ for the pattern pertubation when the perturbation is not a function of the DOA, and $\rho_m=0$ for the sensor position perturbation case), we have

$$
\begin{aligned}
\frac{\partial \mathbf{a}(\theta_r,\boldsymbol{\rho}_0)}{\partial \theta_r} &= \frac{-j2\pi\delta\cos\theta_r}{\lambda} \\
&\quad \left[0,\ 1,\ \ldots,\ M-1\right]^T \odot \mathbf{a}(\theta_r,\boldsymbol{\rho}_0).
\end{aligned}
$$

The Jacobian of $\mathbf{h}$ with respect to $\boldsymbol{\tau}$, $\mathbf{D}_{\boldsymbol{\tau}}$ is

$$\frac{\partial g\left(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c}\right)}{\partial\theta_r}=\frac{\cos\pi\alpha(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})/T_s}{1-4\alpha^2(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})^2/T_s^2}\left(-\frac{\delta\cos\theta_r}{c}\right)\frac{\left(\cos\frac{\pi(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})}{T_s}-\operatorname{sinc}\frac{(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})}{T_s}\right)}{\left(\pi(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})/T_s\right)}$$
$$+\operatorname{sinc}\frac{(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})}{T_s}\frac{\delta}{c}\cos\theta_r\left[\frac{(1-4\alpha^2(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})^2/T_s^2)\frac{\pi\alpha}{T_s}\sin\pi\alpha(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})/T_s}{(1-4\alpha^2(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})^2/T_s^2)^2}\right.$$
$$\left.-\frac{8\alpha^2(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})/T_s^2\cos\pi\alpha(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})/T_s}{(1-4\alpha^2(kT_s-\tau_r-\frac{\delta\sin\theta_r}{c})^2/T_s^2)^2}\right]$$

$$\frac{\partial g(kT_s-\tau_{r,m})}{\partial\tau_{r,m}}=\frac{\cos\frac{\pi\alpha(kT_s-\tau_r)}{T_s}}{1-\frac{4\alpha^2(kT_s-\tau_r)^2}{T_s^2}}\left(\frac{-\pi}{T_s}\right)\frac{\left(\cos\frac{\pi(kT_s-\tau_{r,m})}{T_s}-\operatorname{sinc}\frac{(kT_s-\tau_{r,m})}{T_s}\right)}{\pi(kT_s-\tau_{r,m})/T_s}$$
$$+\operatorname{sinc}\frac{(kT_s-\tau_{r,m})}{T_s}\frac{\left[\left(1-\frac{4\alpha^2(kT_s-\tau_{r,m})}{T_s^2}\right)\sin\frac{\pi\alpha(kT_s-\tau_{r,m})}{T_s}-\frac{8\alpha^2(kT_s-\tau_{r,m})}{T_s^2}\cos\frac{\pi\alpha(kT_s-\tau_{r,m})}{T_s}\right]}{(1-\frac{4\alpha^2(kT_s-\tau_{r,m})^2}{T_s^2})^2} \tag{17}$$

$$\mathbf{D}_{\boldsymbol{\tau}}=\left[\frac{\partial\mathbf{h}}{\partial\tau_{1,1}},\frac{\partial\mathbf{h}}{\partial\tau_{1,2}},\ldots,\frac{\partial\mathbf{h}}{\partial\tau_{R,M-1}},\frac{\partial\mathbf{h}}{\partial\tau_{R,M}}\right],$$
$$\frac{\partial\mathbf{h}}{\partial\tau_{r,m}}=\begin{bmatrix}\mathrm{Re}\left(\mathrm{vec}\left(\frac{\partial\mathbf{H}}{\partial\tau_{r,m}}\right)\right)\\ \mathrm{Im}\left(\mathrm{vec}\left(\frac{\partial\mathbf{H}}{\partial\tau_{r,m}}\right)\right)\end{bmatrix},$$
$$\frac{\partial\mathbf{h}[k]}{\partial\tau_{r,m}}=\gamma_r\left[0\ \ldots\ \frac{\partial g(kT_s-\tau_{r,m})}{\partial\tau_{r,m}}\ \ldots\ 0\right]^T\odot\mathbf{a}(\theta_r,\boldsymbol{\rho}).$$

The derivative of $g\left(kT_s-\tau_{r,m}\right)$ with respect to $\tau_{r,m}$ is shown in (17) at the top of this page. For a ULA, $\boldsymbol{\tau}=[\tau_1,\ldots,\tau_R]^T$, $\tau_{r,m}=\tau_r+\delta(m-1)\sin\theta_r/c$ and the Jacobian can be derived in a straightforward manner. We can derive $\mathbf{D}_{\boldsymbol{\gamma}^{\mathrm{Re}}}$, $\mathbf{D}_{\boldsymbol{\gamma}^{\mathrm{Im}}}$ and $\mathbf{D}_{\boldsymbol{\rho}}$ from the following expressions,

$$\frac{\partial\mathbf{h}[k]}{\partial\gamma_r^{\mathrm{Re}}}=\begin{bmatrix}g(kTs-\tau_{r,1})\\ g\left(kT_s-\tau_{r,2}\right)\\ \vdots\\ g\left(kT_s-\tau_{r,M}\right)\end{bmatrix}\odot\mathbf{a}(\theta_r,\boldsymbol{\rho}),\quad\frac{\partial\mathbf{h}[k]}{\partial\gamma_r^{\mathrm{Im}}}=j\frac{\partial\mathbf{h}[k]}{\partial\gamma_r^{\mathrm{Re}}},$$
$$\frac{\partial\mathbf{h}[k]}{\partial\rho_d}=\sum_{r=1}^{R}\gamma_r\begin{bmatrix}g(kTs-\tau_{r,1})\\ g\left(kT_s-\tau_{r,2}\right)\\ \vdots\\ g\left(kT_s-\tau_{r,M}\right)\end{bmatrix}\odot\frac{\partial\mathbf{a}(\theta_r,\rho)}{\partial\rho_d}.$$

Let $\mathbf{e}_m\in\mathbb{R}^{M\times1}$ be the unit vector with a 1 at the mth index. There are $2MR$ perturbation parameters in the case of pattern perturbation and $M-1$ parameters in the case of position perturbation. If we consider the pattern only perturbations with the pattern perturbation independent of the DOA, $\boldsymbol{\rho}=[\mathrm{Re}(\rho_1),\ldots,\mathrm{Re}(\rho_M),\mathrm{Im}(\rho_1),\ldots,\mathrm{Im}(\rho_M)]^T$ and

$$\left.\frac{\partial\mathbf{h}[k]}{\partial\mathrm{Re}(\rho_m)}\right|_{\boldsymbol{\rho}=\mathbf{0}}=\sum_{r=1}^{R}\gamma_r\begin{bmatrix}g(kTs-\tau_{r,1})\\ g\left(kT_s-\tau_{r,2}\right)\\ \vdots\\ g\left(kT_s-\tau_{r,M}\right)\end{bmatrix}$$
$$\odot\exp\left(-j2\pi(x_m\sin\theta_r+y_m\cos\theta_r)/\lambda\right)\mathbf{e}_m,$$
$$\left.\frac{\partial\mathbf{h}[k]}{\partial\mathrm{Im}(\rho_m)}\right|_{\boldsymbol{\rho}=\mathbf{0}}=j\left.\frac{\partial\mathbf{h}[k]}{\partial\mathrm{Re}(\rho_m)}\right|_{\boldsymbol{\rho}=\mathbf{0}}.$$

Similarly, for nominal uniform linear arrays with position perturbation only we have [32]

$$\left.\frac{\partial\mathbf{h}[k]}{\partial\rho_m}\right|_{\boldsymbol{\rho}=\mathbf{0}}=\sum_{r=1}^{R}\gamma_r\begin{bmatrix}g(kTs-\tau_{r,1})\\ g\left(kT_s-\tau_{r,2}\right)\\ \vdots\\ g\left(kT_s-\tau_{r,M}\right)\end{bmatrix}\odot(j2\pi\delta\cos\theta_r/\lambda)$$
$$\times\mathrm{diag}(\mathbf{a}(\theta_r,\mathbf{0}))\left(\sum_{k=1}^{M-m}k\mathbf{e}_{k+m}\right). \tag{18}$$

For arrays which are not uniform linear arrays, expressions for the derivatives can be derived in a manner similar to (18).

REFERENCES

[1] M. S. Ullah and A. Tewfik, "Pilot aided direction of arrival estimation for mmWave cellular systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 3401–3405.

[2] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[3] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[4] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.
[5] 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.
[6] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
[7] M. Kokshoorn, H. Chen, Y. Li, and B. Vucetic, "RACE: A rate adaptive channel estimation approach for millimeter wave MIMO systems," in *Proc. IEEE Global Commun. Conf.*, 2016, pp. 1–6.
[8] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.
[9] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.
[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[11] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
[12] S. Haghighatshoar and G. Caire, "Enhancing the estimation of mmWave large array channels by exploiting spatio-temporal correlation and sparse scattering," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2016, pp. 1–7.
[13] S. Buzzi and C. D'Andrea, "Subspace tracking algorithms for millimeter wave MIMO channel estimation with hybrid beamforming," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2017, pp. 1–6.
[14] B. Murmann, "Energy limits in A/D converters," in *Proc. IEEE Faible Tension Faible Consommation*, 2013, pp. 1–4.
[15] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
[16] 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.
[17] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: The ADC bottleneck," in *Proc. IEEE Int. Conf. Ultra-Wideband*, 2009, pp. 22–27.
[18] J. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proc. Int. Wireless Commun. Mobile Comput.*, 2006, pp. 1387–1392.
[19] 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.
[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] 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.
[22] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.
[23] M. Cai *et al.*, "Effect of wideband beam squint on codebook design in phased-array wireless systems," in *Proc. IEEE Global Commun. Conf.*, Dec. 2016, pp. 1–6.
[24] R. L. Haupt, "Antenna arrays in the time domain: An introduction to timed arrays," *IEEE Antennas Propag. Mag.*, vol. 59, no. 3, pp. 33–41, Jun. 2017.
[25] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
[26] B. Wang *et al.*, "Spatial-wideband effect in massive MIMO with application in mmWave systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 134–141, Dec. 2018.
[27] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 142–157, Nov. 2017.
[28] M. S. Stein, J. A. Nossek, and K. Barbé, "Fisher information lower bounds with applications in hardware-aware nonlinear signal processing, compression and inference," 2015, *arXiv:1512.03473*.
[29] M. Stein, K. Barbe, and J. A. Nossek, "DOA parameter estimation with 1-bit quantization bounds, methods and the exponential replacement," in *Proc. Int. ITG Workshop Smart Antennas*, Mar. 2016, pp. 1–6.
[30] F. Liu, H. Zhu, J. Li, P. Wang, and P. V. Orlik, "Massive MIMO channel estimation using signed measurements with antenna-varying thresholds," in *Proc. IEEE Statistical Signal Process. Workshop*, Jun. 2018, pp. 188–192.
[31] P. Wang, J. Li, M. Pajovic, P. T. Boufounos, and P. V. Orlik, "On angular-domain channel estimation for one-bit massive MIMO systems with fixed and time-varying thresholds," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 1056–1060.
[32] M. Viberg and A. L. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.
[33] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1991, pp. 3277–3280.
[34] J. X. Zhu and H. Wang, "Effects of sensor position and pattern perturbations on CRLB for direction finding of multiple narrow-band sources," in *Proc. IEEE ASSP Workshop Spectr. Estimation Model.*, 1988, pp. 98–102.
[35] M. Schluter, M. Dorpinghaus, and G. P. Fettweis, "Bounds on channel parameter estimation with 1-bit quantization and oversampling," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2018, pp. 1–5.
[36] C. Mollén, "High-end performance with low-end hardware," Ph.D. dissertation, Linköping Studies in Science and Technology, vol. 1896, 2019.
[37] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. Hoboken, NJ, USA: Wiley, 2004.
[38] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
[39] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[40] A. Mezghani, F. Antreich, and J. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. Int. ITG Workshop Smart Antennas*, 2010, pp. 143–150.
[41] 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.
[42] M. Stein, F. Wendler, A. Mezghani, and J. A. Nossek, "Quantization-loss reduction for signal parameter estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 5800–5804.

Shilpa Rao received the B.E. degree in electronics and communication from Visvesvaraya Technological University, Belgaum, India, in 2012 and the M.Sc. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2015. She is currently working toward the Ph.D. degree with the Department of Electrical Engineering and Computer Science, University of California Irvine, Irvine, CA, USA, under the supervision of Prof. A. Lee Swindlehurst. She held positions at Robert Bosch, India, from 2012 to 2013 and at the Indian Institute of Science from 2015 to 2016. Her research interests are in the area of signal processing and communications with an emphasis on massive MIMO systems.

Amine Mezghani (S'08–M'16) received the Dipl.-Ing. degree in electrical engineering from Technische Universität München, Munich, Germany, the Diplome d'Ingénieur degree from École Centrale Paris, Paris, France, both in 2006, and the Ph.D. degree in electrical engineering from Technische Universität München, Munich, Germany, in 2015. In 2017, he joined The University of Texas at Austin as a Postdoctoral Fellow. He was a Postdoctoral Scholar with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA. He has authored or coauthored more than 100 papers, particularly on the topic of signal processing and communications with low-resolution analog-to-digital and digital-to-analog converters. His current research interests include millimeter-wave massive multi-in multi-out, hardware-constrained radar and communication systems, and the interface between information theory and antenna theory. He was the recipient of the Joint Rohde & Schwarz and EE Department Outstanding Dissertation Award in 2016.

A. Lee Swindlehurst (S'83–M'84–SM'99–F'04) received the B.S. and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, USA, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1991. He was with the Department of Electrical and Computer Engineering, BYU from 1990 to 2007, where he was a Department Chair from 2003 to 2006. From 1996 to 1997, he held a joint appointment as a Visiting Scholar with Uppsala University and the Royal Institute of Technology, Sweden. From 2006 to 2007, he was on leave working as a Vice President of Research for ArrayComm LLC in San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science Department, University of California Irvine, Irvine, CA, USA, where he was an Associate Dean for Research and Graduate Studies in the Samueli School of Engineering from 2013 to 2016. From 2014 to 2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies at the Technical University of Munich. His research focuses on array signal processing for radar, wireless communications, and biomedical applications, and he has authored or coauthored more than 300 publications in these areas. Dr. Swindlehurst was the inaugural Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He received the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, the 2006 and 2010 IEEE Signal Processing Society's Best Paper Awards, and the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award.