Calibration of Phase Shifter Network for Hybrid Beamforming in mmWave Massive MIMO Systems

Xizixiang Wei, Yi Jiang, *Member, IEEE*, Qingwen Liu, *Senior Member, IEEE*, and Xin Wang, *Senior Member, IEEE*

***Abstract*—For the millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems, hybrid beamforming has been proposed to reap a great gain of the large number of antennas with a limited number of radio frequency (RF) chains. The hybrid beamforming relies on a phase shifter network (PSN) in the RF domain to steer the signal power along the desired direction (or subspace). However, the RF circuits connecting the antennas and the RF chains can introduce distinct phase deviations, which need to be calibrated for the effective hybrid beamforming. This paper develops two novel approaches to the estimation and calibration of the PSN in mmWave massive MIMO communication systems under line-of-sight (LOS) and non-LOS channel scenarios. Specifically, we formulate the core phase deviation estimation problem in the calibration task as an optimization program with constant modulus constraints. Efficient algorithms are then developed to estimate the phase deviations that need to be calibrated. To gauge the performance of the proposed schemes, we also derive the Cramer-Rao lower bounds (CRLBs) for the phase estimates. The numerical results validate the effectiveness of our approaches by showing that the proposed algorithms yield estimates whose mean squared errors (MSEs) are close to the CRLBs.**

***Index Terms*—Phased array calibration, massive MIMO, cramer-rao lower bound (CRLB), millimeter wave communication systems.**

I. Introduction

THE millimeter wave (mmWave) bands have the potential for future enhanced mobile broadband (eMBB) service in the 5th generation (5G) wireless communications due to the tens of GigaHertz of available licensed and unlicensed spectrum bands in the range of 30–300 GHz [1], [2]. In particular, the mmWave and massive multiple-input multiple-output (MIMO) make a perfect match. Due to the small wavelength of mmWave, an array of a large number of antennas can be deployed in a moderate size; and the large array gain of massive MIMO can compensate the severe path loss of the mmWave through directional beamforming [3], [4]. The two technologies together have become a key to 5G cellular communication systems [5], [6].

For massive MIMO systems, deployment of fully-digital beamforming schemes would require too many radio frequency (RF) chains, resulting in high cost and high power consumption [7]. One way to reduce the cost and power consumption is to use one-bit analog-to-digital converters (ADCs) and one-bit digital-to-analog converters (DACs) in the RF chains [8], [9]. As another remedy, hybrid beamforming has been proposed to reap the great gain of massive MIMO with a moderate number of RF chains (see [10], [11] and the references therein). The hybrid beamforming scheme employs a phase shifter network (PSN) in the analog domain to steer the signal power along the desired directions. However, as the RF circuits connecting the antennas and the RF chains always have distinct phase drifts, the phase deviations need to be calibrated as a prerequisite of efficient hybrid beamforming designs. Moreover, the phase deviations need to be estimated and calibrated over the air periodically, as they can vary with the temperature and the aging of the hardware [12]. Most existing works on hybrid beamforming presume that no phase deviations are presented in the PSN, i.e., the phase calibration is already done. While [13] first addressed the over-the-air (OTA) calibration in hybrid beamforming transceivers, it mainly focused on the calibration of the RF mismatch between the downlink and uplink channels and did not take into account the imperfection of the PSN. Few work on the OTA calibration of the PSN is available in the literature. The goal of this paper to put together this critical missing piece.

Calibration of the conventional phased array has been investigated for about two decades (see, e.g., [14] and the references therein). The classic calibration methods can be divided into two categories, namely, the amplitude-based approach and the complex-signal based approach, depending on whether the measurement technique is based on the received power (amplitude-only) or the complex signals.

The rotating element electric field vector (REV) method is a well-known amplitude-based approach that was first proposed by Mano *et al.* in [15]. The REV method measures the amplitudes of the received signal as one of the phase shifters changes its phase from 0 to 2π. Measuring the minimal and the

Manuscript received June 10, 2019; revised January 15, 2020; accepted March 27, 2020. Date of publication April 3, 2020; date of current version April 28, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Remy Boyer. This work was supported in part by the National Natural Science Foundation of China Grants 61671154 and 61771005, in part by the Innovation Program of Shanghai Municipal Science and Technology Commission Grant 17510710400, and in part by the research grant from Morningcore Technology Co., Ltd. This article was presented in part at the International Conference on Communications in Shanghai, 2019 and in part at the International Conference on Wireless Communications and Signal Processing in Xi'an, 2019. *(Corresponding author: Yi Jiang.)*

Xizixiang Wei, Yi Jiang, and Xin Wang are with the Shanghai Institute for Advanced Communication and Data Science, Key Laboratory for Information Science of Electromagnetic Waves (MoE), Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China (e-mail: 17210720089@fudan.edu.cn; yijiang@fudan.edu.cn; xwang11@fudan.edu.cn).

Qingwen Liu is with the College of Electronics and Information Engineering, Tongji University, Shanghai 200092, China (e-mail: qliu@tongji.edu.cn).

Digital Object Identifier 10.1109/TSP.2020.2984884

1053-587X © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

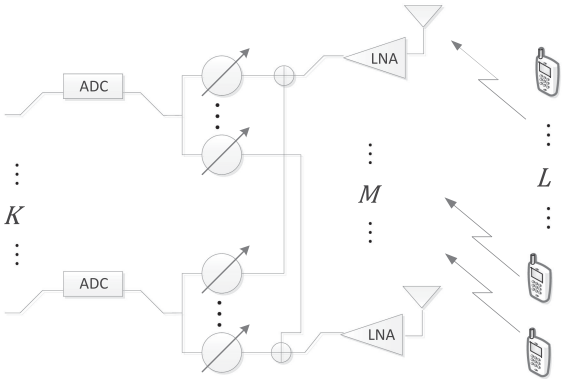

Fig. 1. System configuration with K RF chains and M antenna elements in the base station and L user equipments in the far-field region.

maximal received power along with the phase shifts between them, one can calibrate an individual element of the array. As an improvement to the REV method, a maximum likelihood algorithm was proposed in [16], which can reduce the number of required measurements by switching the phase shifter in four orthogonal states. In [17], an enhanced method was also proposed to further reduce the number of measurements per element through solving a set of nonlinear equations. Another way to improve the measurement efficiency is to calibrate multiple elements simultaneously, as done in [18].

A number of phased array calibration methods based on complex signals were proposed in [14], [19], [20]. In [19], the multi-element phase-toggle (MEP) method calibrates the phase deviations by applying an inverse fast Fourier transform (IFFT) to the received complex signals. This method requires the number of measurements at least twice the number of antenna elements. While [14] proposed a synthetic array-based approach, [20] proposed a parallel calibration method to overcome the limitation with the approach in [18]. All the aforementioned methods aim to calibrate the phase array in the lab environment, where the measurements are almost noise-free.

In this paper, we address phase calibration for the PSN applied to hybrid beamforming. Our problem is different from the conventional phase array calibration in three aspects:

- For a massive MIMO (e.g., uplink) system with M antenna elements and K RF chains as illustrated by Fig. 1, what we need to calibrate is an M-input and K-output *network* rather than an array;
- The calibration is to be done over the air in a functioning communication system in the field with the measurements subject to noise and co-channel interference;
- The mmWave channel may experience multipath fading when taking measurements during the calibration, which is a sharp contrast to the lab environment commonly assumed by the classic methods.

Taking into account these considerations, the aim of this paper is to provide low complexity OTA PSN calibration approaches with as little time-frequency resources as possible. The core of the calibration approaches is to estimate the unknown random phase deviations. To this end, we formulate the relevant estimation task as a non-convex least-squares fitting problem with constant modulus constraints. Building on this formulation, we make the following contributions:

- By applying different training beamformers at the base station (BS) when user equipments (UEs) transmit the training sequences, we propose a novel OTA PSN calibration mechanism based on the measurements of the effective channels of multiple users.
- In line-of-sight (LOS) channel scenario, a low complexity OTA calibration algorithm is developed based on an alternative optimization approach. By jointly leveraging the Fast Fourier Transform (FFT) operation and the iterative power method for unimodular quadratic program (UQP), we can obtain the phase deviation estimates with a near-linear complexity in the number of antennas.
- In multipath channel scenario, we put forth two novel algorithms, repeated orthogonal projection algorithm and extended RELAX algorithm, to estimate channel parameters. Combining with the power method, we can compute the phase deviation estimates with an affordable complexity.
- The Cramer-Rao lower bounds (CRLBs) for the intended phase deviation estimations are derived to benchmark the performance of the proposed algorithms. By CRLB analysis, the number of required time-frequency resources (measurements) for the effective calibration is also delineated.

The numerical results corroborate that the proposed algorithms are efficient as the resultant mean squared errors (MSEs) of the phase deviation estimates are close to the CRLBs. It is also shown that after implementing our algorithm at an uncalibrated PSN, the spectral efficiency and generated power radiation pattern can be significantly improved.

The rest of the paper is organized as follows. The system model and problem formulation are described in Section II. The proposed phased array calibration scheme for LOS channel environments is developed in Section III. In Section IV we generalize our approach to the multipath channel environments. We derive the CRLBs with the intended estimation problem in Section V. Section VI provides the numerical results, followed by the conclusions in Section VII.

Notations: $(\cdot)^H$, $(\cdot)^T$ and $(\cdot)^*$ denote Hermitian transpose, transpose and complex conjucate; $\mathbb{E}(\cdot)$ denotes mathematical expectation; $\mathbb{C}^{M\times K}$ and $\mathbb{R}^{M\times K}$ denote the M-by-K dimensional complex and real space, respectively; $\lceil a \rceil$ denotes the smallest integer larger than a scalar a; $\hat{a}$ denotes the estimation of a parameter a; $\mathcal{D}(\boldsymbol{a})$ denotes a matrix whose diagonal entries are vector $\boldsymbol{a}$; $\mathrm{vec}(\mathbf{A})$ stacks the column vectors of $\mathbf{A}$ into a single column; $\mathrm{Re}(\mathbf{A})$ and $\mathrm{Im}(\mathbf{A})$ stand for taking the real or imaginary part of $\mathbf{A}$; $\mathcal{P}\{\mathbf{A}\}$ is the projector of the range space of matrix $\mathbf{A}$, i.e. $\mathcal{P}\{\mathbf{A}\} \triangleq \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ if $\mathbf{A}$ has full column rank, and $\mathcal{P}^{\perp}\{\mathbf{A}\} \triangleq \mathbf{I} - \mathcal{P}\{\mathbf{A}\}$ is the projector orthogonal to the range space of $\mathbf{A}$; $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius and L-2 norm of a matrix; $\odot$ and $\otimes$ denote the Hadamard and Kronecker products of two matrices, respectively; $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, $\mathbf{1}_N$ and $\mathbf{0}_N$ denote an all-one vector and an all-zero vector of length N; $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for a circularly symmetric complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\sim$ stands for "distributed as".

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

II. System Model and Problem Formulation

Consider an orthogonal frequency division multiplexing (OFDM) based mmWave cellular uplink system as shown in Fig. 1, where UEs transmit data to a BS through different subcarriers. We assume that the transceiver in the BS is equipped with an M-element antenna array and K RF chains, between which a PSN with $K \times M$ phase shifers is deployed to perform hybrid precoding/combining. When a symbol s is transmitted by a single-antenna UE, it would pass the OTA channel $\mathbf{h} \in \mathbb{C}^{M\times 1}$ and be received by M antenna ports. The received signal is then amplified by the M low-noise amplifiers (LNAs) before being combined into K streams by the PSN beamformer $\mathbf{\Phi} \in \mathbb{C}^{K\times M}$ and being digitalized by the K ADCs into baseband signal

$$\mathbf{y}_B = \mathbf{\Phi}\mathbf{h}s + \mathbf{n} \in \mathbb{C}^{K\times 1}, \tag{1}$$

where $\mathbf{n}$ is an additive Gaussian noise vector with the identically and independently distributed (i.i.d.) entries. Here, $\mathbf{\Phi}\mathbf{h} \in \mathbb{C}^{K\times 1}$ can be regarded as the effective (baseband) channel of s. If s is a training pilot, the effective channel can be estimated as $\mathbf{y}$ at the BS.

During the calibration, the BS can engage L UEs and require each of them to send a training sequence to itself via different orthogonal channels, e.g., different resource blocks or symbols. Meanwhile, the BS sets the PSN using N different randomly generated training beamformers $\{\mathbf{\Phi}_n\}_{n=1}^{N}$ to receive the training signals. After the BS measures the received signals of each UE for all training beamformers as $\{\mathbf{y}_n^{(l)}\}_{n=1,l=1}^{N,L}$, it estimates the phase deviations of the PSN. Note that the training beamformers are the phases applied by the PSN during the calibration, while the training sequences are the pilot sequences sent by the UEs when BS applies the training beamformers. With (accurate) phase deviation estimates, the BS can then calibrate the phase deviations of the PSN to achieve effective beamforming.

A. System Model

For mathematical simplicity, we assume that a uniform linear array (ULA) of $M >> 1$ antenna elements is deployed at the BS. The steering vector of such a ULA at direction θ can be denoted by

$$\boldsymbol{a}(\theta) = \left[1, e^{-j\frac{2\pi d}{\lambda}\sin\theta}, e^{-j\frac{4\pi d}{\lambda}\sin\theta}, \ldots, e^{-j\frac{2(M-1)\pi d}{\lambda}\sin\theta}\right]^T, \tag{2}$$

where d and λ represent the inter-element spacing and the wavelength of the carrier, respectively. Without loss of generality, we assume that $d = \lambda/2$ and $\theta \in [-\pi/2, \pi/2)$.

The BS implements a PSN between the K RF chains and an M-element ULA to perform hybrid beamforming. The PSN beamformer can be represented by a matrix $\mathbf{\Phi} \in \mathbb{C}^{K\times M}$ where

$$|\phi_{km}| = 1 \ \text{ and } \ \angle\phi_{km} \in [0, 2\pi) \quad \text{for } \ k = 1, 2, \ldots, K, \ m = 1, 2, \ldots, M. \tag{3}$$

For PSN, the RF electrical wires and components connecting the RF chains and the antenna ports can also incur certain phase deviations denoted by $\mathbf{\Omega} \in \mathbb{R}^{K\times M}$, whose entry $\omega_{km} \in [0, 2\pi)$ represents the extra phase shift induced by the electronic components between the k-th RF chain and the m-th antenna element, including the m-th LNA, the k-th ADC, and the electric wire. We want to stress that only the deviations of the phases rather than the amplitudes are modeled here. It is because the effective electrical length of the circuits is temperature-sensitive, often resulting in remarkable phase variations over a duration of hours. The amplitude attenuations of the circuits, however, can usually be regarded as a constant over their life time. Overall, the actual phase shifts of the PSN are dictated by $\mathbf{W} \odot \mathbf{\Phi}$, where $\mathbf{W} \in \mathbb{C}^{K\times M}$ has elements

$$w_{km} = e^{j\omega_{km}}, \quad \text{for } \ k = 1, 2, \ldots, K, \ m = 1, 2, \ldots, M. \tag{4}$$

In addition to the calibration conducted after manufacturing in the factory, the phase deviations should be calibrated over the air periodically to enable effective hybrid beamforming. The OTA estimation of the phases $\{\omega_{km}\}_{k=1,m=1}^{K,M}$ is the focus of this paper.

As a user often occupies only a small number of resource blocks, its OTA channel can be regarded as a narrow-band and flat-fading channel. Due to the poor scattering nature of mmWave frequency [5], the channel from the l-th user to the BS can be modeled as sum of a few paths from different directions. We adopt a geometric-based channel model in [21] where the channel vector is given by

$$\mathbf{h}^{(l)} = \sum_{i=1}^{d_l} \gamma_l^{(i)} \boldsymbol{a}(\theta_l^{(i)}) \in \mathbb{C}^{M\times 1}. \tag{5}$$

Here, d_l is the number of scatters of the l-th user channel and γ is the random complex-valued small-sale fading gain of the i-th subpath. Given the n-th training beamformer $\mathbf{\Phi}_n$, the effective channel response is the concatenation of the channel and network responses:

$$\mathbf{h}_{\text{eff},n}^{(l)} = (\mathbf{W} \odot \mathbf{\Phi}_n) \sum_{i=1}^{d_l} \gamma_l^{(i)} \boldsymbol{a}(\theta_l^{(i)}) \in \mathbb{C}^{K\times 1}. \tag{6}$$

B. Problem Formulation

As described before, the BS shifts the PSN for N times and estimates/measures the effective channels of L involved UEs based on their training sequences during the calibration. Recall that $\mathbf{h}_{\text{eff},n}^{(l)}$ represents the effective channel of the l-th UE when the PSN at the BS adopts the n-th training beamformer $\mathbf{\Phi}_n$. Denote the measurement of $\mathbf{h}_{\text{eff},n}^{(l)}$ as

$$\mathbf{y}_n^{(l)} = \mathbf{h}_{\text{eff},n}^{(l)} + \mathbf{z}_l, \quad \text{for } \ l = 1, 2, \ldots, L, \tag{7}$$

where $\mathbf{z}_l \sim \mathcal{C}N(0, \sigma_z^2\mathbf{I})$ is the measurement error. Let $\mathbf{Y}_n \triangleq [\mathbf{y}_n^{(1)}, \ldots, \mathbf{y}_n^{(L)}]$ collect the measurements of the effective channel vectors for all L users. Then

$$\mathbf{Y}_n = [\mathbf{W} \odot \mathbf{\Phi}_n]\, \mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma} + \mathbf{Z}_n \in \mathbb{C}^{K\times L} \tag{8}$$

where

$$\mathbf{A}(\mathbf{\Theta}) = \Big[\boldsymbol{a}(\theta_1^{(1)}), \ldots, \boldsymbol{a}(\theta_1^{(d_1)}), \boldsymbol{a}(\theta_2^{(1)}), \ldots, \boldsymbol{a}(\theta_2^{(d_2)}), \ldots, \boldsymbol{a}(\theta_L^{(1)}), \ldots, \boldsymbol{a}(\theta_L^{(d_L)})\Big] \in \mathbb{C}^{M\times(d_1+d_2+\cdots+d_L)}, \tag{9}$$

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

and

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1^{(1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1^{(d_1)} & 0 & \cdots & 0 \\ 0 & \gamma_2^{(1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \gamma_2^{(d_2)} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_L^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_L^{(d_L)} \end{bmatrix} \in \mathbb{C}^{(d_1+d_2+\cdots+d_L)\times L} \quad (10)$$

is a block diagonal matrix. Note that when the channel has LOS and is multipath-free, then $\mathbf{A}(\mathbf{\Theta})$ and $\mathbf{\Gamma}$ can be simplified as

$$\mathbf{A}(\mathbf{\Theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots \boldsymbol{a}(\theta_l)] \in \mathbb{C}^{M\times L}, \quad (11)$$

and

$$\mathbf{\Gamma} = \mathcal{D}([\gamma_1, \gamma_2, \cdots \gamma_L]) \in \mathbb{C}^{L\times L}. \quad (12)$$

Stack the matrices into

$$\mathbf{\Phi} \triangleq \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \vdots \\ \mathbf{\Phi}_N \end{bmatrix} \in \mathbb{C}^{KN\times M} \text{ and } \mathbf{Z} \triangleq \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix} \in \mathbb{C}^{KN\times L}. \quad (13)$$

Similarly, stack $\mathbf{Y}_n, n = 1, \ldots, N$ into a $KN \times L$ matrix $\mathbf{Y}$. Then clearly,

$$\mathbf{Y} = [(\mathbf{1}_N \otimes \mathbf{W}(\mathbf{\Omega})) \odot \mathbf{\Phi}]\, \mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma} + \mathbf{Z} \quad (14)$$

where use the notation $\mathbf{W}(\mathbf{\Omega})$ to emphasize that it is a function of the phases deviations $\mathbf{\Omega}$ as delineated by (3). Here we assume that the directions of arrival $\{\theta_l^{(i)}\}_{l=1,i=1}^{L,d_l}$ and the complex channel gains $\{\gamma_l^{(i)}\}_{l=1,i=1}^{L,d_l}$ of each UE remain unchanged during the N measurements, which is practically true since the measurements can be obtained within a few milli-seconds.

As error matrix $\mathbf{Z}$ is assumed to have i.i.d. Gaussian entries, the maximum likelihood estimates of the phase deviations can be then obtained using least-squares fitting [22]

$$\hat{\mathbf{\Omega}} = \arg\min_{\mathbf{\Omega},\mathbf{\Theta},\mathbf{\Gamma}} \|\mathbf{Y} - [(\mathbf{1}_N \otimes \mathbf{W}(\mathbf{\Omega})) \odot \mathbf{\Phi}]\, \mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma}\|_F^2. \quad (15)$$

Note that, along with the estimation of $\mathbf{\Omega}$, $\mathbf{\Theta}$ and $\mathbf{\Gamma}$ are the nuisance parameters, which also need to be estimated. We can obtain $\hat{\mathbf{W}}$ and $\mathbf{A}(\hat{\mathbf{\Theta}})$ by (4) and (9) once we get $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Theta}}$.

III. PSN Calibration in LOS Channel

Consider the LOS channel scenario; i.e., $\mathbf{A}(\mathbf{\Theta})$ and $\mathbf{\Gamma}$ are given by (11) and (12). We next develop an iterative algorithm to solve the non-convex problem (15) in this case. Specifically, we adopt a block coordinate decent method to optimize $\mathbf{W}(\mathbf{\Omega})$, $\mathbf{A}(\mathbf{\Theta})$ and $\mathbf{\Gamma}$ in an alternating manner to minimize the objective function in (15). The proposed alternating optimization method contains three steps in each loop: first, initialize $\mathbf{W}$ and $\mathbf{A}(\mathbf{\Theta})$ at random and estimate $\hat{\mathbf{\Gamma}}$; second, fix $\mathbf{W}$ and $\hat{\mathbf{\Gamma}}$ to estimate $\mathbf{A}(\hat{\mathbf{\Theta}})$; third, fix $\hat{\mathbf{\Gamma}}$ and $\mathbf{A}(\hat{\mathbf{\Theta}})$ to obtain $\hat{\mathbf{W}}$. The above three steps are detailed as follows.

A. Estimation of Channel Gain Matrix $\mathbf{\Gamma}$

Given the estimates $\hat{\mathbf{W}}$[1] and $\mathbf{A}(\hat{\mathbf{\Theta}})$, $\mathbf{\Gamma}$ in (15) can be estimated via the simple least-squares method. As $\mathbf{\Gamma}$ is a diagonal matrix, its L diagonal elements can be estimated as

$$\hat{\gamma}_l = \arg\min_{\gamma_l} \quad \|\mathbf{y}_l - \mathbf{b}_l\gamma_l\|_2^2, \quad (16)$$

or

$$\hat{\gamma}_l = \frac{\mathbf{b}_l^H\mathbf{y}_l}{\mathbf{b}_l^H\mathbf{b}_l}, \quad (17)$$

where $\mathbf{y}_l$ is the l-th column of $\mathbf{Y}$, and $\mathbf{b}_l$ is given by:

$$\mathbf{b}_l \triangleq \left[(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}\right] \boldsymbol{a}(\theta_l). \quad (18)$$

B. Estimation of User Directions $\mathbf{\Theta}$

Given $\hat{\mathbf{W}}$ and $\hat{\mathbf{\Gamma}}$, the L directions of arrival can be estimated separately. In other words, we can decompose (15) into L sub-problems

$$\hat{\theta}_l = \arg\min_{\theta_l} \|\mathbf{y}_l - \mathbf{Q}_l\boldsymbol{a}(\theta_l)\|_2^2, \quad l = 1, 2, \ldots, L, \quad (19)$$

where

$$\mathbf{Q}_l \triangleq \hat{\gamma}_l \left[(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}\right] \in \mathbb{C}^{KN\times M}. \quad (20)$$

For a ULA, θ_l can be estimated via FFT and a local search algorithm. Define $\nu_l = \frac{M}{\lambda} d\sin(\hat{\theta}_l)$ and insert it into $\boldsymbol{a}(\hat{\theta}_l)$, we then have

$$\mathbf{g}(\nu_l) = \left[1, e^{-j\frac{2\pi}{M}\nu_l}, \cdots e^{-j\frac{2\pi}{M}(M-1)\nu_l}\right]^T, \quad (21)$$

which is a Discrete Fourier Transform (DFT) vector. Each element of vector $\mathbf{Q}_l\boldsymbol{a}(\hat{\theta}_l)$ is actually the DFT of each row of matrix $\mathbf{Q}_l$ at the frequency "ν_l". We can apply FFT to every row of $\mathbf{Q}_l$ and find $\hat{\nu}_l$ that minimizes the objective function in (19). One way to obtain a more accurate estimation of $\hat{\nu}_l$ is to increase the length of FFT at the cost of a higher computational complexity. In fact, a more efficient method is to use a local search algorithm to find a better $\hat{\nu}_l^*$ after the acquisition of the approximate $\hat{\nu}_l$ using FFT. Rewrite the objective function in (19) as

$$f(\nu_l) = (\mathbf{y}_l - \mathbf{Q}_l\mathbf{g}(\nu_l))^H(\mathbf{y}_l - \mathbf{Q}_l\mathbf{g}(\nu_l)). \quad (22)$$

Its first-order derivative is

$$\frac{\partial f}{\partial \nu_l} = -2\text{Re}\left[\mathbf{y}_l^H\mathbf{Q}_l\frac{\partial\mathbf{g}(\nu_l)}{\partial\nu_l}\right] + 2\text{Re}\left[\mathbf{g}(\nu_l)\mathbf{Q}_l^H\mathbf{Q}_l\frac{\partial\mathbf{g}(\nu_l)}{\partial\nu_l}\right], \quad (23)$$

where

$$\frac{\partial\mathbf{g}(\nu_l)}{\partial\nu_l} = -j2\pi\mathcal{D}([1, 2, \cdots M-1]^T)\mathbf{g}(\nu_l). \quad (24)$$

[1]We use $\hat{\mathbf{W}}$ rather than $\mathbf{W}(\hat{\mathbf{\Omega}})$ for notational simplicity.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

Starting from $\hat{\nu}_l$, we can then use back-tracking line search [23] to find the minmum $f(\hat{\nu}_l^*)$ and the desired $\hat{\nu}_l^*$. With $\hat{\nu}_l^*$, the estimation of l-th user direction can be obtained by

$$\hat{\theta}_l = \arcsin\left(\frac{\lambda\hat{\nu}_l^*}{Md}\right). \tag{25}$$

C. Estimation of PSN Deviation Matrix $\mathbf{W}$

Given $\mathbf{A}(\hat{\mathbf{\Theta}})$ and $\hat{\mathbf{\Gamma}}$, the estimation of $\mathbf{W}$ is decoupled row-wise as follows. Take the k-th row of each $\mathbf{Y}_n$ and $\mathbf{\Phi}_n, \forall n = 1, 2, \ldots, N$, and stack them into[2]

$$\widetilde{\mathbf{Y}}_k \triangleq \begin{bmatrix} \mathbf{Y}_1(k,:) \\ \mathbf{Y}_2(k,:) \\ \vdots \\ \mathbf{Y}_N(k,:) \end{bmatrix} \text{ and } \widetilde{\mathbf{\Phi}}_k \triangleq \begin{bmatrix} \mathbf{\Phi}_1(k,:) \\ \mathbf{\Phi}_2(k,:) \\ \vdots \\ \mathbf{\Phi}_N(k,:) \end{bmatrix} \in \mathbb{C}^{N\times M}. \tag{26}$$

Using the notations in (26), we then estimate the k-th row of $\hat{\mathbf{W}}$ by minimizing the following

$$\|\widetilde{\mathbf{Y}}_k - \left[\left(\mathbf{1}_N \otimes \hat{\mathbf{W}}(k,:)\right) \odot \widetilde{\mathbf{\Phi}}_k\right] \mathbf{A}(\hat{\mathbf{\Theta}})\mathbf{\Gamma}\|_F^2 \tag{27a}$$

$$\overset{(a)}{=} \|\widetilde{\mathbf{Y}}_k - \widetilde{\mathbf{\Phi}}_k \mathcal{D}([\hat{\mathbf{W}}(k,:)])\mathbf{P}\|_F^2 \tag{27b}$$

$$\overset{(b)}{=} \|\widetilde{\mathbf{Y}}_k - [\hat{w}_{k1}\mathbf{S}_1 + \hat{w}_{k2}\mathbf{S}_2 + \cdots + \hat{w}_{kM}\mathbf{S}_M]\,\|_F^2 \tag{27c}$$

$$\overset{(c)}{=} \|\mathcal{Y}_k - \mathbf{R}_k[\hat{\mathbf{W}}(k,:)]^T\|_2^2 \tag{27d}$$

where (a) follows as we simplify the Kronecker product and define

$$\mathbf{P} \triangleq \mathbf{A}(\hat{\mathbf{\Theta}})\mathbf{\Gamma}; \tag{28}$$

(b) is due to the usage of vector outer product in replace of matrix product and our definitions

$$\mathbf{S}_m = \widetilde{\mathbf{\Phi}}_k(:,m)\mathbf{P}(m,:) \in \mathbb{C}^{N\times L} \text{ for } m = 1, 2, \cdots M; \tag{29}$$

and (c) follows by defining

$$\mathcal{Y}_k = \text{vec}(\widetilde{\mathbf{Y}}_k) \in \mathbb{C}^{NL\times 1} \tag{30}$$

and

$$\mathbf{R}_k = [\text{vec}(\mathbf{S}_1), \text{vec}(\mathbf{S}_2), \cdots \text{vec}(\mathbf{S}_M)] \in \mathbb{C}^{NL\times M}. \tag{31}$$

The minimization of (27d) can be further converted into to a standard UQP [24]. By denoting

$$\mathbf{U}_k = \begin{bmatrix} -\mathbf{R}_k^H\mathbf{R}_k & \mathbf{R}_k\mathcal{Y}_k \\ \mathcal{Y}_k^H\mathbf{R}_k & 0 \end{bmatrix} \text{ and } \mathbf{v}^{(k)} = \begin{bmatrix} e^{j\xi}\,[\mathbf{W}(k,:)]^T \\ e^{j\xi} \end{bmatrix} \tag{32}$$

where $\xi \in [0, 2\pi)$ is a free variable, we then wish to solve

$$\max_{\mathbf{v}^{(k)}} \ \mathbf{v}^{(k)H}\mathbf{U}_k\mathbf{v}^{(k)}$$
$$\text{s.t. } |v_i^{(k)}| = 1 \text{ for } i = 1, 2, \ldots, M+1. \tag{33}$$

[2]In the subsequent derivations, we use MATLAB type of notation. Namely, $\mathbf{A}(:,k)$ ($\mathbf{A}(k,:)$) stands for the k-th column (row) of $\mathbf{A}$.

Algorithm 1: Calibration of PSN in LOS Channel Case

Input: number of antennas M; number of RF chains K; number of training beamformers N; number of users L; measured effective channel vectors $\mathbf{Y}$;
Output: Phase deviation matrix $\hat{\mathbf{W}} \in \mathbb{C}^{K\times M}$;
1: Random initialization of $\hat{\mathbf{W}}$ and $\hat{\mathbf{\Theta}}$
2: **do**
3: **for** $l = 1 : L$ **do**
4: Estimate $\hat{\gamma}_l$ in (17);
5: **end for**
6: **for** $l = 1 : L$ **do**
7: Estimate $\hat{\theta}_l$ in (19) by FFT and backtracking line search;
8: **end for**
9: **for** $k = 1 : K$ **do**
10: Estimate $\hat{\mathbf{W}}(k,:)$ in (33) via the power method;
11: **end for**
12: $\hat{\mathbf{Y}} = [(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}]\mathbf{A}(\hat{\mathbf{\Theta}})\hat{\mathbf{\Gamma}}$
13: **while** change in $\|\hat{\mathbf{Y}}\|_F$ from the previous iteration is less than ϵ (e.g., $\epsilon = 10^{-5}$)
14: **return** $\hat{\mathbf{W}}$;
15: Perform calibration according to $\hat{\mathbf{W}}$

While the UQP is NP-hard, we can use the computationally efficient power method introduced in [25] to find a local optimum. Specifically, a simple power-method alike iteration follows from

$$\mathbf{v}_{i+1}^{(k)} = e^{j\arg(\mathbf{U}_k\mathbf{v}_i^{(k)})}. \tag{34}$$

Note that the power-method iteration requires $\mathbf{U}_k$ to be positive (semi)definite. To ensure this, when $\lambda_{\min}$, the smallest eigenvalue of $\mathbf{U}_k$, is negative, we can adopt the diagonal loading technique, i.e., $\mathbf{U}_k \leftarrow \mathbf{U}_k - \lambda_{\min}\mathbf{I}_{M+1}$. Such an operation does not change the solution of UQP since $\mathbf{v}^{(k)H}(\mathbf{U}_k - \lambda_{\min}\mathbf{I}_{M+1})\mathbf{v}^{(k)} = \mathbf{v}^{(k)H}\mathbf{U}_k\mathbf{v}^{(k)} - \lambda_{\min}(M+1)$.

Upon convergence of the power-method iteration (34), we obtain

$$\hat{\mathbf{v}}^{(k)} = \begin{bmatrix} e^{j\hat{\xi}}\left[\hat{\mathbf{W}}(k,:)\right]^T \\ e^{j\hat{\xi}} \end{bmatrix}. \tag{35}$$

The estimation $\hat{\mathbf{W}}(k,:)$ can be then obtained by a scalar-vector multiplication where the scalar is the conjugation of the last element of $\hat{\mathbf{v}}^{(k)}$ and the vector consists of the first to the M-th elements of $\hat{\mathbf{v}}^{(k)}$.

D. Alternative Minimization

The aforementioned three alternating steps are summarized in Algorithm 1. In the proposed algorithm, we monotonically decrease the objective function in (15); such a block coordinate decent method is guaranteed to converge to a local optimal solution. We claim the convergence of the algorithm when the change of $\hat{\mathbf{Y}} = [(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}]\mathbf{A}(\hat{\mathbf{\Theta}})\hat{\mathbf{\Gamma}}$ between two consecutive loops is less than a small number, e.g., $\epsilon = 10^{-5}$.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

Remark 1: The three inner for-loops in Algorithm 1 have low computational complexity. Specifically, the complexity with the first loop and FFTs in the second loop are $O(L)$ and $O(LM\log(M))$, respectively. Both the backtracking line search in the second loop and power method-like iteration in the third loop are simple with $O(L)$ and $O(M)$ complexity, and they converge fast. Numerical results also show that the outer do-while loop can converge after about 20 iterations. Hence, the overall complexity with Algorithm 1 is affordable, especially by noting that the calibration needs to be done just a few times per hour, if not per day.

Remark 2: In practice, uniform rectangular array (URA) could be preferred over ULA because it is more compact and can achieve three dimensional beamforming in the BS [26]. The steering vector at direction (φ,θ) for a URA of $M=P\times Q$ elements is

$$\boldsymbol{a}(\varphi,\theta)=\boldsymbol{a}_a(\varphi,\theta)\otimes\boldsymbol{a}_e(\varphi,\theta)\in\mathbb{C}^{PQ\times 1}, \tag{36}$$

where $\boldsymbol{a}_a(\varphi,\theta)$ and $\boldsymbol{a}_e(\varphi,\theta)$ are the $P\times 1$ and $Q\times 1$ steering vectors of azimuth and elevation angles, respectively. Specifically, we have

$$\begin{aligned}&\boldsymbol{a}_a(\varphi,\theta)\\&\quad=[1,e^{-j\frac{2\pi d_x}{\lambda}\sin\varphi\sin\theta},\ldots,e^{-j\frac{2\pi(P-1)d_x}{\lambda}\sin\varphi\sin\theta}]^T,\end{aligned} \tag{37}$$

$$\begin{aligned}&\boldsymbol{a}_e(\varphi,\theta)\\&\quad=[1,e^{-j\frac{2\pi d_y}{\lambda}\sin\varphi\cos\theta},\ldots,e^{-j\frac{2\pi(Q-1)d_y}{\lambda}\sin\varphi\cos\theta}]^T,\end{aligned} \tag{38}$$

where d_x and d_y represent the inter-element spacing for the vertical and horizontal directions of the URA [27]. The PSN calibration in the BS with a URA is similar to that in the ULA case. Since the structures of $\mathbf{W}$ and $\boldsymbol{\Gamma}$ remain the same as in (15), we can use the above methods in (17) and (34) to estimate them. However, the estimation of the steering vector of each user, $\boldsymbol{a}(\varphi_l,\theta_l)$, becomes a two-dimensional (2-D) search problem,

$$\hat{\varphi}_l,\hat{\theta}_l=\arg\min_{\varphi_l,\theta_l}\ \|\mathbf{y}_l-\mathbf{Q}_l\boldsymbol{a}(\varphi_l,\theta_l)\|_2^2,\quad\forall l=1,2,\ldots,L. \tag{39}$$

The estimation of (φ_l,θ_l) can be efficiently obtained by applying 2-D FFT combined with the backtracking line search algorithm. In a nutshell, our approach readily carries over to a BS equipped with a URA and its computational complexity is still affordable.

IV. PSN Calibration in Multipath Channel Case

In this section, we consider the calibration of the PSN in multipath channel scenario. Recall that

$$\mathbf{Y}=[(\mathbf{1}_N\otimes\mathbf{W}(\boldsymbol{\Omega}))\odot\boldsymbol{\Phi}]\,\mathbf{A}(\boldsymbol{\Theta})\boldsymbol{\Gamma}+\mathbf{Z}. \tag{40}$$

We assume that the multipath number of each user channel $d_l,\ \forall l=1\cdots L$, is known. The assumption is reasonable since we can use, e.g., generalized Akaike information criterion (GAIC) to determine the path number of channels [28]. We will see later that the proposed algorithm is quite robust against the erroneous assumption on the d_l's.

Given $\mathbf{A}(\boldsymbol{\Theta})$ and $\boldsymbol{\Gamma}$, it is obvious that the estimation of $\mathbf{W}$ follows the same method as with LOS channel case, i.e., by solving the UQP in (33). However, the estimation of $\mathbf{A}(\boldsymbol{\Theta})$ and $\boldsymbol{\Gamma}$ cannot be done separately, since $\boldsymbol{\Gamma}$ is block diagonal but not diagonal in the multipath case. Instead, we need to estimate $\boldsymbol{\gamma}_l\triangleq[\gamma_l^{(1)},\cdots\gamma_l^{(d_l)}]^T$ and $\boldsymbol{\Theta}_l\triangleq[\theta_l^{(1)},\cdots\theta_l^{(d_l)}]^T$ jointly for the lth user:

$$\hat{\boldsymbol{\gamma}}_l,\hat{\boldsymbol{\Theta}}_l=\arg\min_{\boldsymbol{\gamma}_l,\boldsymbol{\Theta}_l}\ \|\mathbf{y}_l-[(\mathbf{1}_N\otimes\mathbf{W})\odot\boldsymbol{\Phi}]\,\mathbf{A}(\boldsymbol{\Theta}_l)\boldsymbol{\gamma}_l\|_2^2. \tag{41}$$

We next provide two algorithms to solve the problem (41).

A. Repeated Orthogonal Projection Algorithm

With a given $\hat{\boldsymbol{\Theta}}_l$, by the definition of

$$\mathbf{B}\triangleq\left[(\mathbf{1}_N\otimes\hat{\mathbf{W}})\odot\boldsymbol{\Phi}\right]\in\mathbb{C}^{KN\times M}, \tag{42}$$

we can obtain the least-squares (and also the maximum likelihood) estimation of $\boldsymbol{\gamma}_l$ as

$$\hat{\boldsymbol{\gamma}}_l=\left(\mathbf{A}^H(\hat{\boldsymbol{\Theta}}_l)\mathbf{B}^H\mathbf{B}\mathbf{A}(\hat{\boldsymbol{\Theta}}_l)\right)^{-1}\mathbf{A}(\hat{\boldsymbol{\Theta}}_l)^H\mathbf{B}^H\mathbf{y}_l. \tag{43}$$

After inserting (43) into $\boldsymbol{\gamma}_l$ in (41), the estimation of $\hat{\boldsymbol{\Theta}}_l$ becomes

$$\begin{aligned}\hat{\boldsymbol{\Theta}}_l&=\arg\max_{\boldsymbol{\Theta}_l}\\&\quad\mathbf{y}_l^H\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l)\left(\mathbf{A}^H(\boldsymbol{\Theta}_l)\mathbf{B}^H\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l)\right)^{-1}\mathbf{A}^H(\boldsymbol{\Theta}_l)\mathbf{B}^H\mathbf{y}_l\\&=\arg\max_{\boldsymbol{\Theta}_l}\ \ \mathbf{y}_l^H\boldsymbol{\Lambda}_l\mathbf{y}_l.\end{aligned} \tag{44}$$

where

$$\begin{aligned}\boldsymbol{\Lambda}_l&\triangleq\mathcal{P}\left\{\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l)\right\}\\&=\mathbf{B}\mathbf{A}(\boldsymbol{\Theta})[\mathbf{A}^H(\boldsymbol{\Theta})\mathbf{B}^H\mathbf{B}\mathbf{A}(\boldsymbol{\Theta})]^{-1}\mathbf{A}^H(\boldsymbol{\Theta})\mathbf{B}^H\end{aligned} \tag{45}$$

is the projection matrix of $\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l)$.

If $d_l=1$, i.e. $\boldsymbol{\Theta}_l$ becomes a scalar $\theta_l^{(1)}$, program (44) can be solved directly by a one-dimensional (1-D) search:

$$\hat{\theta}_l^{(1)}=\arg\max_{\theta_l^{(1)}}\ \ \frac{\left|\mathbf{y}_l^H\mathbf{B}\boldsymbol{a}(\theta_l^{(1)})\right|^2}{\boldsymbol{a}^H(\theta_l^{(1)})\mathbf{B}^H\mathbf{B}\boldsymbol{a}(\theta_l^{(1)})}, \tag{46}$$

since

$$\mathcal{P}\{\mathbf{B}\boldsymbol{a}(\theta_l^{(1)})\}=\frac{\mathbf{B}\boldsymbol{a}(\theta_l^{(1)})\boldsymbol{a}^H(\theta_l^{(1)})\mathbf{B}^H}{\boldsymbol{a}^H(\theta_l^{(1)})\mathbf{B}^H\mathbf{B}\boldsymbol{a}(\theta_l^{(1)})}. \tag{47}$$

The problem in (44) is clearly NP-hard, when d_l is larger than 1 in multipath case. Yet, we next show that we could approximately estimate each element of $\boldsymbol{\Theta}_l$ separately with an affordable computational complexity. To illustrate this, we establish the following lemma based on the properties of projection matrix.

Lemma 1: Given a j-element subset of $\boldsymbol{\Theta}_l$:

$$\boldsymbol{\Theta}_l^{(j)}\triangleq\left\{\theta_l^{(1)},\theta_l^{(2)}\cdots\theta_l^{(j)}\right\}\quad 1\le j<d_l, \tag{48}$$

$\theta_l^{(j+1)}$ can be obtained by a 1-D search:

$$\hat{\theta}_l^{(j+1)}=\arg\max_{\theta_l^{(j+1)}}\ \ \frac{\left|\mathbf{y}_l^H\boldsymbol{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\theta_l^{(j+1)})\right|^2}{\boldsymbol{a}^H(\theta_l^{j+1})\mathbf{B}^H\boldsymbol{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\theta_l^{j+1})}, \tag{49}$$

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2: Repeated Orthogonal Projection Algorithm

Input: $\mathbf{B}$ in (42); number of multipath d_l
Output: Angle vector $\hat{\boldsymbol{\Theta}}_l$ and gain vector $\hat{\boldsymbol{\gamma}}_l$ of multipath channel;
1: Estimate $\hat{\theta}_l^{(1)}$ in (46);
2: **for** $j = 2 : d_l$ **do**
3: Set $\boldsymbol{\Theta}_l^{(j-1)} \leftarrow \{\hat{\theta}_l^{(1)} \cdots \hat{\theta}_l^{(j-1)}\}$ and use (49) to estimate $\theta_l^{(j)}$
4: Use (49) to update all $\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}, \cdots \hat{\theta}_l^{(j)}$ iteratively until "practical convergence";
5: **end for**
6: Estimate $\hat{\boldsymbol{\gamma}}_l$ with (43);
7: **return** $\hat{\boldsymbol{\Theta}}_l$ and $\hat{\boldsymbol{\gamma}}_l$

where $\boldsymbol{\Lambda}_{l,j}^{\perp} \triangleq \mathcal{P}^{\perp}\{\mathbf{BA}(\boldsymbol{\Theta}_l^{(j)})\} = \mathbf{I} - \mathcal{P}\{\mathbf{BA}(\boldsymbol{\Theta}_l^{(j)})\}$ is the orthogonal projection matrix of $\mathbf{BA}(\boldsymbol{\Theta}_l^{(j)})$.

Proof: See Appendix A. ■

Building on Lemma 1, we now propose a max-likelihood iterative method to estimate $\boldsymbol{\Theta}_l$. Firstly, set $d_l = 1$, and $\theta_l^{(1)}$ can be estimate by (46). Secondly, set $d_l = 2$ and use the obtained $\hat{\theta}_l^{(1)}$ to set $\boldsymbol{\Theta}_l^{(1)} \leftarrow \{\hat{\theta}_l^{(1)}\}$, then $\theta_l^{(2)}$ can be estimate by (49). Update $\theta_l^{(1)}$ by setting $\boldsymbol{\Theta}_l^{(1)} \leftarrow \{\hat{\theta}_l^{(2)}\}$ and using (49) again. Repeat to update $\theta_l^{(2)}$ and $\theta_l^{(1)}$ iteratively until "practical convergence," i.e, the change of $\hat{\theta}_l^{(2)}$ and $\hat{\theta}_l^{(1)}$ between two consecutive iteration are less than some small numbers. Thirdly, set $d_l = 3$ and $\boldsymbol{\Theta}_l^{(2)} \leftarrow \{\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}\}$, then $\theta_l^{(3)}$ can be estimated by (49). Repeat to update the obtained $\theta_l^{(1)}, \theta_l^{(2)}$ and $\theta_l^{(3)}$ iteratively in a similar manner until "practical convergence". Continue similarly until all desired $\theta_l^{(j)}$ are obtained. The practical convergences are guaranteed since each update monotonously decreases the objective function which is bounded below. Note that the 1-D search in (49) is repeatedly used for quite a few times, but due to its special structure we can adapt FFTs to accelerate the algorithm. We summarize the aforementioned procedure in Algorithm 2.

B. Extended RELAX Algorithm

Note that $\mathbf{A}(\boldsymbol{\Theta}_l)\boldsymbol{\gamma}_l$ in (41) can be considered as a sum of complex sinusoids, i.e., a mixed-spectrum signal. As a result, we can adopt a RELAX algorithm to jointly estimate $\mathbf{A}(\boldsymbol{\Theta})$ and $\boldsymbol{\Gamma}$. RELAX was first proposed in [29] which iteratively estimates complex sinusoidal parameters (frequencies and amplitudes). Different from Algorithm 2, RELAX estimates each pair of parameters, i.e, $\gamma_l^{(j)}$ and $\theta_l^{(j)}$, separately. We next provide an extend RELAX algorithm to solve (41).

Insert (42) into (41), we have:

$$\hat{\boldsymbol{\gamma}}_l, \hat{\boldsymbol{\Theta}}_l = \arg \min_{\boldsymbol{\gamma}_l, \boldsymbol{\Theta}_l} \quad \|\mathbf{y}_l - \mathbf{BA}(\boldsymbol{\Theta}_l)\boldsymbol{\gamma}_l\|_2^2 . \tag{50}$$

Due to the existence of $\mathbf{B}$, the objective function in (50) is slightly different from that with the original RELAX; yet, it does not prevent us from using the core idea of the RELAX algorithm.

Algorithm 3: Extended RELAX Algorithm

Input: $\mathbf{y}_l$ in (50); number of multipath d_l
Output: Angle vector $\hat{\boldsymbol{\Theta}}_l$ and gain vector $\hat{\boldsymbol{\gamma}}_l$ of multipath channel;
1: Set subsets $\boldsymbol{\Theta}_l^{(j)}$ and $\boldsymbol{\gamma}_l^{(j)}$ to be empty, estimate $\hat{\theta}_l^{(1)}$ and $\hat{\gamma}_l^{(1)}$ with (52) and (53);
2: **for** $j = 2 : d_l$ **do**
3: Set $\boldsymbol{\Theta}_l^{(j-1)} \leftarrow \{\hat{\theta}_l^{(1)} \cdots \hat{\theta}_l^{(j-1)}\}$ and $\boldsymbol{\gamma}_l^{(j-1)} \leftarrow \{\hat{\gamma}_l^{(1)} \cdots \hat{\gamma}_l^{(j-1)}\}$ and calculate $\mathbf{y}_l^{(j)}$ with (51);
4: Estimate $\hat{\theta}_l^{(j)}$ and $\hat{\gamma}_l^{(j)}$ with (52) and (53);
5: Use (52) and (53) to update all $\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}, \cdots \hat{\theta}_l^{(j)}$ and $\hat{\gamma}_l^{(1)}, \hat{\gamma}_l^{(2)}, \cdots \hat{\gamma}_l^{(j)}$ iteratively until "practical convergence";
6: **end for**
7: **return** $\hat{\boldsymbol{\Theta}}_l$ and $\hat{\boldsymbol{\gamma}}_l$

Assume that two j-element subsets $\boldsymbol{\Theta}_l^{(j)}$ and $\boldsymbol{\gamma}_l^{(j)}$ of $\boldsymbol{\Theta}_l$ and $\boldsymbol{\gamma}_l$ are given. By the definition of

$$\mathbf{y}_l^{(j)} \triangleq \mathbf{y}_l - \mathbf{BA}(\boldsymbol{\Theta}_l^{(j)})\boldsymbol{\gamma}_l^{(j)}, \tag{51}$$

the estimations of $\theta_l^{(j+1)}$ and $\gamma_l^{(j+1)}$ could be obtained by:

$$\hat{\theta}_l^{(j+1)} = \arg \max_{\theta_l^{(j+1)}} \quad \frac{\left|\boldsymbol{a}^H(\theta_l^{(j+1)})\mathbf{B}\mathbf{y}_l^{(j)}\right|^2}{\boldsymbol{a}^H(\theta_l^{(j+1)})\mathbf{B}^H\mathbf{B}\boldsymbol{a}(\theta_l^{(j+1)})}, \tag{52}$$

and

$$\hat{\gamma}_l^{(j+1)} = \frac{\mathbf{y}_l^{(j)^H}\mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\boldsymbol{a}(\hat{\theta}_l^{(j+1)})}{M}. \tag{53}$$

Note that (52) can be efficiently computed by using FFT while (53) is closed-form. Based on (52) and (53), we summarize the extended RELAX in Algorithm 3. Since RELAX can also be applied to the 2-D data [29], Algorithm 3 can be easily extended to estimate channel parameters when a URA is deployed at the BS. Hence, our proposed calibration approach can carry over to the URA case.

Remark 3: RELAX is computationally simpler than Alogrihtm 2, since it avoids the projection matrix calculation. However, as an sub-optimum method, it tends to have significantly larger error if $\boldsymbol{\Theta}_l$ includes two or more quite close elements when M is small. Hence, we should use Algorithm 2 when a relatively small antenna array is deployed in the BS. However, in massive MIMO scenarios, a large M increases the linear independence of each $\boldsymbol{a}(\theta_l^{(i)})$ vector; the estimation errors can be then significantly reduced. Numerical results in Section VI show that RELAX performs as good as Algorithm 2 when M is large (e.g., $M \geq 64$). Therefore, we can use RELAX to replace Algorithm 2 when M is large enough to reduce computational complexity.

C. Overall Algorithm

Combining the power-method iterations with Algorithm 2 or Algorithm 3, we can iteratively minimize the objective function

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

Algorithm 4: Calibration of PSN in Multipath Channel Case

Input: number of antennas M; number of RF chains K; number of beamformers N; number of users L; measured effective channel vectors $\mathbf{Y}$;
Output: Phase deviation matrix $\hat{\mathbf{W}} \in \mathbb{C}^{K\times M}$;
1: Random initialization of $\hat{\mathbf{W}}$ and $\hat{\mathbf{\Theta}}$
2: **do**
3: **for** $l = 1 : L$ **do**
4: Calculate $\mathbf{B}$ with (42);
5: Use Algorithm 2 or Algorithm 3 to estimate $\hat{\gamma}_l$ and $\hat{\mathbf{\Theta}}_l$;
6: **end for**
7: **for** $k = 1 : K$ **do**
8: Estimate $\hat{\mathbf{W}}(k,:)$ in (33) via the power method;
9: **end for**
10: $\hat{\mathbf{Y}} = [(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}]\mathbf{A}(\hat{\mathbf{\Theta}})\hat{\mathbf{\Gamma}}$
11: **while** change in $\|\hat{\mathbf{Y}}\|_F$ from the previous iteration is less than a small number ϵ
12: **return** $\hat{\mathbf{W}}$;
13: Perform calibration according to $\hat{\mathbf{W}}$

in (15) to obtain $\mathbf{A}(\hat{\mathbf{\Theta}})$, $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{W}}$. The PSN calibration method in multipath channel case is summarized in Algorithm 4.

Remark 4: Compared to Algorithm 1, Algorithm 4 does not incur significantly increased complexity. In particular, the two inner for-loops in Algorithm 4 have affordable computational complexity. The power-method in the second for-loop has a $O(M)$ complexity and converges fast. In the first for-loop, the complexity of projection operation in Algorithm 2 is $O(M^3)$; yet, notice that we only need to use it when M is relatively small. For large M, we employ RELAX which is more efficient as it avoids the inverse operation in projection matrix calculation and takes advantage of FFT with $O(M\log(M))$ complexity. Moreover, poor scattering nature of mmWave implies that the number d_l of scatters would not be large, e.g, typically $d_l = 2$ or 3, which ensures that the number of iterations inside Algorithm 2 or RELAX is limited. Numerical results also verify that the outer do-while-loop can converge after about 30 iterations. Hence, the overall complexity with Algorithm 4 is still affordable as the over-the-air calibration only needs to be implemented just a few times per hour at the BS.

V. Cramer-Rao Lower Bound

In this section, we derive the CRLBs to gauge the performance for the phase deviation estimations with the proposed algorithms. To this end, we first show that ambiguities cannot be avoided in the phase deviation estimation regardless of the number of measurements (training beamformers) N taken during the calibration.

A. Ambiguities in Phased Array Calibration

Checking the cost function in (15), we can easily see that $\mathbf{W}$, $\mathbf{\Theta}$, and $\mathbf{\Gamma}$ cannot be uniquely determined. If $\hat{\mathbf{W}}$ and $\hat{\mathbf{\Gamma}}$ are solutions to (15), so are $e^{j\beta}\hat{\mathbf{W}}$ and $e^{-j\beta}\hat{\mathbf{\Gamma}}$, $\forall\beta \in [0, 2\pi)$. Moreover, if $\hat{\mathbf{W}}$ and $\mathbf{A}(\hat{\mathbf{\Theta}})$ are solutions to (15), so are $\hat{\mathbf{W}}\mathbf{T}$ and $\mathbf{T}^{-1}\mathbf{A}(\hat{\mathbf{\Theta}})$ for a diagonal matrix

$$\mathbf{T} = \mathcal{D}([1,\ e^{j\alpha},\ \ldots\ , e^{j(M-1)\alpha}]), \forall\alpha \in [0, 2\pi) \tag{54}$$

as $\hat{\mathbf{W}}\mathbf{T}$ remains unimodular and $\mathbf{T}^{-1}\mathbf{A}(\hat{\mathbf{\Theta}})$ is still a Vandermonde matrix. These ambiguities exist in both multipath and LOS channel cases, regardless of the number of measurements N.

Fortunately, the phase ambiguity does not impact the performance of hybrid beamforming, since the power-radiation pattern for each user remains the same with ambiguities:

$$||(\mathbf{W} \odot \mathbf{\Phi})\mathbf{A}(\mathbf{\Theta}_l)\boldsymbol{\gamma}_l||^2 = ||(e^{-j\beta}\mathbf{W}\mathbf{T} \odot \mathbf{\Phi})\mathbf{T}^{-1}\mathbf{A}(\mathbf{\Theta}_l)\boldsymbol{\gamma}_l||^2 \tag{55}$$

Hence we can, without loss of generality, set $\omega_{11} = 0$ and $\theta_1 = 0$. It is worth noting that such constraints need not to be imposed when running Algorithm 1 or 4. Instead, after obtaining a solution, we can simply remove the ambiguity by setting $\omega_{km} \leftarrow (\omega_{km} - \omega_{11})$ and $\theta_l^{(i)} \leftarrow (\theta_l^{(i)} - \theta_1^{(1)})$.

B. Derivation of CRLB

We next derive the CRLB for the intended phase deviation estimations. To this end, define

$$\boldsymbol{\eta} \triangleq [\mathbf{\Omega}^T, \mathbf{\Theta}^T, \mathrm{Re}\{\boldsymbol{\gamma}^T\}, \mathrm{Im}\{\boldsymbol{\gamma}^T\}]^T \in \mathbb{R}^{MK+3(d_1+\cdots+d_L)-2}, \tag{56}$$

where

$$\mathbf{\Omega} = [\omega_{21}, \ldots, \omega_{K1}, \omega_{12}, \ldots, \omega_{KM}] \in \mathbb{R}^{MK-1}, \tag{57}$$

are the phases of $\mathbf{W}$,

$$\mathbf{\Theta} = [\theta_1^{(2)}, \ldots, \theta_1^{(d_1)}, \theta_2^{(1)}, \ldots, \theta_L^{(d_L)}] \in \mathbb{R}^{d_1+\cdots+d_L-1}, \tag{58}$$

and

$$\boldsymbol{\gamma} = \left[\gamma_1^{(1)}, \ldots, \gamma_1^{(d_1)}, \gamma_2^{(1)}, \ldots, \gamma_L^{(d_L)}\right] \in \mathbb{R}^{d_1+\cdots+d_L}. \tag{59}$$

Note that ω_{11} and $\theta_1^{(1)}$ are not included since they can be set, without loss of optimality, to be 0 to avoid the estimation ambiguity as mentioned in SubSection V-A. Using the fact that $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, we vectorize $\mathbf{Y}_n^T$ in (8) into

$$\begin{aligned}\mathbf{y}_{\mathrm{vec}}^{(n)} &\triangleq \mathrm{vec}(\mathbf{Y}_n^T) \\ &= \left[\mathbf{I}_K \otimes \left(\mathbf{\Gamma}^T\mathbf{A}^T(\mathbf{\Theta})\right)\right]\mathcal{D}(\mathbf{\Phi}_{\mathrm{vec}}^{(n)})\mathbf{w}_{\mathrm{vec}} + \mathbf{z}_n \in \mathbb{C}^{KL\times 1}\end{aligned} \tag{60}$$

where $\mathbf{w}_{\mathrm{vec}} = \mathrm{vec}(\mathbf{W}^T)$ and $\mathbf{\Phi}_{\mathrm{vec}}^{(n)} = \mathrm{vec}({\mathbf{\Phi}_n}^T)$. Clearly, we have

$$\mathbf{y}_{\mathrm{vec}}^{(n)} \sim \mathcal{C}N(\boldsymbol{\mu}_n, \sigma_z^2\mathbf{I}), \tag{61}$$

where

$$\boldsymbol{\mu}_n = \left[\mathbf{I}_K \otimes \left(\mathbf{\Gamma}^T\mathbf{A}^T(\mathbf{\Theta})\right)\right]\mathcal{D}(\mathbf{\Phi}_{\mathrm{vec}}^{(n)})\mathbf{w}_{\mathrm{vec}} \in \mathbb{C}^{KL\times 1}. \tag{62}$$

The Fisher information matrix (FIM) can be derived as [30]

$$\mathbf{F} = \frac{2}{\sigma_z^2}\sum_{n=1}^{N}\mathrm{Re}\left[\frac{\partial\boldsymbol{\mu}_n^H(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\frac{\partial\boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right]. \tag{63}$$

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

We are now set to compute different parts of $\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$ separately. First, using the fact that $\frac{\partial e^{j\omega}}{\partial \omega} = j e^{j\omega}$, we have

$$\frac{\partial \mathbf{w}_{\text{vec}}(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega}} = j\mathcal{D}(\mathbf{w}_{\text{vec}}). \tag{64}$$

It readily follows that

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}} = j\left[\mathbf{I}_K \otimes \left(\boldsymbol{\Gamma}^T \mathbf{A}^T(\boldsymbol{\Theta})\right)\right] \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n) \times \left[\mathcal{D}(\mathbf{w}_{\text{vec}})(:, 2:KM)\right] \in \mathbb{C}^{KL\times(MK-1)}, \tag{65}$$

where $\mathcal{D}(\mathbf{w}_{\text{vec}})(:, 2:KM)$ stands for taking all but the first column of the matrix $\mathcal{D}(\mathbf{w}_{\text{vec}})$.

Second, based on the fact that

$$\frac{\partial \boldsymbol{a}(\theta)}{\partial \theta} = j\pi \mathbf{D}_\theta \boldsymbol{a}(\theta), \tag{66}$$

where

$$\mathbf{D}_\theta = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \cos\theta & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & (M-1)\cos\theta \end{bmatrix}, \tag{67}$$

we have

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}} = j\pi \times \left[\mathbf{q}_1^{(2)}, \ldots, \mathbf{q}_1^{(d_1)}, \mathbf{q}_2^{(1)}, \ldots, \mathbf{q}_2^{(d_2)}, \ldots, \mathbf{q}_L^{(1)}, \ldots, \mathbf{q}_L^{(d_L)}\right] \in \mathbb{C}^{KL\times(d_1+\cdots+d_L-1)}, \tag{68}$$

and each column of which is

$$\mathbf{q}_i^{(j)} = \mathbf{I}_K \otimes \begin{bmatrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{bmatrix} i-1 \\ \gamma_i^{(j)} \boldsymbol{a}^T(\theta_i^{(j)}) \mathbf{D}_{\theta_i^{(j)}} \\ \begin{bmatrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{bmatrix} L-i \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n)\mathbf{w}_{\text{vec}}$$

$$\text{for} \quad i = 1, 2, \cdots L \quad \text{and} \quad j = 1, 2, \ldots, d_i. \tag{69}$$

Third, we have

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Re}\{\boldsymbol{\gamma}\}} = \left[\mathbf{r}_1^{(1)}, \ldots, \mathbf{r}_1^{(d_1)}, \mathbf{r}_2^{(1)}, \ldots, \mathbf{r}_2^{(d_2)}, \ldots, \mathbf{r}_L^{(1)}, \ldots, \mathbf{r}_L^{(d_L)}\right] \in \mathbb{C}^{KL\times(d_1+\cdots+d_L)}, \tag{70}$$

where

$$\mathbf{r}_i^{(j)} = \mathbf{I}_K \otimes \begin{bmatrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{bmatrix} i-1 \\ \boldsymbol{a}^T(\theta_i^{(j)}) \\ \begin{bmatrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{bmatrix} L-i \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n)\mathbf{w}_{\text{vec}}$$

$$\text{for} \quad i = 1, 2, \cdots L \quad \text{and} \quad j = 1, 2, \ldots, d_i. \tag{71}$$

and similarly

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Im}\{\boldsymbol{\gamma}\}} = j\frac{\partial \boldsymbol{\mu}(\boldsymbol{\eta})}{\partial \operatorname{Re}\{\boldsymbol{\gamma}\}}. \tag{72}$$

Based on (65), (69), (72), and (74), we obtain

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Re}\{\boldsymbol{\gamma}\}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Im}\{\boldsymbol{\gamma}\}}\right] \times \in \mathbb{C}^{KL\times(MK+3(d_1+\cdots+d_L)-2)}. \tag{73}$$

The FIM follows from inserting (75) into (63), by which we could directly establish Lemma 2.

Lemma 2: The Fisher information of $\{\omega_{km}\}_{k=1,m=1,km\neq 1}^{K,M}$ based on the observation $\mathbf{Y}$ is the first $KM-1$ diagonal elements of the FIM, which is given by

$$\mathcal{I}(\omega_{km}) = \frac{2N}{\sigma_z^2} \sum_{l=1}^{L} \left| \sum_{i=1}^{d_l} \gamma_l^{(i)} e^{-j\frac{2\pi(m-1)d}{\lambda}\sin\theta_l^{(i)}} \right|^2, \tag{74}$$

and it can be simplified as

$$\mathcal{I}(\omega_{km}) = \frac{2N}{\sigma_z^2} \sum_{l=1}^{L} |\gamma_l|^2, \tag{75}$$

in LOS channel case.

Proof: See Appendix B. ■

Note that Fisher information measures the amount of information that the observable/measurable random variable $\mathbf{Y}$ carries about the unknown parameters $\boldsymbol{\Gamma}$, $\mathbf{A}(\boldsymbol{\Theta})$ and $\mathbf{W}$. By Lemma 2, we can see that the Fisher information of ω_{km} grows up linearly as N increases regardless of the value of $\boldsymbol{\Phi}_n$. This justifies the intuition that finer estimations can be achieved with a larger number of measurements (or training beamformers in use).

Having obtained the FIM, the CRLBs for the estimates in (15) are given by its inverse:

$$\mathbf{C}_{\hat{\boldsymbol{\eta}}} = \mathbf{F}^{-1}. \tag{76}$$

In particular, the first $KM-1$ diagonal elements of $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ dictate the lower bounds to the variances of estimates $\hat{\boldsymbol{\Omega}}$. We could use it as a benchmark to evaluate the effectiveness of the proposed algorithms.

C. Number of Required Measurements

Lemma 2 states that the larger number of PSN training beamformers (i.e., the number N) employed for the calibration measurements, the better performance of the phase estimation and the ensuing calibration. Yet, the number N is also desired to be as small as possible to spare scarce time-frequency resources. By counting the degrees of freedom (DOF) of the variables to estimate, we can readily derive a lower bound for N to ensure the phase deviation $\boldsymbol{\Omega}$ is estimable. Note that if $\boldsymbol{\Omega}$ is estimable, the FIM in (63) would be non-singular. On the contrary, the FIM would be singular if $\boldsymbol{\Omega}$ is not estimable, since the CRLBs obtained from the inverse become FIM is infinity.

In the LOS channel case, the DOF of $\boldsymbol{\Omega}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Gamma}$ are

$$\text{DOF}(\boldsymbol{\Omega}) = MK-1, \ \text{DOF}(\boldsymbol{\Theta}) = L-1, \ \text{DOF}(\boldsymbol{\Gamma}) = 2L, \tag{77}$$

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

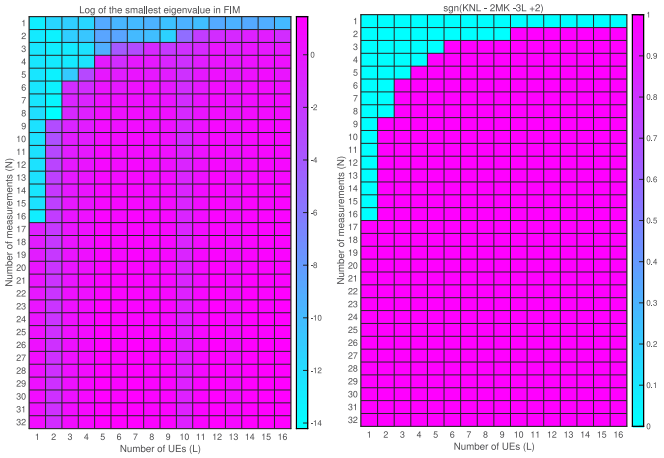

Fig. 2. The left picture shows the logarithm of the smallest eigenvalue of FIM when L and N vary for a ULA with $M = 32$ antennas and $K = 4$ RF chains; the right picture visualizes the theoretical lower bound of N, where if $2KNL - MK - 3L + 2 \geq 0$, sign function returns 1 (purple), otherwise it returns 0 (blue).

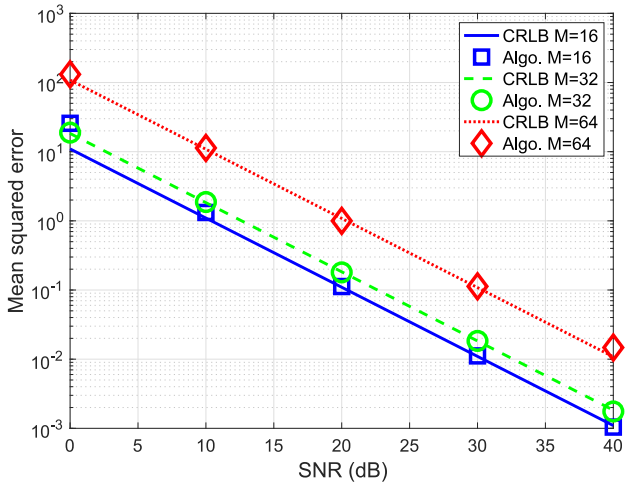

Fig. 3. $\text{MSE}_{\mathbf{\Omega}}$ and CRLB vs. SNR with $N = M, L = 4, K = M/8$ in LOS channel environments.

respectively, while the number of observations in the matrix $\mathbf{Y}$ is $2KNL$. To ensure that the problem (15) is feasible, i.e., the phase deviation $\mathbf{\Omega}$ is estimable, we need to have

$$2KNL \geq MK + 3L - 2. \tag{78}$$

Hence, the lower bound of N is

$$N \geq \left\lceil \frac{M}{2L} + \frac{3}{2K} - \frac{1}{KL} \right\rceil. \tag{79}$$

Similarly, in multipath channel case, the DOF of $\mathbf{\Theta}$ and $\mathbf{\Gamma}$ change to

$$\text{DOF}(\mathbf{\Theta}) = \sum_{l=1}^{L} d_l - 1, \ \text{DOF}(\mathbf{\Gamma}) = 2\sum_{l=1}^{L} d_l. \tag{80}$$

The lower bound of N then becomes

$$N \geq \left\lceil \frac{M}{2L} + \frac{3\sum_{l=1}^{L} d_l - 2}{2KL} \right\rceil. \tag{81}$$

To further illustrate the impact of N and L, we consider a LOS channel scenario where a BS is equipped with a ULA of $M = 32$ antennas and $K = 4$ RF chains. We calculate the FIM as N and L vary from 1 to 32 and 1 to 16. Fig. 2 (left) shows the logarithm of the smallest eigenvalue of every FIM. Blue pixels represent the cases with the smallest eigenvalues less than 10^{-10}, which can be essentially regarded as numerical zero. In such cases, the FIM is essentially singular and the phase deviation estimation problem in (15) is infeasible. Purple pixels indicate the cases with the smallest eigenvalues much greater than 10^{-10}, implying that the corresponding problem is feasible. In Fig. 2 (right), we visualize the lower bound of N in (81) by using a signum (sgn) function (i.e, if the inequality holds, the function returns 1 otherwise returns 0). It is clearly observed that the two sub-figures are perfectly matched for each pair of N and L. This demonstrates that the lower bound in (81) is tight; i.e., it indeed provides the minimum number for the required measurements to ensure feasible phase calibration. Similar results can be obtained for (83) in non-LOS channel case.

Note that the lower bounds in (81) and (83) are only necessary conditions of having the problem (15) estimable; they do not ensure that the obtained estimates are accurate. In practice, Algorithm 1 and Algorithm 4 may converge to *local* optimums, when N is small even though it satisfies (81) and (83). The numerical results in Section VI show that the phase deviation estimation errors with the proposed algorithms would be close to the CRLBs when the number of measurements (N) is larger than about 5 times the lower bounds provided in (81) and (83).

VI. Numerical Results

We provide numerical results to evaluate the performance of the proposed algorithms. In the simulations, we assume a BS equipped with an M-element ULA as shown in Fig. 1. We consider four cases: $M = 16, 32, 64$ and 128. The number of RF chains is set to be $K = M/8$, and the number of UEs involved for calibration is $L = 4$. Suppose that angles of arrival (AoA) of all UEs follow the uniform distribution over $[0, 2\pi)$ for both LOS and non-LOS channel cases, and the angular spread is 10 degrees around the AoA for non-LOS channel case. The complex-valued gains, $\{\gamma_l\}_{l=1}^{L}$, are set to be unit-modulus with random phases. We use the mean squared error of $\mathbf{\Omega}$

$$\text{MSE}_{\mathbf{\Omega}} = \mathbb{E}[\|\mathbf{\Omega} - \hat{\mathbf{\Omega}}\|_F^2] \tag{82}$$

to measure the calibration performance. Each $\text{MSE}_{\mathbf{\Omega}}$ value is obtained as the average over 1000 Monte Carlo trials. The CRLB for $\text{MSE}_{\mathbf{\Omega}}$ is given by the sum of the first $KM - 1$ diagonal elements of $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ in (78). The signal-to-noise ratio (SNR) is defined as $\text{SNR} \triangleq \frac{\|[(\mathbf{1}_N \otimes \mathbf{W}) \odot \mathbf{\Phi}]\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma}\|_F^2}{KNL\sigma_z^2}$, where σ_z^2 is the variance of the measurement noise.

Fig. 3 and Fig. 4 compare $\text{MSE}_{\mathbf{\Omega}}$ to the CRLB under different M with SNR varying from 0 dB to 40 dB, where $N = M$ randomly generated (different) PSN training beamformers are applied to obtain the measurements. Fig. 3 provides results for LOS channel cases, while Fig. 4 provides results for multipath

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

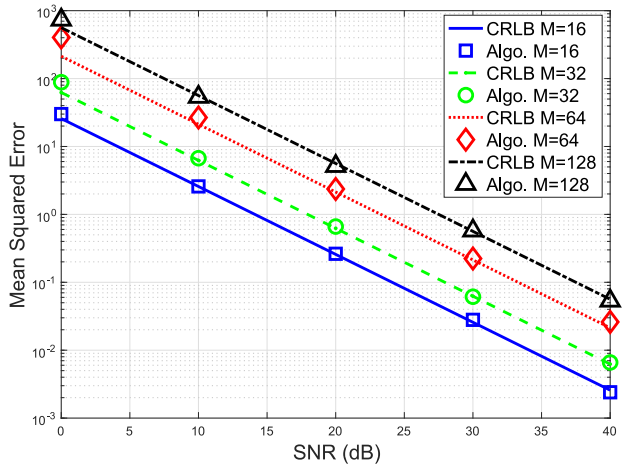

Fig. 4. $\text{MSE}_{\mathbf{\Omega}}$ and CRLB vs. SNR with $N = M, L = 4, K = M/8$ in multipath channel environments.

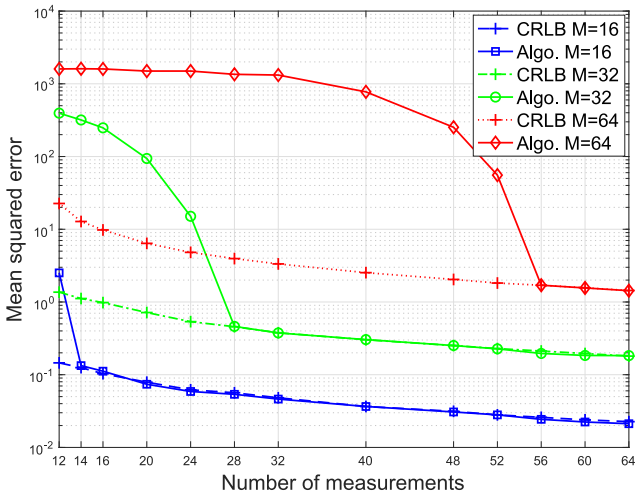

Fig. 5. Effect on calibration by N value when $M = 32, K = 4, L = 4$, and SNR $= 20$ dB in LOS channel environments.

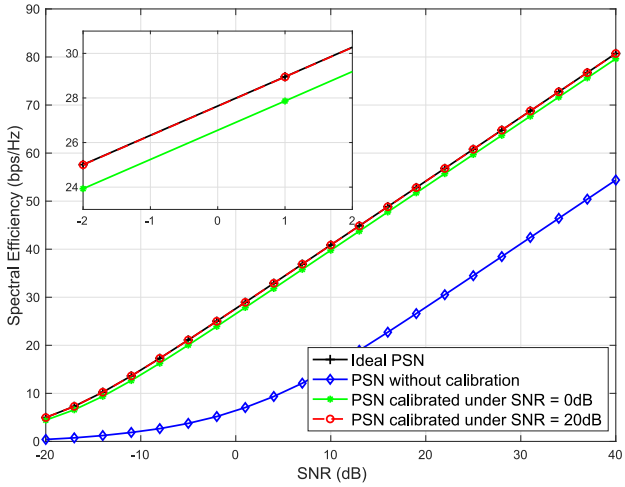

Fig. 6. Spectral efficiency when $M = 256$ and $K = 4$. The measurements for calibration are taken through $L = 4$ UEs in LOS channel environments.

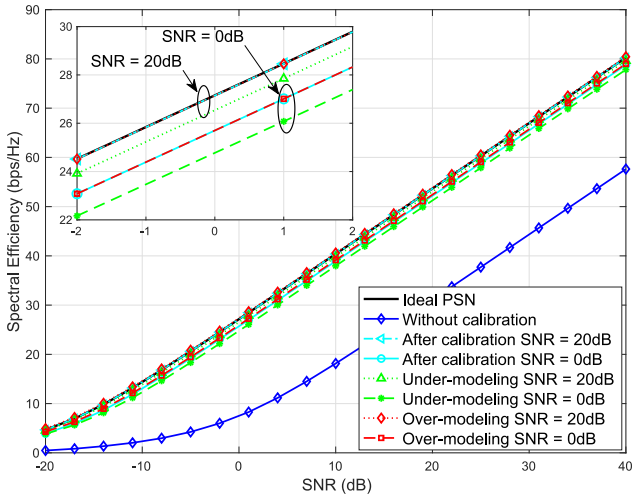

Fig. 7. Spectral efficiency when $M = 256$ and $K = 4$. The measurements for calibration are taken through $L = 4$ UEs in multipath channel environments.

channel cases where we set $\{d_l\}_{l=1}^{L} = 3$. We can observe that for all different M, $\text{MSE}_{\mathbf{\Omega}}$ is close to the CRLB even when SNR is low. This verifies the effectiveness of the proposed Algorithm 1 and 4. Note that for multipath channel cases, we use Algorithm 2 to estimate channel parameters when $M = 16$ and 32, and replace it by Algorithm 3 when $M = 64$ and 128 to reduce computational complexity.

In the second test case, we set $M = 16, 32,$ and $64, K = M/8$, $L = 4$, and SNR $= 20$ dB in LOS channel scenario. We change N, the number of the PSN training beamformers (i.e., the number of measurements), to check its impact on the performance of the phase deviation estimation. Fig. 5 shows the comparison of $\text{MSE}_{\mathbf{\Omega}}$ and CRLB. It is observed that the $\text{MSE}_{\mathbf{\Omega}}$ converges to the CRLB as N increases, and a "threshold effect" exists: the $\text{MSE}_{\mathbf{\Omega}}$ is much higher than the CRLB when N is less than 14, 28 and 56 for $M = 16, 32$ and 64, respectively. This indicates that although the lower bound of N in (81) ensures the phase deviation is estimable, it cannot guarantee the estimation errors close to the CRLBs. It is because the proposed iterative algorithms tend to converge to a *local* optimum when N is small. Simulations show that we may need about 5 times the lower bound of N to ensure the estimation errors close to the CRLBs. An improved method that can mitigate the "threshold effect" is left to future research.

We next show the necessity of PSN calibration. In the simulations, we assume that $M = 256$ antenna elements and $K = 4$ RF chains are deployed at the BS and the channel measurements for calibration are taken through $L = 4$ UEs when the BS applies $N = M$ training beamformers in LOS channel environments. We use spectral efficiency to evaluate the system capacity before and after the calibration. The ideal precoder is calculated via the hybrid beamforming method proposed in [31]. As shown in Fig. 6, the achieved spectral efficiency with uncalibrated random phase deviations in the PSN has about 20 dB performance loss when compared to ideal PSN. On the other hand, after the PSN is calibrated by the estimated phase deviations, the spectral efficiency is improved dramatically. With the measurements taken under SNR $= 20$ dB, the spectral efficiency with the calibrated PSN performs almost the same as in the ideal case. Even when the measurements are taken under SNR $= 0$ dB, the

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

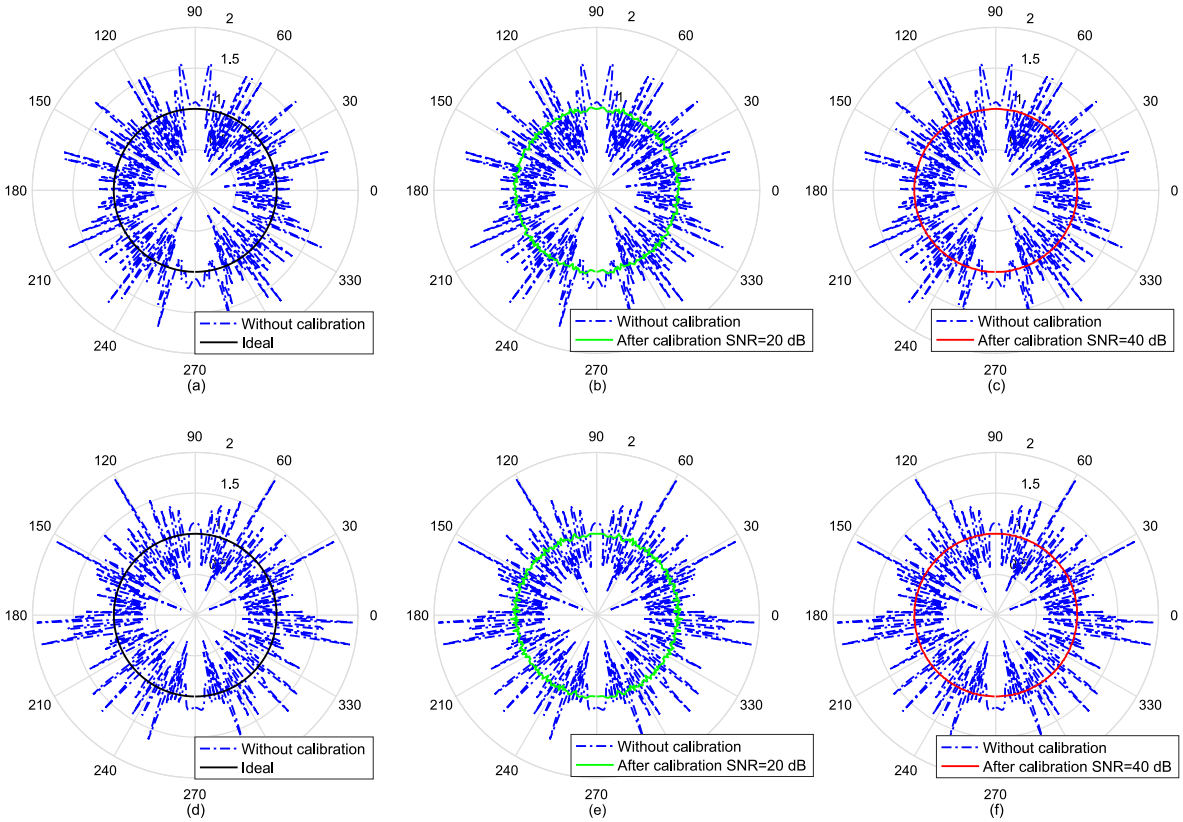

Fig. 8. Power-radiation pattern of omnidirectional beamforming when $M = 256$ and $K = 2$. The measurements for calibration are taken through $L = 4$ UEs in LOS channel environments for (b) and (c) and in non-LOS channel environments for (e) and (f).

spectral efficiency performance has only about 0.8 dB loss when compared to the ideal case.

Given the same BS configuration, similar performance comparisons are shown in Fig. 7 when the measurements are taken in the multipath channel environments. The spectral efficiency performance is almost the same or has 1 dB loss when compared to the ideal case, after the calibration under SNR $= 20$ dB and SNR $= 0$ dB, respectively. In multipath channel case, it is possible that we may estimate the multipath number of each UE channel incorrectly due to the noise. We consider both over-modeling and under-modeling cases to test the robustness of the proposed method. Suppose that the actual number of multipaths $d_l = 3 \ \ \forall l = 1 \cdots L$ for each UE. In the over-modeling case, we adopt $d_l = 4 \ \ \forall l = 1 \cdots L$ in the proposed Algorithm 4. Simulation results demonstrate that over-modeling hardly affects the algorithm performance, as the achieved spectral efficiency is very closed to the ideal one. This is because the final estimates of over-modeling parameters, i.e., $\boldsymbol{\gamma}_l^{(4)} \ \ \forall l = 1 \cdots L$ are close to 0, so that over-modeling does not harm the algorithm performance. On the other hand, in the under-modeling case, we (mistakenly) use $d_l = 2 \ \ \forall l = 1 \cdots L$ in the algorithm. Although we miss some channel state information when estimating $\mathbf{A}(\hat{\boldsymbol{\Theta}})$ and $\hat{\boldsymbol{\Gamma}}$, the performance with under-modeling calibration is only $0.5 - 0.8$ dB worse than the ideal case.

Finally, we use an illustrative example to show the impact of PSN calibration. Consider a ULA with $M = 256$ antennas and $K = 2$ RF chains using a pair of complementary sequences as beamforming vectors in the PSN to achieve omnidirectional transmission [27], [32], [33]. An ideal power radiation pattern is perfectly omnidirectional shown by the solid lines in Fig. 8(a) and Fig. 8(d). However, without calibration, the power radiation pattern exhibits very large fluctuations due to the existence of phase deviations, as shown by the dotted lines. On the other hand, after calibration using the proposed Algorithm 1 and Algorithm 4 with $L = 4$ UEs in the LOS and the non-LOS channel cases, respectively, we can perform effective omnidirectional transmission again. Specifically, given the measurements with SNR of 20 dB, the fluctuations of the power radiation pattern are greatly reduced as shown in Fig. 8(b) for LOS channel and Fig. 8(e) for non-LOS channel; and given the measures with SNR of 40 dB, the power radiation pattern becomes almost ideally omnidirectional again as shown in Fig. 8(c) for LOS channel and Fig. 8(f) for non-LOS channel.

VII. Conclusion

We addressed the calibration of the PSN for mmWave massive MIMO systems. Two novel iterative algorithms were developed to estimate the phase deviations with the PSN in LOS and non-LOS channel cases, respectively. In addition, the CRLBs for the phase deviation estimates were derived. From the CRLB analysis, we also revealed the minimum number of measurements to ensure that the phase calibration is achievable. The simulation result validates the effectiveness of the proposed

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

algorithms by showing that the MSEs of the estimates can be close to the CRLBs.

APPENDIX

A. Proof of Lemma 1

Suppose that a matrix $\mathbf{X}$ can be decomposed by columns as $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$. We have

$$\mathcal{P}\{\mathbf{X}\} = \mathcal{P}\{\mathbf{X}_1\} + \mathcal{P}\{\mathcal{P}^{\perp}\{\mathbf{X}_1\}\mathbf{X}_2\} \tag{83}$$

based on the blockwise formula of projection matrix [34]. Decompose matrix $\mathbf{BA}(\mathbf{\Theta}_l^{(j+1)})$ into $[\mathbf{BA}(\mathbf{\Theta}_l^{(j)}) \quad \mathbf{B}\boldsymbol{a}(\boldsymbol{\theta}_l^{(j+1)})]$ and insert it into problem (44). We have:

$$\arg \max_{\mathbf{\Theta}_l^{(j)}, \boldsymbol{\theta}_l^{(j+1)}} \mathbf{y}_l^H \left[\mathcal{P}\{\mathbf{BA}(\mathbf{\Theta}_l^{(j)})\} + \mathcal{P}\{\mathcal{P}^{\perp}\{\mathbf{BA}(\mathbf{\Theta}_l^{(j)})\}\mathbf{B}\boldsymbol{a}(\boldsymbol{\theta}_l^{(j+1)})\}\right] \mathbf{y}_l. \tag{84}$$

Given $\mathbf{\Theta}_l^{(j)}$, (86) becomes

$$\arg \max_{\theta_l^{(j+1)}} \quad \mathbf{y}_l^H \mathcal{P}\{\mathcal{P}^{\perp}\{\mathbf{BA}(\mathbf{\Theta}_l^{(j)})\}\mathbf{B}\boldsymbol{a}(\boldsymbol{\theta}_l^{(j+1)})\}\mathbf{y}_l. \tag{85}$$

By the definition of $\mathbf{\Lambda}_{l,j}^{\perp} \triangleq \mathcal{P}^{\perp}\{\mathbf{BA}(\mathbf{\Theta}_l^{(j)})\}$, we have:

$$\mathcal{P}\{\mathcal{P}^{\perp}\{\mathbf{BA}(\mathbf{\Theta}_l^{(j)})\}\mathbf{B}\boldsymbol{a}(\boldsymbol{\theta}_l^{(j+1)})\} = \frac{\mathbf{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\theta_l^{(j+1)})\boldsymbol{a}^H(\theta_l^{(j+1)})\mathbf{B}^H\mathbf{\Lambda}_{l,j}^{\perp}}{\boldsymbol{a}^H(\boldsymbol{\theta}_l^{j+1})\mathbf{B}^H\mathbf{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\boldsymbol{\theta}_l^{j+1})}. \tag{86}$$

Inserting (88) into (87), the estimation of $\theta_l^{(j+1)}$ becomes a 1-D search:

$$\hat{\theta}_l^{(j+1)} = \arg \max_{\theta_l^{(j+1)}} \quad \frac{\left|\mathbf{y}_l^H\mathbf{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\theta_l^{(j+1)})\right|^2}{\boldsymbol{a}^H(\theta_l^{j+1})\mathbf{B}^H\mathbf{\Lambda}_{l,j}^{\perp}\mathbf{B}\boldsymbol{a}(\theta_l^{j+1})}. \tag{87}$$

B. Proof of Lemma 2

The Fisher information of $\{\omega_{km}\}_{k=1,m=1,km\neq 1}^{K,M}$ is given by the diagonal elements of

$$\mathbf{F}_{\mathbf{\Omega}} = \frac{2}{\sigma_z^2}\sum_{n=1}^{N} \mathrm{Re}\left[\frac{\partial \boldsymbol{\mu}_n^H(\boldsymbol{\eta})}{\partial \mathbf{\Omega}}\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathbf{\Omega}}\right]. \tag{88}$$

Recall

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathbf{\Omega}} = j\left[\mathbf{I}_K \otimes \left(\mathbf{\Gamma}^T\mathbf{A}^T(\mathbf{\Theta})\right)\right]\mathcal{D}(\mathbf{\Phi}_{\text{vec}}^n) \times [\mathcal{D}(\mathbf{w}_{\text{vec}})(:, 2:KM)] \in \mathbb{C}^{KL\times(MK-1)}, \tag{89}$$

Inserting (91) into (90), we have

$$\mathbf{F}_{\mathbf{\Omega}} = \frac{2}{\sigma_z^2}\sum_{n=1}^{N} \mathrm{Re}\left\{\mathcal{D}^H(\mathbf{\Phi}_{\text{vec}}^n)\left[\mathcal{D}^H(\mathbf{w}_{\text{vec}})(:, 2:KM)\right] \times \mathbf{M}\mathcal{D}(\mathbf{\Phi}_{\text{vec}}^n)\left[\mathcal{D}(\mathbf{w}_{\text{vec}})(:, 2:KM)\right]\right\}. \tag{90}$$

where

$$\mathbf{M} = \left[\mathbf{I}_K \otimes (\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma})\right]\left[\mathbf{I}_K \otimes \left(\mathbf{\Gamma}^H\mathbf{A}^H(\mathbf{\Theta})\right)\right] \times \in \mathbb{C}^{(KM-1)\times(KM-1)} \tag{91}$$

Since $\mathcal{D}(\mathbf{\Phi}_{\text{vec}}^n)$ and $\mathcal{D}(\mathbf{w}_{\text{vec}})$ are unimodular diagonal matrix, $\mathcal{D}^H(\mathbf{\Phi}_{\text{vec}}^n)\mathbf{M}\mathcal{D}(\mathbf{\Phi}_{\text{vec}}^n)$ or $\mathcal{D}^H(\mathbf{w}_{\text{vec}})\mathbf{M}\mathcal{D}(\mathbf{w}_{\text{vec}})$ does not change the diagonal elements of $\mathbf{M}$. Moreover, $\mathbf{M}$ is a block diagonal matrix, and all K blocks in the diagonal are $\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma}\mathbf{\Gamma}^H\mathbf{A}^H(\mathbf{\Theta})$. By checking the diagonal elements of $\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma}\mathbf{\Gamma}^H\mathbf{A}^H(\mathbf{\Theta})$, we have

$$m_{ii} = \sum_{l=1}^{L}\left|\sum_{i=1}^{d_l}\gamma_l^{(i)}e^{-j\frac{2\pi(m-1)d}{\lambda}\sin\theta_l^{(i)}}\right|^2. \quad \forall i = 1, 2, \cdots KM-1 \tag{92}$$

Hence, the Fisher information of ω_{km} is given by

$$\mathcal{I}(\omega_{km}) = \frac{2N}{\sigma_z^2}\sum_{l=1}^{L}\left|\sum_{i=1}^{d_l}\gamma_l^{(i)}e^{-j\frac{2\pi(m-1)d}{\lambda}\sin\theta_l^{(i)}}\right|^2. \tag{93}$$

For LOS channel case, we have $d_l = 1, \ \forall l = 1, \cdots L$; therefore, (95) can be simplified as

$$\mathcal{I}(\omega_{km}) = \frac{2N}{\sigma_z^2}\sum_{l=1}^{L}|\gamma_l|^2. \tag{94}$$

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 3, pp. 1617–1655, Jul.–Sep. 2016.

[3] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surv. Tuts.*, vol. 17, no. 3, pp. 1738–1760, Jul.–Sep. 2015.

[4] M. H. Alsharif and R. Nordin, "Evolution towards fifth generation (5G) wireless networks: Current trends and challenges in the deployment of millimetre wave, massive MIMO, and small cells," *Telecommun. Syst.*, vol. 64, no. 4, pp. 617–637, Apr. 2017.

[5] A. Alkhateeb, O. E. 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, 2014.

[6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[7] A. Li and C. Masouros, "Energy-efficient SWIPT: From fully digital to hybrid analog digital beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3390–3405, Apr. 2018.

[8] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7559–7575, Nov. 2018.

[9] C. Kong, A. Mezghani, C. Zhong, A. L. Swindlehurst, and Z. Zhang, "Multipair Massive MIMO relaying systems with one-bit ADCs and DACs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2984–2997, 2018.

[10] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[11] I. Ahmed *et al.*, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 4, pp. 3060–3097, Oct.–Dec. 2018.

[12] A. J. Fenn, D. H. Temme, W. P. Delaney, and W. E. Courtney, "The development of phased-array radar technology," *Lincoln Lab. J.*, vol. 12, no. 2, pp. 321–340, 2000.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.

[13] X. Jiang and F. Kaltenberger, "Channel reciprocity calibration in TDD hybrid beamforming massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 422–431, Jun. 2018.
[14] W. P. M. N. Keizer, "Fast and accurate array calibration using a synthetic array approach," *IEEE Trans. Antennas Propag.*, vol. 59, no. 11, pp. 4115–4122, Nov. 2011.
[15] S. Mano and T. Katagi, "A method for measuring amplitude and phase of each radiating element of a phased array antenna," *Electron. Commun. Japan (Part I: Commun.)*, vol. 65, no. 5, pp. 58–64, 1982.
[16] R. Sorace, "Phased array calibration," *IEEE Trans. Antennas Propag.*, vol. 49, no. 4, pp. 517–525, Apr. 2001.
[17] R. Long, J. Ouyang, F. Yang, W. Han, and L. Zhou, "Fast amplitude-only measurement method for phased array calibration," *IEEE Trans. Antennas Propag.*, vol. 65, no. 4, pp. 1815–1822, Apr. 2017.
[18] T. Takahashi, Y. Konishi, S. Makino, H. Ohmine, and H. Nakaguro, "Fast measurement technique for phased array calibration," *IEEE Trans. Antennas Propag.*, vol. 56, no. 7, pp. 1888–1899, Jul. 2008.
[19] G. A. Hampson and A. B. Smolders, "A fast and accurate scheme for calibration of active phased-array antennas," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, vol. 2, Jul. 1999, pp. 1040–1043.
[20] C. He, X. Liang, J. Geng, and R. Jin, "Parallel calibration method for phased array with harmonic characteristic analysis," *IEEE Trans. Antennas Propag.*, vol. 62, no. 10, pp. 5029–5036, Oct. 2014.
[21] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.
[22] P. Stoica *et al.*, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[24] J. Jalden, C. Martin, and B. Ottersten, "Semidefinite programming for detection in linear systems - optimality conditions and space-time decoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2003, vol. 4, pp. 9–12.
[25] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221–1234, Mar. 2014.
[26] H. Ji *et al.*, "Overview of full-dimension MIMO in LTE-advanced pro," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 176–184, Feb. 2017.
[27] D. Su, Y. Jiang, X. Wang, and X. Gao, "Omnidirectional precoding for massive MIMO with uniform rectangular array—Part I: Complementary codes based schemes," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4761–4771, Sep. 2019.
[28] H. Bozdogan, "Akaike's information criterion and recent developments in information complexity," *J. Math. Psychol.*, vol. 44, no. 1, pp. 62–91, 2000.
[29] 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.
[30] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[31] Y. Jiang, Y. Feng, and M. K. Varanasi, "Hybrid beamforming for massive MIMO: A unified solution for both phase shifter and switch networks," in *Proc. IEEE 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Hangzhou, Oct. 2018, pp. 1–5.
[32] X. Xia and X. Gao, "A space-time code design for omnidirectional transmission in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 512–515, Oct. 2016.
[33] D. Su, Y. Jiang, X. Wang, and X. Gao, "Omnidirectional precoding for massive MIMO with uniform rectangular array—Part II: Numerical optimization based schemes," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4772–4781, Sep. 2019.
[34] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: SIAM, 2000.

Xizixiang Wei received the B.E. degree in communication engineering from Tongji University, Shanghai, China, in 2017. He is currently working toward the M.E. degree in communication and information systems with Fudan University, Shanghai, China. His research interests include massive MIMO and signal processing for communications.

Yi Jiang (Member, IEEE) received the B.S. degree in electrical engineering and information science from the University of Science and Technology of China, Hefei, China, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, FL, USA, in 2003 and 2005, respectively. In 2005, he was a Research Consultant with Information Science Technologies Inc., Fort Collins, CO, USA. From 2005 to 2007, he was a Postdoctoral Researcher with the University of Colorado, Boulder CO, USA. He moved to San Diego, CA, USA, in 2007, and worked for multiple companies, including NextWave Wireless from May 2007 to July 2008, Qualcomm Corporate Research and Development from August 2008 to May 2012, IAA Incorporated from June 2012 to January 2013, and Silvus Technologies from February 2013 to July 2016. He also worked as a Part-Time Researcher with the Electrical Engineering Department, University of California at Los Angeles, Los Angeles, CA from October 2014 to July 2016. He joined Fudan University in August 2016, and is currently a Professor with the Department of Communication Science and Engineering, Fudan University, China. His research interests include array signal processing and mobile ad hoc networks.

Qingwen Liu (Senior Member, IEEE) received the B.S. degree in electrical engineering and information science from the University of Science and Technology of China, Hefei, China, in 2001, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, in 2003 and 2006, respectively. He is currently a Professor with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. His research interests lie in the areas of simultaneous wireless power and data transfer, and Internet of Things.

Xin Wang (Senior Member, IEEE) received the B.Sc. and M.Sc.degrees from Fudan University, Shanghai, China, in 1997 and 2000, respectively, and the Ph.D. degree from Auburn University, Auburn, AL, USA, in 2004, all in electrical engineering. From September 2004 to August 2006, he was a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. In August 2006, he joined the Department of Electrical Engineering, Florida Atlantic University, Boca Raton, FL, USA, as an Assistant Professor, then was promoted to a tenured Associate Professor in 2010. He is currently a Distinguished Professor and the Chair of the Department of Communication Science and Engineering, Fudan University, China. His research interests include stochastic network optimization, energy-efficient communications, cross-layer design, and signal processing for communications. He was as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, as an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He currently serves as a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is an IEEE Distinguished Lecturer for the Vehicular Technology Society.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on May 25,2020 at 15:22:25 UTC from IEEE Xplore. Restrictions apply.