Channel Parameter Estimation for Millimeter-Wave Cellular Systems with Hybrid Beamforming

Fazal-E-Asim[a,*], Felix Antreich[c], Charles C. Cavalcante[a], André L. F. de Almeida[a], Josef A. Nossek[a,b]

[a] *Wireless Telecommunications Research Group (GTEL), Universidade Federal do Ceará (UFC), Campus do Pici, s/n, Bloco 722, CP 6005, 60455-760, Fortaleza, Brazil*
[b] *Department of Electrical and Computer Engineering, Technical University of Munich (TUM), Germany*
[c] *Department of Telecommunications, Aeronautics Institute of Technology (ITA), Brazil.*

Abstract

To achieve high data rates defined in 5G, the use of millimeter-waves and massive-MIMO are indispensable. To benefit from these technologies, an accurate estimation of the channel parameters is crucial. We propose a novel two-stage algorithm for channel parameters estimation. In the first stage, coarse estimation is accomplished by applying parameter estimation via interpolation based on a DFT grid (PREIDG) with a fixed look-up table (LUT), while the second stage refines the estimates by means of the space-alternating generalized expectation maximization (SAGE) algorithm. The two-stage algorithm uses discrete Fourier transform beamforming vectors which are efficiently implemented by a Butler matrix in the analog domain. We found that this methodology improves the estimates compared to the auxiliary beam pair (ABP) method. The two-stage algorithm shows efficient performance in the low signal to noise ratio regime for the channel parameters i.e. angles of departure, complex path gains and delays of the multipaths. Finally, we derived the Cramér-Rao lower bound (CRLB) to assess the performance of our two-stage estimation algorithm.

Keywords: AoD estimation, DFT beamforming, Butler matrix, Maximum likelihood, Hybrid beamforming, Millimeter waves, space-alternating generalized expectation maximization.

1. Introduction

Millimeter-Waves and massive multiple-input-multiple-output (MIMO) are the two essential candidates to achieve the high promising data rates for the fifth generation (5G) cellular systems [1, 2]. On one hand, introducing high frequencies will facilitate the introduction of large antenna arrays at the base station (BS), which will end up with high beamforming gains but on the other hand, the large bandwidth in mmWaves with large antenna arrays impose the challenge of hardware implementation [3]. Therefore, energy efficiency with addition to spectral efficiency becomes an important design goal.

In order to fully exploit the benefit of using large antenna arrays both at the transmitter and receiver, full digital precoding is used, which means all signal processing is done in the baseband, introducing the cost of one radio frequency (RF)-chain per antenna, which turns out with high implementation complexity in addition to power consumption. Multiple solutions are proposed to cater for this challenge including less number of RF-chains termed as hybrid beamforming [4, 5, 6, 7] and low resolution analog-to-digital converter (ADC)s [8, 9, 10]. Another practical example in mobile communication is hybrid beamforming designed to overcome the implementation and energy challenges in mmWave massive-MIMO precoding. In hybrid beamforming, a small number of RF-chains are introduced to excite a large number of antenna elements both at the transmitter and receiver. Unlike full digital beamforming, the hybrid beamforming is separated into two different parts, one is the baseband digital precoding and the other is the analog beamforming implemented at RF domain [11, 12, 13].

Therefore, beamforming needs information about the random wireless channel. Recently, many techniques have been studied to solve the channel estimation problem in mmWave systems by exploiting the sparsity of the mmWave channel. The works [14, 15, 16, 17] discuss channel parameter estimation based on compressed sensing (CS) theory by exploiting the sparsity, especially in mmWave channels. In [14], a CS-based channel parameter estimation is applied in a practical scenario and in a controlled environment, which turns out to be more attractive as compared to the exhaustive search methods. In [15], the frequency selective case is considered and a CS-based algorithm is proposed to estimate the channel parameters assuming a hybrid architecture with quantized phase shifters. In [16], a two-stage algorithm is introduced for position and orientation estimation. In the first stage, a modified version of the so-called distributed compressed sensing-simultaneous orthogonal matching pursuit (DCS-SOMP) algorithm [18] is used on a predefined grid, while in the second stage the space-alternating generalized expectation maximization (SAGE) algorithm [19] is used for the refinement of the channel parameter estimates. However, the paper does not shed light on the selection of beamformer, which is an important issue in the design of efficient hybrid architectures. In [17], a low complexity orthogonal matching pursuit (OMP) algorithm is introduced by exploiting the correlation in the angular domain from *a priori* statistical information about scattering. The algorithm, however, only considers flat fading channels.

In [20, 21, 22, 23], the auxiliary beam pair (ABP) method is proposed and

discussed, which is based on a set of beam pairs to estimate the AoD. In the ABP method, by forming many beam pairs, a set of many ratios is calculated to estimate the corresponding AoD. The authors showed that the ABP method outperforms the grid of beams (GoB) method [24]. Furthermore, authors in [25] showed that the ABP method outperforms the standard CS techniques such as OMP [26] and adaptive codebook method [27]. Furthermore, in our previous work [28], the DFT beams are implemented in the analog domain using a Butler matrix (BM) to estimate the channel via a maximum likelihood (ML) approach for a single path (frequency flat) scenario.

In this paper, we extend the previous work [28] and generalize it to the frequency selective channel model, for which the time-delay of each path is also taken into account in the algorithm design. This turns the channel parameter estimation into a non-linear optimization problem, which is solved with the SAGE algorithm as an approximation to the ML estimator in our scenario. In addition, we propose the column vectors of a DFT matrix, to probe the channel with constant amplitude zero autocorrelation (CAZAC) sequences [29, 30, 31]. Their constant amplitude property allows us to operate the power amplifier (PA) near to the saturation region and the set of beamforming vectors can be implemented at RF with the Butler matrix. This structure has an improved energy-efficiency and avoids the need for adaptive RF- phase shifters [32]. Our contributions can be summarized as follows:

- We perform an ad-hoc estimation based on our proposed algorithm, herein referred to as PREIDG (parameter estimation via interpolation based on a DFT grid) to coarsely estimate the model order (number of paths) and the parameters of these paths using a fixed look-up table (LUT). The proposed PREIDG algorithm outperforms the ABP method [20, 23].

- The accuracy of the estimates can be further improved/refined if necessary, by using them as initialization of the SAGE algorithm to obtain the ML estimates. New expressions are derived for channel parameter estimation in the multiple-input-single-output (MISO) case.

- We derive the Cramér-Rao lower bound (CRLB) to assess the performance of our two-stage estimation algorithm.

Notation: a (lower case italic letters) denote scalars, $\mathbf{a}$ (bold lowercase letters) denote vectors and $\mathbf{A}$ (bold upper case letters) denote matrices; $\mathbf{I}_M$ represents identity matrix of size $M \times M$; $\mathbf{1}_M$ denote all ones matrix of $M \times M$ dimensions; $\odot$ represents Hadamard product; $(\cdot)^{\mathrm{H}}$ represents conjugate transpose; $\mathbb{E}[\cdot]$ represents expectation; $\lfloor . \rfloor$ represents floor operator; $\mathrm{tr}\{.\}$ represents trace of the matrix and $\mathrm{mod}(.)$ represents modulus, respectively.

2. System Model

2.1. System Architecture

To overcome the energy consumption and hardware complexity due to the use of massive MIMO and mmWaves, hybrid beamforming is introduced to cater
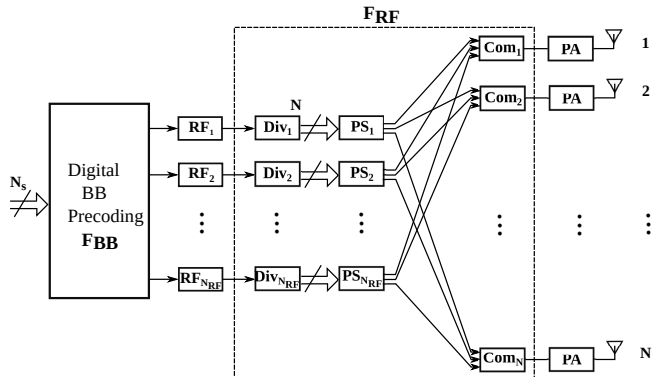
Figure 1: Fully connected analog phase shifting network.

for the reduction of RF-chains. In practice, there are two approaches in hybrid beamforming, one is based on a fully connected analog phase shifting (FCAPS) network and the other is based on a partially connected analog phase shifting (PCAPS) network. Both methodologies have their own pros and cons.

In both the strategies dividers are used which divide the signals from the RF-chains to the many phase shifters (PSs). The combiner is only used in the FCAPS network. Although the power dividers are theoretically lossless, for instance, Wilkinson splitter, the combiners are not.

In the FCAPS network, $N_{RF}N$ adaptive phase shifters are used, where N_{RF} is the total number of RF-chains and N is the total number of transmit antennas as shown in Figure 1. In FCAPS network, at the input of each power amplifier and antenna, N_{RF} signals have to be combined which introduces losses. The losses depend on the mutual correlation of the precoded signals to be combined which obviously depends on the estimated propagation channel. On one hand the benefit of using FCAPS network, we can form narrow beams using all N antennas simultaneously, but on the other hand, there are quite significant losses in the analog domain by the combination of signals which are channel dependent. For instance, if we have $N_{RF} = N_s = 2^q$, data streams, the combination of N_s uncorrelated data streams will have a power loss of $q \times 3$ dB. On top of that, there come losses due to the parasitics especially at mmWaves.

To avoid these significant losses, we use the PCAPS approach as in Figure 2, where the combination of the signals in the RF-domain is avoided. In the PCAPS network, we divide the total number of transmit antennas N into N_{RF} sub-arrays, where each sub-array has M antenna elements leading to $N = N_{RF}M$. In this approach, each sub-array gets its signal from one RF-chain. In this architecture, no combiner is required before the input of each antenna. With this, there will be no losses due to the combination of signals in the analog domain but also the number of adaptive phase shifters will be reduced from $N_{RF}N$ to $N_{RF}M$.

Furthermore, to avoid the implementation of adaptive phase shifters, the PCAPS
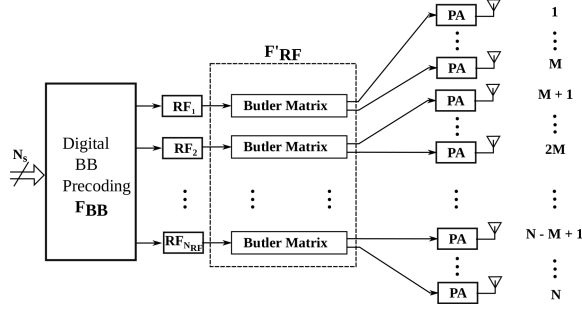
4

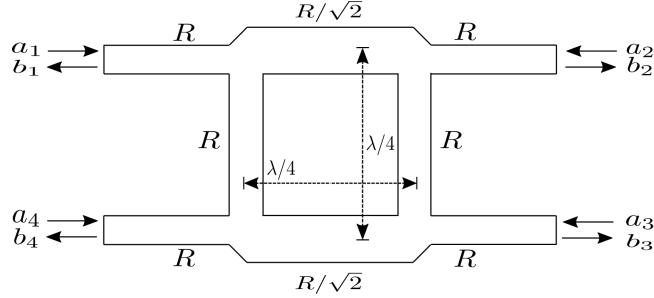Figure 2: Partially connected analog phase shifting network.

Figure 3: 90° hybrid coupler.

network, can be implemented by introducing the BM, which is the analog implementation of the DFT matrix as shown in Figure 2.

The M beams of the DFT matrix are implemented using $\frac{M}{2}\log_2 M$ (90° hybrids) and a number of fixed phase shifters. The 90° hybrid is theoretically a lossless 4 - port as shown in Figure 3, which is described by the following scattering matrix

$$
\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -j & -1 & 0 \\ -j & 0 & 0 & -1 \\ -1 & 0 & 0 & -j \\ 0 & -1 & -j & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}. \tag{1}
$$

Properly terminating all ports and with no incoming waves at ports 2 and 3 ($a_2 = 0, a_3 = 0$), resulting with no reflected waves at ports 1 and 4 ($b_1 = 0, b_4 = 0$), therefore (1) can be reduced to

$$
\begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \frac{-1}{\sqrt{2}} \begin{bmatrix} j & 1 \\ 1 & j \end{bmatrix} \begin{bmatrix} a_1 \\ a_4 \end{bmatrix}. \tag{2}
$$

An example of the BM structure with $M = 8$ is shown in Figure 4, the output signals of the BM excited by one RF-chain connected to one input will be supplied to the M power amplifiers and the M antenna elements of each sub-array, which will result in utilizing one column of DFT matrix as beamforming
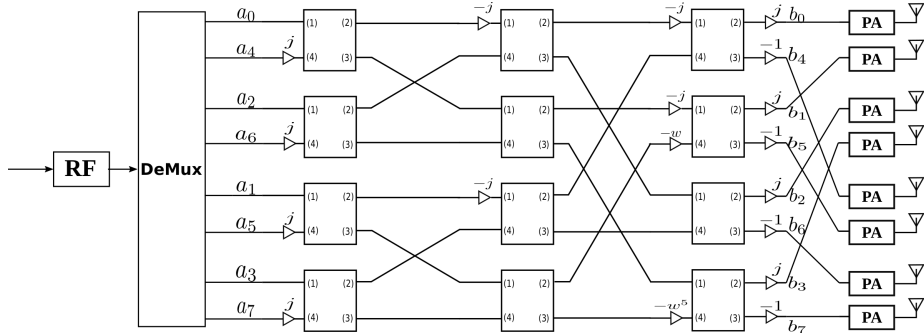
Figure 4: A $M \times M$, BM with $\frac{M}{2} \log_2 M$, 90° hybrids and fixed PSs only, where $w = \exp(-j\frac{2\pi}{M})$ for $M = 8$.

vector.

Now let us consider the input power P_{in} to a 90° hybrid at input port 1 and port 4 as P_1 and P_4 respectively. The 90° hybrid will divide the power to the output port 2 and port 3, where only one of them will be used, as in Figure 3. The input power can be written as

$$P_{in} = P_1 + P_4 = \mathbb{E}\left[|a_1|^2\right] + \mathbb{E}\left[|a_4|^2\right], \tag{3}$$

and the output power at port 2 as

$$P_{out_2} = \mathbb{E}\left[|b_2|^2\right] = \frac{1}{2}\left(\mathbb{E}\left[|a_1|^2\right] + \mathbb{E}\left[|a_4|^2\right] - 2\mathrm{Im}\{\rho\}\sqrt{\mathbb{E}\left[|a_1|^2\right]\mathbb{E}\left[|a_4|^2\right]}\right), \tag{4}$$

where ρ is the correlation coefficient and is given as

$$\rho = \frac{\mathbb{E}\left[a_1 a_4^*\right]}{\sqrt{\mathbb{E}\left[|a_1|^2\right]\mathbb{E}\left[|a_4|^2\right]}}. \tag{5}$$

If both the signals at port 1 and port 4 are uncorrelated, i.e., $\rho = 0$, then $P_{out_2} = \frac{1}{2}P_{in}$, which means half of the power is lost due to the uncorrelated signals. However, in the BM architecture as shown in Figure 4, always both the outputs of each 90° hybrid are used and no power is lost.

2.2. Signal Model

We consider a single-user downlink scenario, where the user is equipped with a single antenna. The transmitter has PCAPS network architecture as shown in Figure 2 and equipped with a total number of N antennas having M number of antennas at each sub-array. The sub-arrays are controlled by N_{RF} chains following $N = N_{\mathrm{RF}}M$. The signal model based on the sub-array design and without loss of generality can be given as

$$\mathbf{y}_k^{\mathrm{T}} = \sqrt{P_T} \sum_{r=1}^{R} \alpha_r \, \mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\mathbf{c}_k^{\mathrm{T}}(\tau_r) + \mathbf{n}_k^{\mathrm{T}} \in \mathbb{C}^{1 \times L}, \tag{6}$$

6

(6) is the received vector for the user based on one sub-array with the kth beamforming vector. P_T represents the transmit power, α_r is the complex path coefficient of the given path r and $\mathbf{n}$ is the random noise vector with Gaussian distribution $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{L\times 1}, \sigma_n^2 \mathbf{I}_L)$, where σ_n^2 is the noise variance. $\mathbf{c}_k(\tau_r)$ is the pilot sequence for k-th beamforming vector $\mathbf{w}(\Phi_k)$, where τ_r is the delay of each path-r. The channel steering vector $\mathbf{a}(\mu_r)$ for uniform linear array (ULA) and beamforming vector $\mathbf{w}(\Phi_k)$ are given as

$$\mathbf{a}(\mu_r) = [1, e^{-j\mu_r}, \dots, e^{-j(M-1)\mu_r}]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}, \tag{7}$$

$$\mathbf{w}(\Phi_k) = \frac{1}{\sqrt{M}} \left[1, e^{-j\Phi_k}, \dots, e^{-j(M-1)\Phi_k}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}, \tag{8}$$

The $\mathbf{w}(\Phi_k)$ is $k+1$ th column of a DFT matrix of size M, where $\Phi_k = \frac{2\pi}{M}k, k = 0, \dots, M-1$. μ_r is the spatial frequency of each path r where $\mu_r = 2\pi\frac{d}{\lambda}\sin\theta_r$. θ_r is the angle of departure (AoD) of each path r and d is the distance between antenna elements respectively.

The CAZAC sequence specific for each beamforming vector is $\mathbf{c}_k$, where each of the CAZAC sequence symbol is constructed as

$$c(n) = e^{\left(j\frac{2\pi}{\sqrt{L}}\left(\mathrm{mod}\{n,\sqrt{L}\}+1\right)\left(\lfloor \frac{n}{\sqrt{L}}\rfloor+1\right)+j\frac{\pi}{4}\right)}, \tag{9}$$

where $n \in \{0, 1, \dots, L-1\}$ and $\mathbf{c}_k(n) = \mathbf{c}(n-k)$. By considering the length of the CAZAC sequence $L = 16$ helps in forming the CAZAC symbols as QPSK symbols $c(n) \in \left\{\frac{1}{\sqrt{2}}(\pm 1 \pm j)\right\}$, which is useful because of the constant modulus allow us to operate the PA near the saturation region. The CAZAC sequence for first beamforming vector is represented as

$$\mathbf{c}_0(0) = [c(0), c(1), \dots, c(L-1)]^{\mathrm{T}} \in \mathbb{C}^{L\times 1}. \tag{10}$$

$\mathbf{c}_k$, where $k = 0, \dots, M-1$ are shifted wrap around versions of $\mathbf{c}_0$ and assigned each wrap around to a specific beamforming vector.

By probing the channel with each beamforming vector $\mathbf{w}(\Phi_k)$ with its corresponding CAZAC sequence $\mathbf{c}_k$ and then collecting all the received vectors $\mathbf{y}_k$ in matrix $\mathbf{Y}$ can be represented as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_0^{\mathrm{T}} \\ \mathbf{y}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{y}_{M-1}^{\mathrm{T}} \end{bmatrix} =$$

$$= \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\underbrace{\begin{bmatrix} \mathbf{c}_0^{\mathrm{T}}(\tau_r) \\ \mathbf{c}_1^{\mathrm{T}}(\tau_r) \\ \vdots \\ \mathbf{c}_{M-1}^{\mathrm{T}}(\tau_r) \end{bmatrix}}_{=\mathbf{C}(\tau_r)} + \underbrace{\begin{bmatrix} \mathbf{n}_0^{\mathrm{T}} \\ \mathbf{n}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{n}_{M-1}^{\mathrm{T}} \end{bmatrix}}_{=\mathbf{N}} \in \mathbb{C}^{M\times L}, \tag{11}$$

where $\mathbf{A}(\mu_r) = \mathrm{diag}\{\mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1}$ and the noise covariance matrix is $\mathbf{R} = \mathbb{E}\left[\mathrm{vec}\{\mathbf{N}\}\,\mathrm{vec}\{\mathbf{N}\}^{\mathrm{H}}\right] = \sigma_n^2\mathbf{I}_{ML}$.

3. Parameter Estimation via Interpolation based on a DFT Grid (PREIDG) based coarse estimation

We are now probing the channel with all M beamforming vectors, one at a time with one specific CAZAC sequence of length L, preferably $L = M$. There is a strict correspondence between $\mathbf{w}(\Phi_k)$ and $\mathbf{c}_k$. This way the user equipment (UE) observes M consecutive receive sequences $\mathbf{y}_k$ (6) and multiplies each of them with the already stored CAZAC sequence $\mathbf{c}_k^*$. This can be cast in a matrix

$$\mathbf{Z} = \mathbf{Y}\mathbf{C}^{\mathrm{H}}(0) = \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\mathbf{C}^{\mathrm{H}}(0) + \mathbf{N}\mathbf{C}^{\mathrm{H}}(0), \tag{12}$$

For simplicity, let us first assume that each of the R AoD's, μ_r is equal to one of the Φ_k. Then we get

$$\mathbf{A}(\mu_r) = \mathrm{diag}\left\{\mathbf{a}^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\right\}_{k=0}^{M-1}\Bigg|_{\mu_r=\Phi_{k_r}}$$
$$= \sqrt{M}\,\mathbf{e}_{k_r+1}\mathbf{e}_{k_r+1}^{\mathrm{T}}, \tag{13}$$

where $\mathbf{e}_{k_r+1}$ is the M- dimensional $(k_r + 1)$ canonic unit vector. In addition, let us assume that each of the R delays are integer multiples of the symbol period leading to

$$\mathbf{C}(\tau_r)\mathbf{C}^{\mathrm{H}}(0)\Bigg|_{\tau_r=i_r} = M\mathbf{P}_{i_r}, \tag{14}$$

where $\mathbf{P}_{i_r}$ is the $M \times M$, permutation matrix of the following form, by using the canonic unit vector $\mathbf{e}$,

$$\mathbf{P}_{i_r} = \sum_{j=1}^{M}\mathbf{e}_j\mathbf{e}_{j+i_r}^{\mathrm{T}}$$

$$= \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \ddots & \dots & \vdots \\ \vdots & \dots & \vdots & \vdots & \dots & \dots & \ddots & \vdots \\ 0 & \dots & \vdots & \vdots & \dots & \dots & \dots & 1 \\ 1 & \dots & \vdots & \vdots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & \dots & \dots & \vdots \\ 0 & \dots & 1 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}, \tag{15}$$
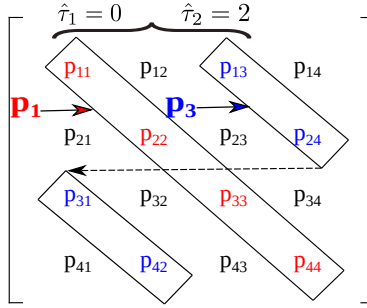
Figure 5: An example of integer delay estimation $\hat{\tau}_i$ using power matrix $\mathbf{P}$ as in (17) where $\mathbf{p}_1$ and $\mathbf{p}_3$ are the two diagonals, for which (19) is fulfilled.

with $\mathbf{P}_0 = M\mathbf{I}_M$. In this simplified case, the matrix $\mathbf{Z}$ is written as

$$\mathbf{Z} = \sqrt{P_T}M\sqrt{M}\sum_{r=1}^{R}\alpha_r\left(\mathbf{e}_{k_r+1}\mathbf{e}_{k_r+1}^{\mathrm{T}}\sum_{j=1}^{M}\mathbf{e}_j\mathbf{e}_{j+i_r}^{\mathrm{T}}\right) + \mathbf{NC}^{\mathrm{H}}(0)$$

$$= \sqrt{P_T}M\sqrt{M}\sum_{r=1}^{R}\alpha_r\mathbf{S}_r + \mathbf{NC}^{\mathrm{H}}(0). \tag{16}$$

where $\mathbf{S}_r$ is a matrix, where only the entry in the $(k_r + 1)$ row and $\mod(k_r + i_r + 1, M)$ column is equal to one, while all the other entries are zero. The integers k_r and i_r identify the AoD and τ of the rth multipath component.
Finally, we compute the post correlation power matrix for the general form as

$$\mathbf{P} = \mathbb{E}\left[\mathbf{Z} \odot \mathbf{Z}^*\right] = P_T M^3 \sum_{r=1}^{R}|\alpha_r|^2\mathbf{S}_r \odot \mathbf{S}_r^* + \sigma_n^2 M\mathbf{1}_M. \tag{17}$$

The post correlation receive signal to noise ratio (SNR) is enhanced by a process gain M from correlation and by an antenna gain of M from the array, which is valid for our simplistic assumption, that both μ_r and τ_r are on the grid of the beamforming vectors as well as of the symbol timing. In any realistic scenario, this will not be true and the matrices $\mathbf{S}_r$ will not be strictly sparse with only one non-zero entry. But the power matrix, which is available at the user, will still provide useful information about the model order, i.e. the number of the paths/wavefronts and the parameters $\sqrt{P_T}\alpha_r, \mu_r, \tau_r$ of each path. We exploit the power matrix by searching the (wrap-around) diagonals $\mathbf{p}_i, i = 1, \ldots, M$ of $\mathbf{P}$ for the largest entry as shown in Figure 5 for $M = 4$.
We have

$$\mathbf{p}_i^{\mathrm{T}} = \left[p_{1,i}, p_{2,\mod(i,M)+1}, \ldots, p_{M,\mod(i+M-2,M)+1}\right]^{\mathrm{T}} \in \mathbb{R}^M, \tag{18}$$

where $i = 1, \ldots, M$, with $\mathbf{p}_1$ being the main diagonal of $\mathbf{P}$. For every diagonal $\mathbf{p}_i$ we check, whether the largest element is above a certain threshold G,

$$\max_{k=1,\ldots,M}\left(p_{k,\mod(i+(k-2),M)+1}\right) \geq G, \ i = 1, \ldots, M. \tag{19}$$
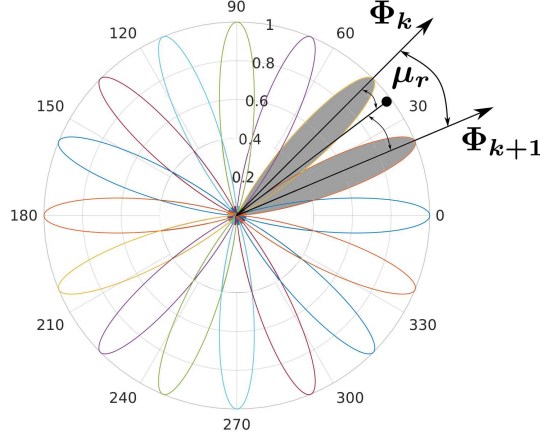
Figure 6: An example of AoD estimation using a DFT matrix based beamforming via PREIDG.

G should make sure that we have a signal above the noise floor, which is $\sigma_n^2 M$ (17). The integer delay τ_{i_r} for each path r can be found as shown in Figure 5,

$$\hat{\tau}_{i_r} = i_r - 1. \tag{20}$$

For the main diagonal $\mathbf{p}_1$, this will always be fulfilled assuming a LOS path, which has relative delay $\tau_i = i_1 - 1\big|_{i_1=1} = 0$. The number of diagonals $\mathbf{p}_i$, which fulfill (19) is the model order $\hat{R}$ and we have a coarse estimate $\hat{\tau}_{i_r}$ for each path, where we drop index r for simplicity.

Since the real spatial frequency of each detected path will be somewhat in between two spatial frequencies $\Phi_k < \mu_r < \Phi_{k+1}$ as in Figure 6, there will be two significant adjacent entries along the diagonal, which we denote P_k and P_{k+1}.

LUT: We get an estimate of μ_r, by interpolation with $K+1$ spatial frequencies μ_l, generated as follows

$$\mu_l = \Phi_k + l\Delta_\mu, \ \ l = 0\ldots, K, \tag{21}$$

$$\Delta_\mu = \frac{\Phi_{k+1} - \Phi_k}{K} = \frac{2\pi}{MK}, \tag{22}$$

and compute the hypothetical noise free normalized power for those angles, μ_l

$$P_{k,l} = \left|\mathbf{a}^{\mathrm{H}}(\mu_l)\mathbf{w}(\Phi_k)\right|^2, \tag{23}$$

$$P_{k+1,l} = \left|\mathbf{a}^{\mathrm{H}}(\mu_l)\mathbf{w}(\Phi_{k+1})\right|^2, \tag{24}$$

with ratios

$$\Delta_l = \sqrt{\frac{P_{k,l}}{P_{k+1,l}}}. \tag{25}$$

10

Since $P_{k,l}$ and $P_{k+1,l}$ are independent of k, we need only $K+1$ ratios to provide Δ_l in a LUT. Note that the proposed LUT is different from the idea discussed in [33], which is based on a re-weighted l-1 minimization problem. The entries of the measurement matrix are independent and identically distributed (i.i.d) Gaussian random variables, and besides the observed data is contaminated by noise. This approach is significantly different from what we have considered in the generation of the fixed LUT.

Coarse estimation of $\hat{\mu}_r$***:*** Now we are selecting those indices l and $l+1$, where the corresponding ratios Δ_l and Δ_{l+1} are closest to the ratio (26)

$$\Delta = \sqrt{\frac{P_k}{P_{k+1}}}, \tag{26}$$

and get the estimated spatial frequencies

$$\hat{\mu}_r = \mu_l + b\Delta_\mu = \Phi_k + \Delta_\mu(l+b), \tag{27}$$

with

$$b = \frac{\Delta_l - \Delta}{\Delta_l - \Delta_{l+1}}. \tag{28}$$

Finally, (27) can be converted to an estimated azimuth angle in degrees

$$\hat{\theta}_r = \begin{cases} \arcsin(\frac{\hat{\mu}_r}{\pi}), \ 0 \le \hat{\mu}_r \le \pi, \\ \arcsin(\frac{\hat{\mu}_r - 2\pi}{\pi}), \ \pi < \hat{\mu}_r \le 2\pi. \end{cases} \tag{29}$$

The interpolation between Φ_k and Φ_{k+1} can go wrong in some cases especially if the signal level is weak and μ_r is close to letting say Φ_k. In such a case P_k may be quite large, but P_{k+1} may be close to the noise floor. Therefore it is not clear whether P_{k+1} or rather P_{k-1} is the second largest power from the signal, which is masked by the noise. Therefore we check whether

$$|P_{k+1} - P_{k-1}| \le \frac{\sigma_n^2}{v}. \tag{30}$$

If (30) is fulfilled, then it is not worthwhile to interpolate at all, but simply choose $\hat{\mu}_r = \Phi_k$. The value of $v = 3$, has been heuristically chosen based on numerical experiments. Keep in mind, that the LUT once generated is fixed for every spatial frequency μ_r.

After having estimated all AoDs, the model order estimation may be refined, because of the integer estimation of the delays. One non-integer delay may have lead to two adjacent integer delays, and both of them will have the same AoD estimate. If this occurs, we drop one of the two delays.

Now we can put together the whole coarse estimation approach for AoD estimation in Algorithm 1. The only thing the UE has to feedback to the BS is Δ in (26) which is one real number, and one index with $\log_2 M$ bits for each path r. The rest of the computation can be done by BS for each AoD.

11

Algorithm 1 Proposed coarse estimation based on the PREIDG algorithm.

Require: Y (11)
 1: The UE received **Y** and get $\mathbf{p}_i$ for each path r (17)
 2: Calculate Δ as in (26)
 3: Find l such that $\Delta_l \geq \Delta \geq \Delta_{l+1}$
 4: Calculate constant b as in (28)
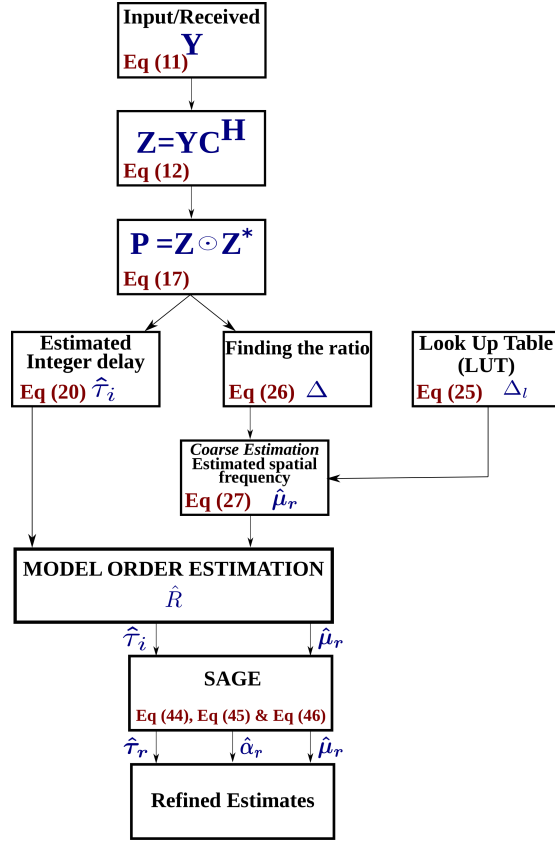 5: **return** $\hat{\mu}_r$ and $\hat{\theta}_r$ as in (27) and (29).

Figure 7: A Flowchart for the two-stage estimation algorithm.

4. Maximum Likelihood Estimation using SAGE

The channel parameters obtained using coarse estimation based on the PREIDG algorithm can be further improved using an iterative procedure which is initialized by the ad-hoc estimates found by PREIDG. This is a multidimensional, non-linear optimization problem (36), which can be solved by space-alternating generalized expectation maximization (SAGE) algorithm. We use the standard formulation of the SAGE algorithm [19] to solve and derive new expressions for

our non-linear problem considering important hardware constraints.

To make sure that the SAGE algorithm converges to the global optimum with low computational complexity. Initializing SAGE with the coarse estimation is indispensable. Furthermore, SAGE will also help us to estimate the non-integer delay, unlike the coarse estimation in addition to the more refined complex path gain for each path r.

We treat the observed data $\mathbf{Y}$ as a random variable, parametrized by a Gaussian probability density function (pdf) with an unknown channel parameter vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = \left[\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\}^\mathrm{T}, \sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\}^\mathrm{T}, \boldsymbol{\mu}^\mathrm{T}, \boldsymbol{\tau}^\mathrm{T} \right]^\mathrm{T}, \tag{31}$$

where

$$\sqrt{P_T}\mathrm{Re}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Re}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Re}\{\alpha_r\}, \ldots, \sqrt{P_T}\mathrm{Re}\{\alpha_R\} \right]^\mathrm{T}, \tag{32}$$

$$\sqrt{P_T}\mathrm{Im}\{\boldsymbol{\alpha}\} = \left[\sqrt{P_T}\mathrm{Im}\{\alpha_1\}, \ldots, \sqrt{P_T}\mathrm{Im}\{\alpha_r\}, \ldots, \sqrt{P_T}\mathrm{Im}\{\alpha_R\} \right]^\mathrm{T}, \tag{33}$$

$$\boldsymbol{\mu} = \left[\mu_1, \ldots, \mu_r, \ldots, \mu_R \right]^\mathrm{T}, \tag{34}$$

$$\boldsymbol{\tau} = \left[\tau_1, \ldots, \tau_r, \ldots, \tau_R \right]^\mathrm{T}, \tag{35}$$

with the likelihood function given as

$$\mathrm{L}(\mathbf{Y}; \boldsymbol{\eta}) = \frac{1}{\pi^{ML}\det\mathbf{R}} \ \exp\left(-\mathrm{vec}\left\{ \mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r) \right\}^\mathrm{H} \right.$$
$$\left. \mathbf{R}^{-1}\ \mathrm{vec}\left\{ \mathbf{Y} - \sqrt{P_T}\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\mathbf{C}(\tau_r) \right\} \right). \tag{36}$$

SAGE uses the observable but incomplete data space mentioned in (11) to estimate the parameters of the superimposed R wavefronts.

$$\mathbf{Y} = \boldsymbol{f}(\mathbf{X}) = \boldsymbol{f}\left([\mathbf{X}_1, \ldots, \mathbf{X}_R]\right) = \sum_{r=1}^{R}\mathbf{X}_r, \tag{37}$$

where $\mathbf{X}$ is the complete but unobservable data space. $\mathbf{X}_r$ is called a hidden data space.

In each iteration the following expectation and maximization steps are performed. The $\hat{\tau}_i$ and $\hat{\mu}_r$ are the ad-hoc estimates obtained using (20) and (27) which will be used to initialize SAGE, while $\hat{\alpha}_r = 0$ is assumed.

Expectation Step: The conditional expectation of the hidden data space can be calculated based on the incomplete data space $\mathbf{Y}$ and the previous estimation $\hat{\boldsymbol{\eta}}$.

$$\hat{\mathbf{X}}_r = \mathbb{E}_{\mathbf{X}_r}[\mathbf{X}_r|\mathbf{Y}; \hat{\boldsymbol{\eta}}] = (1 - \beta_r)\mathbf{S}_r(\hat{\boldsymbol{\eta}}_r) + \beta_r\left(\mathbf{Y} - \sum_{\substack{r'=1 \\ r'\neq r}}^{R}\mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}) \right), \tag{38}$$

13

where

$$\mathbf{S}_r(\boldsymbol{\eta}_r) = \sqrt{P_T}\alpha_r \mathbf{A}(\mu_r)\mathbf{C}(\tau_r), \tag{39}$$

and

$$\boldsymbol{\eta}_r = \left[\sqrt{P_T}\mathrm{Re}\{\alpha_r\}, \sqrt{P_T}\mathrm{Im}\{\alpha_r\}, \mu_r, \tau_r \right], \tag{40}$$

and β_r controls the convergence rate. Assuming $\beta_r = 1$ the estimated hidden data space is estimated as

$$\hat{\mathbf{X}}_r = \mathbf{Y} - \sum_{\substack{r'=1 \\ r' \neq r}}^{R} \mathbf{S}_{r'}(\hat{\boldsymbol{\eta}}_{r'}). \tag{41}$$

Maximization Step: To compute a refined $\hat{\boldsymbol{\eta}}_r$, the following optimization problem has to be solved,

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\boldsymbol{\eta}_r} \mathbb{E}_{\mathbf{X}_r} \left[\ell(\mathbf{X}_r; \boldsymbol{\eta}_r) | \mathbf{Y}; \hat{\boldsymbol{\eta}} \right], \tag{42}$$

where

$$\ell(\mathbf{X}_r; \boldsymbol{\eta}_r) = \ln\left(\mathrm{L}\left(\mathbf{X}_r; \boldsymbol{\eta}_r\right) \right), \tag{43}$$

is the log-likelihood function. This non-linear optimization problem (42) can be solved iteratively and sequentially. The new derivation of the below equations are given in Appendix A.

The delay estimation $\hat{\tau}_r$ can be iteratively maximized as

$$\hat{\tau}_r = \arg\max_{\tau_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \, \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\tau_r)\}} \right\}, \tag{44}$$

similarly the spatial frequency $\hat{\mu}_r$ can be iteratively maximized as

$$\hat{\mu}_r = \arg\max_{\mu_r} \left\{ \frac{\left| \mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \, \mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\hat{\tau}_r)\}} \right\}, \tag{45}$$

while in the end $\sqrt{\hat{P_T}}\alpha_r$ can be analytically found as

$$\sqrt{\hat{P_T}}\alpha_r = \frac{\mathrm{tr}\left\{ \mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r \right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\hat{\tau}_r)\}}. \tag{46}$$

One iteration of the SAGE algorithm is defined as a full update of all parameters of the parameter vector $\boldsymbol{\eta}$. To make sure that SAGE converges to the global optimum, it is important to initialize the SAGE algorithm with a good coarse estimation. We initialize SAGE with the coarse estimated spatial frequency $\hat{\mu}_r$

and integer delay $\hat{\tau}_i$ as shown in Figure 7.

The stopping thresholds for convergence of SAGE are defined as

$$T_1 = \frac{|\hat{\mu}_{r_p} - \hat{\mu}_r|}{|\hat{\mu}_r|}, \tag{47}$$

$$T_2 = \frac{|\hat{\tau}_{r_p} - \hat{\tau}_r|}{|\hat{\tau}_r|}, \tag{48}$$

$$T_3 = \frac{|\sqrt{\hat{P}_T}\alpha_{r_p} - \sqrt{\hat{P}_T}\alpha_r|}{|\sqrt{\hat{P}_T}\alpha_r|}, \tag{49}$$

where $\hat{\mu}_{r_p}, \hat{\tau}_{r_p}, \sqrt{\hat{P}_T}\alpha_{r_p}$, are the previous estimates of spatial frequency, delay time and complex path gain. The stopping criteria for SAGE is satisfied, when (50) is fulfilled

$$\max\{T_1, T_2, T_3\} \leq \Gamma, \tag{50}$$

where Γ is the stopping threshold. The coarse parameters are fed to the SAGE algorithm after refinement of the model order estimation as shown in Figure 7. All these computations will be done by the UE and for each path r four real numbers i.e. $\hat{\mu}_r, \sqrt{\hat{P}_T}\alpha_r, \hat{\tau}_r$ will be quantized and feedback to the BS for further hybrid beamforming. Of course, the single antenna UE has to have enough power to feedback this information to the BS. The flowchart for the two-stage algorithm is shown in Figure 7.

5. Derivation of the CRLB

In this section, we derive the CRLB for spatial frequency (μ_r), complex path gain ($\sqrt{P_T}\alpha_r$) and delay (τ_r) estimation.

Assuming $\hat{\boldsymbol{\eta}}$ as an unbiased estimate of $\boldsymbol{\eta}$, then the variance, var of the estimation error for the different parameters can be lower-bounded by the diagonal elements of the inverse of Fisher information matrix (FIM) represented as $\mathbf{F}(\boldsymbol{\eta})$ [34]

$$\text{var}(\hat{\eta}_i) \geq \left[\mathbf{F}^{-1}(\boldsymbol{\eta})\right]_{ii}. \tag{51}$$

The bound on the error is calculated as

$$\sqrt{\text{CRLB}(\hat{\eta}_i)} = \sqrt{[\mathbf{F}^{-1}(\boldsymbol{\eta})]_{ii}}. \tag{52}$$

The FIM for complex data is given as [34],

$$[\mathbf{F}(\boldsymbol{\eta})]_{ij} = \frac{2}{\sigma_n^2} \text{ Re } \left\{\text{tr}\left\{\frac{\partial \mathbf{S}^{\text{H}}(\boldsymbol{\eta})}{\partial \eta_i} \frac{\partial \mathbf{S}(\boldsymbol{\eta})}{\partial \eta_j}\right\}\right\}, \tag{53}$$

with $\mathbf{S}(\boldsymbol{\eta})$ is defined as

$$\mathbf{S}(\boldsymbol{\eta}) = \sum_{r=1}^{R} \mathbf{S}(\boldsymbol{\eta}_r) = \sqrt{P_T} \sum_{r=1}^{R} \alpha_r \, \mathbf{A}(\mu_r) \, \mathbf{C}(\tau_r). \tag{54}$$

15

The FIM can be structured as

$$\mathbf{F}(\boldsymbol{\eta}) =$$

$$\begin{bmatrix} \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Re}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}}^{\mathrm{T}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\mathrm{Im}\{\boldsymbol{\alpha}\}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}} \\ \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}^{\mathrm{T}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}^{\mathrm{T}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\mu}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}} \\ \mathbf{F}_{\mathrm{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}^{\mathrm{T}} & \mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}^{\mathrm{T}} & \mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}}^{\mathrm{T}} & \mathbf{F}_{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix}. \tag{55}$$

The block matrices of $\mathbf{F}(\boldsymbol{\eta})$ are derived in Appendix B.

6. Numerical Results

In this section, the performance of the proposed two-stage algorithm is evaluated, assessed with CRLB and compared with ABP. The transmitter is deployed with a subarray antenna structure with $M = 16$ and a total number of antennas $N = N_{\mathrm{RF}} M$ as a ULA with $\lambda/2$ inter-element spacing.
We assume the bandwidth of a system, $B = 200\,\mathrm{MHz}$, from which one symbol time can be calculated as $T_s = 5\,\mathrm{ns}$. The system is operating at a carrier frequency of $f_c = 28\,\mathrm{GHz}$. 32 pilot symbols are used per beamforming vector which ends up with a total number of 512 pilot symbols. The distance for LOS is uniformly distributed as [35]

$$D_{\mathrm{los}} \sim \mathrm{U}(30\,\mathrm{m}, 60\,\mathrm{m}), \tag{56}$$

while the NLOS distances are distributed as

$$D_{\mathrm{nlos}} = D_{\mathrm{los}} + \Delta_{\mathrm{nlos}}, \tag{57}$$

where the Δ_{nlos} is the difference of NLOS and LOS paths and can be distributed as $\Delta_{\mathrm{nlos}} \sim \mathrm{U}(4.5\,\mathrm{m}, 24\,\mathrm{m})$, which gives the delay between 3 and 16 symbols. The length of CAZAC sequence restricts the maximum delay difference estimated. The path loss (PL) is calculated as [35]

$$PL(\mathrm{dB}) = 10\bar{n}\log_{10}\left(\frac{D}{D_0}\right), \tag{58}$$

where $\bar{n}$ is the path loss exponent which is assumed 2.1 for LOS and 2.4 for NLOS paths and $D_0 = 1\,\mathrm{m}$. The complex path gain for LOS is assumed such that $\alpha_1 = 1$. To calculate the complex path gains for NLOS, we need to use the path loss (58) for LOS (PL_{los}) and NLOS (PL_{nlos}),

$$\gamma_r = \sqrt{\frac{PL_{\mathrm{los}}}{PL_{\mathrm{nlos}}}}, \tag{59}$$

where γ_r is the ratio of LOS and NLOS paths and is the magnitude of the complex path coefficient. The complex path gain, α_r for each path r can be obtained as

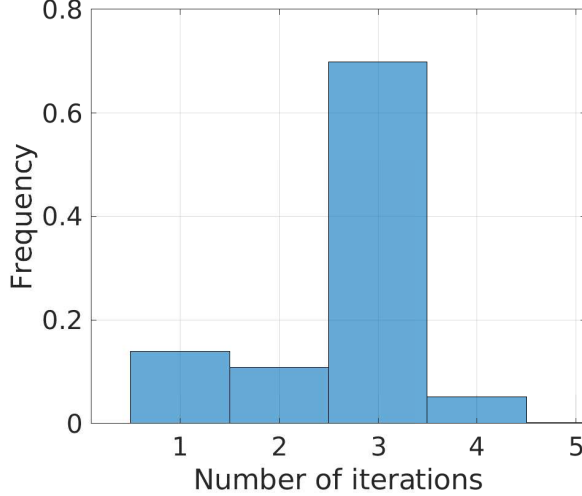$$\frac{\alpha_r}{\alpha_1} = \gamma_r e^{j\phi_r}, \tag{60}$$

Figure 8: Histogram for the number of iterations of SAGE.

where ϕ_r is the phase of the complex coefficient of path r and is uniformly distributed as $\phi_r \sim \mathrm{U}(0, 2\pi)$ [35]. α_1 is assumed as 1, because what matters is the ratio of the LOS path as compared to the NLOS path.

The AoD for both LOS path and NLOS paths are generated from a uniform distribution, i.e. $\theta_r^\circ \sim \mathrm{U}(-60°, +60°)$ (one sector of a cell) as given in [35]. The noise variance is kept as $\sigma_n^2 = 1$.

The SNR is defined as

$$\mathrm{SNR} = \frac{P_T |\alpha_1|^2}{\sigma_n^2}, \tag{61}$$

The root mean square error (RMSE) for θ_r, $\sqrt{P_T}\alpha_r$ and τ_r are calculated as

$$\mathrm{RMSE}(\hat{\theta}_r) = \sqrt{\mathbb{E}\left[|\theta_r - \hat{\theta}_r|^2\right]}, \tag{62}$$

$$\mathrm{RMSE}(\sqrt{\hat{P_T}}\alpha_r) = \sqrt{\mathbb{E}\left[\left|\frac{\sqrt{P_T}\alpha_r - \sqrt{\hat{P_T}}\alpha_r}{\sqrt{P_T}\alpha_r}\right|^2\right]}, \tag{63}$$

$$\mathrm{RMSE}(\hat{\tau}_r) = \sqrt{\mathbb{E}\left[|\tau_r - \hat{\tau}_r|^2\right]}. \tag{64}$$

In calculating RMSE, we are not calculating true expected values, but we used numerical averaging over 10000 realizations.

We are calculating CRLB for all realizations as given in (52) and then obtain a numerical average over 10000 channel scenarios. Finally, we take the square root to compare with RMSE respectively. The length of the LUT is chosen as $K = 101$.

In the ABP approach, the formation of auxiliary pairs of beams is important
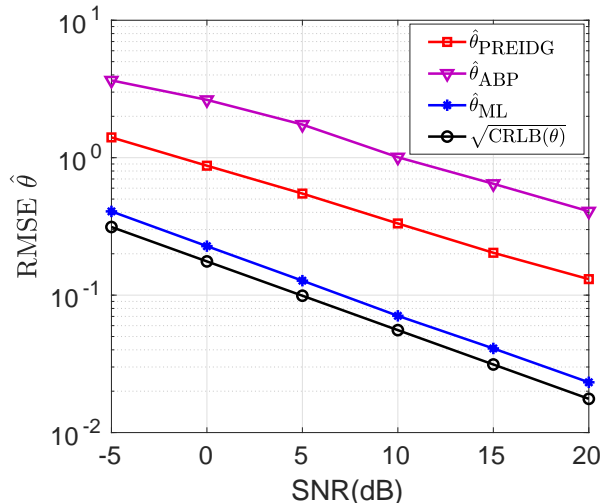
Figure 9: Performance comparison of two-stage and ABP algorithm for LOS AoD only.

to estimate the corresponding AoD. For this, a specific spacing difference between the beams named as δ is important and need to be kept constant as in [20, 23]. The spacing is given as, $\delta = \frac{2m\pi}{M}$, where m can be chosen from the set as $m = 1, \ldots, \frac{M}{4}$. We choose $m = 1$ and get $\delta = \frac{\pi}{8}$. To keep $\delta = \frac{\pi}{8}$ constant in all the beam pairs of DFT beams, we form 16 beam pairs as $(1, 3), (2, 4), \ldots, (15, 1), (16, 2)$ respectively. The criteria for choosing the auxiliary beam pair is the pair which gives the largest average power out of all the probed beam pairs.

Figure 8 shows the convergence behavior of SAGE. By properly initializing SAGE with coarse estimation of spatial frequency $\hat{\mu}_r$, integer delay $\hat{\tau}_i$ assuming $\tau_1 = 0$ and $\hat{\alpha}_r = 0$, the maximum number of iterations that SAGE needs to converge is 4 considering $\Gamma = 10^{-3}$. 70% of all channel realizations, SAGE took 3 iterations to converge to the global optimum.

In Figure 9, the two-stage estimation algorithm is evaluated and compared with ABP for LOS AoD estimation, given the aforementioned scenario. Simulation results based on 10 thousands of channel realizations show that the proposed PREIDG method performs better than ABP. After using coarse estimation based on PREIDG, as an ad-hoc estimation to initialize the SAGE algorithm gives the improved ML performance which nearly satisfies the theoretical bound.

The AoDs performance of the two-stage method for NLOS paths are compared with ABP and is assessed with theoretical CRLB in Figure 10. The performance of PREIDG still performs better than ABP for NLOS paths. The performance of the ML approach performs better than both PREIDG and ABP method. The ML method approaches the CRLB closely.
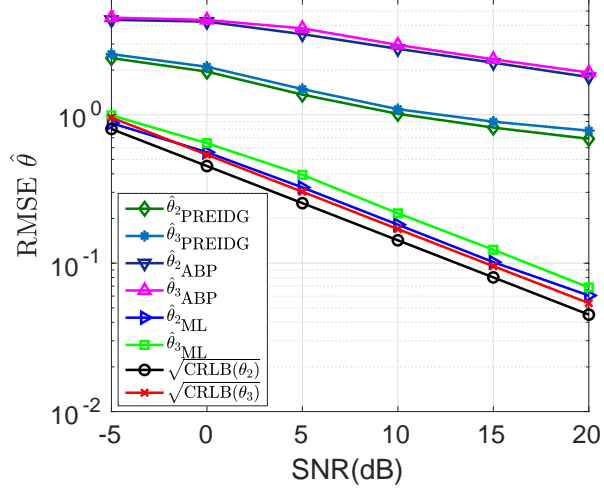
18

Figure 10: Performance comparison of two-stage and ABP algorithm for NLOS AoDs only.
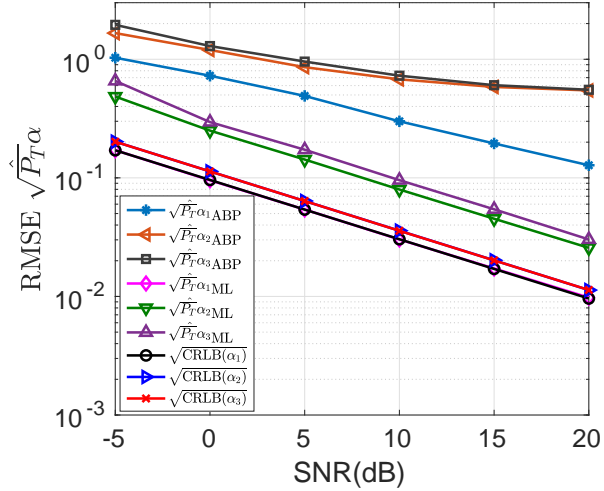

Figure 11: Performance comparison of two-stage and ABP algorithm for α_r.

Figure 11 shows the performance comparison of complex path gain, $\sqrt{\hat{P_T}}\alpha_r$ via two-stage estimation and ABP method. The performance of the two-stage estimation method for $\sqrt{\hat{P_T}}\alpha_r$ performs better compared to the ABP approach. For the LOS, the two-stage estimation algorithm achieved CRLB because of $\alpha_1 = 1$, while for NLOS paths estimates are not good because of the fact that the model order estimation does not always detect all the NLOS paths.
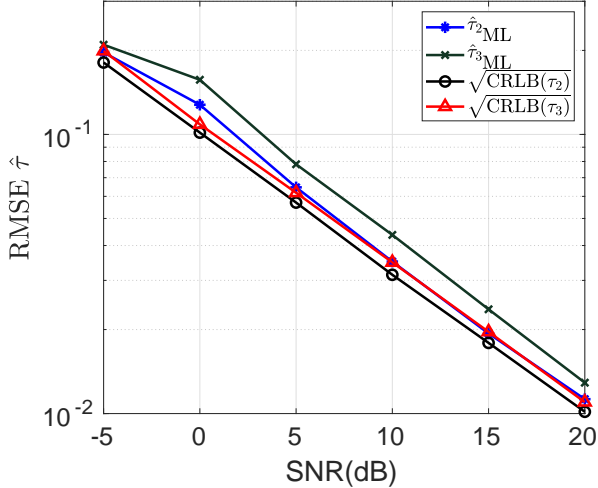
Figure 12: Performance comparison of two-stage algorithm for τ_r.

Finally in Figure 12, the delay time in fractions of symbol period are estimated for NLOS paths assuming $\tau_1 = 0$ for LOS path. The ABP algorithm is unable to estimate the delay time. The two-stage estimation algorithm performs efficiently and achieves the theoretical bound closely.

Combining both coarse estimation based on PREIDG and using it as an initial guess for SAGE, enhances the estimation accuracy of θ_r, $\sqrt{P_T}\alpha_r$ and τ_r drastically, especially in the low SNR regime, which is an important design goal for 5G mmWave systems.

7. Conclusion

In this paper, a novel two-stage estimation method is proposed for channel parameter estimation. The coarse estimation is achieved by the novel proposed PREIDG method which is based on interpolation with a fixed LUT. In a second stage, the SAGE algorithm is applied to refine the estimates of the spatial frequency $\hat{\mu}_r$, obtain the complex path gain $\sqrt{\hat{P}_T}\alpha_r$ and, the non-integer delay $\hat{\tau}_r$, for every path.

The two-stage estimation method used DFT beams for the estimation of μ_r, $\sqrt{P_T}\alpha_r$ and τ_r which is efficiently implemented using a Butler matrix in the analog domain and avoid the use of adaptive phase shifters. The proposed two-stage method remarkably reduces the implementation complexity in the analog domain and improves the estimation accuracy and energy efficiency, especially in the low SNR regime which is interesting for 5G mmWave systems.

We have derived Cramér-Rao lower bound (CRLB) on estimation uncertainty for the spatial frequencies, complex path gains and the delay time between the

line of sight and non-line of sight paths.

Through simulations, we analyzed and compared the performance of our proposed algorithm with ABP, which proved that the proposed two-stage estimation algorithm has lower implementation complexity and efficient estimation performance.

Although the method has been described in detail only for single antenna UE's, it can be extended to multi-antenna UE's, two-dimensional arrays at the BS and usage of orthogonal polarization.

Appendix A. Derivation of ML estimates

The log-likelihood function used in (42) is

$$\ell(\mathbf{X}_r; \boldsymbol{\eta}_r) = \ln\left(\frac{1}{(\pi \beta_r \sigma_n^2)^{ML}} \exp\left(\frac{-||\mathbf{X}_r - \mathbf{S}_r(\boldsymbol{\eta}_r)||_{\mathrm{F}}^2}{\beta_r \sigma_n^2}\right)\right). \qquad (A.1)$$

By taking the expectation with respect to $\mathbf{X}_r$, we get

$$\mathbb{E}_{\mathbf{X}_r}\left[\ell(\mathbf{X}_r; \boldsymbol{\eta}_r)|\mathbf{Y}; \hat{\boldsymbol{\eta}}\right] = -ML \ln\left(\pi \beta_r \sigma_n^2\right) -$$
$$\frac{1}{\beta_r \sigma_n^2}\left(\mathrm{tr}\left\{\hat{\mathbf{X}}_r^{\mathrm{H}}\hat{\mathbf{X}}_r\right\} - \mathrm{tr}\left\{\mathbf{S}_r^{\mathrm{H}}(\boldsymbol{\eta}_r)\hat{\mathbf{X}}_r\right\} -$$
$$\mathrm{tr}\left\{\mathbf{S}_r^{\mathrm{T}}(\boldsymbol{\eta}_r)\hat{\mathbf{X}}_r^*\right\} + ||\mathbf{S}_r(\boldsymbol{\eta}_r)||_{\mathrm{F}}^2\right). \qquad (A.2)$$

by simplifying (A.2) using the following

$$||\mathbf{S}_r^{\mathrm{T}}(\boldsymbol{\eta}_r)||_{\mathrm{F}}^2 = P_T \alpha_r \alpha_r^* \mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_{\mathrm{r}})\mathbf{A}^{\mathrm{H}}(\mu_{\mathrm{r}})\mathbf{A}(\mu_{\mathrm{r}})\mathbf{C}(\tau_{\mathrm{r}})\right\}, \qquad (A.3)$$

by putting (A.3) into (A.2) can lead us to the following concentrated cost function

$$\hat{\boldsymbol{\eta}}_r = \arg\max_{\boldsymbol{\eta}_r} \boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r), \qquad (A.4)$$

where the concentrated $\boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)$ is represented as

$$\boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r) = \left(\sqrt{P_T}\alpha_r^* \mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\hat{\mathbf{X}}_r\right\} +$$
$$\sqrt{P_T}\alpha\, \mathrm{tr}\left\{\mathbf{C}^{\mathrm{T}}(\tau_r)\mathbf{A}^{\mathrm{T}}(\mu_r)\hat{\mathbf{X}}_r^*\right\} -$$
$$\sqrt{P_T}\alpha_r \alpha_r^* \mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\mathbf{A}(\mu_r)\mathbf{C}(\tau_r)\right\}\right). \qquad (A.5)$$

Taking the derivative of $\boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)$ with respect to $\sqrt{P_T}\alpha_r^*$ and setting $\frac{\partial \boldsymbol{\Lambda}_r(\boldsymbol{\eta}_r)}{\partial \sqrt{P_T}\alpha_r^*} = 0$, leads to

$$\sqrt{\hat{P_T}}\alpha_r = \frac{\mathrm{tr}\left\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\hat{\mathbf{X}}_r\right\}}{\mathrm{tr}\{\mathbf{C}^{\mathrm{H}}(\hat{\tau}_r)\mathbf{A}^{\mathrm{H}}(\hat{\mu}_r)\mathbf{A}(\hat{\mu}_r)\mathbf{C}(\hat{\tau}_r)\}}, \qquad (A.6)$$

and finally to

$$(\hat{\tau}_r, \hat{\mu}_r) = \arg\max_{\tau_r, \mu_r} \left\{ \frac{\left| \text{tr} \left\{ \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \hat{\mathbf{X}}_r \right\} \right|^2}{\beta_r \sigma_n^2 \, \text{tr} \{ \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \mathbf{A}(\mu_r) \mathbf{C}(\tau_r) \}} \right\}. \qquad \text{(A.7)}$$

which is used to iteratively and sequentially solve for $\hat{\tau}_r$ and $\hat{\mu}_r$ in (44) and (45).

Appendix B. Entries of the FIM $\mathbf{F}(\boldsymbol{\eta})$

The entries of the block matrices of the FIM (55) are derived as

$$[\mathbf{F}_{\text{Re}\{\boldsymbol{\alpha}\}\text{Re}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\text{Re} \left(\text{tr} \left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \text{Re}\{\alpha_i\}} \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \right. \right.$$
$$\left. \left. \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \text{Re}\{\alpha_j\}} \mathbf{A}(\mu_r) \mathbf{C}(\tau_r) \right\} \right), \qquad \text{(B.1)}$$

$$[\mathbf{F}_{\text{Re}\{\boldsymbol{\alpha}\}\text{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\text{Re} \left(\text{tr} \left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \text{Re}\{\alpha_i\}} \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \right. \right.$$
$$\left. \left. \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \text{Im}\{\alpha_j\}} \mathbf{A}(\mu_r) \mathbf{C}(\tau_r) \right\} \right), \qquad \text{(B.2)}$$

$$[\mathbf{F}_{\text{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\text{Re} \left(\text{tr} \left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \text{Re}\{\alpha_i\}} \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \right. \right.$$
$$\left. \left. \sum_{r=1}^{R} \alpha_r \frac{\partial \mathbf{A}(\mu_r)}{\partial \mu_j} \mathbf{C}(\tau_r) \right\} \right), \qquad \text{(B.3)}$$

$$[\mathbf{F}_{\text{Re}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\text{Re} \left(\text{tr} \left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \text{Re}\{\alpha_i\}} \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \right. \right.$$
$$\left. \left. \sum_{r=1}^{R} \alpha_r \mathbf{A}(\mu_r) \frac{\partial \mathbf{C}(\tau_r)}{\partial \tau_j} \right\} \right), \qquad \text{(B.4)}$$

$$[\mathbf{F}_{\text{Im}\{\boldsymbol{\alpha}\}\text{Im}\{\boldsymbol{\alpha}\}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\text{Re} \left(\text{tr} \left\{ \sum_{r=1}^{R} \frac{\partial \alpha_r^*}{\partial \text{Im}\{\alpha_i\}} \mathbf{C}^{\text{H}}(\tau_r) \mathbf{A}^{\text{H}}(\mu_r) \right. \right.$$
$$\left. \left. \sum_{r=1}^{R} \frac{\partial \alpha_r}{\partial \text{Im}\{\alpha_j\}} \mathbf{A}(\mu_r) \mathbf{C}(\tau_r) \right\} \right), \qquad \text{(B.5)}$$

22

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R}\frac{\partial\alpha_r^*}{\partial\mathrm{Im}\{\alpha_i\}}\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\right.\right.$$
$$\left.\left.\sum_{r=1}^{R}\alpha_r\frac{\partial\mathbf{A}(\mu_r)}{\partial\mu_j}\mathbf{C}(\tau_r)\right\}\right), \tag{B.6}$$

$$[\mathbf{F}_{\mathrm{Im}\{\boldsymbol{\alpha}\}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R}\frac{\partial\alpha_r^*}{\partial\mathrm{Im}\{\alpha_i\}}\mathbf{C}^{\mathrm{H}}(\tau_r)\mathbf{A}^{\mathrm{H}}(\mu_r)\right.\right.$$
$$\left.\left.\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\frac{\partial\mathbf{C}(\tau_r)}{\partial\tau_j}\right\}\right), \tag{B.7}$$

$$[\mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\mu}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R}\alpha_r^*\mathbf{C}^{\mathrm{H}}(\tau_r)\frac{\partial\mathbf{A}^{\mathrm{H}}(\mu_r)}{\partial\mu_i}\right.\right.$$
$$\left.\left.\sum_{r=1}^{R}\alpha_r\frac{\mathbf{A}(\mu_r)}{\partial\mu_j}\mathbf{C}(\tau_r)\right\}\right), \tag{B.8}$$

$$[\mathbf{F}_{\boldsymbol{\mu}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R}\alpha_r^*\mathbf{C}^{\mathrm{H}}(\tau_r)\frac{\partial\mathbf{A}^{\mathrm{H}}(\mu_r)}{\partial\mu_i}\right.\right.$$
$$\left.\left.\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\frac{\partial\mathbf{C}(\tau_r)}{\partial\tau_j}\right\}\right), \tag{B.9}$$

$$[\mathbf{F}_{\boldsymbol{\tau}\boldsymbol{\tau}}]_{ij} = \frac{P_T}{\sigma_n^2} 2\mathrm{Re}\left(\mathrm{tr}\left\{\sum_{r=1}^{R}\alpha_r^*\frac{\partial\mathbf{C}^{\mathrm{H}}(\tau_r)}{\partial\tau_i}\mathbf{A}^{\mathrm{H}}(\mu_r)\right.\right.$$
$$\left.\left.\sum_{r=1}^{R}\alpha_r\mathbf{A}(\mu_r)\frac{\partial\mathbf{C}(\tau_r)}{\partial\tau_j}\right\}\right). \tag{B.10}$$

The partial derivative of α_r and α_r^* is calculated as

$$\frac{\partial\alpha_r}{\partial\mathrm{Re}\{\alpha_i\}} = \frac{\partial\alpha_r^*}{\partial\mathrm{Re}\{\alpha_i\}} = \begin{cases} 1 \text{ if } (r=i) \\ 0 \text{ if } (r\neq i), \end{cases} \tag{B.11}$$

$$\frac{\partial\alpha_r}{\partial\mathrm{Im}\{\alpha_i\}} = -\frac{\partial\alpha_r^*}{\partial\mathrm{Im}\{\alpha_i\}} = \begin{cases} j \text{ if } (r=i) \\ 0 \text{ if } (r\neq i). \end{cases} \tag{B.12}$$

The partial derivative of $\mathbf{A}(\mu_r)$ is calculated as

$$\frac{\partial\mathbf{A}(\mu_r)}{\partial\mu_r} = \mathrm{diag}\{\mathbf{a}'^{\mathrm{H}}(\mu_r)\mathbf{w}(\Phi_k)\}_{k=0}^{M-1}, \tag{B.13}$$

23

with

$$\mathbf{a}'(\mu_r) = \begin{bmatrix} 0, & (-j)\mathrm{e}^{-j\mu_r}, \ldots, (-j(M-1))\mathrm{e}^{-j(M-1)\mu_r} \end{bmatrix}^{\mathrm{T}}. \qquad \text{(B.14)}$$

The partial derivative of the sequence $c(t)$ with respect to τ_r is calculated as

$$c(t) = \sum_{n=-\infty}^{+\infty} c(n)\, h(t - nT_s), \qquad \text{(B.15)}$$

where $h(t)$ is the raised cosine (RC) pulse

$$h(t) = \frac{\sin\left(\pi \frac{t}{T_s}\right)}{\pi \frac{t}{T_s}} \frac{\cos\left(\rho\pi \frac{t}{T_s}\right)}{1 - (2\rho\frac{t}{T_s})^2}, \qquad \text{(B.16)}$$

where $\rho \in [0,1]$, represents the roll-off factor. The delayed sequence by τ_r is represented as

$$c(t - \tau_r) = \sum_{n=-\infty}^{+\infty} c(n)h(t - nT_s - \tau_r). \qquad \text{(B.17)}$$

The partial derivative with respect to τ_r can be written as

$$\begin{aligned}
\frac{\partial c(t - \tau_r)}{\partial \tau_r} &= \sum_{n=-\infty}^{+\infty} c(n)\frac{\partial h(t - nT_s - \tau_r)}{\partial \tau_r} \\
&= -\sum_{n=-\infty}^{+\infty} \left(c(n)\,\frac{\partial h(\tilde{t})}{\partial \tilde{t}} \bigg|_{\tilde{t}=t-nT_s-\tau_r} \right).
\end{aligned} \qquad \text{(B.18)}$$

References

References

[1] Z. Pi, F. Khan, An introduction to Millimeter-Wave mobile broadband systems, IEEE Communications Magazine 49 (6) (2011) 101–107. doi:10.1109/MCOM.2011.5783993.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, F. Gutierrez, Millimeter wave mobile communications for 5G cellular: It will work!, IEEE Access 1 (2013) 335–349. doi:10.1109/ACCESS.2013.2260813.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays, IEEE Signal Processing Magazine 30 (1) (2013) 40–60. doi:10.1109/MSP.2011.2178495.

[4] C. Lee, W. Chung, Hybrid RF-baseband precoding for cooperative multiuser massive MIMO systems with limited RF chains, IEEE Transactions on Communications 65 (4) (2017) 1575–1589. `doi:10.1109/TCOMM.2016.2646359`.

[5] J. Zhu, W. Xu, N. Wang, Secure massive MIMO systems with limited RF chains, IEEE Transactions on Vehicular Technology 66 (6) (2017) 5455–5460. `doi:10.1109/TVT.2016.2615885`.

[6] L. Yang, Y. Zeng, R. Zhang, Efficient channel estimation for Millimeter Wave MIMO with limited RF chains, in: 2016 IEEE International Conference on Communications (ICC), 2016, pp. 1–6. `doi:10.1109/ICC.2016.7510952`.

[7] L. Lian, A. Liu, V. K. N. Lau, Optimal-tuned weighted LASSO for Massive MIMO channel estimation with limited RF chains, in: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017, pp. 1–6. `doi:10.1109/GLOCOM.2017.8254796`.

[8] A. Mezghani, J. A. Nossek, Analysis of rayleigh-fading channels with 1-bit quantized output, in: 2008 IEEE International Symposium on Information Theory, 2008, pp. 260–264. `doi:10.1109/ISIT.2008.4594988`.

[9] A. Mezghani, J. A. Nossek, On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization, in: 2007 IEEE International Symposium on Information Theory, 2007, pp. 1286–1289. `doi:10.1109/ISIT.2007.4557400`.

[10] J. Mo, R. W. Heath, Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information, IEEE Transactions on Signal Processing 63 (20) (2015) 5498–5512. `doi:10.1109/TSP.2015.2455527`.

[11] Y. Jiang, Y. Feng, M. K. Varanasi, Hybrid beamforming for Massive MIMO: A unified solution for both phase shifter and switch networks, in: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), 2018, pp. 1–5. `doi:10.1109/WCSP.2018.8555527`.

[12] X. Yu, J. Zhang, K. B. Letaief, Hybrid precoding in millimeter wave systems: How many phase shifters are needed?, in: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017, pp. 1–6. `doi:10.1109/GLOCOM.2017.8254864`.

[13] S. Suh, A. Basu, C. Schlottmann, P. E. Hasler, J. R. Barry, Low-power Discrete Fourier Transform for OFDM: A programmable analog approach, IEEE Transactions on Circuits and Systems I: Regular Papers 58 (2) (2011) 290–298. `doi:10.1109/TCSI.2010.2071950`.

[14] D. E. Berraki, S. M. D. Armour, A. R. Nix, Application of compressive sensing in sparse spatial channel recovery for beamforming in mmwave outdoor

systems, in: 2014 IEEE Wireless Communications and Networking Conference (WCNC), 2014, pp. 887–892. `doi:10.1109/WCNC.2014.6952205`.

[15] K. Venugopal, A. Alkhateeb, N. Gonzalez Prelcic, R. W. Heath, Channel estimation for hybrid architecture-based wideband millimeter wave systems, IEEE Journal on Selected Areas in Communications 35 (9) (2017) 1996–2009. `doi:10.1109/JSAC.2017.2720856`.

[16] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, H. Wymeersch, Position and orientation estimation through Millimeter-Wave MIMO in 5G systems, IEEE Transactions on Wireless Communications 17 (3) (2018) 1822–1835. `doi:10.1109/TWC.2017.2785788`.

[17] X. Wu, G. Yang, F. Hou, S. Ma, Low-complexity downlink channel estimation for Millimeter-Wave FDD Massive MIMO systems, IEEE Wireless Communications Letters 8 (4) (2019) 1103–1107. `doi:10.1109/LWC.2019.2907955`.

[18] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, R. G. Baraniuk, Distributed compressed sensing of jointly sparse signals, in: Conference Record of the Thirty-Ninth Asilomar Conference onSignals, Systems and Computers, 2005., 2005, pp. 1537–1541. `doi:10.1109/ACSSC.2005.1600024`.

[19] B.H.Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, K. I. Pedersen, Channel parameter estimation in mobile radio environments using the SAGE algorithm, IEEE Journal on Selected Areas in Communications 17 (3) (1999) 434–450. `doi:10.1109/49.753729`.

[20] D. Zhu, J. Choi, R. W. Heath, Auxiliary beam pair design in mmwave cellular systems with hybrid precoding and limited feedback, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 3391–3395. `doi:10.1109/ICASSP.2016.7472306`.

[21] D. Zhu, J. Choi, R. W. Heath, Auxiliary beam pair enabled AoD and AoA estimation in mmwave FD-MIMO systems, in: 2016 IEEE Global Communications Conference (GLOBECOM), 2016, pp. 1–6. `doi:10.1109/GLOCOM.2016.7841616`.

[22] D. Zhu, J. Choi, R. W. Heath, Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems, IEEE Transactions on Wireless Communications 16 (7) (2017) 4770–4785. `doi:10.1109/TWC.2017.2702617`.

[23] D. Zhu, J. Choi, R. W. Heath, Two-dimensional AoD and AoA acquisition for wideband millimeter-wave systems with dual-polarized MIMO, IEEE Transactions on Wireless Communications 16 (12) (2017) 7890–7905. `doi:10.1109/TWC.2017.2754369`.

[24] J. Singh, S. Ramakrishna, On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays, IEEE Transactions on Wireless Communications 14 (5) (2015) 2670–2683.

[25] C. Hu, L. Dai, T. Mir, Z. Gao, J. Fang, Super-Resolution Channel Estimation for MmWave Massive MIMO With Hybrid Precoding, IEEE Transactions on Vehicular Technology 67 (9) (2018) 8954–8958.

[26] J. Lee, G. Gil, Y. H. Lee, Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications, IEEE Transactions on Communications 64 (6) (2016) 2370–2386.

[27] A. Alkhateeb, O. El Ayach, G. Leus, R. W. Heath, Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems, IEEE Journal of Selected Topics in Signal Processing 8 (5) (2014) 831–846.

[28] F. Asim, J. Nossek, F. Antreich, C. C. Cavalcante, A. L. F. d. Almeida, Maximum likelihood channel estimation for Millimeter-Wave MIMO systems with hybrid beamforming, in: WSA 2019; 23rd International ITG Workshop on Smart Antennas, 2019, pp. 1–6.

[29] B. M. Popovic, Generalized chirp-like polyphase sequences with optimum correlation properties, IEEE Transactions on Information Theory 38 (4) (1992) 1406–1409. `doi:10.1109/18.144727`.

[30] U. H. Rohrs, L. P. Linde, Some unique properties and applications of perfect squares minimum phase CAZAC sequences, in: Proceedings of the 1992 South African Symposium on Communications and Signal Processing, 1992, pp. 155–160. `doi:10.1109/COMSIG.1992.274294`.

[31] R. L. Frank, S. A. Zadoff, Phase shift pulse codes with good periodic correlation properties (correspondence), IRE Trans. on Information Theory 8 (Oct 1962) 381–382.

[32] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, C. Masouros, Hybrid analog ;Digital precoding revisited under realistic RF modeling, IEEE Wireless Communications Letters 5 (5) (2016) 528–531. `doi:10.1109/LWC.2016.2598777`.

[33] E. J. Candes, M. B. Wakin, N. S. P. Boyd, Enhancing sparsity by re-weighted l_1 minimization, Journal of Fourier Analysis and Applications, Springer 14 (2008) 877–905.

[34] S.M.Kay, Fundamentals of statistical signal processing: Estimation theory., in: New York,NY, USA: prentice-Hall, 2010, 2010.

[35] M. K. Samimi, T. S. Rappaport, 3-D millimeter-wave statistical channel model for 5G wireless system design, IEEE Transactions on Microwave Theory and Techniques 64 (7) (2016) 2207–2225. `doi:10.1109/TMTT.2016.2574851`.