2-15-2017

Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment

Song Noh
Purdue University, songnoh@purdue.edu

Michael D. Zoltowski
Purdue University, michael.d.zoltowski.1@purdue.edu

David J. Love
Purdue University, djlove@purdue.edu

Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment

Song Noh, Michael D. Zoltowski, and David J. Love

***Abstract*—Millimeter wave (mmWave) communication is expected to be widely deployed in fifth generation (5G) wireless networks due to the substantial bandwidth available at mmWave frequencies. To overcome the higher path loss observed at mmWave bands, most prior work focused on the design of directional beamforming using analog and/or hybrid beamforming techniques in large-scale multiple-input multiple-output (MIMO) systems. Obtaining potential gains from highly directional beamforming in practical systems hinges on sufficient levels of channel estimation accuracy, where the problem of channel estimation becomes more challenging due to the substantial training overhead needed to sound all directions using a high-resolution narrow beam. In this work, we consider the design of multi-resolution beamforming sequences to enable the system to quickly search out the dominant channel direction for single-path channels. The resulting design generates a multilevel beamforming sequence that strikes a balance between minimizing the training overhead and maximizing beamforming gain, where a subset of multilevel beamforming vectors is chosen adaptively to provide an improved average data rate within a constrained time. We propose an efficient method to design a hierarchical multi-resolution codebook utilizing a Butler matrix, a generalized *discrete Fourier transform* (DFT) matrix implemented using analog RF circuitry. Numerical results show the effectiveness of the proposed algorithm.**

***Index Terms*—Millimeter wave, MIMO, multi-resolution beam alignment, Butler matrix**

I. Introduction

The use of millimeter wave (mmWave) bands is considered a key technology for future wireless cellular communications due to the tens of GigaHertz of available licensed and unlicensed spectrum bands in the range of 20–100 GHz. The propagation properties of each of the mmWave bands are characterized by severe path loss and sparse scattering conditions [2]. Compared to the frequency bands used in most current wireless systems, directional beamforming design using large-scale MIMO systems is crucial to overcome the higher path loss at mmWave bands [3], [4].

Due to the radio frequency (RF) hardware limitations found in mmWave implementations, there is growing interest in the design of channel estimation and precoding algorithms [5]. To reduce hardware complexity, the works in [6]–[9] propose analog beam alignment techniques in which phase shifters or discrete lens arrays are employed. To provide increased multiplexing and beamforming gains while using a limited number of RF chains, [10]–[14] present the design of hybrid beam alignment using a combination of analog and digital beamforming.

Obtaining potential gains from highly directional beamforming using a large number of antennas hinges on sufficient levels of channel estimation accuracy. In large-scale MIMO, much of the prior work exploits *a priori* knowledge of the underlying channel statistics such as the channel covariance matrix and the spatial sparsity along with a specific antenna arrangement [15]–[20]. Without the prior information, which requires additional learning algorithms [16], [21]–[23], the problem of channel estimation in the mmWave case becomes more challenging because narrow pencil beams capable of supporting high beamforming gain should be aligned to dominant channel directions located randomly within a cell.

To tackle the problem of channel estimation in a mmWave system, there has been some work on channel estimation using a hierarchical multi-resolution beam codebook design [8], [10], [11], [24]. When the angles of departure or arrival (AoDs/AoAs) of the base station and the user are divided into multiple partitions, the beamforming vectors are sounded at increasing angular resolution to further divide either azimuth or elevation angular regions of interest. Ideally, the beamforming vectors should have beamwidths that grow narrow as the search level increases.

Since the entire area spanned by a beam pattern is bounded given a fixed transmit power, a wide beamwidth beamformer (i.e., low angular resolution) has relatively lower main-lobe directivity gain than a narrow beamwidth beamformer capable of supporting high angular resolution. Therefore, the probability of successfully estimating the AoA/AoD ranges of the previous level becomes lower and lower, and the initial estimation of angular partition can be the bottleneck to minimize the beam misalignment probability.

Most prior work on multilevel beamforming [8], [10], [11], [24] is premised on similar estimation performance at each training stage, which is possible at high signal-to-noise ratio (SNR). When the total signal power used for the training period is large, power allocation can

S. Noh, M. D. Zoltowski, and D. J. Love are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA (e-mail:songnoh@purdue.edu and {mikedz,djlove}@ecn.purdue.edu). A preliminary version of this work will be presented in [1].

improve the estimation performance during early training stages [11], which requires an increased dynamic range for the power amplifiers. Since the performance achievable using a multilevel beamforming framework is closely related to both the training overhead and the beamforming gain, it is critical to understand how to design the multi-resolution beamforming sequences to improve the data rate performance.

In this work, we consider the design of a multilevel beamforming sequence that properly allocates a multi-resolution codebook to the training phase using a fixed signal power. Given a set of multi-resolution codebooks, a subset of the available codebooks is adaptively selected to provide an improved average data rate within a constrained time. We propose an efficient method to design a hierarchical multi-resolution codebook utilizing a Butler matrix implemented using analog RF circuitry.

This manuscript is organized as follows: The system model and background are described in Section II. Section III describes the proposed multilevel beamforming sequence design method. A DFT-based codebook design is discussed in Section IV. Numerical results are provided in Section V, followed by conclusions in Section VI.

Notation Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ indicate the transpose and Hermitian transpose of $\mathbf{A}$, respectively. $[\mathbf{A}]_{i,:}$ denotes the i-row of $\mathbf{A}$. $\text{diag}(d_1, \ldots, d_n)$ is the diagonal matrix composed of elements $d_1, \ldots, d_n$. For a vector $\mathbf{a}$, $a(i)$ denotes the i-th element and $\|\mathbf{a}\|_p$ represents the p-norm. For a real-valued scalar a, $\lceil a \rceil$ and $\lfloor a \rfloor$ denote the smallest integer greater than or equal to a and the largest integer less than or equal to a, respectively. $\mathbb{E}\{\cdot\}$ represents statistical expectation. $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ denote the sets of natural, real, and complex numbers, respectively.

II. System Model

A. System Setup

We consider a mmWave MIMO link composed of uniform linear arrays (ULAs) of N_t transmit antennas and N_r receive antennas. We assume block transmission with K successive symbols for one block where the block length K is smaller than the channel coherence time, and K_p out of K channel uses are set aside for channel sounding. For a unit norm transmit beamforming vector $\mathbf{f}_k \in \mathbb{C}^{N_t}$, the received signal $\mathbf{y}_k \in \mathbb{C}^{N_r}$ at a discrete symbol time k is given by

$$\mathbf{y}_k = \sqrt{P}\mathbf{H}\mathbf{f}_k s_k + \mathbf{z}_k, \quad \text{for } k = 1, 2, \ldots, K \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the MIMO channel matrix, an $N_r \times 1$ Gaussian noise vector $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_{N_r})$, P denotes the transmit power, and $s_k \in \mathbb{C}$ represents either the pilot symbol during the training period (i.e., $1 \le k \le K_p$) or the data symbol during the data transmission period ($K_p < k \le K$) with $\mathbb{E}\{|s_k|^2\} = 1$. In (1), the transmitter sounds K_p beamforming vectors and chooses a vector from them that archives the largest received signal power, referred to as hard beam alignment [8]. The selected beamformer is then used for data transmission (i.e., channel uses satisfying $K_p < k \le K$).

Transmission over mmWave frequencies using narrow transmit and receive beams can limit the multipath components of the effective mmWave channel [25]. Thus, we adopt a geometric-based channel model with L scatterers [26] where the channel matrix $\mathbf{H}$ for a ULA is given by

$$\mathbf{H} = \sqrt{N_t N_r} \sum_{\ell=1}^{L} g_\ell \mathbf{u}_r(\theta_\ell^r) \mathbf{u}_t^H(\theta_\ell^t), \tag{2}$$

where $\mathbf{u}_t(\cdot)$ and $\mathbf{u}_r(\cdot)$ denote the transmitter steering vector and the receiver response vector in the normalized steering phases, i.e.,

$$\mathbf{u}_s(\theta_\ell^s) = \frac{1}{\sqrt{N_s}} \left[1, e^{-j2\pi \frac{d}{\lambda}\theta_\ell^s}, \cdots, e^{-j(N_s-1)2\pi \frac{d}{\lambda}\theta_\ell^s}\right]^T, \tag{3}$$

where $j = \sqrt{-1}$, $s \in \{t, r\}$, λ is the carrier wavelength, and d is the antenna spacing given by half-wavelength spacing $d = \lambda/2$. For the possible physical angles $\phi_\ell^s \in (-\pi/2, \pi/2)$ that cover the entire (one-sided) spatial horizon, we define the *spatial* frequency $\theta_\ell^s \in (-1, 1)$ as $\theta_\ell^s = \sin(\phi_\ell^s)$ [27] where there is a one-to-one mapping between θ_ℓ^s and ϕ_ℓ^s. Denote the complex small-scale fading gain of the ℓ-th subpath by g_ℓ modeled as independent and identically distributed (i.i.d.) complex Gaussian, i.e., $g_\ell \sim \mathcal{CN}(0, \sigma_g^2)$ [25].

We first consider the case of quantized AoAs/AoDs assuming $\theta_\ell^s \in \{-1 + (2n-1)/N_s : 1 \le n \le N_s\}$ for $s \in \{t, r\}$ as in [10], [11] and then extend to the case of continuous angles in Sections IV-A and V. In this scenario, the MIMO channel matrix $\mathbf{H}$ in (2) can be rewritten as

$$\mathbf{H} = \mathbf{U}_r \tilde{\mathbf{H}} \mathbf{U}_t^H \tag{4}$$

where $\mathbf{U}_s = [\mathbf{u}_s(\theta_1^s) \cdots \mathbf{u}_s(\theta_{N_s}^s)] \in \mathbb{C}^{N_s \times N_s}$ for $s \in \{t, r\}$. Here, $\tilde{\mathbf{H}}$ has only L non-zero elements, i.e., $[\tilde{\mathbf{H}}]_{m,n} = \sqrt{N_t N_r} g_\ell$ where the indices (m, n) are associated with the angles of θ_m^r and θ_n^t of the ℓ-th path. Thus, we may require at *most* $N_r N_t$ channel measurements to find the non-zero entries of $\tilde{\mathbf{H}}$.

To partition the set of AoAs into subintervals as in [24], [28], we assume that the receiver employs the filter bank $\mathbf{U}_r^H$ to the received signal in (1) given by

$$\tilde{\mathbf{y}}_k = \mathbf{U}_r^H \mathbf{y}_k = \sqrt{P} \tilde{\mathbf{H}} \mathbf{U}_t^H \mathbf{f}_k s_k + \tilde{\mathbf{z}}_k, \tag{5}$$

where $\tilde{\mathbf{z}}_k = \mathbf{U}_r^H \mathbf{z}_k$. For a ULA, the receive filter bank can be implemented using a DFT-based filter [15]. We convert a joint detection of the non-zero elements of $\tilde{\mathbf{H}}$ in the whole range of AoAs to separate detection problems

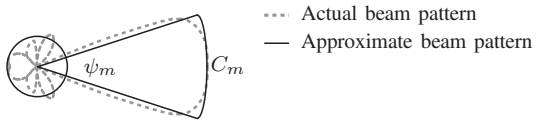


Fig. 1: Sectored beam model.

in small angle intervals.[1]

B. Review of Prior Work on Multi-Resolution Beam Alignment

Much of the prior work considers a multi-resolution beam alignment framework [8], [10], [11] where the hierarchical multi-resolution codebook consists of M levels to achieve better angular resolution. For $1 \leq m \leq M$, the m-th level codebook $\mathcal{F}_m = \left\{\mathbf{f}_1^{(m)}, \ldots, \mathbf{f}_{\text{card}(\mathcal{F}_m)}^{(m)}\right\}$ is designed to uniformly cover all the spatial frequency range $(-1, 1)$ (i.e., one-sided physical angles $(-\pi/2, \pi/2)$) and to satisfy the relation $\text{card}(\mathcal{F}_1) < \cdots < \text{card}(\mathcal{F}_M)$, where $\text{card}(\mathcal{F})$ denotes the cardinality of $\mathcal{F}$. (An example of a multi-resolution codebook with $[\text{card}(\mathcal{F}_1), \text{card}(\mathcal{F}_3), \text{card}(\mathcal{F}_3)] = [8, 16, 64]$ is shown in Fig. 5.) The beamwidth of each of the codewords in $\mathcal{F}_m$ is given by

$$\psi_m = 2/\text{card}(\mathcal{F}_m), \qquad \text{for } 1 \leq m \leq M, \tag{6}$$

and satisfies $\psi_1 > \cdots > \psi_M$. Accordingly, the angular domains are partitioned into $\text{card}(\mathcal{F}_m)$ angular intervals for the m-th level codebook where we denote each angular interval covered by $\mathbf{f}_v^{(m)}$ by $\text{AoD}_{m,v} \subset (-1, 1)$ such that $(-1, 1) = \cup_{v=1}^{\text{card}(\mathcal{F}_m)} \text{AoD}_{m,v}$ and $\text{AoD}_{m,v} \cap \text{AoD}_{m,w} = \emptyset$ for $v \neq w$. Thus, the length of the interval $\text{AoD}_{m,v}$ is given by $|\text{AoD}_{m,v}| = \psi_m$ for $1 \leq v \leq \text{card}(\mathcal{F}_m)$ where $|\text{AoD}_i|$ denotes the length of an angular interval AoD_i. The training period is partitioned into M stages where the m-th training stage is allocated for sounding a subset of beamforming vectors in $\mathcal{F}_m$.[2]

The codebook $\mathcal{F}_m$ is designed to provide an approximately constant main-lobe beamforming gain over the angle interval, which is mathematically described as

$$\left|\mathbf{u}_t(\theta)^H \mathbf{f}_v^{(m)}\right| = \begin{cases} \sqrt{C_m} & \text{if } \theta \in \text{AoD}_{m,v} \\ 0 & \text{if } \theta \notin \text{AoD}_{m,v} \end{cases}, \tag{7}$$

where C_m is a normalization constant that satisfies $\|\mathbf{f}_v^{(m)}\|_2^2 = 1$. Therefore, the beam patterns of $\mathcal{F}_m$ are characterized by the main-lobe directivity gain C_m and the main-lobe beamwidth ψ_m with negligible side-lobe gain, in a manner similar to [29], [30], which is illustrated in Fig. 1.

Denote the angular interval to be searched at the m-th training stage by $\text{AoD}_m \subset (-1, 1)$ and the number of distinct beamforming vectors in $\mathcal{F}_m$ required for searching AoD_m by $V_m \in \mathbb{N}$ given by $V_m = \lceil |\text{AoD}_m|/\psi_m \rceil$. By abuse of notation, we write $\left\{\mathbf{f}_1^{(m)}, \ldots, \mathbf{f}_{V_m}^{(m)}\right\} \in \mathcal{F}_m$ to denote the selected V_m codewords to cover the collection of disjoint and mutually exclusive AoD_m, i.e., $\text{AoD}_m = \cup_{v=1}^{V_m} \text{AoD}_{m,v}$.

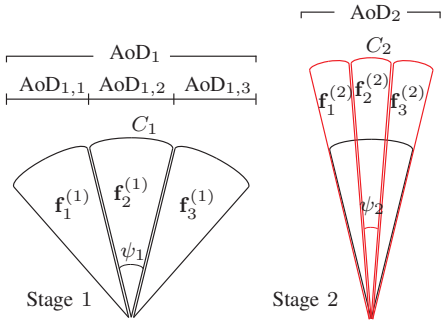


Fig. 2: Approximate beamforming patterns in a multilevel beamforming structure where $V_1 = V_2 = 3$, $M = 2$, and $\text{AoD}_{1,2} = \text{AoD}_2$.

At the m-th training stage, the system model of (5) can be rewritten as

$$\left[\tilde{\mathbf{y}}_1^{(m)}, \ldots, \tilde{\mathbf{y}}_{V_m}^{(m)}\right] = \sqrt{P}\tilde{\mathbf{H}}\mathbf{U}_t^H \left[\mathbf{f}_1^{(m)}, \ldots, \mathbf{f}_{V_m}^{(m)}\right] + \left[\tilde{\mathbf{z}}_1^{(m)}, \ldots, \tilde{\mathbf{z}}_{V_m}^{(m)}\right], \tag{8}$$

where the pilot symbols are assumed normalized to unit power and omitted. From (8), the receiver selects a beam index v' that achieves the largest received signal power given by

$$v' = \underset{1 \leq v \leq V_m}{\text{argmax}} \left|\tilde{y}_v^{(m)}(n)\right|^2, \tag{9}$$

where $\tilde{y}_v^{(m)}(n)$ denotes the n-th element of the vector $\tilde{\mathbf{y}}_v^{(m)}$ and the receiver selects the index n that yields the largest average magnitude of $\left[\tilde{y}_1^{(m)}(n), \ldots, \tilde{y}_{V_m}^{(m)}(n)\right]$. For the sake of simplicity, we assume that the receiver knows the index n and will omit the dependence on (n). The beam index v' in (9) which corresponds to $\mathbf{f}_{v'}^{(m)}$ and $\text{AoD}_{m,v'}$ is fed back to enable the transmitter to choose the next-level codewords in $\mathcal{F}_{m+1}$. (A hierarchical beam alignment system requires feedback of the codebook index across the training stages.)

At training stage $(m+1)$, the angle interval of interest is set to $\text{AoD}_{m+1} = \text{AoD}_{m,v'}$ and the V_{m+1} codewords required for probing AoD_{m+1} is selected from $\mathcal{F}_{m+1}$. Note from *(A.1)* that the value of V_{m+1} is given by

$$V_{m+1} = \left\lceil \frac{|\text{AoD}_{m+1}|}{\psi_{m+1}} \right\rceil = \left\lceil \frac{\psi_m}{\psi_{m+1}} \right\rceil, \tag{10}$$

where $\psi_m = |\text{AoD}_{m,v'}|$ for $1 \leq m < M$ given the constrained training period length $K_p = \sum_{m=1}^{M} V_m$ within a block.

An example of a multilevel beamforming framework with $V_1 = V_2 = 3$ and $M = 2$ is shown in Fig. 2.

[1] The receiver can use a combiner for the received signals as the transmitter uses the beamformer in the same way.

[2] We use the term "level" to refer the codebook and the term "stage" to refer the training phase.

Given AoD_1, $\mathbf{f}_2^{(1)}$ is chosen at stage 1 and then higher-resolution beamformers $\{\mathbf{f}_1^{(2)}, \ldots, \mathbf{f}_3^{(2)}\}$ on $\text{AoD}_2 = \text{AoD}_{1,2}$ are sounded at stage 2.

III. Proposed Multilevel Beamforming Framework

In this section, We first explain the design of a single-level beam alignment sequence that use a single codebook chosen from a set of multilevel codebooks during the training period. We then extend the single-level beam alignment framework to design a multilevel beam alignment sequence where a subset of the set of multilevel codebooks used for channel sounding is selected adaptively aimed at improving the data rate performance.

A. Adaptive Single Level Beam Alignment and Sequence Design

To estimate an AoD partition during the training period, the receiver chooses a codeword achieving the largest received signal power as in (9). However, if the desired signal power is less than the uncertainty in the noise power at moderate-to-low SNR, the receiver cannot robustly detect the desired signal. In contrast to the previous methods explained in Section II-B, we do not use the expression in (9). Instead we employ a threshold technique that is adapted to the SNR.

In this subsection, we consider a single-level beam alignment in which a single codebook selected from $\{\mathcal{F}_m : 1 \leq m \leq M\}$ is exclusively used during the training period. For the m^{th}-level codebook $\mathcal{F}_m$, we replace the alignment decision in (9) with a hypothesis test given by

$$\begin{cases} \mathcal{H}_0^m : \tilde{y}_v^{(m)} \sim \mathcal{CN}(0, \sigma_z^2) \\ \mathcal{H}_1^m : \tilde{y}_v^{(m)} \sim \mathcal{CN}(G_m P |g|^2, \sigma_z^2) \end{cases}, \tag{11}$$

where $G_m := N_t N_r C_m$ and the path gain $|g|^2$ may not be known a priori but follows a central chi-square distribution with 2ℓ degrees of freedom, i.e., $|g|^2 \sim (\sigma_g^2/2)\chi_{2\ell}^2$. Here, $\ell \leq L$ represents the number of paths within the corresponding beamwidth.[3] For tractability of the analysis, we consider the case of single dominant path channels (i.e., $L = 1$ in (2)), then we make some comments regarding the multi-path case.

Under a fixed signal power, the beamformer $\mathbf{f}_v^{(m)}$ has a lower main-lobe directivity gain than $\mathbf{f}_v^{(m+1)}$ due to its broader beamwidth ($\psi_m > \psi_{m+1}$). Thus, the probability of successfully estimating the angular interval becomes lower as the index m decreases. However, using high-resolution beamforming vectors during the training period requires substantial training overhead due to its narrow beamwidth which can degrade the data rate performance. This tradeoff indicates that the index m specifying the codebook level $\mathcal{F}_m$ used during the training period is a suitable design parameter.

[3]For the multi-path case, the number of paths covered by a beamforming vector depends on the propagation geometry, which makes the closed-form expression of the data rate performance required for a sequence design intractable.

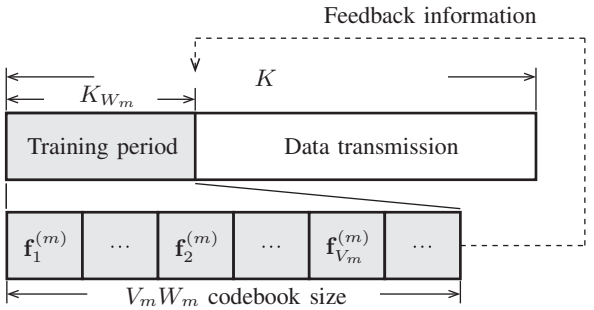


Fig. 3: Training transmission where the codewords $\{\mathbf{f}_1^{(m)}, \ldots, \mathbf{f}_{V_m}^{(m)}\} \in \mathcal{F}_m$ are retransmitted over W_m channel uses.

Given the codebook $\mathcal{F}_m$, we introduce a new variable $W_m \in \mathbb{N}$ to specify the number of repetitions corresponding to each of the codewords in $\mathcal{F}_m$. Then, the codewords $\{\mathbf{f}_1^{(m)}, \ldots, \mathbf{f}_{V_m}^{(m)}\} \in \mathcal{F}_m$ selected to cover the sector of interest denoted by AoD_1 are sounded, where the received signal $\tilde{y}_v^{(m)} \in \mathbb{C}$ in (9) corresponding to $\mathbf{f}_v^{(m)}$ is changed to $\left[\tilde{y}_{1,v}^{(m)} \ldots, \tilde{y}_{W_m,v}^{(m)}\right]^T \in \mathbb{C}^{W_m}$ given by

$$\begin{aligned}\left[\tilde{y}_{1,v}^{(m)}, \ldots, \tilde{y}_{W_m,v}^{(m)}\right] &= \sqrt{P}\tilde{\mathbf{H}}_{n,:}\mathbf{U}_t^H\left(\mathbf{1}_{W_m}^T \otimes \mathbf{f}_v^{(m)}\right) \\ &\quad + \left[\tilde{z}_{1,v}^{(m)}, \ldots, \tilde{z}_{W_m,v}^{(m)}\right], \end{aligned} \tag{12}$$

where $\mathbf{1}_N \in \mathbb{R}^N$ denotes the all-ones vector, $\otimes$ denotes the Kronecker product, and recall that we omit the index (n) for a concise notation. Here, the variable V_m is redefined as

$$V_m = \left\lceil \frac{|\text{AoD}_1|}{\psi_m} \right\rceil, \tag{13}$$

as to represent the size of the subset of $\mathcal{F}_m$ used to cover the angular interval AoD_1. The training overhead becomes

$$K_m = V_m W_m. \tag{14}$$

In Fig. 3, we illustrate the process of training and data transmission in a single level beam alignment framework. We will formulate an optimization problem over the variables m and W_m where the objective is to provide an improved average data rate.

For the present model in (11) and (12), we adopt a Neyman-Pearson detector for $\mathbf{f}_v^{(m)}$ given by [31]

$$\left| \frac{1}{W_m} \sum_{n=1}^{W_m} \tilde{y}_{n,v}^{(m)} \right|^2 \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma_m, \tag{15}$$

where the threshold is set to $\gamma_m = -\sigma_z^2 \ln(P_{FA})/W_m$

given a target false alarm probability P_{FA}, i.e., $P_{FA} = P\{\left|\frac{1}{W_m}\sum_{n=1}^{W_m}\tilde{y}_{n,v}^{(m)}\right|^2 > \gamma_m | \mathcal{H}_0^m\}$. If there are multiple codewords that satisfy (15), the receiver selects the single codeword index among them that achieves the largest received signal power (i.e., the left-hand side of (15)). The combined signal on the left-hand side of (15) can be rewritten as

$$\frac{1}{W_m}\sum_{n=1}^{W_m}\tilde{y}_{n,v}^{(m)} = \sqrt{G_m P}g + \frac{1}{W_m}\sum_{n=1}^{W_m}\tilde{z}_{n,v}^{(m)} =: x_m + z_m, \tag{16}$$

where $x_m = \sqrt{G_m P}g$ and $z_m = (1/W_m)\sum_{n=1}^{W_m}\tilde{z}_{n,v}^{(m)}$ denote the desired signal and noise, respectively, i.e., $x_m \sim \mathcal{CN}(0, G_m P\sigma_g^2)$ and $z_m \sim \mathcal{CN}(0, \sigma_z^2/W_m)$, and the variables are assumed to be mutually independent. That is, the magnitude of x_m represents the instantaneous effective channel gain achievable after a decision $\mathcal{H}_1^m$, while we operate as if $|x_m| = 0$ for a decision $\mathcal{H}_0^m$ because there is no transmission. The beam misalignment analysis, similar to [8], [11], can be performed using the expressions of false alarm and miss detection probabilities of (15). However, this approach cannot be applied to the design of the optimal training length because reducing the probability of beam misalignment requires the maximum training length. Instead, we consider the average data rate as a performance metric and focus on designing a beam alignment sequence and the training length within a constrained time.

When a single propagation path is covered by a beamformer $\mathbf{f}_v^{(m)} \in \mathcal{F}_m$, an instantaneous data rate for $\mathcal{H}_1^m$ is obtained given by

$$R(x_m) = \frac{K - K_m}{K}\cdot\log\left(1 + \frac{|x_m|^2}{\tilde{\sigma}_z^2}\right), \tag{17}$$

where $\tilde{\sigma}_z^2 = \sigma_z^2/W_m$ and the pre-log factor in (17) represents the cost of channel sounding within K channel uses. Note that since a signal is detected when the squared absolute value of the (combined) received signal is above the threshold, consideration of joint variations for the desired signal and noise is needed to provide an exact average data rate, i.e.,

$$\begin{aligned} R_m &= \mathbb{E}\left\{R(x_m)|A_0\right\}p(A_0) + \mathbb{E}\left\{R(x_m)|A_1\right\}p(A_1) \\ &= \frac{K - K_m}{K}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\log\left(1 + \frac{|x_m|^2}{\tilde{\sigma}_z^2}\right)\cdot \\ &\quad p\left(x_m, z_m|A_1\right)dx_m dz_m \times p(A_1), \end{aligned} \tag{18}$$

where A_0 and A_1 denotes mutually exclusive events described in terms of the random variables x_m and z_m (i.e., $A_0 = \{|x_m + z_m|^2 < \gamma_m\}$ and $A_1 = \{|x_m + z_m|^2 \geq \gamma_m\}$) and $p\left(x_m, z_m|A_1^m\right)$ is a conditional joint probability distribution function (pdf) of (x_m, z_m) given A_1. Here, $\mathbb{E}\left\{R(x_m)|\mathcal{H}_0^m\right\} = 0$ because we set $|x_m|^2 = 0$ for $(x_m, z_m) \in A_0$ from a consideration of no transmission. To the best of the authors' knowledge, the integral in (18) does not admit a closed-form expression. Instead, we evaluate an average magnitude of the effective channel gain (i.e., $|x_m|^2$) to obtain a reasonable criterion for approximating the expression of (18). To gain a clear insight to the main factor, we now provide a simple expression for the desired signal power in the high SNR regime, while an exact expression without any approximation is given in Appendix A.

Lemma 1: The desired signal power in the high SNR regime (i.e., $G_m P\sigma_g^2 \gg \sigma_z^2/W_m$) is approximated as

$$\begin{aligned} \mathbb{E}\left\{|x_m|^2\right\} \approx\ & e^{-\frac{\gamma_m}{G_m P\sigma_g^2}}\left(G_m P\sigma_g^2 + \gamma_m\right) + \frac{P_{FA}}{2} \\ & - f_{\sigma_z^2, P_{FA}, W_m}, \end{aligned} \tag{19}$$

where $f_{\sigma_z^2, P_{FA}, W_m} = \frac{\sigma_z^2\ln(1/P_{FA})}{W_m}\left(\sqrt{P_{FA}} - P_{FA}\right)$.

Proof: See Appendix A.

Note that the average signal power given $\mathcal{H}_1^m$ is characterized by the first term on the right-hand side of (19), which is the expected value of $|x_m|^2$ integrated over (γ_m, ∞), because the second term is negligible due to a small target false alarm rate and we numerically find that the magnitude of $f_{\sigma_z^2, P_{FA}, W_m}$ is less than $0.53\sigma_z^2$ for all $P_{FA} \in (0, 1)$ and $W_m \in \{1, \ldots, 10^2\}$.

From Lemma 1, we consider the case that an instantaneous data rate is given by (17) when $|x_m|^2 > \gamma_m$ (i.e., $G_m P|g|^2 > \gamma_m$), otherwise we have $R(x_m) = 0$. Then, R_m in (18) is approximated as

$$R_m \approx \frac{K - K_m}{K}\int_{\gamma_m}^{\infty}\log\left(1 + \frac{|x_m|^2}{\tilde{\sigma}_z^2}\right)\frac{e^{-|x_m|^2/\sigma_m^2}}{\sigma_m^2}d|x_m| \tag{20}$$

$$= \frac{K - K_m}{K}\left(e^{\frac{\tilde{\sigma}_z^2(1-\beta_m)}{\sigma_m^2}}\log\beta_m + \frac{e^{\frac{\tilde{\sigma}_z^2}{\sigma_m^2}}}{\ln 2}\Gamma\left(0, \frac{\tilde{\sigma}_z^2\beta_m}{\sigma_m^2}\right)\right), \tag{21}$$

where $\sigma_m^2 := G_m P\sigma_g^2$, $\beta_m = 1 + \gamma_m/\tilde{\sigma}_z^2$, and $\Gamma(\cdot,\cdot)$ denotes the incomplete gamma function [32, Eq. 8.350.2]. Maximization of R_m in (21) gives the design of the codebook level m and the number of retransmissions of the corresponding codewords W_m.

Problem 1: Given the codebooks $\{\mathcal{F}_m : 1 \leq m \leq M\}$, the variables m and W_m must be chosen to maximize

$$\begin{aligned} \max_{\{m, W_m\}}\ & R_m \\ \text{subject to}\ & 1 \leq m \leq M \text{ and } K_m \leq K, \end{aligned} \tag{22}$$

where K_m and R_m are defined in (14) and (21), respectively.

Considering the integer constraint on the optimization

variables, we tackle Problem 1 by exhaustive search. Due to space limitations, we omit the details of our algorithm for Problem 1. Instead, in the next subsection, we present an algorithm for a multilevel beam alignment sequence design that can be used to solve Problem 1.

Remark 1: By applying Alzer's inequality to the incomplete gamma function [33], R_m in (21) can be lower bounded in a closed form given by

$$R_m \geq \underline{R}_m =: \frac{K-K_m}{K}\left(e^{\frac{\tilde{\sigma}_z^2(1-\beta_m)}{\alpha_m}}\log\beta_m - e^{\frac{\tilde{\sigma}_z^2}{\alpha_m}}\log\left(1-e^{-e^{C}\left(\frac{\tilde{\sigma}_z^2\beta_m}{\alpha_m}\right)}\right)\right), \tag{23}$$

where C is Euler's contant, i.e., $C = 0.5772....$

To reduce the computational complexity required for numerical calculation of $\Gamma(\cdot,\cdot)$, we can use the lower bound of R_m in (23) for a beam alignment sequence design. Simulation results in Section V show that the proposed algorithm using the lower bound in (23) performs very closely to those results obtained using the expression in (21).

Note that the proposed beam alignment design can be directly used for the multi-path case where the selected single codeword is expected to cover at *least* a single dominant path. Alternatively, the codeword indices that achieve the received signal power above the threshold value in (15) can be fed back to the transmitter to enable multi-stream transmission using the combined codewords, which requires additional feedback. Joint processing of the multiple training snapshots for the multi-path case is beyond the scope of the current work.

B. Adaptive Multilevel Beam Alignment and Sequence Design

In this subsection, we focus on the design of a multilevel beam alignment sequence where a subset of codebooks (i.e., $\{\mathcal{F}_m : m_F \leq m \leq m_L\}$) can be used during the training period. Here, we define the indices specifying the codeword levels used at the first and last training stages as m_F and m_L for $1 \leq m_F \leq m_L \leq M$, respectively.

Since the codewords in $\mathcal{F}_{m_F}$ sounded at the first training stage have relatively lower directivity gain due to its broader beamwidth than those in $\{\mathcal{F}_m : m_F < m \leq m_L\}$, the received signal power at the first training stage can be easily swayed by noise impact. Thus, we assume retransmission of the codewords in $\mathcal{F}_{m_F}$ sounded during the first training stage where W_{m_F} represents the number of retransmissions. The system model and the detection process at the first training stage can be described by replacing the index of the codebook in (12) and (15) with (m_F), respectively. During the intermediate

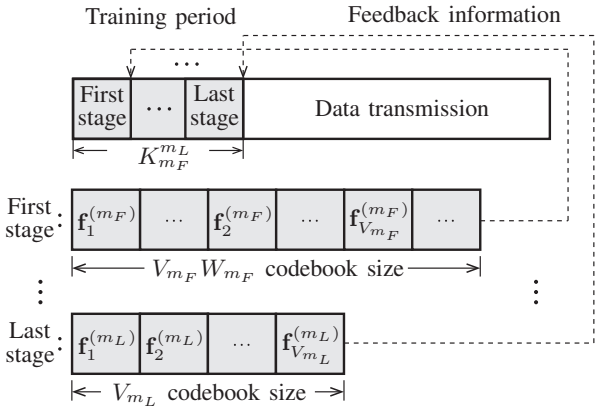


Fig. 4: Training and data transmissions where the codewords sounded at the first training stage are retransmitted over W_{m_F} channel uses, otherwise each codeword is transmitted once.

training stages allocated for sounding the codewords in $\{\mathcal{F}_m : m_F < m \leq m_L\}$, we assume that each codeword is transmitted once, i.e., we set $W_m = 1$ in (12) and (15) for $m_F < m \leq m_L$.

Given this, the training overhead becomes

$$K_{m_F}^{m_L} = V_{m_F}W_{m_F} + \sum_{m=m_F+1}^{m_L} V_m, \tag{24}$$

where the variable V_m corresponding to the codebook $\mathcal{F}_m$ is redefined as

$$V_{m_F} = \lceil |\text{AoD}_1|/\psi_{m_F} \rceil \text{ and } V_m = \lceil \psi_{m-1}/\psi_m \rceil, \tag{25}$$

for $m_F < m \leq m_L$. Here, AoD_1 denotes the sector angle to be searched at the first training stage. The overall process of training and data transmission is illustrated in Fig. 4. As a result, we will optimize the codeword levels (m_F, m_L) and W_{m_F} used during the consecutive training stages to provide an improved average data rate, which will be shown shortly.

We observe that the performance achievable by the multilevel beam alignment framework is characterized by the detection performance at the first stage because the codewords in $\mathcal{F}_{m_F}$ have relatively low directivity gain than those in $\{\mathcal{F}_m : m_F < m \leq m_L\}$. For the purpose of analysis, we assume that if a beamforming vector in $\mathcal{F}_{m_F}$ is properly selected, then a beamforming vector in $\mathcal{F}_m$ for $m_F < m \leq m_L$ that covers the dominant channel path can be perfectly selected due to its relatively higher beamforming gain. We will refer to this assumption as *(A.1)*. (As the comparison with numerical results without *(A.1)*, this approximation does not significantly affect our results.)[4]

[4]This assumption is not very restrictive in our view because for a single level codebook (i.e., $m_F = m_L$), *(A.1)* is not involved in a beam alignment sequence design, and for a multilevel codebook (i.e., $m_F < m_L$) we also optimize the codeword levels (m_F, m_L) to improve the performance.

From *(A.1)*, an (approximate) instantaneous data rate for $|x_{m_F}|^2 > \gamma_{m_F}$ can be written, similar to (17), as

$$R(x_{m_F}, m_L) \approx \frac{K - K_{m_F}^{m_L}}{K} \cdot \log\left(1 + \frac{G_{m_L}}{G_{m_F}} \frac{|x_{m_F}|^2}{\tilde{\sigma}_z^2}\right), \tag{26}$$

where the operand in the log function denotes the beamforming gain achievable at the last training stage under *(A.1)*. We then evaluate the (approximate) average throughput in the multilevel beamforming framework in semi-closed form given by

$$\begin{aligned} R_{m_F}^{m_L} &\approx \mathbb{E}\left\{R(x_{m_F}, m_L)\right\} \qquad (27)\\ &= \frac{K - K_{m_F}^{m_L}}{K}\left(e^{\tilde{\sigma}_z^2\left(1-\beta_{m_F}^{m_L}\right)/\alpha^{m_L}} \cdot \log \beta_{m_F}^{m_L}\right.\\ &\quad \left. + \left(e^{\tilde{\sigma}_z^2/\alpha^{m_L}}/\ln 2\right) \cdot \Gamma\left(0, \tilde{\sigma}_z^2 \beta_{m_F}^{m_L}/\alpha^{m_L}\right)\right) \qquad (28)\end{aligned}$$

where $\alpha^{m_L} = G_{m_L} P \sigma_g^2$, $\beta_{m_F}^{m_L} = 1 + \frac{G_{m_L}\gamma_{m_F}}{G_{m_F}\tilde{\sigma}_z^2}$, and the expectation in (27) is performed over $|x_{m_F}|^2 \in (\gamma_{m_F}, \infty)$, similar to (20).

Therefore, using the objective function in (28), we design a multilevel beam alignment sequence optimized over the variables (m_F, m_L, W_{m_F}) that strikes a balance between minimizing the training overhead and maximizing beamforming gain.

Problem 2: Given the codebooks $\{\mathcal{F}_m : 1 \le m \le M\}$ and *(A.1)*, the level indices (m_F, m_L) and W_{m_F} must be chosen to maximize

$$\max_{\{m_F, m_L, W_{m_F}\}} R_{m_F}^{m_L} \tag{29}$$

$$\text{subject to } 1 \le m_F \le m_L \le M, \quad K_{m_F}^{m_L} \le K, \quad \text{and } 1 \le W_{m_F} \le \lfloor G_{m_F+1}/G_{m_F} \rfloor, \tag{30}$$

where $K_{m_F}^{m_L}$ and $R_{m_F}^{m_L}$ are defined in (24) and (28), respectively.

The upper bound on W_{m_F} in (30) guarantees that the effective SNR at level m_F is smaller than that of the next level, i.e., $G_{m_F} P|g|^2/(\sigma_z^2/W_{m_F}) \le G_{m_F+1} P|g|^2/\sigma_z^2$ because of $W_{m_F} G_{m_F} \le G_{m_F+1}$. The corresponding algorithm is summarized in Algorithm 1 with reasonable computational complexity of $\mathcal{O}(M^2 K)$ as only three to five levels are typically used in the multilevel beamforming framework.

Remark 2: Similar to Remark 1, the expression of (28) can be lower bounded by $R_{m_F}^{m_L} \ge \underline{R}_{m_F}^{m_L}$ where

$$\begin{aligned} \underline{R}_{m_F}^{m_L} = \frac{K - K_{m_F}^{m_L}}{K}\left(e^{\frac{\tilde{\sigma}_z^2\left(1-\beta_{m_F}^{m_L}\right)}{\alpha^{m_L}}} \log \beta\right.\\ \left. - e^{\frac{\tilde{\sigma}_z^2}{\alpha^{m_L}}} \log\left(1 - e^{-e^C \frac{\tilde{\sigma}_z^2 \beta_{m_F}^{m_L}}{\alpha^{m_L}}}\right)\right). \end{aligned} \tag{31}$$

Algorithm 1 Multilevel Beam Sequence Design

```
Set R = 0
for m_L = M : −1 : 1 do
    for m_F = m_L : −1 : 1 do
        [V_{m_F}, ..., V_{m_L}] = [⌈AoD_1/ψ_{m_F}⌉, ⌈ψ_{m_F}/ψ_{m_F+1}⌉, ..., ⌈ψ_{m_L−1}/ψ_{m_L}⌉]
        W_max = ⌊(K − Σ_{m=m_F+1}^{m_L} V_m)/V_{m_F}⌋
        if m_F < m_L then
            W_max = min{W_max, ⌊G_{m_F+1}/G_{m_F}⌋}
        end if
        for W_{m_F} = 1 : 1 : W_max do
            Compute R_{m_F}^{m_L}                                   (Step †)
            if R_{m_F}^{m_L} > R then
                R = R_{m_F}^{m_L}
                [m̂_F, m̂_L] ← [m_F, m_L]
                Ŵ_{m̂_F} ← W_{m_F}
            end if
        end for
    end for
end for
```

(In Step †, the notations follow those of (28) or (31). Here, the hat notation denotes the output of the algorithm.)

IV. Proposed Codebook Design Technique

Throughout the manuscript, we assumed that the beam pattern is approximated by constant directivity gain C_m within the main-lobe beamwidth ψ_m. To design accurate approximations to the ideal beam patterns, much of the prior work considers analog beamforming [6]–[8] and hybrid beamforming techniques [10]–[12] where the codebook is assumed to be designed off-line constructed using a limited number of RF chains.

However, when a user or scatterers are around the boundary of two adjacent (pre-designed) beamforming patterns, it might be detected in both angular partitions. Then, shifting the main-lobe beam direction to the target angle or redesigning the codebook is needed to provide an improved beamforming gain. Considering many prior works on DFT-based codebooks in [7], [34]–[36], we present an efficient method to design a multi-resolution codebook using the DFT vectors.

A. DFT-based Multilevel Codebook Design

Considering the quantized spatial angles $\{\theta_n = -1 + (2n-1)/N_t : 1 \le n \le N_t\}$, the m-th level codebook $\mathcal{F}_m$ can be designed by combining the array steering vectors corresponding to the subset of quantized angles

given by

$$\mathcal{F}_m = \left\{ \mathbf{f}_1^{(m)}, \mathbf{f}_2^{(m)}, \ldots, \mathbf{f}_{N/N_m}^{(m)} \right\} \tag{36}$$

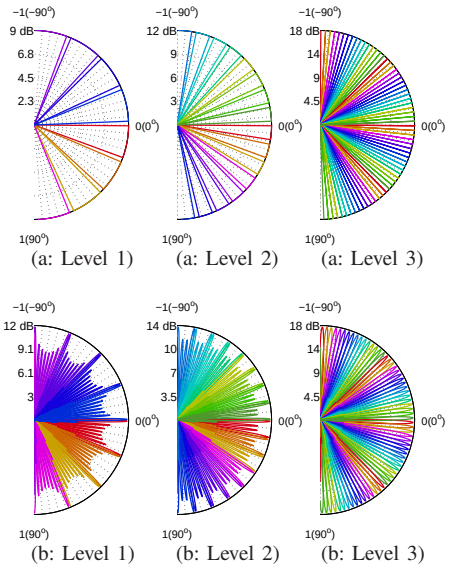
$$= \left\{ \frac{1}{\sqrt{N_m}} \sum_{p=1}^{N_m} \mathbf{u}_t(\theta_p) e^{j\omega_m p}, \frac{1}{\sqrt{N_m}} \sum_{p=N_m+1}^{2N_m} \mathbf{u}_t(\theta_p) e^{j\omega_m p}, \right.$$
$$\left. \ldots, \frac{1}{\sqrt{N_m}} \sum_{p=N_t-N_m+1}^{N_t} \mathbf{u}_t(\theta_p) e^{j\omega_m p} \right\}, \tag{37}$$

where we introduce a new variable ω_m to enable the beam patterns of $\mathbf{f}_v^{(m)}$ to have a little to no variation in main-lobe directivity gain as best as possible. Here, we assume that N_m is a divisor of N_t for simplicity. The array steering vector on the quantized angles $\{\theta_n\}$ in (37) becomes a *generalized* DFT vector with offset $-(1+N_t)/2$, i.e.,

$$\mathbf{u}_t(\theta_n) = \frac{1}{\sqrt{N_t}} \left[1, e^{-j\frac{2\pi}{N_t}\left(n-\frac{N_t+1}{2}\right)}, e^{-j\frac{2\pi}{N_t}2\left(n-\frac{N_t+1}{2}\right)}, \right.$$
$$\left. \ldots, e^{-j\frac{2\pi}{N_t}(N_t-1)\left(n-\frac{N_t+1}{2}\right)} \right]^T, \tag{38}$$

which is the n-th column of the Butler matrix [37] and can be easily implemented using millimeter microstrip technology [38]–[40]. We denote the number of combined vectors at level m by N_m and the number of active RF chains by N_{RF}, respectively. From (37), we need at *least* $N_{RF} = N_m$ RF chains for combining the vectors.

The codewords in (37) form *ideal* beam patterns over the quantized angles such that the codebook with arbitrary $\omega_m \in \mathbb{R}$ achieves the (constant) maximum beamforming gain N_t/N_m on the quantized angles $\{\theta_n\}$. For $N_t = 64$, $M = 3$, and $\{N_1, N_2, N_3\} = \{8, 4, 1\}$, Fig. 5(a) shows the corresponding beam patterns over the quantized angles, which is useful for theoretical analysis and can be used as a benchmark showing what can be achieved under idealized assumptions. For realizing this scenario in practice, large-scale antenna arrays can be applied in the transmitter and/or the receiver required for finer spatial sampling of continuous angles.



(a: Level 1) (a: Level 2) (a: Level 3)

(b: Level 1) (b: Level 2) (b: Level 3)

Fig. 5: Phase-shifted DFT beam patterns where $N_t = 64$, $M = 3$, $[\text{card}(\mathcal{F}_1), \text{card}(\mathcal{F}_3), \text{card}(\mathcal{F}_3)] = [8, 16, 64]$, and $[N_1, N_2, N_3] = [8, 4, 1]$: (a) ideal case on the quantized spatial domain with an arbitrary $\omega_m \in \mathbb{R}$ and (b) non-ideal case on the continuous spatial domain with $\omega_m = 0$ for $1 \le m \le M$.

Note that the codebook $\mathcal{F}_m$ in (37) is parameterized by the phase shift ω_m, and we refer to this scheme as a *phase-shifted* DFT method. Optimizing the single parameter ω_m is important to achieve constant main-lobe directivity gain as possible. An example of the codebook for $\omega_m = 0$ (i.e., a simple addition of the array steering vectors as used in [24]) is shown in Fig. 5(b). Especially at lower level codewords, there are a large number of null points within the main beamwidth and picking points

$$\mathbf{u}_t(\theta)^H \mathbf{f}_i^{(m)} = \mathbf{u}_t(\theta)^H \left(\frac{1}{\sqrt{N_m}} \sum_{p=(i-1)N_m+1}^{iN_m} \mathbf{u}_t(\theta_p) \right) \tag{32}$$

$$= \frac{1}{N_t\sqrt{N_m}} \sum_{p=(i-1)N_m+1}^{iN_m} e^{j\omega_m p} e^{-j\pi\frac{N-1}{N}\left(p-\frac{N_t(\theta+1)+1}{2}\right)} \frac{\sin\left(\pi\left(p-\frac{N_t(\theta+1)+1}{2}\right)\right)}{\sin\left(\frac{\pi}{N_t}\left(p-\frac{N_t(\theta+1)+1}{2}\right)\right)} \tag{33}$$

$$f_{\theta,i,\omega_m}^{(m)} = \left| \mathbf{u}_t(\theta)^H \mathbf{f}_i^{(m)} \right|^2 \tag{34}$$

$$= \frac{1}{N_t^2 N_m} \left\{ \sum_{p=(i-1)N_m+1}^{iN_m} \frac{\sin^2\left(\pi(p-\tilde{\theta})\right)}{\sin^2\left(\frac{\pi}{N_t}(p-\tilde{\theta})\right)} \right.$$
$$\left. + \sum_{p=(i-1)N_m+1}^{iN_m} \sum_{p'=p+1}^{iN_m} 2\cos\left((p'-p)\left(\frac{N-1}{N}\pi-\omega_m\right)\right) \times \frac{\sin\left(\pi(p-\tilde{\theta})\right)\sin\left(\pi(p'-\tilde{\theta})\right)}{\sin\left(\frac{\pi}{N_t}(p-\tilde{\theta})\right)\sin\left(\frac{\pi}{N_t}(p'-\tilde{\theta})\right)} \right\}, \tag{35}$$

where $\tilde{\theta} := \frac{N_t(\theta+1)+1}{2}$.

near the boundary of beam patterns due to the overlapped DFT beam patterns on the continuous angles.

For the i-th codeword $\mathbf{f}_i^{(m)} \in \mathcal{F}_m$ in (36), we evaluate a closed-form expression for the array response over the continuous angle θ in (33) where $\mathbf{u}_t(\theta)^H \mathbf{u}_t(\theta_p)$ represents the discrete Fourier transform using the p-th vector in (38) evaluated at the spatial location parameter θ. Given this, the array response magnitude can be described by a function $f_{\theta,i,\omega_m}^{(m)}$ defined in (34). As a result, we focus on an exhaustive search for the proper phase shift ω_m that minimizes the variation of the main-lobe directivity gain evaluated with $\mathbf{f}_i^{(m)}$ over the quantized version of main beam range given by

$$\left[-1+\frac{2((i-1)N_m+1)-1}{N_t}, -1+\frac{2(iN_m)-1}{N_t}\right]. \tag{39}$$

In this case, it is important to reduce the computational complexity of the exhaustive search, and thus we derive a property of the objective function $f_{\theta,i,\omega_m}^{(m)}$.

Theorem 1: Given the level m and $1 \le i \le N/N_m$, the function $f_{\theta,i,\omega_m}^{(m)}$ in (35) satisfies the following properties:
(P.1) for ω_m, 2π-periodic and an even function at $\pi(N_t-1)/N_t$,
(P.2) for θ, an even function at $-1+(2i-1)N_m/N_t$,
(P.3) independent of the index i where $1 \le i \le N/N_m$.

Proof: See Appendix B.

Remark 3: Theorem 1 yields a much simpler implementation of the exhaustive search without loss in performance because: the search space for finding the phase shift ω_m is reduced to the interval $[-\pi/N_t, \pi(1-1/N_t)]$ of length π by *(P.1)*; the continuous angular interval for θ can be set to one-half of the main beam interval of (39) by *(P.2)*, e.g., $[-1+(2i-1)N_m/N_t, -1+(2iN_m-1)/N_t]$; the optimized phase shift ω_m for the i-th codeword $\mathbf{f}_i^{(m)}$ is also an optimal design for all other codewords in $\mathcal{F}_m$ by *(P.3)*.

Therefore, the problem of the codebook design at level m in (37) is formulated as

$$\min_{\omega_m \in \left[-\frac{\pi}{N_t}, \pi\left(1-\frac{1}{N_t}\right)\right]} \operatorname{var}\left(f_{\boldsymbol{\theta}_b,i,\omega_m}^{(m)}\right), \tag{40}$$

for any fixed index $i \in \{1,\ldots,N/N_m\}$. Here, $\boldsymbol{\theta}_b \in \mathbb{R}^{2^b}$ denotes the uniformly quantized points of the half of the main beam interval $[-1+(2i-1)N_m/N_t, -1+(2iN_m-1)/N_t]$ using b bits where we set $b=9$. Since the objective function of (35) can operate on the vector input $\boldsymbol{\theta}_b$, the design parameter ω_m is swept over the possible range to minimize the sample variance in (40).

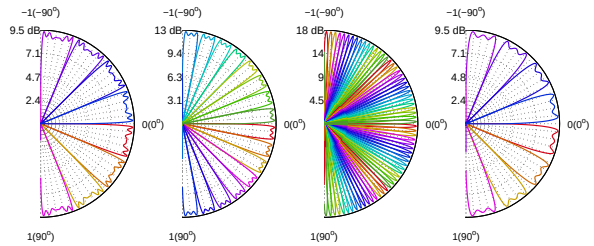
Remark 4: Since each codeword $\mathbf{f}_i^{(m)} \in \mathcal{F}_m$ should provide the maximal constant level of concentration of energy over the spatial interval $[\theta_{(i-1)N_m+1}, \theta_{iN_m}]$, one can relate the codebook design problem in (40) to the design of bandpass filters for estimating the power spectral density (PSD) over $(-\pi,\pi)$. The finite impulse response (FIR) bandpass filter that concentrates its energy to the interval $[\pi\theta_{(i-1)N_m+1}, \pi\theta_{iN_m}]$ can be designed by using N_m dominant eigenvectors of $\boldsymbol{\Gamma} \in \mathbb{C}^{N_t \times N_t}$, referred to as Slepian sequences [41], given by

$$\boldsymbol{\Gamma} = \frac{1}{2\pi}\int_{\pi\theta_{(i-1)N_m+1}}^{\pi\theta_{iN_m}} \mathbf{u}_t(\vartheta)\mathbf{u}_t(\vartheta)^H d\vartheta, \tag{41}$$

where $\mathbf{u}_t(\cdot)$ denotes the steering vector defined in (3) and (p,q) element of $\boldsymbol{\Gamma}$ in (41) is given by

$$[\boldsymbol{\Gamma}]_{p,q} = e^{-j\pi(p-q)\left(-1+\frac{N_m(2i-1)}{N_t}\right)} \frac{\sin\left(\frac{\pi(p-q)(N_m-1)}{N_t}\right)}{\pi(p-q)}.$$

Since the power of the filtered signal is approximately equal to the PSD value at θ, the auto-PSD estimation over the interval is given by $\frac{1}{N_m}\sum_{p=1}^{N_m}\left|\mathbf{u}_t(\theta)^H\mathbf{v}_p\right|^2$ where $\mathbf{v}_p \in \mathbb{C}^{N_t}$ is the p-th dominant eigenvector of $\boldsymbol{\Gamma}$, referred to as the p-th dominant Slepian sequence. However, simulations of a codebook design using the Slepian sequences as the basis vectors and optimizing over a single phase-shift design parameter, ω_m, similar to the codebook design for generalized DFT (Butler) beams, yielded distorted beamforming patterns. Thus, multi-dimensional optimization is needed to linearly combine Slepian sequences to achieve a maximal constant level of energy concentration over the sector of interest as best as possible. However, with respect to practical implementation, our proposed codebook using DFT beams according to (37) can be implemented via the well-known Butler Matrix Beamformer [37] in analog RF circuitry which, as discussed previously, can be implemented using millimeter microstrip technology [38]–[40]. In contrast, the values of the Slepian sequences are not constant magnitude and also their phase values are arbitrary, i.e. are not confined to a finite alphabet.



(a: Level 1) (a: Level 2) (a: Level 3) (a′: Level 1)

Fig. 6: Phase-shifted DFT beam patterns on the continuous spatial domain where: (a) $N_t = 64$, $M = 3$, $\{N_1, N_2, N_3\} = \{8, 4, 1\}$, and $[\omega_1, \omega_2] = [1.93, 2.24]$ with any $\omega_3 \in \mathbb{R}$, and (a′) $N_t = 32$, $N_1 = 4$, and $\omega_1 = 2.2$.

The proposed design obtained from (40) yields good performance over the continuous angles as shown in Fig. 6 for $\{N_1, N_2, N_3\} = \{8, 4, 1\}$, $M = 3$, and $N_t = 64$. The proposed codebook has constant main-

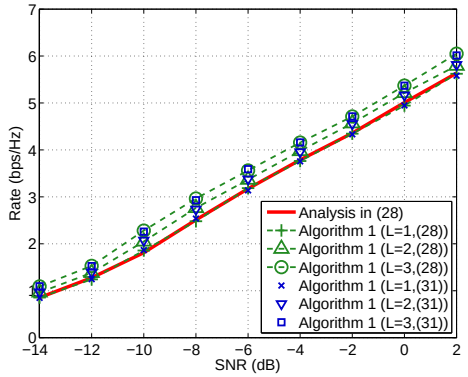


Fig. 7: Data rate versus SNR for different L where $R_{m_F}^{m_L}$ in (28) and $\underline{R}_{m_F}^{m_L}$ in (31) are used in Algorithm 1 for the same setup in Fig. 6(a).

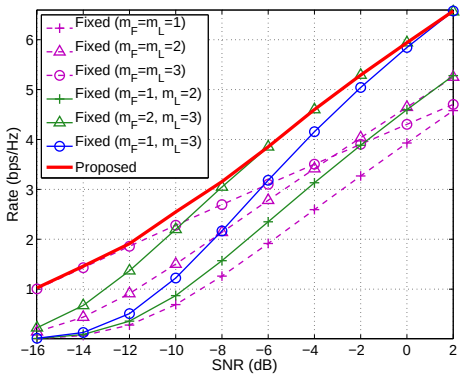


Fig. 8: Performance comparison with the methods that use the fixed levels (i.e., $W_{m_F} = 1$ and $1 \le m_F \le m_L \le M$) where $N_t = 64$, $M = 3$, $\{N_1, N_2, N_3\} = \{8, 4, 1\}$, and $\text{AoD}_1 = (-30°, 30°)$.

lobe directivity gain per beam similar to the ideal beam patterns in Fig. 5(a). For the codebook design in Fig. 6(a), we use at *most* $N_{RF} = 8$ RF chains at the first level codebook because of $N_1 = 8$. To accommodate the case that we cover the same angular interval with the small number of RF chains, we can reduce the number of active antenna elements by turning off a subset of successive antenna elements while having a coarse spatial resolution [7]. In comparison to the beam patterns in Fig. 6(a: Level 1), Fig. 6(a′) shows the corresponding beam patterns obtained with $N_{RF} = N_1 = 4$ and $N_t = 32$. Although the estimation accuracy of the angular interval is restricted by the reduced number of antenna elements, the proposed beam patterns still yield reasonably good performance, which will be further demonstrated in Section V.

V. Numerical Results

In this section, we provide some numerical results to evaluate the proposed beam sequence design and the corresponding algorithm. Throughout the simulation, we fixed the noise power $\sigma_z^2 = 1$ and $K = 40$. The sector angle at the transmitter is given by $\text{AoD}_1 \in \{(-0.5, 0.5), (-0.77, 0.77)\}$ (i.e., $\{(-30°, 30°), (-50°, 50°)\}$) and the angular interval at the receiver is set to the entire range of quantized angles. The performance was measured by averaging the instantaneous data rate in (26) over 10^6 Monte Carlo runs, where the false alarm probability for each AoD direction is set to $P_{FA} = 0.01$ and the SNR is defined as $\sigma_g^2 P$ to account for the propagation path gain and the transmit power. Note that the SNR values in the abscissa of each plot does not include the beamforming gain, which is substantial in a Massive MIMO setting. Thus, in examining the results, it is important to look at SNRs significantly below 0 dB in consideration of path loss attenuation.

We consider a MIMO system with $N_t = 64$, $N_r = 4$, $M = 3$, and $\text{AoD}_1 = (-30°, 30°)$ where we adopt Laplacian distributed angles of channel paths with the mean angle $\mu_\theta = 0$ and the standard deviation $\sigma_\theta = 7$ [42]. In Fig. 7, we examine the data rate performance based on the proposed method along with the codebook shown in Fig. 6(a) on the quantized angles. It is seen that the proposed algorithms using $R_{m_F}^{m_L}$ in (28) and the lower bound $\underline{R}_{m_F}^{m_L}$ in (31) yield almost the same performance. Thus, the simpler algorithm with (31) can be used without much performance loss. For $L = 1$, simulation result without *(A.1)* matches well the analytic result, i.e., the result in (28) gives expressions for the average data rate for a single path channel $L = 1$. For $L > 1$, simulation results are obtained by choosing a single codeword that achieves the largest received signal power.

For purpose of comparison, we consider the case that the transmitter fixes the codebook levels $\{\mathcal{F}_m : m_F \le m \le m_L\}$ for $1 \le m_F \le m_L \le M$ with $W_{m_F} = 1$. With the same setup used in Fig 7, we compared the data rate performance of the proposed method to those of others using the analytic result of (28) in Fig. 8. It appears that the proposed method using adaptively selected codebooks outperforms other methods. The optimized beamforming sequence obtained by Algorithm 1 is summarized in Table I. At low SNR, the proposed method retransmits the codewords sounded at the first training stage by two times aimed at improving the effective SNR. As the operating SNR increases, the proposed method uses all codebook levels with a single codeword transmission to decrease the training overhead. We observed that the use of a multilevel codebook (i.e., $m_F < m_L$) yields good performance compared to the use of a single level codebook (i.e., $m_F = m_L$) for most of the SNR range.

Next, we compare the performance of the proposed methods to those of several codebook design techniques [8], [11]. For $M = 3$ and $N_t = 64$, we considered the channel model with a number of path $L = 1$ and the

SNR(dB)	M	$\{m_F, \ldots, m_L\}$	$\{W_{m_F}, \ldots, W_{m_L}\}$
-16 $\sim$ -12	2	$\{2, 3\}$	$\{2, 1\}$
-10 $\sim$ -2	2	$\{2, 3\}$	$\{1, 1\}$
0 $\sim$ 2	3	$\{1, 2, 3\}$	$\{1, 1, 1\}$

TABLE I: Optimized beam alignment sequence

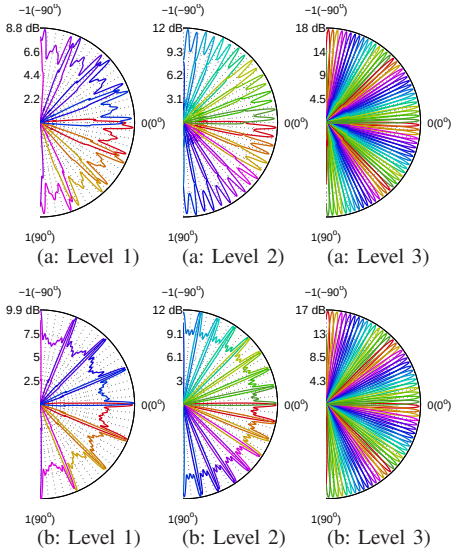


Fig. 9: Beam patterns on the continuous spatial domain where $N_t = 64$ and $M = 3$: (a) analog beamforming ($N_{RF} = 1$) [8] and (b) hybrid beamforming ($N_{RF} = 4$) [11]. For each codebook level, there are eight, sixteen, and sixty-four codewords, respectively.

AoDs are assumed to take continuous values, i.e., not quantized and uniformly distributed in AoD_1 given by $\text{AoD}_1 = (-50°, 50°)$. Fig. 9 shows the beam patterns obtained by analog beamforming [8] (i.e., $N_{RF} = 1$) and hybrid beamforming [11] where we set $N_{RF} = 4$ and adopt enough phase quantization bits so that the phase quantization error on the performance is negligible. These should be compared to that obatined with our proposed method in Fig. 6. Given $\{N_1, N_2, N_3\} = \{8, 4, 1\}$, the proposed codebook design (i.e., the phase-shifted DFT method shown in Fig. 6(a)) needs at *most* $N_{RF} = 8$ RF chains because of $N_1 = 8$. For a fair comparison in terms of N_{RF}, we also considered a modified method where we replace the codebook in Fig. 6(a: Level 1) by the codebook in Fig. 6(a′) in order to maintain $N_{RF} = 4$.

To evaluate the performance of the phase-shifted DFT method without beam sequence optimization, we first fix the number of codewords levels as $M = 3$ where $m_F = 1$, $m_L = 3$, and $W_{m_F} = 1$, marked (a) in Fig. 10. Hybrid beamforming design achieves better performance than analog beamforming design due to the well-designed codebooks especially at level 1. For the fixed levels, the phase-shifted DFT method is superior to the other methods over the whole SNR range relative to data rate performance, even when we keep the same number of RF chains $N_{RF} = 4$. Using more RF chains

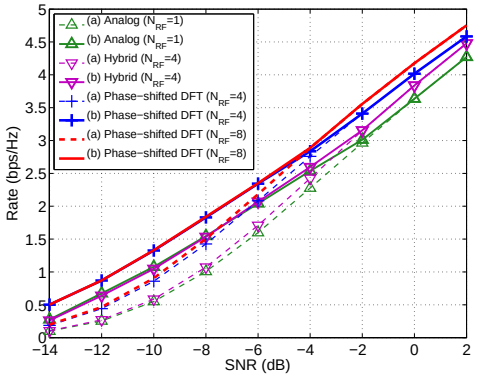


Fig. 10: Performance comparison with the previous designs where $N_t = 64$, $M = 3$, $L = 1$, and $\text{AoD}_1 = (-50°, 50°)$: (a) fixed beamforming sequence (i.e., $m_F = 1$, $m_L = M$, and $W_{m_F} = 1$) and (b) optimized beamforming sequence by Algorithm 1.

$N_{RF} = 8$ in the phase-shifted DFT method yields slightly better performance than using $N_{RF} = 4$ due to the different codebooks at level 1 shown in Fig. 6(a, a′).

By averaging the main-lobe directivity gain of the multilevel codebook over the interval AoD_1, we investigate the performance of all the methods including analog, hybrid, and the proposed methods in conjunction with Algorithm 1. The results marked (b) in Fig. 10 show that the data rate performance is further enhanced especially at moderate-to-low SNR due to beam sequence optimization. The methods for analog and hybrid beamforming yield similar performance up to the SNR around -4dB since only the codebooks at levels (2, 3) are used at low SNR and its beam patterns are comparable to each other shown in Fig. 9. Similarly, the phase-shifted DFT methods that use $N_{RF} \in \{4, 8\}$ also show the same performance up to mid-range SNR.

VI. Conclusions

We considered multi-resolution beam alignment sequence design to improve the data rate performance achievable in mmWave systems. The resulting multilevel beamformer alignment design generates a beamformer sequence that counterbalances minimizing training overhead and maximizing beamforming gain. We then provided a phase-shifted DFT method that efficiently designs the multi-resolution codebook. The proposed codebook design and beam alignment sequence design provide an efficient means to improve the data rate performance in large-scale mmWave MIMO systems.

Appendix

A. Proof of Lemma 1

Define random variables X and Y by $X \sim \mathcal{CN}(0, \sigma^2)$ and $Z \sim \mathcal{CN}(0, \tilde{\sigma}_z^2)$ where $\sigma^2 := G_m P \sigma_g^2$ and $\tilde{\sigma}_z^2 :=$

σ_z^2/W_m. We will omit the codebook index (m) for notational simplicity. The average power of X is given by

$$\mathbb{E}\{|X|^2\} = \mathbb{E}\{|X|^2|A_0\}p(A_0) + \mathbb{E}\{|X|^2|A_1\}p(A_1) \tag{42}$$

$$= p(A_1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x|^2 \cdot p_{XZ}(x,z|A_1)\,dxdz, \tag{43}$$

where A_0 and A_1 denotes mutually exclusive events described in terms of the random variables X and Z, i.e., $A_0 = \{|X+Z|^2 < \gamma\}$ and $A_1 = \{|X+Z|^2 \geq \gamma\}$. $p_{XZ}(x,z|A_1)$ denotes a conditional joint probability distribution function (pdf) given by [43]

$$p_{XZ}(x,z|A_1) = \frac{p_{XZ}(x,z)}{p(A_1)}, \tag{44}$$

for $(X,Z) \in A_1$, otherwise $p_{XZ}(x,z|A_1) = 0$, where $p(A_1) = \iint_{(X,Z)\in A_1} p_{XZ}(x,z)dxdz$. (When there is no ambiguity, we omit the subscript X from $p_X(x)$ and write $p(x)$.) Note that $\mathbb{E}\{|X|^2|A_0\} = 0$ in (42) because we set $|X|^2 = 0$ for $(X,Z) \in A_0$ from a consideration of no transmission.

For computational purposes, we separate the complex random variables X and Z into real part and imaginary part as

$$X = X_{\mathcal{R}} + jX_{\mathcal{I}} \quad \text{and} \quad Z = Z_{\mathcal{R}} + jZ_{\mathcal{I}},$$

and change the random variable from X to $R = \sqrt{X_{\mathcal{R}}^2 + X_{\mathcal{I}}^2} \in (0,\infty)$ and $\Theta = \tan^{-1}(X_{\mathcal{I}}/X_{\mathcal{R}})$ where $p_R(r) = \frac{r}{\sigma^2/2}e^{-r^2/\sigma^2}$ and $p_\Theta(\theta) = 1/(2\pi)$. Then, the event of A_1 is rewritten as

$$\begin{aligned}&\{|X+Z|^2 \geq \gamma\}\\ &\Leftrightarrow \{R^2 + 2r(Z_{\mathcal{R}}\cos\Theta + Z_{\mathcal{I}}\sin\Theta) + (Z_{\mathcal{R}}^2 + Z_{\mathcal{I}}^2) \geq \gamma\}\\ &\Leftrightarrow \{(R+\tilde{Z}_{\mathcal{R}})^2 \geq \gamma - \tilde{Z}_{\mathcal{I}}^2\} =: \tilde{A}_1,\end{aligned} \tag{45}$$

where $N_{\mathcal{R}} := Z_{\mathcal{R}}\cos\Theta + Z_{\mathcal{I}}\sin\Theta$ and $N_{\mathcal{I}} := -Z_{\mathcal{R}}\sin\Theta + Z_{\mathcal{I}}\cos\Theta$. When $(X_{\mathcal{R}}, X_{\mathcal{I}})$ and $(Z_{\mathcal{R}}, Z_{\mathcal{I}})$ are initially aligned with the real and imaginary axes, $\tilde{Z}_{\mathcal{R}}$ and $\tilde{Z}_{\mathcal{I}}$ denote the noise components in the direction of $Re^{j\Theta}$ and perpendicular to those of $Re^{j\Theta}$, respectively. Note that $\tilde{Z}_{\mathcal{R}}$ and $\tilde{Z}_{\mathcal{I}}$ are two independent Gaussian variables, i.e., $\tilde{Z}_{\mathcal{R}} \sim \mathcal{N}(0, \tilde{\sigma}_z^2/2)$ and $\tilde{Z}_{\mathcal{I}} \sim \mathcal{N}(0, \tilde{\sigma}_z^2/2)$ due to circular symmetry assumed for noise modeling. Therefore, (43) can be rewritten in terms of R, $\tilde{Z}_{\mathcal{R}}$, and $\tilde{Z}_{\mathcal{I}}$ given by

$$\mathbb{E}\{|X|^2\} = p(\tilde{A}_1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} r^2 \times p\big(r,\tilde{z}_{\mathcal{R}},\tilde{z}_{\mathcal{I}}|\tilde{A}_1\big)drd\tilde{z}_{\mathcal{R}}d\tilde{z}_{\mathcal{I}} \tag{46}$$

$$= \iiint_{(r,\tilde{z}_{\mathcal{R}},\tilde{z}_{\mathcal{I}})\in\tilde{A}_1} r^2 \cdot p(r,\tilde{z}_{\mathcal{R}},\tilde{z}_{\mathcal{I}})\,drd\tilde{z}_{\mathcal{R}}d\tilde{z}_{\mathcal{I}}, \tag{47}$$

where $p\big(r,\tilde{z}_{\mathcal{R}},\tilde{z}_{\mathcal{I}}|\tilde{A}_1\big)$ is defined similar to (44) and the nuisance variable Θ is integrated out in the analysis. Since the variables R, $\tilde{Z}_{\mathcal{R}}$, and $\tilde{Z}_{\mathcal{I}}$ are mutually independent, the joint pdf for R, $\tilde{Z}_{\mathcal{R}}$, and $\tilde{Z}_{\mathcal{I}}$ can be factored as $p(r,\tilde{z}_{\mathcal{R}},\tilde{z}_{\mathcal{I}}) = p(r)p(\tilde{z}_{\mathcal{R}})p(\tilde{z}_{\mathcal{I}})$. Then, (47) can be rewritten as

$$\begin{aligned}\mathbb{E}\{|X|^2\} = &\int_{-\infty}^{\infty} p(\tilde{z}_{\mathcal{I}})\int_{-\infty}^{\infty} p(\tilde{z}_{\mathcal{R}})\times\\ &\int_{(r+\tilde{z}_{\mathcal{R}})^2>\gamma-\tilde{z}_{\mathcal{I}}^2} r^2\cdot p(r)drd\tilde{z}_{\mathcal{R}}d\tilde{z}_{\mathcal{I}}\\ = &\int_{\tilde{z}_{\mathcal{I}}^2>\gamma} p(\tilde{z}_{\mathcal{I}})\int_{-\infty}^{\infty} p(\tilde{z}_{\mathcal{R}})\times\\ &\int_{0}^{\infty} r^2\cdot p(r)drd\tilde{z}_{\mathcal{R}}d\tilde{z}_{\mathcal{I}}\\ &+\int_{\tilde{z}_{\mathcal{I}}^2<\gamma} p(\tilde{z}_{\mathcal{I}})\int_{-\infty}^{\infty} p(\tilde{z}_{\mathcal{R}})\times\\ &\left(\int_0^{I_1} r^2p(r)dr + \int_{I_2}^{\infty} r^2p(r)dr\right)d\tilde{z}_{\mathcal{R}}d\tilde{z}_{\mathcal{I}},\end{aligned} \tag{48}$$

where $I_1 := \max\{0, -\tilde{z}_{\mathcal{R}} - \sqrt{\gamma - \tilde{z}_{\mathcal{I}}^2}\}$ and $I_2 := \max\{0, -\tilde{z}_{\mathcal{R}} + \sqrt{\gamma - \tilde{z}_{\mathcal{I}}^2}\}$.

In (48), we consider two separated cases for $\tilde{z}_{\mathcal{I}}$: if $\tilde{z}_{\mathcal{I}}^2 > \gamma$, all values of $\tilde{z}_{\mathcal{R}}$ and r satisfy A_1 (i.e., the first term on the right-hand side of (48)), and if $\tilde{z}_{\mathcal{I}}^2 < \gamma$, there is a range of r for each $\tilde{z}_{\mathcal{R}}$ as to satisfy A_1 (i.e., the second term on the right-hand side of (48)). The first term of (48) can be evaluated as $2\sigma^2\left(1-\Phi\left(\sqrt{\gamma/(\tilde{\sigma}_z^2/2)}\right)\right)$. If τ denotes the second term of (48), we provide a complete form of τ in (49). After some tedious integration, we obtain the analytical formula for the third term of (49), denoted as τ', shown in (50).

Note that equation (48) depends on σ^2, $\tilde{\sigma}_z^2$, and γ as seen from equations (49) and (50). The result may look complicated, so we apply an approximation in high SNR regime (i.e., $\sigma^2 \gg \tilde{\sigma}_z^2$) to derive a simple formula of σ^2, $\tilde{\sigma}_z^2$, and γ, where we use $\sigma^2 = G_m P\sigma_g^2$ and

$$\begin{aligned}\tau &= 2\sigma^2 \int_{-\sqrt{\gamma}}^{\sqrt{\gamma}} p(\tilde{z}_{\mathcal{I}}) \cdot \left(1 - \Phi\left(\sqrt{\frac{\gamma - \tilde{z}_{\mathcal{I}}^2}{\tilde{\sigma}_z^2/2}}\right)\right) d\tilde{z}_{\mathcal{I}} \\ &\quad + \int_{-\sqrt{\gamma}}^{\sqrt{\gamma}} p(\tilde{z}_{\mathcal{I}}) \left[- \int_{-\infty}^{-\sqrt{\gamma-\tilde{z}_{\mathcal{I}}^2}} p(\tilde{z}_{\mathcal{R}}) \cdot e^{-\left(-\tilde{z}_{\mathcal{R}} - \sqrt{\gamma-\tilde{z}_{\mathcal{I}}^2}\right)^2} \left(\sigma^2 + \left(-\tilde{z}_{\mathcal{R}} - \sqrt{\gamma - \tilde{z}_{\mathcal{I}}^2}\right)^2\right) d\tilde{z}_{\mathcal{R}} \right. \\ &\quad \left. + \int_{-\infty}^{\sqrt{\gamma-\tilde{z}_{\mathcal{I}}^2}} p(\tilde{z}_{\mathcal{R}}) \cdot e^{-\left(-\tilde{z}_{\mathcal{R}} + \sqrt{\gamma-\tilde{z}_{\mathcal{I}}^2}\right)^2} \left(\sigma^2 + \left(-\tilde{z}_{\mathcal{R}} + \sqrt{\gamma - \tilde{z}_{\mathcal{I}}^2}\right)^2\right) d\tilde{z}_{\mathcal{R}} \right] d\tilde{z}_{\mathcal{I}} \\ &= 2\sigma^2 \left(-1 + \Phi\left(\sqrt{\frac{\gamma}{\tilde{\sigma}_z^2/2}}\right) + \frac{1}{2} e^{-\frac{\gamma}{2\tilde{\sigma}_z^2}}\right) + \frac{e^{-\gamma/\tilde{\sigma}_z^2}}{2\left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^2} \\ &\quad + \frac{1}{\tilde{\sigma}_z \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tilde{\sigma}_z^2}}} \int_{-\sqrt{\gamma}}^{\sqrt{\gamma}} P(\tilde{z}_{\mathcal{I}}) \cdot e^{-\frac{\gamma - \tilde{z}_{\mathcal{I}}^2}{\sigma^2 + \tilde{\sigma}_z^2}} \left(\frac{\gamma - \tilde{z}_{\mathcal{I}}^2}{\left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^2} + \sigma^2 + \frac{1}{2\left(\frac{1}{\sigma^2} + \frac{1}{\tilde{\sigma}_z^2}\right)} \right) \cdot \operatorname{erf}\left(\sqrt{\frac{\gamma - \tilde{z}_{\mathcal{I}}^2}{\tilde{\sigma}_z^2 \left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)}} \right) d\tilde{z}_{\mathcal{I}}, \end{aligned} \tag{49}$$

where $\operatorname{erf}(\cdot)$ denotes the error function [32, Eq. 8.250.1].

$$\begin{aligned}\tau' = \frac{1}{\tilde{\sigma}_z \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tilde{\sigma}_z^2}}} \Bigg[& \left(\gamma \left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^{-2} + \sigma^2 + \frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tilde{\sigma}_z^2}\right)^{-1} \right) e^{-\frac{\gamma}{\sigma^2 + \tilde{\sigma}_z^2}} \sqrt{1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}} \left(1 - e^{-\frac{\gamma}{2\tilde{\sigma}_z^2\left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)}}\right) \\ & - \left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^{-2} \left(\frac{\tilde{\sigma}_z^2}{2}\left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^{\frac{3}{2}} e^{-\frac{\gamma}{\sigma^2 + \tilde{\sigma}_z^2}} - \frac{\tilde{\sigma}_z^2}{2}\left(1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}\right)^{\frac{3}{2}} e^{-\frac{\gamma}{2}\left(\frac{1}{\tilde{\sigma}_z^2} + \frac{1}{\sigma^2 + \tilde{\sigma}_z^2}\right)} - \gamma \sqrt{1 + \frac{\tilde{\sigma}_z^2}{\sigma^2}} e^{-\frac{\gamma}{\tilde{\sigma}_z^2}} \right) \Bigg]. \end{aligned} \tag{50}$$

$\tilde{\sigma}_z^2 = \sigma_z^2/W_m$. In this case, (48) is given by

$$\begin{aligned}\mathbb{E}\{|X|^2\} &= e^{-\frac{\gamma}{\sigma^2}}\left(\sigma^2 + \gamma\right) + \frac{1}{2} e^{-\frac{\gamma}{\sigma_n^2}} \\ &\quad - \gamma\left(e^{-\frac{\gamma}{2\sigma_n^2}} - e^{-\frac{\gamma}{\sigma_n^2}}\right) \\ &= e^{-\frac{\gamma}{G_m P \sigma_g^2}}\left(G_m P \sigma_g^2 + \gamma\right) + \frac{P_{FA}}{2} \\ &\quad - \frac{\sigma_z^2 \ln(1/P_{FA})}{W_m}\left(\sqrt{P_{FA}} - P_{FA}\right), \end{aligned} \tag{51}$$

where $\gamma = \sigma_z^2 \ln(1/P_{FA})/W_m$ and $\int_\gamma^\infty r^2 \cdot p(r) dr = e^{-\frac{\gamma}{G_m P \sigma_g^2}}\left(G_m P \sigma_g^2 + \gamma\right)$. ■

B. Proof of Theorem 1

(P.1) Since only the cosine function in the second term in (35) depends on the phase shift ω_n, $f_{\theta,i,\omega_m}^{(m)}$ becomes periodic with the period 2π. Define $c^+ = \pi(N-1)/N + \epsilon$ and $c^- = \pi(N-1)/N - \epsilon$ for $\epsilon > 0$. Then, we show that the values of cosine function evaluated at c^+ and c^- are the same, i.e., $\cos\left((p'-p)\left(\pi(N-1)/N - c^+\right)\right) = \cos\left((p'-p)\epsilon\right) = \cos\left((p'-p)\left(\pi(N-1)/N - c^-\right)\right)$.

(P.2) For the i-th codeword $\mathbf{f}_i^{(m)} \in \mathcal{F}_m$, recall that the summation index p in (35) is given by $p = (i-1)N_m + q$ for $1 \le q \le N_m$. From the center of the main beam range in (39) (i.e., $-1 + (2i-1)N_m/N_t$), we set the continuous angle as

$$\theta^+ = -1 + (2i-1)N_m/N_t + \epsilon \tag{52}$$

for $\epsilon > 0$. In (33), we consider the argument of the sin and exponential functions that correspond to θ given by

$$p - \frac{N_t(\theta^+ + 1) + 1}{2} = q - \frac{1 + N_m + \epsilon N_t}{2}, \tag{53}$$

where $1 \le q \le N_m$. Then, we define $p' := N_m(2i-1) + 1 - p = -q + iN_m + 1 \in \{(i-1)N_m + 1, \ldots, iN_m\}$ that gives a one-to-one correspondence between p and p'. As in (53), we evaluate the same term for the angle $\theta^- = -1 + (2i-1)N_m/N_t - \epsilon$ and the index p' as

$$\begin{aligned} p' - \frac{N_t(\theta^- + 1) + 1}{2} &= -\left(q - \frac{1 + N_m + \epsilon N_t}{2}\right) \\ &= -\left(p - \frac{N_t(\theta^+ + 1) + 1}{2}\right). \end{aligned} \tag{54}$$

Note from (54) that for each term in (53) derived at θ^+, there is the same term with a negative sign derived at

θ^-. Then, we have

$$\mathbf{u}_t(\theta^-)^H \mathbf{f}_i^{(m)} = e^{j\omega_m(N_m(2i-1)+1)} \left(\mathbf{u}_t(\theta^+)^H \mathbf{f}_i^{(m)}\right)^H,$$

because for the summand in (33), two minus signs on the sin function cancel each other and the phase of the exponential function changes from negative to positive. Since the function $f_{\theta,i,\omega_m}^{(m)}$ is obtained from the square of (33), $f_{\theta^-,i,\omega_m}^{(m)} = f_{\theta^+,i,\omega_m}^{(m)}$ holds.

(P.3) Given any $i \in \{1,\ldots,N/N_m\}$, the function $f_{\theta,i,\omega_m}^{(m)}$ is evaluated on the spatial angle in one-half of the main beam interval in (39) by *(P.2)*, e.g., set $\theta = -1 + (2i-1)N_m/N_t + \epsilon \in [-1 + (2i-1)N_m/N_t, -1 + (2iN_m - 1)/N_t]$ as in (52). For the i-th codeword, we consider the indices $p = (i-1)N_m + q$ and $p' = (i-1)N_m + q'$ for $q, q' \in \{1,\ldots,N_m\}$. Then, $f_{\theta,i,\omega_m}^{(m)}$ can be rewritten as

$$f_{\theta,i,\omega_m}^{(m)} = \frac{1}{N_t^2 N_m} \left\{ \sum_{q=1}^{N_m} \frac{\sin^2(\pi\tilde{q})}{\sin^2\left(\frac{\pi}{N_t}\tilde{q}\right)} + \sum_{q=1}^{N_m} \sum_{q'=q+1}^{N_m} 2\cos\left((q'-q)\left(\frac{N-1}{N}\pi - \omega_m\right)\right) \times \frac{\sin(\pi\tilde{q})\sin(\pi\tilde{q}')}{\sin\left(\frac{\pi}{N_t}\tilde{q}\right)\sin\left(\frac{\pi}{N_t}\tilde{q}'\right)} \right\}, \quad (55)$$

where $\tilde{q} = q - (N_m + \epsilon N_t - 1)/2$ and $\tilde{q}' = q' - (N_m + \epsilon N_t - 1)/2$. Since the function $f_{\theta,i,\omega_m}^{(m)}$ does not depend on the index i, we have the claim. ■

REFERENCES

[1] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook based beamforming sequence design in millimeter-wave systems," in *Proc. IEEE Global Commun. Conf.*, to be published, 2015.

[2] S. Sun, T. S. Rappaport *et al.*, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110 – 121, Dec. 2014.

[3] T. S. Rappaport, S. Sun *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" in *Proc. IEEE Access*, vol. 1, no. 1, 2014, pp. 335 – 349.

[4] W. Roh, J.-Y. Seol *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106 – 113, Feb. 2014.

[5] M. D. Renzo, H. Haas *et al.*, "Spatial modulation for generalized MIMO: Challenges, opportunities and implementation," in *Proc. IEEE*, vol. 102, no. 1, Jan. 2014, pp. 56 – 103.

[6] J. Wang, Z. Lan *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390 – 1399, Oct. 2009.

[7] J. Brady, N. Behdad, and A. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 2931 – 2941, Jul. 2013.

[8] S. Hur, T. Kim *et al.*, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391 – 4403, Oct. 2013.

[9] J. Song, J. Choi *et al.*, "Adaptive millimeter wave beam-alignment for dual-polarized broadcast MIMO systems," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1 – 12, Jul. 2015.

[10] O. E. Ayach, S. Rajagopal *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499 – 1513, Mar. 2014.

[11] A. Alkhateeb, O. E. Ayach *et al.*, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831 – 846, Oct. 2014.

[12] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *Proc. IEEE Int. Conf. on Commun.*, London, UK, Jun. 2015.

[13] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505 – 4515, Sep. 2014.

[14] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," submitted for publication. [Online]. Available: http://arxiv.org/abs/1407.1786.

[15] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546 – 557, Feb. 2004.

[16] S. Noh, M. D. Zoltowski *et al.*, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787 – 801, Oct. 2014.

[17] J. Choi, Z. Chance *et al.*, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016 – 5029, Dec. 2013.

[18] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802 – 814, Oct. 2014.

[19] X. Rao and V. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261 – 3271, Jun. 2014.

[20] J. So, D. Kim *et al.*, "Pilot signal design for massive MIMO systems: A received signal-to-noise-ratio-based approach," *IEEE Signal Process. Lett.*, vol. 52, no. 5, pp. 549 – 553, May 2015.

[21] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222 – 232, Feb. 2001.

[22] B. M. Hochwald and T. L. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 642 – 653, Mar. 2001.

[23] B. Gopalakrishnan and N. D. Sidiropoulos, "Cognitive transmit beamforming from binary CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 895 – 906, Feb. 2015.

[24] J. Seo, Y. Sung *et al.*, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," submitted for publication. [Online]. Available: http://arxiv.org/abs/1410.3711.

[25] M. R. Akdeniz, Y. Liu *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164 – 1179, Jun. 2014.

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

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

[28] A. Tadaion, M. Derakhtian *et al.*, "A fast multiple-source detection and localization array signal processing algorithm using the spatial filtering and ML approach," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1815 – 1827, May 2007.

[29] A. M. Hunter, J. G. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5058 – 5071, Dec. 2008.

[30] T. Bai, A. Alkhateeb, and R. W. H. Jr., "Coverage and capacity

of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70 – 77, Sep. 2014.

[31] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1998.

[32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York, NY: 7th ed., Academic Press, 2007.

[33] H. Alzer, "On some inequalities for the incomplete Gamma function," *Math. Comput.*, vol. 66, no. 218, pp. 771 – 778, Apr. 1997.

[34] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102 – 1110, Jul. 2003.

[35] 3GPP TSG RAN WG1 #62, R1-105011, "Way forward on 8Tx codebook for Rel. 10 DL MIMO," 2010.

[36] C. Lim, T. Yoo *et al.*, "Recent trend of multiuser MIMO in LTE-Advanced," *IEEE Wireless Commun. Mag.*, vol. 51, no. 3, pp. 127 – 135, Mar. 2013.

[37] J. Butler and R. Lowe, "Beamforming matrix simplifies design of electrically scanned antennas," *Electron. Design*, no. 9, pp. 170 – 173, Apr. 1961.

[38] K.-C. Huang and D. J. Edwards, *Millimetre Wave Antennas for Gigabit Wireless Communications: A Practical Guide to Design and Analysis in a System Context*. New York, NY: Wiley, 2008.

[39] C. Chang, R. Lee, and T. Shih, "Design of a beam switching/steering Butler matrix for phased array system," *IEEE Trans. Antennas Propag.*, vol. 58, no. 2, pp. 367 – 374, Feb. 2010.

[40] K. Wu, Y. J. Cheng *et al.*, "Substrate-integrated millimeter-wave and teraherz antenna technologies," *Proc. IEEE*, vol. 100, no. 7, pp. 2219 – 2232, Jul. 2012.

[41] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.

[42] A. Forenza, D. J. Love, and R. W. Heath Jr., "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924 – 1934, Jul. 2007.

[43] L. C. Andrews and R. L. Phillips, *Mathematical Techniques for Engineers and Scientists*. Bellingham, Washington: SPIE Press, 2003.