IEEE *Access*

Multidisciplinary ┊ Rapid Review ┊ Open Access Journal

Beam Codebook Design for 5G mmWave Terminals

JIANHUA MO[1], **(Member, IEEE), BOON LOONG NG**[1], **(Member, IEEE), SANGHYUN CHANG**[3],
PENGDA HUANG[1], **(Member, IEEE), MANDAR N. KULKARNI**[1], **(Student Member, IEEE),**
AHMAD ALAMMOURI[2], **(Student Member, IEEE), JIANZHONG CHARLIE ZHANG**[1],
(Fellow, IEEE), JEONGHEUM LEE[3], **(Member, IEEE), AND WON-JOON CHOI**[3]

[1]Samsung Research America, Plano, TX 75023, USA
[2]Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712, USA
[3]Samsung Electronics Company Ltd., Suwon 16677, South Korea

Corresponding author: Jianhua Mo (jianhua.m@samsung.com)

ABSTRACT A beam codebook of 5G millimeter wave (mmWave) for data communication consists of multiple high-peak-gain beams to compensate the high pathloss at the mmWave bands. These beams also have to point to different angular directions, such that by performing beam searching over the codebook, a good mmWave signal coverage over the full sphere around the terminal (spherical coverage) can be achieved. A model-based beam codebook design that assumes ideal omni-directional antenna pattern, and neglects the impact of terminal housing around the antenna, does not work well because the radiation pattern of a practical mmWave antenna combined with the impact of terminal housing is highly irregular. In this paper, we propose a novel and efficient data-driven method to generate a beam codebook to boost the spherical coverage of mmWave terminals. The method takes as inputs the measured or simulated electric field response data of each antenna and provides the codebook according to the requirements on the codebook size, spherical coverage, and so on. The method can be applied in a straightforward manner to different antenna type, antenna array configuration, placement, and terminal housing design. Our simulation results show that the proposed method generates a codebook better than the benchmark and 802.15.3c codebooks in terms of the spherical coverage.

INDEX TERMS Millimeter wave, beamforming, beam codebook, 5G handsets, spherical coverage, K-means, unsupervised machine learning.

I. INTRODUCTION

In 5G cellular networks, beamforming is necessary for overcoming large channel pathloss when a user equipment (UE) tries to establish a connection with a base station (BS) in millimeter wave (mmWave) bands such as the 28 GHz, 39 GHz, or 60 GHz bands [1]–[3]. To compensate for the smaller angular coverage due to the narrow analog beamwidth in mmWave, beam sweeping can be employed to enable wider angular signal reception or transmission coverage for the UE [4]–[6]. A beam codebook comprises a set of beams or codewords, where a codeword is a set of analog phase shift values, or a set of magnitude plus phase shift values, applied to the antenna elements, in order to form an analog beam. The 3rd Generation Partnership Project (3GPP) is the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenyu Xiao.

5G standardization body that specifies the minimum peak equivalent isotropic radiated power (EIRP) and the spherical coverage requirements of UE defined as a certain percentile of thecumulative distribution function (CDF) over the full sphere around the UE. There are a total of four UE power classes defined for various use cases or deployment scenarios; and the minimum peak EIRP and the spherical coverage requirements are different for different UE power classes. For example, it has been specified for the first generation (Release 15) of 5G mmWave handheld UE (power class 3) that the minimum peak EIRP is 22.4 dBm (20.6 dBm) and the minimum EIRP at the 50th percentile CDF over the full sphere around the UE is 11.5 dBm (8 dBm) for 28 GHz bands (39 GHz band) [7, Table 6.2.1.3-3]. This paper describes a novel codebook generation procedure and algorithms to obtain a beam codebook given a set of requirements and performance criteria.

A. RELATED WORK

Beam codebook design has been extensively considered in both academia and industry [5], [6]. A beam codebook design was provided in 802.15.3c [8, Ch. 13], assuming 1-D and 2-D arrays with uniform spacing of half-wavelength. The beam searching or training process is divided into three stages, namely, link-level device discovery, sector-level alignment, and beam-level refinement. The omni or quasi-omni radiation pattern, wide beam, and narrow beam are designed respectively to fulfill the requirements of these three phases [9], [10]. The same 3-stage training process was adopted in [11] where codebooks for a 2-ring circular array were proposed. The inner small ring generates quasi-omni and sector radiation patterns while the outer larger ring generates the last-stage directional beam patterns.

The idea of 3-stage codebook design was extended to general hierarchical codebook design where the number of stages or layer are not limited to be three. The analog codebook design was considered in [12] where the sub-array method used to generate a "flatted" wide beam. The paper [13] proposed a heuristic method where the uniform linear array is divided into 2, 3, or 4 sub-arrays and the steering direction and length of the sub-arrays are numerically optimized to maximize the minimum beamforming gain in the required coverage region. The work [14] proposed a deactivation approach where the antenna elements are adaptively deactivated to create beam with various beamwidth. Combined with the deactivation method, the sub-array based hierarchical codebook generation was optimized in [15] where either all or a half of the antenna elements are activated. The method of [15] was further enhanced in [16] where the deactivation method is dropped, and all the antenna are always activated to increase the maximal total transmission power. The sub-array method was adopted in [17] to design a 3-D wide beam for uniform planar array. To design beams with a small ripple in both the main and side lobes, a beam pattern optimization problem was formulated in [18]. However, the optimization problem considered the total power constraints rather than the individual antenna power constraint, thereby the resultant beam has a large peak-to-average-power ratio, which implies low power efficiency.

Besides analog precoding, hierarchical codebook design for hybrid analog-digital precoding was considered in [19]–[21]. Given the required angular region to cover, the analog and digital precoder design was formulated in [19] as a sparse approximation problem, and solved by a variant of of orthogonal matching pursuit algorithms. The authors of [21] proposed a DFT-based multilevel codebook design where the adjacent phase-shifted DFT beams are summed up to construct wide beams. Last, the sub-array method was altered in [20] to support the hybrid precoding.

Considering the high cost, power consumption and form factor of radio frequency (RF) chain, the mmWave terminals are not likely to adopt hybrid or fully digital beamforming where more than one RF chains are needed for a single array.

Therefore, analog beamforming for each antenna array is assumed throughout this paper.

In this paper, we focus on the beam codebook for data transmission, i.e, the third stage codebook in 802.15.3c for beam searching [8], or the bottom layer fine codebook in a hierarchical codebook design [14], [15], [21]. In 802.15.3c, the codebooks are generated with 2-bit phase shifters without amplitude adjustments for the consideration of the hardware complexity. To reduce the gain loss at the intersections of two beams, the number of beams should be twice the number of array elements [22]. In the 802.11ad document [23, Section 6.6], the beam codebook design is formulated as a geometric problem to cover the sphere sector with circles by assuming that the main lobe of the beam has a circular shape. The assumption of circular shape, however, does not hold when a beam is beamforming towards directions away from the broadside direction and therefore results in coverage gaps between beams. In [14], [15], [19], and [21], the last layer beams are pointing to directions uniformly distributed in the angular domain or spatial frequency. In such cases, the beamforming vector is just the steering vector for a given beamforming direction (or an approximation of it if there is a phase shifters resolution constraint). For example, for a simple linear array with spacing d, the beamforming weights would have the progressive phases as $\frac{2\pi i}{\lambda} d \cos\theta$ where λ is the wavelength, i is the antenna index, and θ is the beamforming direction with respect to the axis of the array. The codeword for 2-D planar array is then the Kronecker product of two codewords for 1-D linear arrays, as done in [17] and [24].

All these work [8]–[24] assumed an ideal isotropic radiation pattern and considered rather regular antenna setup, i.e., uniform linear, uniform planar or uniform circular array. We call these designs, which are based on simple theoretical assumptions, as **model-based** approach hereafter. Such designs, however, ignore many practical issues as described next.

Antenna for mmWave bands is intrinsically directional. For example, the patch antenna usually has a high front-to-back ratio and consequently can cover at most half-sphere [25]. The directional element radiation pattern will also result in the drift of the peak gain direction from the intended one if the beamforming vector is merely designed based on the steering vector. In addition, when placed inside mobile handsets, the radiation gain of the mmWave antenna is less than the free-space case due to blockage loss and the radiation pattern shape is also changed [26].

Antenna placement and antenna spacing may not be regular. For example, the planar array may not have the half-wavelength spacing between adjacent elements due to form-factor constraints. Another reason is related to the multi-frequency bands that the mmWave terminal has to support. The mmWave bands for 5G deployment in US will include 24 GHz, 28 GHz and 39 GHz, etc. [1] The same antenna

---

[1]Federal Communications Commission's Facilitate America's Superiority in 5G Technology (the 5G FAST Plan).

arrays, however, are likely to be used at all these carrier frequency bands. Therefore, a half-wavelength spacing at a frequency band will result in less than (or more than) half-wavelength spacing at other lower (or higher) frequency bands.

A 5G mmWave capable UE is typically equipped with multiple antenna arrays. For example, in a design given in [27], there are at most four mmWave modules mounted on the top, bottom, left and right edges of the phone, respectively. Multiple mmWave antenna arrays are necessary to enable a good spherical coverage over the whole sphere and to circumvent human body blocking. In a benchmark codebook design, the beam codewords are designed independently for each array as assumed in [27], which is a suboptimal solution since the interaction and coordination between the arrays are ignored.

B. CONTRIBUTIONS

A practical beam codebook design should at least take into account the following factors,
1) Antenna element type and gain (e.g. isotropic, dipole, microstrip patch);
2) Array layout (e.g. linear, rectangular, circular, cylinder) and placement if there are multiple arrays;
3) Requirements of codebook (e.g. codebook size, required coverage regions, phase shifter resolution);
4) Consideration about UE housing (e.g., display screen, battery);
5) The coordination among different arrays mounted on the same terminal.

Although it might be able to model the antenna element type by approximation models [25], it is difficult, if not impossible, to analytically model the other factors, including the housing effects caused by a plurality of components inside the mmWave terminal with various size, shape and electromagnetic properties. Faced with aforementioned challenges, it is generally difficult to find an analytical method to generate the codebook. It is also impossible to find the optimal codebook by an exhaustive search because of its exponential complexity as $\mathcal{O}\left(2^{bLK}\right)$ where b is the phase shifter resolutions, L is the antenna array size, and K is the codebook size. For instance, for a small array where $b = 2$, $L = 4$ and $K = 4$, there are 2^{32} possible codebooks assuming that the analog beamforming codewords are ordered.

In this paper, we present a **data-driven** codebook design method. An important advantage of our method is that it can be applied agnostically with any antenna type, array layout and placement. The antenna information required for our method is simply the electrical field (E-field) response of each antenna element in a given layout, which can be obtained through electromagnetic simulation software (for example, high-frequency structure simulator (HFSS) by Ansys) or through measurements.

Our first algorithm is a *Greedy* algorithm, which sequentially selects the beam codewords to augment the spherical coverage. The performance of this algorithm relies on the quality of the candidate codewords pool as well as the codeword selection criterion. Our second algorithm is based on an unsupervised machine learning algorithm, namely, *K-Means*. In this algorithm, the angular directions are clustered based on their E-field response and then the beam codewords are optimized to improve the average gain of the clustered points. This clustering and optimization procedure is repeated until convergence.

The contributions of this paper are summarized as follows.
1) We formulate a beam codebook design problem from the perspective of maximizing the spherical coverage. The optimization problem takes into account the amplitude and phase resolution constraint, as well as the codebook size. In addition, compared with the previous work (e.g., [27]), the two polarization components are both considered in the design.
2) We propose a data-driven approach for codebook design. The proposed approach, which takes as inputs the E-field response data from simulations or measurements, automatically generate the codebook without request of modeling the antenna element pattern, the housing effects, etc.
3) An upper bound of the composite radiation pattern is derived. The upper bound provides a reference for evaluating the performance of the designed codebook.
4) Comprehensive numerical simulations are provided to confirm the effectiveness and superiority of the proposed codebook design.

The paper is organized as follows. In Section II, we formulate the problem of beam codebook design. In Section III, we present the design of a single beam with power and phase constraints, which lays the foundation for discussions of our algorithms. The upper bound of the composite radiation pattern is discussed in Section IV. Our two heuristic algorithms are provided in Section V and Section VI. The simulation results are shown in Section VII. Further discussions on the additional advantages of our algorithms are provided in Section VIII. A comparison with other model-based method based on simplified E-field response data is given in Section IX. The paper is concluded in Section X.

Notation: Bold uppercase letter **A** and bold lowercase letter **a** represents a matrix and a column vector, respectively. $\mathbf{A} \geq 0$ implies that $\mathbf{A}$ is a positive semi-definite matrix. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ denotes the transpose, conjugate and Hermitian of a vector or matrix, respectively. $\|\mathbf{a}\|$ is the norm of the vector **a**. $\arg(\cdot) \in [0, 2\pi)$ denotes the phase of a complex-valued input. $\mod(a, n)$ stands for the remainder of a divided by n.

II. BEAM CODEBOOK DESIGN PROBLEM

The coordinate system used throughout this paper is shown in Fig. 1. The UE is placed around the origin. θ (ϕ) is defined as the zenith (azimuth) angle. Since the electrical field is a vector field, it is represented by three orthogonal components, denoted as (E_R, E_Θ, E_Φ), at each observation point on the surface of a sphere. We consider the E-field response in the
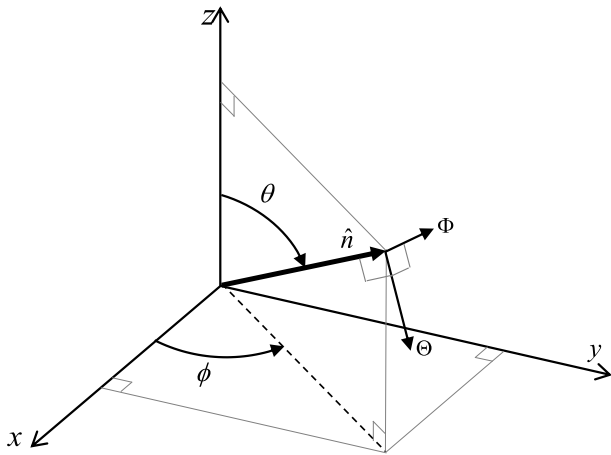
FIGURE 1. The coordinate system for E-field analysis.

far-field (Fraunhofer) region where the electromagnetic wave appears locally as a plane wave in any specified direction. As a result, the radial component, i.e., E_R, is zero or vanishingly small compared to the other two components, i.e., E_Θ and E_Φ [25]. Therefore, for a given direction $\hat{\mathbf{n}}$, we only consider the E-field Θ component and Φ component, which are perpendicular to $\hat{\mathbf{n}}$ as shown in Fig. 1.

Assume there are L antenna elements in an array. Let $e_\ell^\Theta(\theta, \phi)$ and $e_\ell^\Phi(\theta, \phi)$, $\ell \in \{1, 2, \cdots, L\}$, denote the complex-valued E-field response of the ℓ-th antenna element for the Θ component and the Φ component, respectively, at the direction (θ, ϕ).[2] Denote the E-field data in vectors as,

$$\mathbf{e}_\Theta(\theta, \phi) \triangleq [e_1^\Theta(\theta, \phi), e_2^\Theta(\theta, \phi), \cdots, e_L^\Theta(\theta, \phi)]^T, \quad (1)$$

$$\mathbf{e}_\Phi(\theta, \phi) \triangleq [e_1^\Phi(\theta, \phi), e_2^\Phi(\theta, \phi), \cdots, e_L^\Phi(\theta, \phi)]^T. \quad (2)$$

As mentioned in the introduction, the E-field data can be obtained through electromagnetic simulation or measurement, which is usually sampled on a mesh grid, for example, $[\theta, \phi] = [0° : q_\theta : 180°] \times [0° : q_\phi : 360°)$, where q_θ, q_ϕ are the simulation or measurement step sizes.

Let $\mathbf{w} \triangleq [w_1, w_2, \cdots, w_L]^T$ denote the complex-valued weights applied on the antenna elements. Without loss of generality, we will assume the beamforming codeword $\mathbf{w}$ always has unit-norm throughout this paper, i.e., $\|\mathbf{w}\|^2 = 1$. According to the superposition principle, the E-fields for the Θ and Φ components after applying the beamforming weights are given by

$$\mathcal{E}_{\Theta,\mathbf{w}}(\theta, \phi) = \sum_{\ell=1}^{L} w_\ell^* e_\ell^\Theta(\theta, \phi) = \mathbf{w}^H \mathbf{e}_\Theta(\theta, \phi), \quad (3)$$

$$\mathcal{E}_{\Phi,\mathbf{w}}(\theta, \phi) = \sum_{\ell=1}^{L} w_\ell^* e_\ell^\Phi(\theta, \phi) = \mathbf{w}^H \mathbf{e}_\Phi(\theta, \phi). \quad (4)$$

---

[2]The E-field response in this paper denotes a product, i.e., $r \cdot E_X$, where E_X ($X = \Theta, \Phi$) is the E-field strength measured at a distance r to the origin when the incident power to the antenna element is 1 Watt. Since $E_X \propto \frac{1}{r}$ in the far-field region [25], the E-field response is independent of the distance r. Note that the E-field response in this paper corresponds to the term 'rE' in HFSS.

The realized beamforming gain is the sum of the realized gains of the Θ and Φ components [25],

$$G_\mathbf{w}(\theta, \phi)$$
$$= \frac{4\pi}{\|\mathbf{w}\|^2} \cdot \frac{1}{2\eta_0} \left(|\mathcal{E}_{\Theta,\mathbf{w}}(\theta, \phi)|^2 + |\mathcal{E}_{\Phi,\mathbf{w}}(\theta, \phi)|^2 \right) \quad (5)$$
$$= \frac{2\pi}{\eta_0} \left(\mathbf{w}^H \left(\mathbf{e}_\Theta(\theta, \phi) \mathbf{e}_\Theta^H(\theta, \phi) + \mathbf{e}_\Phi(\theta, \phi) \mathbf{e}_\Phi^H(\theta, \phi) \right) \mathbf{w} \right) \quad (6)$$
$$= \frac{2\pi}{\eta_0} \mathbf{w}^H \mathbf{M}(\theta, \phi) \mathbf{w}. \quad (7)$$

where $\eta_0 \approx 377 \ \Omega$ is the impedance of the free space, $\mathbf{M}(\theta, \phi) \triangleq \mathbf{e}_\Theta(\theta, \phi)\mathbf{e}_\Theta^H(\theta, \phi) + \mathbf{e}_\Phi(\theta, \phi)\mathbf{e}_\Phi^H(\theta, \phi)$, and (6) is obtained by plugging in (3) and (4) and noticing the unit-norm assumption of $\mathbf{w}$.

We assume that the phase shifters are constrained to b bits, and the codebook $\mathcal{W}_c$ has a size limitation K, i.e., $\mathcal{W}_c \triangleq \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K\}$. A codebook of small size will help reduce the beam sweeping time, power consumption as well as the memory space in the modem. There is also a requirement on the composite radiation gain pattern, which is the maximum over all the gain patterns of the codewords and is denoted as $S(\mathcal{W}_c, \theta, \phi)$. The composite radiation pattern indicates the wellness of the spherical coverage of the codebook. Specifically, it can be used to identify coverage holes. The beam codebook design problem is formulated as below.

$$(\textbf{P1}) \max_{\mathcal{W}_c} U\left(S(\mathcal{W}_c, \theta, \phi) \right) \quad (8a)$$

$$s.t. \ S(\mathcal{W}_c, \theta, \phi) = \frac{2\pi}{\eta_0} \max_{\mathbf{w}_k \in \mathcal{W}_c} \mathbf{w}_k^H \mathbf{M}(\theta, \phi)\mathbf{w}_k, \quad (8b)$$

$$\left(\sqrt{L} w_{k\ell} \right)^{2^b} = 1, \quad \forall k, \ell, \quad (8c)$$

where the last equation (8c) encapsulates the magnitude constraint $|w_{k\ell}| = \frac{1}{\sqrt{L}}$ as well as the phase constraint $\arg(w_{k\ell}) \in \left\{ 0, \frac{2\pi}{2^b}, \cdots, (2^b - 1) \frac{2\pi}{2^b} \right\}$.

When there are multiple arrays, we assume that only one of the antenna arrays is activated at a given time, which is a typical implementation assumption. As a result, the problem formulation is similar to (**P1**) with the exception that the composite radiation pattern is the maximum over all the codewords of all the arrays.

The utility function $U(\cdot)$ can be defined as the average gain across the whole sphere, or the xth-percentile of the gain over the unit-sphere (e.g., $x = 20, 50$). As mentioned in the introduction, the 3GPP specifies the spherical converage requirement for handheld UE (power class 3) in terms of the 50th percentile EIRP [7, Table 6.2.1.3-3].[3] Throughout this paper, the utility function is defined over a uniform sampling over the sphere or a specified angular region. In particular,

---

[3]Note that EIRP is equal to the sum of the realized beamforming gain and the incident power towards the antennas in the log scale. We normalize the incident power as one throughout this paper and thus the optimization of EIRP is equivalent to the optimization of beamforming gain.

the CDF of the gain over the sphere is defined as,

$$F_S(s) = \frac{1}{4\pi} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \mathbb{1}\left\{S\left(\mathcal{W}_c, \theta, \phi\right) \leq s\right\} \sin\theta d\theta d\phi \quad (9)$$

$$\approx \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbb{1}\left\{S\left(\mathcal{W}_c, \theta_i, \phi_i\right) \leq s\right\}. \quad (10)$$

Maximizing a particular percentile value (e.g., 50% requested by 3GPP) of the distribution is not an easy task when considering the mathematical tractability. A more tractable utility function is the average gain over the sphere, which is defined as

$$\mathbb{E}_{\theta,\phi}\left[S\left(\mathcal{W}_c, \theta, \phi\right)\right]$$
$$= \frac{1}{4\pi} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} S\left(\mathcal{W}_c, \theta, \phi\right) \sin\theta d\theta d\phi \quad (11)$$

$$\approx \frac{1}{N_p} \sum_{i=1}^{N_p} S(\mathcal{W}_c, \theta_i, \phi_i), \quad (12)$$

where the approximation in (12) comes from a set of N_p uniformly distributed sampling points on the sphere, which can be obtained through a Fibonacci grid [28].

The problem **P1** is non-convex and NP-hard due to the constraint (8c). In this paper, we provide two heuristic algorithms. As verified through our simulation based on practical phone design, the proposed heuristic algorithms have low complexities and provide satisfactory performance.

In the problem formulation **P1**, the flatness of each beam is ignored for several reasons. First, we are designing high-peak-gain narrow beams for data transmission instead of quasi-omni or wide beam for initial device discovery and sector-level searching, therefore there is no flatness issue in our designed narrow beams. Second, to establish a successful mmWave connection, the spherical coverage of the composite radiation pattern is a more relevant and effective metric than the flatness of the individual beam. Last, but not least, if the radiation pattern is severely irregular due to element pattern or strong housing effects, then it is impossible to design a wide beam with flat gain.

III. DATA-DRIVEN DESIGN OF A SINGLE BEAM

Before presenting our beam codebook design, we first present the approach on the design of a single beam, which will be used in the Greedy and K-Means algorithms.

It is noteworthy that the beam design in the data-driven codebook is quite different from the conventional model-based method in two-fold. First, the beam codeword is designed based on the E-field data, instead of being a steering vector,[4] or a weighted sum of a few steering vectors [21], or a concatenation of steering vectors on sub-arrays [12], [13], [15], [17]. Due to the directional element pattern and housing

---

[4]A steering vector of an array at a direction (θ, ϕ) has the form $w_\ell = \frac{1}{\sqrt{L}} \exp\left(j\frac{2\pi}{\lambda}\hat{\mathbf{n}}^T \mathbf{x}_\ell\right)$ where $\hat{\mathbf{n}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T$ (see the definition of θ and ϕ in Fig. 1), λ is the wavelength, and $\mathbf{x}_\ell$ is the 3-D coordinate of the ℓ-th element.

effects, the beam radiation pattern may point away from the intended direction. Second, the beam is carefully designed to take into account two polarization components, which is not considered in the prior work, e.g., [27].

Given the E-field response matrix $\mathbf{M}(\theta, \phi)$ at a given direction (θ, ϕ), or the sum E-field response over a set of directions, i.e., $\mathbf{M} = \sum_{(\theta,\phi)\in\mathcal{A}} \mathbf{M}(\theta, \phi)$, we want to design a beamforming vector to maximize the beamforming gain $\mathbf{w}^*\mathbf{M}\mathbf{w}$ under different constraints.

First, consider a simple case with sum power constraint. The optimization problem is as follows.

$$B_1(\mathbf{M}) \triangleq \max_{||\mathbf{w}||\leq 1} \mathbf{w}^H \mathbf{M}\mathbf{w} \quad (13a)$$
$$= \lambda_{\max}(\mathbf{M}), \quad (13b)$$

where $\lambda_{\max}$ represents the maximal eigenvalue and the optimal $\mathbf{w}$ is the eigenvector corresponding to the largest eigenvalue. The solution value is denoted as B_1 and $\frac{2\pi}{\eta_0}B_1$ is the maximum achievable beamforming gain.

A. CONTINUOUS-PHASE UNIMODULAR BEAM DESIGN

In our beam codebook design, the beam codeword is subject to per-element constant power constraints. The problem with such constraint can be formulated as,

$$B_2(\mathbf{M}) \triangleq \max_{\mathbf{w}:|w_i|\leq\frac{1}{\sqrt{L}}, \forall i} \mathbf{w}^H \mathbf{M}\mathbf{w}. \quad (14)$$

First, it is not hard to see that the optimal $\mathbf{w}$ should fully utilize the power, i.e., $|w_i| = \frac{1}{\sqrt{L}}$. A proof can be found in [29, Corollary 2]. Second, if rank$(\mathbf{M}) = 1$, i.e., $\mathbf{M} = \mathbf{m}\mathbf{m}^H$, then the optimal solution is the co-phasing beamforming, i.e., $w_i^\star = \frac{m_i}{|m_i|\sqrt{L}}$ and $B_2 = \frac{1}{L}\left(\sum_i |m_i|\right)^2$ [27]. However, since $\mathbf{e}_\Theta(\theta, \phi)$ is not a scaled vector of $\mathbf{e}_\Phi(\theta, \phi)$ almost surely, rank$(\mathbf{M})$ is larger than one almost surely and thus there is no closed-form solution.

In fact, since both the objective function and the constraints ($w_i^* w_i = \frac{1}{L}, \forall i$) are quadratic functions and $\mathbf{M}$ is positive semi-definite, (14) is a non-convex quadratically constrained quadratic program (QCQP), which is in general an NP-hard problem proved by reducing an NP-complete matrix partitioning problem [30]. An approximate solution can be found by using the prevailing semi-definite relaxation (SDR) method [31] as follows.

Denote $\mathbf{D}_i$ as an $L \times L$ all-zero matrix except that the i-th diagonal element is 1. We relax (14) as a semi-definite programming (SDP) as follows,

$$\overline{B}_2(\mathbf{M}) \triangleq \max_{\mathbf{W}} \text{tr}(\mathbf{M}\mathbf{W}) \quad (15a)$$

$$s.t. \ \text{tr}(\mathbf{D}_i\mathbf{W}) = \frac{1}{L}, \quad 1 \leq i \leq L, \quad (15b)$$

$$\mathbf{W} \geq \mathbf{0}. \quad (15c)$$

A standard interior point method [32] or a more efficient row-by-row block coordinate descend method [33], [34] can be applied to solve this convex SDP problem. The worst-case complexity to solve a SDP is $\mathcal{O}(L^{4.5})$, while the customized

Algorithm 1 Gaussian Randomization Procedure (GRP) [35]

Inputs: SDP solution $\mathbf{W}_0$ with rank($\mathbf{W}_0$) > 1, and the number of randomizations N_G.

1) Compute the eigenvalue decomposition of $\mathbf{W}_0$, i.e., $\mathbf{W}_0 = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$.
2) For $1 \le n \le N_G$, generate $\mathbf{w}^{(n)} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\xi^{(n)}$, where $\xi^{(n)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ are complexed-valued Gaussian random vectors.
3) Construct N_G feasible solutions,

$$\widetilde{\mathbf{w}}^{(n)} = \frac{1}{\sqrt{L}}\exp\left(\mathrm{j}\arg\left(\mathbf{w}^{(n)}\right)\right). \quad (16)$$

4) Determine $\widetilde{\mathbf{w}}_0 = \arg\max_{\widetilde{\mathbf{w}}^{(n)}}\left(\widetilde{\mathbf{w}}^{(n)}\right)^H\mathbf{M}\widetilde{\mathbf{w}}^{(n)}$.

Algorithm 2 Iterative Coordinate Descent Algorithm for Beamforming Design With Per-Element Power Constraint [29], [37]

1) Initialize $\mathbf{w}$ and $i \leftarrow 1$.
2) Update w_i as

$$w_i \leftarrow \frac{1}{\sqrt{L}}\exp\left(\mathrm{j}\arg\left(\sum_{k \neq i}M_{ik}w_k\right)\right). \quad (17)$$

3) Check convergence of the beamforming gain. If yes, stop; if not, let $i \leftarrow \mathrm{mod}(i, L) + 1$ and go back to Step 2.

---

row-by-row method has a complexity of $\mathcal{O}(L^3)$ [35]. If the obtained optimal solution $\mathbf{W}_0$ is of rank one, then we can write $\mathbf{W}_0 = \mathbf{w}_0\mathbf{w}_0^H$, and $\mathbf{w}_0$ is a feasible optimal solution. On the other hand, if the rank of $\mathbf{W}_0$ is larger than one, then a random approximation procedure [30], [35], [36] can be used to find an approximate optimal solution. The details of the procedure is shown in Algorithm 1, where N_G realizations of $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_0)$ are generated and the best one is selected and denoted as $\widetilde{\mathbf{w}}_0$. The theoretical approximation accuracy is $\frac{\pi}{4}$, i.e., the expectation of $\widetilde{\mathbf{w}}_0^H\mathbf{M}\widetilde{\mathbf{w}}_0$ is no less than $\frac{\pi}{4}$ of the global optimum [30], [36]. In our simulation setup where $L = 4$ and rank($\mathbf{M}$) = 2, a rank-one solution is obtained in more than 99% of the cases. That is to say, SDR provides the optimal unimodular beam in more than 99% of simulation cases.

Besides SDP-GRP, another sub-optimal but more time-efficient algorithm is to sequentially optimize the phase of each element [29, Table 1] [37, Algorithm 2]. The details of the iterative algorithm is given in Algorithm 2 for completeness. The solution of Algorithm 2 is guaranteed to converge to a stationary local optimal solution satisfying the Karush-Kuhn-Tucker (KKT) condition [29], but may not be the optimal solution. The performance of the iterative algorithm depends on the choice of the initialization in Step 1. As an option, we can choose the initial $\mathbf{w}$ as the eigenvector associated with the largest eigenvalue of $\mathbf{M}$ to increase the likelihood of convergence to the global optimum. Another option is to set the solution from SDP-GRP as the initial $\mathbf{w}$, i.e., concatenate SDP-GRP and Algorithm 2. By doing so, the sequential optimization method is used to further improve the quality of the SDP-GRP solution and therefore the chance of finding the global optimum. The complexity of the iterative algorithm is $\mathcal{O}(L^2)$ [37]. Another alternative of Algorithm 2 is the power method-like approach given in [38, Section III] which, however, has a higher complexity as $\mathcal{O}(L^3)$.

B. DISCRETE-PHASE UNIMODULAR BEAM DESIGN

In practice, there is also a resolution constraint on the phase shifters. By taking into account both the per-element power

and phase constraints, the beam design problem is,

$$B_3(\mathbf{M}) \triangleq \max_{\mathbf{w}:\ \left(\sqrt{L}w_i\right)^{2^b}=1, \forall i} \mathbf{w}^H\mathbf{M}\mathbf{w}. \quad (18)$$

If $\mathbf{M}$ is not rank-deficient, it is proven that (18) is also a NP-hard problem [30], since it includes the max-cut problem and max-3-cut problem which are known to be NP-hard. An approximate solution can be obtained by applying the SDP-GRP technique shown above but with minor modifications. First, in this discrete-phase case, we have to run the Gaussian randomization procedure even if $\mathbf{W}_0$ is a rank-one matrix because of the phase constraints. Second, when constructing feasible solutions inside the Gaussian randomization procedure, the phases need be quantized. This can be done by replacing (16) with the following equation,

$$\widetilde{\mathbf{w}}^{(n)} = \frac{1}{\sqrt{L}}\exp\left(\mathrm{j}\mathcal{Q}_b\left(\arg\left(\mathbf{w}^{(n)}\right)\right)\right), \quad (19)$$

where the function $\mathcal{Q}_b(\cdot)$ quantizes the phase from $[0, 2\pi)$ to $\left\{0, \frac{2\pi}{2^b}, \cdots, (2^b - 1)\frac{2\pi}{2^b}\right\}$. The approximation accuracy of the SDP-GRP solution is $\frac{\left(2^b\sin\left(\frac{\pi}{2^b}\right)\right)^2}{4\pi}$ [36], which is same as the case of continuous phase, i.e., $\frac{\pi}{4}$, when $b \to \infty$.

Similar to the case of continuous phase, the sequential optimization method can be used to improve the quality of the SDP-GRP solution. The iterative process is done in Step 2-3 in Algorithm 3. It is not hard to see that Algorithm 3 will definitely converge since in each iteration, the phase of w_i is assigned as the optimal value from the discrete set $\left\{0, \frac{2\pi}{2^b}, \cdots, (2^b - 1)\frac{2\pi}{2^b}\right\}$ to maximize $\mathbf{w}^H\mathbf{M}\mathbf{w}$. The overall complexity of Algorithm 3 is $\mathcal{O}(L^3)$.

On the other hand, when $\mathbf{M}$ is rank-deficient, the problem (18) can be solved with polynomial complexity of L, i.e., $\mathcal{O}\left(\left(\frac{2^bL}{2}\right)^{2\,\mathrm{rank}(\mathbf{M})}\right)$ [39]. Unfortunately, such algorithms for discrete phases have exponential complexity with respect to b and cannot be extended to the continuous phase case in (14) where 2^b is approaching infinity. In addition, the runtime of the algorithm when b is larger, e.g., $b = 5$, is much longer than solving a SDP problem. Therefore, we do not adopt this method in this paper.

Algorithm 3 Algorithm for Beamforming Design With Per-Element Power and Phase Constraints

1) Solve the SDP given in (15) and perform the Gaussian randomization procedure shown in Algorithm 1 where (16) is replaced with (19) to obtain a discrete-phase solution **w**.

2) Update w_i as

$$w_i \leftarrow \frac{1}{\sqrt{L}} \exp\left(j\mathcal{Q}_b\left(\arg\left(\sum_{k \neq i} M_{ik} w_k\right)\right)\right). \quad (20)$$

3) Check convergence of the beamforming gain. If yes, stop; if not, let $i \leftarrow \mod(i, L) + 1$ and go back to Step 2.

Last, it is not hard to see that $B_1(\mathbf{M}) \geq B_2(\mathbf{M}) \geq B_3(\mathbf{M})$ since the set of feasible solutions shrinks from (13), to (14) and (18).

IV. UPPER BOUND OF THE COMPOSITE GAIN PATTERN

In the prior work, the upper bound has a uniform gain across the whole sphere by assuming all the elements are omni-directional. However, this is not the case at mmWave band where the antenna element has an inherent directional radiation pattern.

In this section, we provide an upper bound for the composite radiation pattern. The upper bound is directly derived from antenna element E-field response data (i.e., $\mathbf{e}_\Theta$ and $\mathbf{e}_\Phi$) and independent of codebook size K. It provides a good reference for evaluation of codebooks. For example, the number of beams required for the composite pattern to approach the upper bound can be evaluated.

Mathematically, the upper bound is obtained by solving the following problem over the whole sphere $(\theta, \phi) \in [0°, 180°] \times [0, 360°)$.

$$\overline{S}(\theta, \phi) \triangleq \frac{2\pi}{\eta_0} \max_{\|\mathbf{w}\| \leq 1} \mathbf{w}^H \mathbf{M}(\theta, \phi) \mathbf{w} \quad (21a)$$

$$= \frac{2\pi}{\eta_0} \lambda_{\max}\left(\mathbf{M}(\theta, \phi)\right). \quad (21b)$$

The upper bound can only be achieved by a beam codebook consisting of the maximal eigenvector of $\mathbf{M}(\theta, \phi)$ for every direction. In other words, we have to remove codebook size limitation, per-element power constraint, and the discrete-phase constraint to construct a codebook being able to attain the upper bound.

Last, the upper bound for a multi-array setup is simply taken as the maximum over the upper bounds of each individual array.

V. GREEDY ALGORITHM

In this section, we present a Greedy algorithm for the beam codebook design. The proposed algorithm greedily selects codewords from a candidate set as shown in Algorithm 4.

In Step 1 of the Greedy algorithm, candidate beam codewords are generated. We provide two possible methods to

Algorithm 4 Greedy Algorithm for Beam Codebook Design

1) Generate the candidate beam set, denoted as $\mathcal{W}_d$. Initialize the beam codebook as an empty set, i.e., $\mathcal{W}_c = \varnothing$.

2) Find a beam codeword maximizing the utility function of the spherical coverage, i.e.,

$$\mathbf{w}^\star = \underset{\mathbf{w} \in \mathcal{W}_d \setminus \mathcal{W}_c}{\arg\max} \ U\left(S\left(\mathcal{W}_c \cup \mathbf{w}, \theta, \phi\right)\right). \quad (22)$$

Insert the selected beam codeword into the beam codebook,

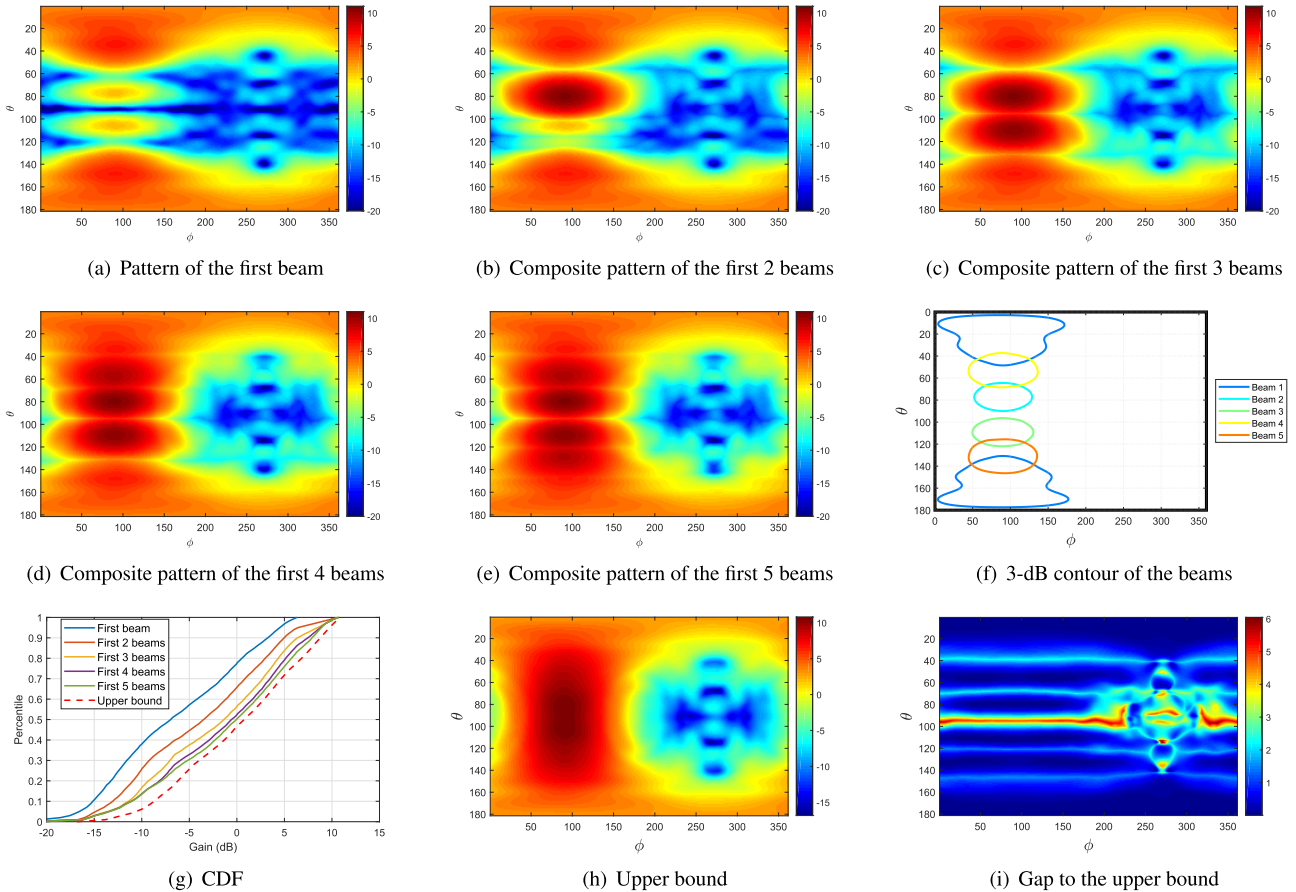$$\mathcal{W}_c \leftarrow \mathcal{W}_c \cup \mathbf{w}^\star. \quad (23)$$

3) Stop if a certain stopping criterion is met; go back to Step 2 otherwise.

---

generate the candidate beam codewords in Section V-A. In Step 2, given certain performance criteria, a beam from the candidate set is selected. In Step 3, check if the stopping condition is met. If the answer is yes, the algorithm is terminated and the selected beam codewords constitute the final codebook. Otherwise, Step 2 is repeated.

Denote the size of the candidate beam set as N_d and the codebook size as K. The complexity of the Greedy algorithm is mainly from Step 2, whose runtime is proportional to $\sum_{i=0}^{K-1}(N_d - i)$. Along with the complexity of $\mathcal{O}(N_d L^3)$ to generate N_d candidate beams, the total complexity is $\mathcal{O}(N_d(L^3 + K))$.

Fig. 2 shows an example of the Greedy algorithm operation. A linear 1×4 patch antenna array along the z-axis with broadside direction being $(\theta = 90°, \phi = 90°)$, according to the coordinate system in Fig. 1, is simulated by HFSS. The 5 codewords are selected one at a time from a candidate set of 363 codewords to boost the composite radiation pattern (see Fig. 2(a)-Fig. 2(e)). It is important to note that the main lobe of the selected beams shown in Fig. 2(f) are naturally pointing to different directions without explicit or manual enforcement. In each step, the selected codeword naturally targets the region with the poorest coverage thus far. It is observed in Fig. 2(g) that the spherical coverage improves with increasing number of beams. The composite radiation pattern is compared with the upper bound in Fig. 2(i). The gap is less than 2 dB excluding a coverage hole located around $\theta = 95°$. The codewords selected in the next iterations are expected to cover this hole.

Note that for this simulated patch array, the antenna elements do not assume omni-directional radiation patterns. As clearly seen from the upper bound shown in Fig. 2(h), the array can cover only a half sphere $(0° \leq \phi \leq 180°)$ because of the high front-to-back ratio of the patch antenna. As seen in the CDF curve shown in Fig. 2(g), the dynamic range of the upper bound is from -15 dB to 10 dB, namely, the front-to-back ratio is around 25 dB. In addition, the 3-dB contour of Beam 1, which is pointing to the

(a) Pattern of the first beam

(b) Composite pattern of the first 2 beams

(c) Composite pattern of the first 3 beams

(d) Composite pattern of the first 4 beams

(e) Composite pattern of the first 5 beams

(f) 3-dB contour of the beams

(g) CDF

(h) Upper bound

(i) Gap to the upper bound

FIGURE 2. Greedy algorithm operation on a linear 1 × 4 patch antenna array with 5-bit phase shifters. In this example, the candidate codewords are phase-quantized and magnitude-normalized eigenvectors corresponding to the maximal eigenvalues, and the selection criterion is the mean gain. Note that the beamforming gain is plotted in the dB scale.

directions away from the broadside, does not have a regular circular or ellipsoid shape.

In the following subsections, we discuss more details on the candidate codewords generation in Step 1, selection criteria in Step 2 and the stopping condition in Step 3.

A. CANDIDATE CODEWORDS GENERATION

We provide two methods to generate the candidate codewords based on the E-field data. We first pick a set of sampling points on the sphere. This can be done by generating a Fibonacci grid $(\tilde{\theta}, \tilde{\phi})$ [28] and rounding the points to the nearest simulated (or measured) ones (θ, ϕ). Then we generate the beamforming vectors pointing to these points. One option is to find the optimal or near-optimal codewords according to Algorithm 3. The second option is to first find the eigenvector of $\mathbf{M}(\theta, \phi)$ corresponding to the maximal eigenvalue and then obtain the beamforming vector by scaling the magnitude and quantizing the phase of the eigenvector to meet the per-element power and phase constraint. We will compare the performances of these two candidate codewords generation methods in Section VII.

B. CODEWORD SELECTION CRITERIA

Codeword selection is performed based on the performance optimization criterion which defines the utility function the algorithm tries to maximize.

A possible design goal is to maximize the mean gain, i.e. the average gain of the composite radiation pattern over a given spatial coverage region. In each step, the codeword that maximizes the improvement of the mean composite gain is selected, i.e.,

$$\mathbf{w}^\star = \underset{\mathbf{w} \in \mathcal{W}_d \setminus \mathcal{W}_c}{\operatorname{argmax}} \ \mathbb{E}_{(\theta, \phi) \in \mathcal{A}} \left[S \left(\mathcal{W}_c \cup \mathbf{w}, \theta, \phi \right) \right]. \quad (24)$$

where $\mathcal{A}$ is the coverage region of interest on the sphere.

Another option can be the maximization of gain value at one or more percentile points. For optimization with multiple percentile points, weighted average of percentile points of interest can be considered, i.e.,

$$\sum_i \beta_i F_S^{-1} \left(X_i \% \right), \quad (25)$$

where β_i is the weight. Single percentile point optimization is a special case with all but one percentile point set to be zero weight.

In the simulations shown in Fig. 2, the mean gain over the whole sphere is assumed as the selection criterion. Generally, the performance optimization criterion is a design choice which depends on the spherical coverage CDF requirements and link budget analysis, etc.

C. ALGORITHM STOPPING CONDITION

The algorithm stopping condition can be taken from the codebook requirements. If there is a limitation on the codebook size, then the algorithm stops picking new codeword once enough codewords are selected.

The Greedy algorithm can also generate codebooks with variable size. In this case, stopping conditions are based on the spherical coverage performance, i.e., the algorithm is terminated once a required spherical coverage has been reached. For example, the requirement could be the average gain over an angular region $\mathcal{A}$,

$$\mathbb{E}_{(\theta,\phi)\in\mathcal{A}}\left[S\left(\mathcal{W}_c,\theta,\phi\right)\right] > Y, \tag{26}$$

where Y is a threshold that can be assigned. Another example of stopping condition is related to the spherical coverage CDF requirement, where the selection stops when the gain value at one (or more) $X\%$-tile is larger than a threshold Y',

$$F_S^{-1}\left(X\%\right) > Y'. \tag{27}$$

VI. K-MEANS ALGORITHM

For the Greedy algorithm, care is needed to ensure that the candidate codewords sufficiently cover the whole sphere (or the angular region of interest). If the set of candidate beam codewords offered for selection does not cover certain directions well, the resulting codebook performance can be poor, for example, coverage holes at certain directions. In this section, we propose another algorithm called *K-Means*, which does not require careful constructions of the candidate codewords. As the name suggests, the core idea of this algorithm is based on the K-Means clustering, which is an unsupervised machine learning algorithm [40].

Given an initial set of K beams $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K\}$ and a set of N_p directions of interest

$$\mathcal{D} = \left\{(\theta_1,\phi_1), (\theta_2,\phi_2), \cdots, \left(\theta_{N_p},\phi_{N_p}\right)\right\}, \tag{28}$$

the algorithm proceeds by alternating between two steps:

1) *Assignment step*: Assign each direction to the beam, which provides the largest gain. Mathematically, this means partitioning the set of directions $\mathcal{D}$ into K subsets, denoted as $\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_K$. The set of directions $\mathcal{D}_k$ is served by the beam $\mathbf{w}_k$ and is defined as follows,

$$\mathcal{D}_k = \left\{(\theta,\phi)\,\middle|\,k = \operatorname*{argmax}_{1\le i\le K}\mathbf{w}_i^H\mathbf{M}(\theta,\phi)\mathbf{w}_i\right\}. \tag{29}$$

2) *Update step*: Optimize the beams to serve the directions in their associated subsets. This is done by solving the

following optimization problem for $1 \le k \le K$,

$$\mathbf{w}_k = \operatorname*{argmax}_{\mathbf{w}:\left(\sqrt{L}w_\ell\right)^{2^b}=1,\forall\ell} \mathbf{w}^H\left(\sum_{(\theta,\phi)\in\mathcal{D}_k}\mathbf{M}(\theta,\phi)\right)\mathbf{w}. \tag{30}$$

Problem (30) is similar to the optimization problem (18) discussed in Section III. Hence, Algorithm 3 is adopted to solve this problem.

The algorithm is terminated when the average gain of the composite pattern no longer improves or assignments no longer changes.

The complexity of the K-Means algorithm concentrates on the *Update step* where the Algorithm 3 is run for K times in each iteration. Overall, the K-Means has a complexity of $\mathcal{O}(L^3 KN_pN_I)$ where N_I is the number of iterations needed until convergence [41]. As found in our simulations, the K-Means algorithm converges very quickly and N_I is usually very small, i.e., less than 20.

Fig. 3 shows an example of the K-Means algorithm. The same linear 1×4 patch array considered in Fig. 2 is assumed. Each colored point in Fig. 3(a)-Fig. 3(b) represents one direction to cover. Note that for a uniform distribution of points on the sphere, there are less points around the polar regions than the equator region. The five different colors correspond to the five codewords. As seen in Fig. 3(a)-Fig. 3(b), the coverage regions change as the K-Means algorithm updates the codebook iteratively. Compared to the Greedy algorithm example in Fig. 2(i), there is no deep coverage hole in Fig. 3(f). There are some directions with gap as large as 4.5 dB. However, the directions fall within the back-of-the-panel regions, which has less gain as well as interest.

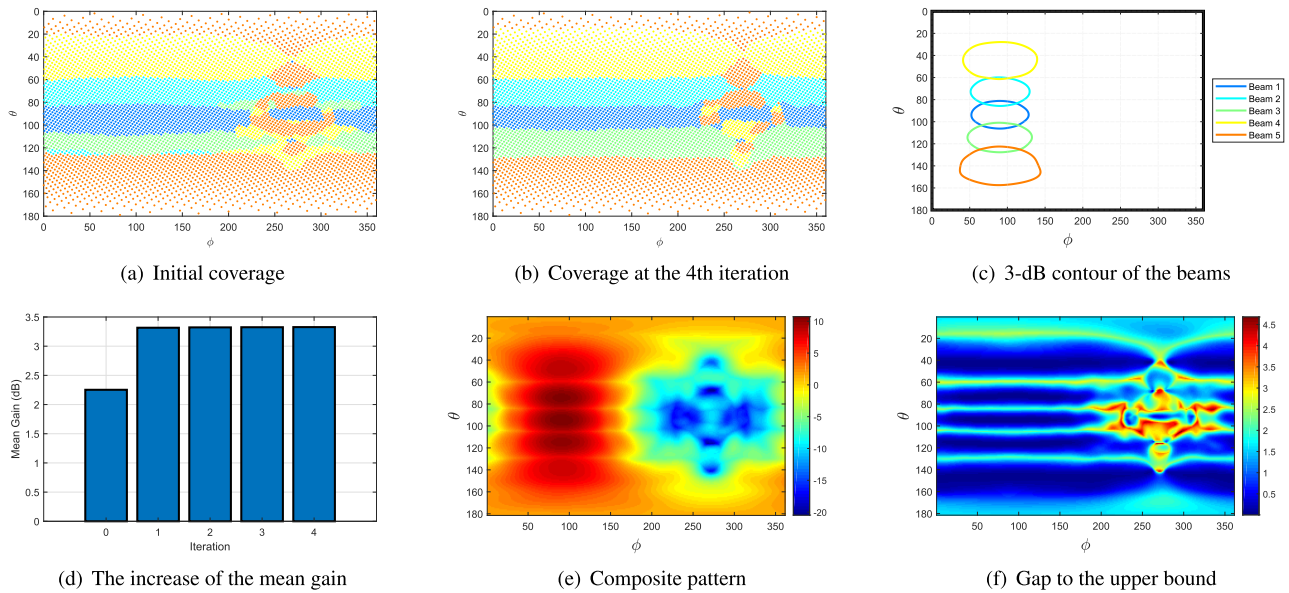## A. CONVERGENCE OF THE K-MEANS ALGORITHM

The proposed K-Means algorithm is guaranteed to converge. This can be seen as follows. In the first step, the algorithm finds the best beam for each direction. In other words, the best assignment for a given beam codebook is obtained. Hence, the average gain increases (or keeps same) in this step. In the second step, for the directions served by the same beam, the algorithm finds out an optimal (or local optimal) beam to maximize the average gain over these directions. The average gain increases (or keeps same) in the second step as well.

Since the mean gain is monotonically nondecreasing in each iteration, and there is an upper bound on the mean gain, we can conclude that the K-Means algorithm always converges.

In the above example, Fig. 3(d) shows the convergence of the mean gain of the codebook. It is seen that the algorithm converges quickly within 4 iterations.

B. INITIALIZATION OF THE K-MEANS ALGORITHM

For the initialization of the codebook $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K\}$, two options are considered.

(a) Initial coverage

(b) Coverage at the 4th iteration

(c) 3-dB contour of the beams

(d) The increase of the mean gain

(e) Composite pattern

(f) Gap to the upper bound

FIGURE 3. K-Means algorithm operation on a linear 1×4 patch antenna array with 5-bit phase shifters. The 'Uniform' initialization is used in this example.
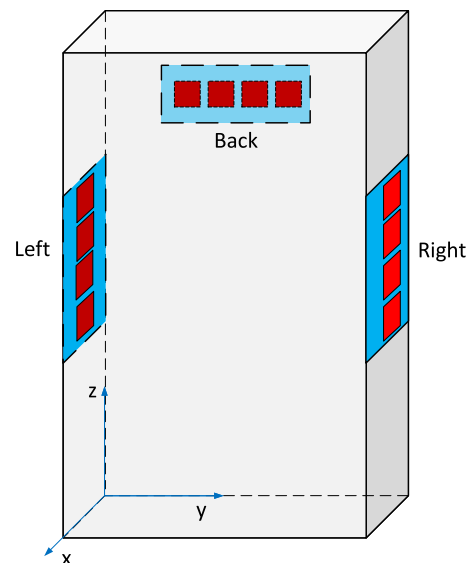
1) 'Greedy': The initial codebook is generated from the Greedy algorithm shown in Section V. In other words, we concatenate the two algorithms. We first run the Greedy algorithm and then take the output of the Greedy algorithm as the initialization of the K-Means algorithm.

2) 'Uniform': First generate K uniformly distributed points on the sphere or the required coverage region. Then compute the codewords by normalizing the magnitude and quantizing the phase of the maximal eigenvector of the $\mathbf{M}$ matrix at these directions. This procedure is similar to one of the methods of generating candidate codewords given in Section V-A, but a small number of codewords are generated. The idea underlying this option is to ensure that the initial codewords are pointing to different and well-separated directions.

The 'Uniform' initialization is employed in the example shown in Fig. 3. Although the mean gain of this initial codebook is not good (see the mean gain at the 0-th iteration in Fig. 3(d)), the mean gain increases substantially with a single iteration comprising an assignment step and an update step. We will compare these two initializations in Section VII.

VII. SIMULATION RESULTS

A. SIMULATION SETUP AND DATA GENERATION

In the simulation, we consider a terminal operating at 28 GHz with three antenna arrays, where the first is placed on the left edge, the second is placed at the right edge and the third is placed at the back of the terminal as shown in Fig. 4. All the three arrays are 1×4 linear patch antenna arrays with half-wavelength spacing. Assume that the terminal is placed
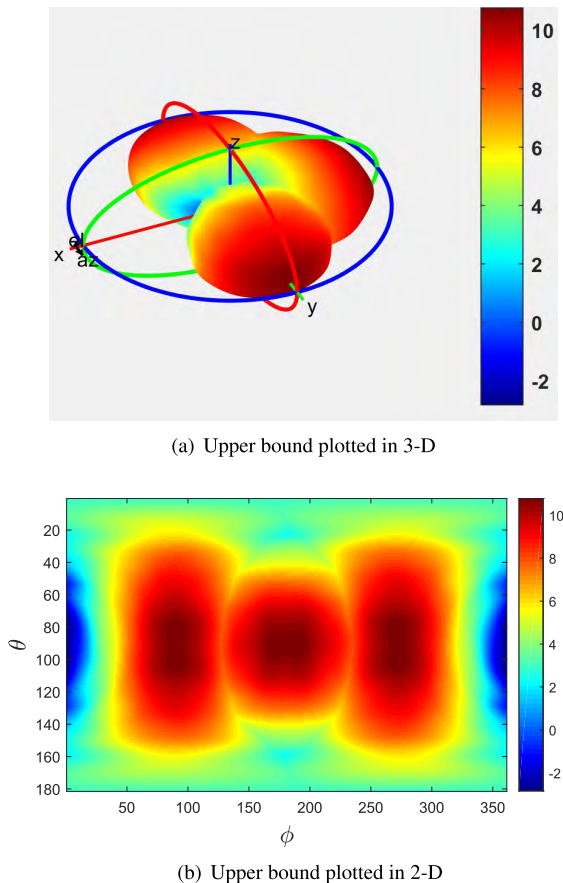

FIGURE 4. A terminal with three patch arrays on the left edge, right edge and the back, respectively.

vertically in the y-z plane with the front facing $+x$ direction. The three arrays are pointing to the $-y$, $+y$, $-x$ directions, respectively.

We assume that each antenna element is supplied with the same power, i.e. the per-element power constraint holds. The resolution of the phase shifters is assumed to be 5 bits. In addition, we assume that only one of the antenna arrays can be activated at a given time, which is a common practice.

The E-field data used in simulations is generated using finite-element electromagnetic simulator HFSS by Ansys. We assume the E-field data of each antenna element are available, i.e., $\mathbf{e}_\Theta(\theta, \phi)$, $\mathbf{e}_\Phi(\theta, \phi)$, in the form of discrete

(a) Upper bound plotted in 3-D

(b) Upper bound plotted in 2-D

FIGURE 5. Upper bound of the radiation pattern. The three panels are pointing to −*y*, +*y*, −*x* directions, respectively.

samples on a mesh grid, e.g. $(\theta, \phi) = [0° : q_\theta : 180°] \times [0° : q_\phi : 360°)$. We assume $q_\theta = q_\phi = 1°$ for results illustration in this paper; however it should be noted that this is not a necessary assumption for the algorithms, and other values of q can also be assumed, such as 5° or 15°.

Fig. 5(a) and Fig. 5(b) show the upper bound of the composite pattern in 3-D and 2-D, respectively. The three arrays cover the angular regions around $(\theta = 90°, \phi = 270°)$, $(\theta = 90°, \phi = 90°)$ and $(\theta = 90°, \phi = 180°)$, respectively. The mean and median gain of the upper bound is 7.56 dB and 7.41 dB, respectively.

B. PERFORMANCE OF THE PROPOSED ALGORITHMS

In this subsection, we compare the algorithms proposed in this paper. Depending on different initializations, we consider five implementations of the proposed algorithms as follows.

1) **Greedy(Eigen)**: Greedy algorithm where the candidate codewords are phase-quantized and magnitude-normalized eigenvectors corresponding to the maximal eigenvalues.
2) **Greedy(Iterative)**: Greedy algorithm where the candidate codewords are generated using Algorithm 3.
3) **K-Means(Greedy(Eigen))**: K-Means algorithm initialized by the 'Greedy(Eigen)' algorithm.

4) **K-Means(Greedy(Iterative))**: K-Means algorithm initialized by the 'Greedy(Iterative)' algorithm.
5) **K-Means(Uniform)**: K-Means algorithm where the initial codewords are beamforming to K uniformly distributed directions.

For the two implementations of the Greedy algorithm listed above, the selection criterion is assumed to be the mean gain over the whole sphere, which is aligned with the optimization metric of the K-Means implementations for the sake of a fair comparison. 363 candidate codewords pointing to quasi-uniformly distributed 363 directions are generated by either 'Eigen' or 'Iterative' approach. The angle separation of adjacent directions is around 10°. The Greedy algorithms stop selecting new codewords when the codebook size limitation is reached. For the K-Means algorithm, the beams are updated by Algorithm 3 where the number of randomization in the first step is chosen as $N_G = 1000$.

In Fig. 6(a), we compare the mean beamforming gain over the sphere. Our first observation is that the choice of the candidate codewords, i.e.,'Iterative' and 'Eigen', does not result in a significant performance difference. This can be explained by noting that, although an 'Iterative' codeword may be slightly better than the 'Eigen' codeword in a given direction, it may be worse than the 'Eigen' codeword when considering the average gain of the surrounding region of this direction. Our second observation is that there is nearly no performance difference across these implementations. Nevertheless, there may be meaningful performance difference for other antenna and terminal designs. Therefore, it is expected that the choice of the algorithm needs to be considered on a case-by-case basis. Finally, the mean gain increases with the codebook size and saturates when the codebook size is larger than 24. Actually, when the codebook size is 32, the mean gain is around 7.39 dB, which is very close to the mean gain of the upper bound, i.e., 7.56 dB.

In Fig. 6(b), we show the median gains produced by the algorithms since the 3GPP has defined the requirement of spherical coverage for handheld UE in terms of the median gain. Unlike the case with mean gains, there are more variations among different algorithms. This is reasonable since we set the mean gain as the common optimization metric and distributions with the same mean value could have very different median values. Furthermore, the difference in the median gains is very small when the codebook size is larger than 20, implying that all algorithms are converging to similar spherical coverage as the codebook size increases.

In our simulations, we find that the K-Means algorithm generally provides slightly better performance than the Greedy algorithm, but not in all cases. In particular, it is less likely to find coverage holes in the codebooks generated by K-Means algorithm than those by Greedy algorithm. However, the Greedy algorithm is much more flexible than K-Means algorithm. First, there are many possible choices about the utility function and stopping condition in the Greedy algorithm, while the K-Means has limited options. In this paper, we only consider the metric of mean gain for
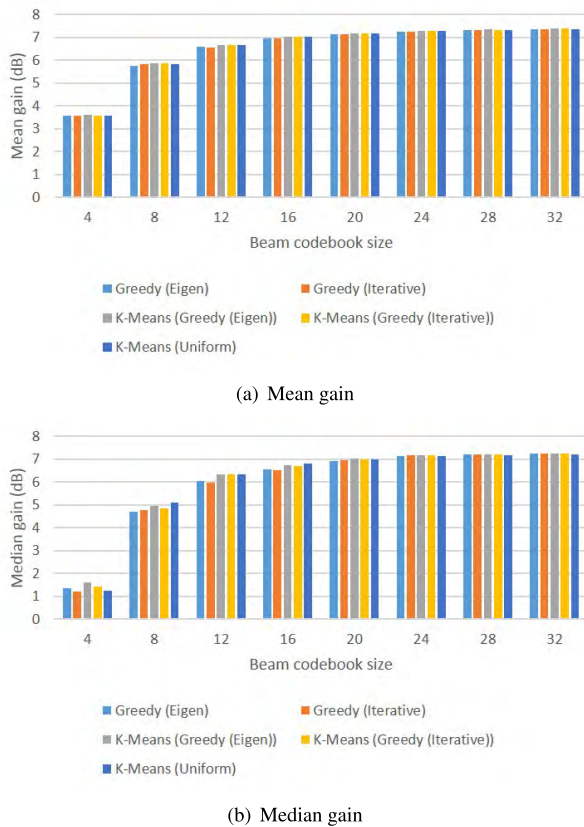
(a) Mean gain

(b) Median gain

FIGURE 6. The mean gain and median gain of the codebooks generated by different implementations.

TABLE 1. Comparison of the different codebooks in terms of mean gain and median gain. The three arrays are mounted at the left edge, right edge and back of the terminal.

Beamforming gain (dB)	Codebook size K				
	12	15	18	21	24
Mean Gain of Benchmark.	6.556	6.854	7.045	7.171	7.248
Mean Gain of 802.15.3c.	6.423	6.816	7.030	7.156	7.246
Mean Gain of Proposed.	**6.665**	**6.945**	**7.102**	**7.216**	**7.279**
Median gain of Benchmark.	6.305	**6.744**	**6.934**	7.047	7.120
Median gain of 802.15.3c.	5.804	6.362	6.745	6.964	7.128
Median gain of Proposed.	**6.322**	6.659	6.870	**7.118**	**7.171**

the K-Means algorithm. Second, the Greedy algorithm can generate variable-sized codebooks, whereas the codebook size has to be determined before running the K-Means algorithm. To sum up, the choice of the algorithm depends on the codebook requirements as well as the E-field data.

VIII. ADVANTAGES OF THE PROPOSED METHOD
Our proposed method can be easily configured to deal with different antenna setups and to meet a variety of requirements on the beam codebook. In this section, we show several such cases.

In this subsection, we compare the proposed method with two other designs. The K-Means(Greedy(Iterative)) algorithm is assumed here for the performance comparison purpose.

The benchmark beams are designed to point to certain directions. In this paper, we assume that a benchmark codebook for a single linear array consist of K' beams pointing to directions $\arccos\left(-1 + \frac{2k-1}{K'}\right)$, where $k = 1, 2, \ldots, K'$ [15], [16], [20]. For example, a beam codebook of size 4 consists of codewords pointing to 138.6°, 104.5°, 75.5°, 41.4° with respect to the array axis, respectively. The beam codewords are computed as,

$$w(\ell, k) = \frac{1}{\sqrt{L}} \exp\left(j\mathcal{Q}_b\left(\frac{2\pi d\ell}{\lambda}\left(-1 + \frac{2k-1}{K'}\right)\right)\right), \quad (31)$$

where the function $\mathcal{Q}_b(\cdot)$ quantizes the phase from $[0, 2\pi)$ to $\left\{0, \frac{2\pi}{2^b}, \cdots, (2^b - 1)\frac{2\pi}{2^b}\right\}$.

In additional to the benchmark method, the 802.15.3c codebook is also included here for comparison. The original 802.15.3c codebooks are designed for 2-bit phase-shifters. To have a fair comparison, we adopt its generalization to 2^b-phase codebook shown as follows [42],

$$w(\ell, k)$$
$$= \frac{1}{\sqrt{L}} \exp\left(j\frac{2\pi}{2^b}\lfloor\frac{(\ell-1) \times \mod\left(k - 1 + \frac{K'}{2}, K'\right)}{K'/2^b}\rfloor\right),$$
$$(32)$$

for $1 \leq \ell \leq L$, $1 \leq k \leq K'$, where $\lfloor x \rfloor$ rounds x to the nearest integer less than or equal to x. The 802.15.3c codebook basically consists of K' codewords having (approximately) progressive phases.

A. JOINT DESIGN OF MULTI-ARRAY CODEBOOK
When there are multiple arrays mounted on the terminal, such as the three patch arrays as shown in Fig. 4, a conventional design may assume the same set of codewords for each array [27]. This assumption restricts the codebook size to be an integer multiples of the number of arrays. In contrast, there is no limitation on the choice of codebook size in our proposed algorithm, as seen in Fig. 6. More importantly, the conventional design does not take into account the possible overlapping of the coverage regions of different arrays and therefore the generated codebook may include codewords pointing to similar directions.

As shown in Table 1, the proposed codebook is better than benchmark and 802.15.3c codebooks in terms of the mean and median gains, in most of the cases. The advantage of the proposed algorithm is large especially when the codebook size is small, i.e, $K = 12$. When the codebook size increases, the performance of all the algorithms approach the upper bound and thus are similar to each other. The advantage of the proposed algorithm is clearly seen in Fig. 7 where the 3-dB beam contours are shown. In the benchmark codebook, beam 1 and beam 4 associated with the back array is pointing to the regions which are also partially covered by the beams from the other two arrays, i.e. beam 10-11 and beam 6-7. Similarly for the 802.15.3c codebook, beam 1 associated with the back array is pointing to the regions which are also
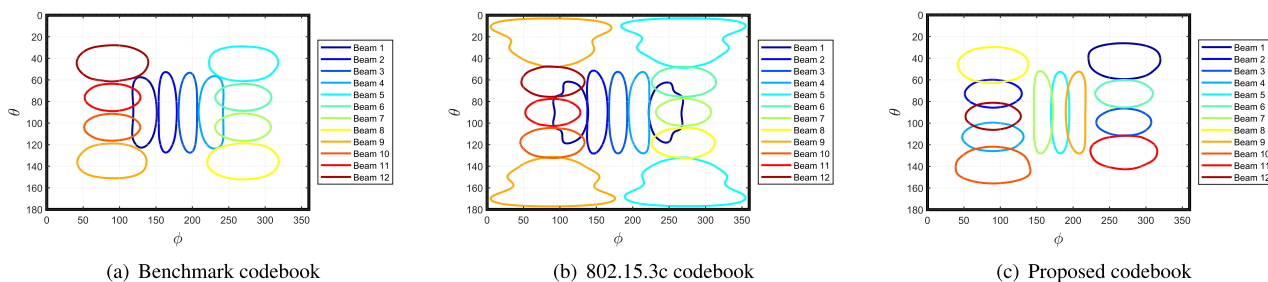
(a) Benchmark codebook

(b) 802.15.3c codebook

(c) Proposed codebook

FIGURE 7. The 3-dB beam contours of the benchmark and proposed codebooks when $K = 12$. The three arrays are mounted at the left edge, right edge and back of the terminal.
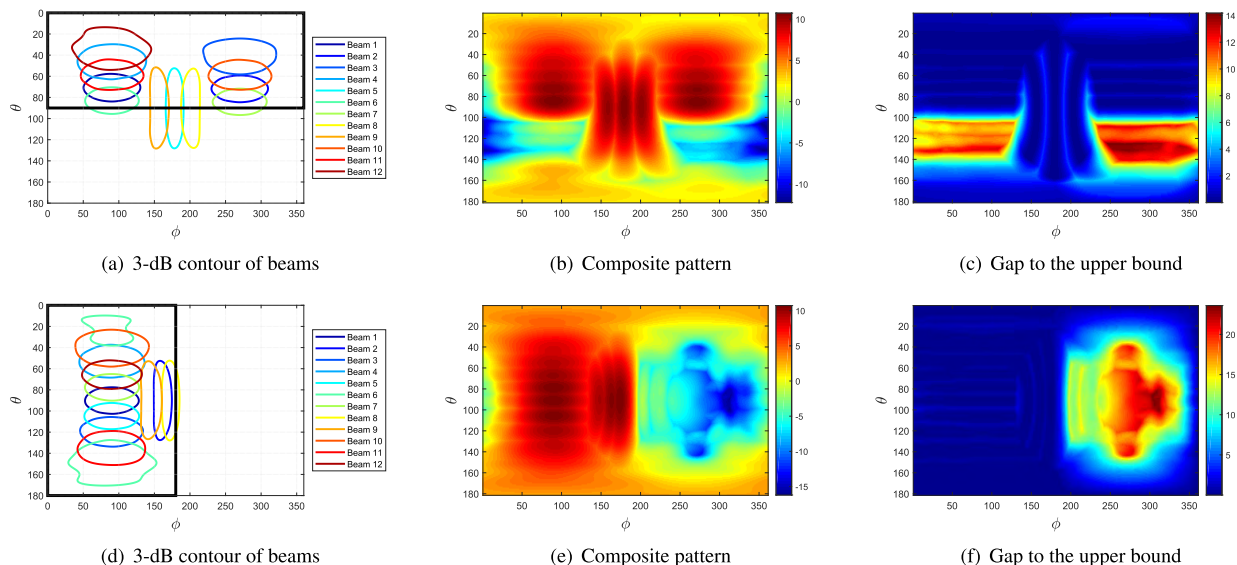

(a) 3-dB contour of beams

(b) Composite pattern

(c) Gap to the upper bound

(d) 3-dB contour of beams

(e) Composite pattern

(f) Gap to the upper bound

FIGURE 8. The beam codebook adaptation to the coverage region requirements. In Fig. 8(a)-Fig. 8(c), the required coverage region is $(\theta, \phi) \in [0°, 90°] \times [0°, 360°]$ whereas in Fig. 8(d)-Fig. 8(f), the required coverage region is $(\theta, \phi) \in [0°, 180°] \times [0°, 180°]$.

(a) Benchmark codebook

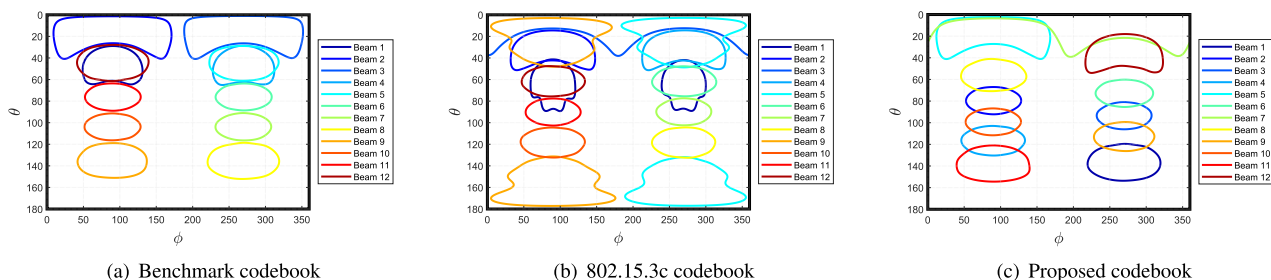(b) 802.15.3c codebook

(c) Proposed codebook

FIGURE 9. The 3-dB beam contours of the benchmark and proposed codebooks when $K = 12$. The three arrays are mounted at the left, right and top edge of the terminal.

covered by the beams from the other two arrays, i.e. beam 6-8 and beam 10-12. By contrast, the proposed codebook displays a much better coordination among different arrays by automatically allocating different number of beams to the arrays and avoiding the beam overlapping.

B. FLEXIBLE ADAPTATION TO REQUIRED COVERAGE REGION

In certain scenarios, the beam codebook is required to cover a part of the sphere rather than the whole sphere. For instance, when the user is holding the phone next to the head to make a call, the phone should not beam towards user's head, because of the high blockage loss of human body to mmWave signals and the radio frequency exposure compliance [43]–[45].

Our proposed method is capable of adapting to varying coverage region requirement. In the Greedy algorithm, we can define the utility function U over the region of interest instead of the whole sphere. For example, the mean gain over a region could be optimized as shown in (24). For the K-Means algorithm, we can filter the set of directions $\mathcal{D}$ to keep only the directions within the required coverage region.
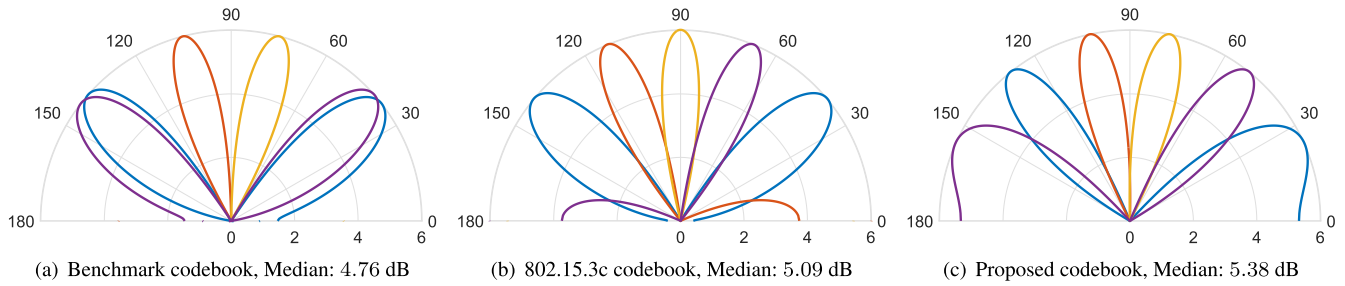
(a) Benchmark codebook, Median: 4.76 dB (b) 802.15.3c codebook, Median: 5.09 dB (c) Proposed codebook, Median: 5.38 dB

FIGURE 10. The beam patterns of a 1×4 uniform linear array where $d = 0.65\lambda$, $K = 4$.

TABLE 2. Comparison of the different codebooks in terms of mean gain and median gain. The three arrays are mounted at the left edge, right edge and top edge of the terminal.

Beamforming gain (dB)	Codebook size K				
	12	15	18	21	24
Mean Gain of Benchmark.	6.320	6.646	6.847	6.983	7.062
Mean Gain of 802.15.3c.	6.277	6.649	6.842	6.963	7.052
Mean Gain of Proposed.	**6.635**	**6.887**	**7.016**	**7.088**	**7.150**
Median gain of Benchmark.	5.921	6.334	6.584	6.733	6.768
Median gain of 802.15.3c.	5.735	6.133	6.376	6.592	6.748
Median gain of Proposed.	**6.131**	**6.551**	**6.768**	**6.839**	**6.885**

FIGURE 11. CDF curves of the benchmark, 802.15.3c and the proposed codebooks. It simulated a 1×4 uniform linear array with $d = 0.65\lambda$, $K = 4$.

Fig. 8 illustrates two cases with required coverage region being $(\theta, \phi) \in [0°, 90°] \times [0° 360°)$ and $(\theta, \phi) \in [0°, 180°] \times [0° 180°]$, respectively. The regions are highlighted by black boxes drawn in Fig. 8(a) and Fig. 8(d). As seen in the figures, the resulting beam codewords are naturally concentrated in the required region. As a result, the composite patterns in these two cases have a less than 2 dB gap to the upper bound in the required region in contrast to a more than 10 dB gap out of the required region. When the coverage region is the half-sphere $0° \leq \phi \leq 180°$, we find that the array on the left edge of the phone is turned off automatically as shown Fig. 8(d), since the required coverage region is at the back of it.

C. STRAIGHTFORWARD EXTENSION TO DIFFERENT MODULE PLACEMENT

Our proposed method can deal with any kind of antenna type and placement on the terminal. Here, we show an example with the same antenna arrays as Fig. 4 but a different placement. Specifically, we assume the arrays are distributed on the left edge, right edge and top edge. As shown in Table 2, the proposed codebook provides better mean and median gains than the benchmark and 802.15.3c codebooks in all the cases. Fig. 9 compares the 3-dB beam contours when $K = 12$. We find that the beams of benchmark and 802.15.3c codebook are largely overlapping in the upper half-sphere $(0° \leq \theta \leq 90°)$, which explains why their spherical coverage is worse than the proposed codebook which maintains a much better angle separation among 12 beams.

Last, comparing Table 1 and Table 2, we find that the first module placement results in better beamforming gains that the second placement. In other words, it is better to put the third array on the back than on the top edge, if the optimization target is the spherical coverage.
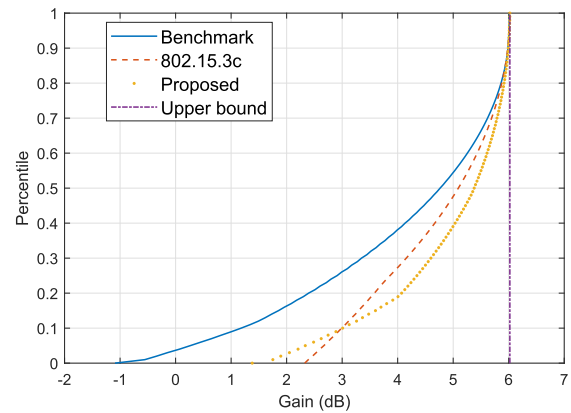
IX. FURTHER COMPARISONS WITH SIMPLIFIED E-FIELD DATA

To provide a comprehensive comparison with conventional model-based approach and illustrate the effectiveness of the proposed method in general cases, we perform more simulations based on simplified E-field response data besides HFSS data.
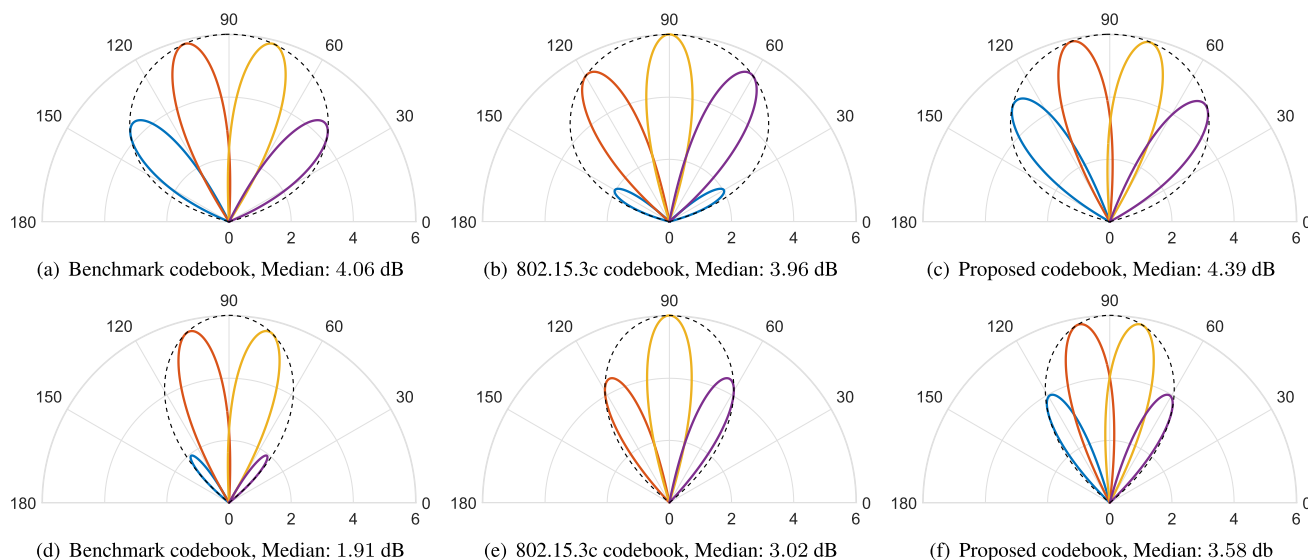
The uniform antenna array (ULA) with exactly same antenna element is assumed in the comparison. The E-field data are generated for the angular directions uniformly distributions over the spatial frequency, i.e., $\theta = \arccos(x)$, where $x = -1, -(a-1)/a, \ldots, (a-1)/a, 1$. In this section where $1 \times L$ ULA is assumed, we choose $a = 30 L$. The E-field data at the angle θ is,

$$\mathbf{e}_\Theta(\theta) = \sqrt{p(\theta)} \exp\left(j\frac{2\pi \cos(\theta)}{d}[0, 1, \ldots, L-1]^T\right), \quad (33)$$

and $\mathbf{e}_\Phi(\theta) = \mathbf{0}$, where θ is angle with respect to the axis of the linear array, $p(\theta)$ is the element radiation pattern.

The K-Means algorithm is used for the comparison, and the K-Means algorithm is initialized by the benchmark codebook. In addition, 5-bit analog beamforming is assumed.

In the ideal case of omni-directional antenna and half-wavelength spacing, the codebook design has been well studied and our proposed method does not bring further improvements. However, our proposed method does bring large gain when designing a codebook for a practical antenna array where the ideal assumptions do not hold.

FIGURE 12. The beam patterns of a 1 × 4 uniform linear array where $d = 0.5\lambda$, $K = 4$. The element pattern is $p(\theta) = \sin(\theta)$ in (a)-(c), while the element pattern is $p(\theta) = \sin^3(\theta)$ in (d)-(f). The dashed envelope stands for the upper bound.

A. IRREGULAR ANTENNA SPACING

We first consider a scenario where the antenna array is not half-wavelength, which results from form-factor constraints or the multi-frequency bands the array has to support. In particular, a 1×4 ULA with $d = 0.65\lambda$ is simulated. The number 0.65 is chosen by assuming that the antenna array has the antenna spacing of 5 mm (i.e., half wavelength at 30 GHz), and operates at the 39 GHz band. We want to generate a codebook of 4 beams.

Fig. 10 illustrates the radiation pattern of the 4 beams of each codebook. As seen in Fig. 10, there are strong side lobes when the main lobe is pointing away from the broadside direction. The proposed method can adjust the beamforming direction of the beams to fully utilize the side lobes to achieve a better spherical coverage. It is worthy to note that the adjustment is done automatically by the proposed algorithm based on E-field response data. By contrast, the benchmark and 802.15.3c codebooks do not adapt well to the spacing change. The CDF curves of these three codebook is illustrated in Fig. 11. It is clear that the proposed codebook shows much better spherical coverage than the other two codebooks. The median gain values are 4.76, 5.09, 5.38 dB, respectively.

Last, the side lobes are also used in [10], [11] to achieve a good coverage. However, their handcrafted approach requires careful design of the beamforming weights and cannot apply to arbitrary antenna spacing.

B. THE DIRECTIONAL ANTENNA RADIATION PATTERN

Now we consider another case where the antenna is directional. A simple model of directional radiation pattern is as follows.

$$p(\theta) = \sin^q(\theta), \qquad (34)$$

where q controls the directionality of the radiation pattern. A 1×4 ULA with half-wavelength spacing is considered here.

Fig. 12 shows the comparison of the beam patterns when $q = 1$ in (a)-(c) and $q = 3$ in (d)-(f). The dashed envelope represents the upper bound. As the parameter q becomes larger, the element pattern as well as the upper bound becomes more directional. As seen in the figure, as q increases, two out of the four beams in the benchmark codebook have diminishing gains and provide negligible contribution to the spherical coverage. Similarly, one of the 802.15.3c codewords has relatively small gain. By contrast, the proposed codebook is capable of tweaking the beam direction automatically. The beams are moving towards the broadside direction as the antenna element becomes more directional. The median gain of each codebook is listed below the figures. The proposed codebooks have the largest median gains.

X. CONCLUSION

In this paper, we have formulated the beam codebook design problem to enhance the spherical coverage of the mmWave terminals. The codebooks designed based on the isotropic antenna assumptions will not work well for mmWave terminals, due to the inherent directional radiation pattern of the mmWave antenna element and the impact from housing components of the terminals, such as coupling, blockage, absorption, reflection, etc. We proposed a novel approach to automatically design the beam codebook solely based on the E-field response of each element. First, a flexible Greedy algorithm is proposed to choose a subset of the candidate codeword pool to form the final codebook according to any given criterion. Second, a machine learning based iterative algorithm is proposed to generate the codebook. Through simulations, we find out that the composite radiation pattern of the proposed codebook is better than the benchmark and 802.15.3c codebook. Actually, the performance of the proposed beam codebook is shown to be able to approach the upper bound as codebook size increases. Furthermore,

the proposed data-driven method can be used for any kind of array layouts, placement and antenna type. It is a very generic method capable of designing good codebooks for a wide range of practical scenarios where the conventional model-based method does not work well.

Note that the proposed method depends closely on the E-field data. There are several possible factors distorting the far-filed E-field response. For example, the protection case of the phone and the hand grip of the users [26], [46]. We model the hand grip impact on the E-field response and propose an adaptive beam codebook generation method in [47].

Last, even without the distortions by hand grip, the E-field response data from simulations or measurements may be different from the true response. For instance, there may be deviations between the antenna and phone model used in the simulations and the manufactured ones. The measurement data may also be inaccurate because of the heating of the phone in the measurement process. A future direction to improve the proposed method is to design a robust beam codebook taking into account these deviations.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[3] V. Raghavan, A. Partyka, A. Sampath, S. Subramanian, O. H. Koymen, K. Ravid, J. Cezanne, K. Mukkavilli, and J. Li, "Millimeter-wave MIMO prototype: Measurements and experimental results," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 202–209, Jan. 2018.

[4] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 63–69, Sep. 2014.

[5] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.

[6] W. Hong, Z. H. Jiang, C. Yu, J. Zhou, P. Chen, Z. Yu, H. Zhang, B. Yang, X. Pang, M. Jiang, Y. Cheng, M. K. T. Al-Nuaimi, Y. Zhang, J. Chen, and S. He, "Multibeam antenna technologies for 5G wireless communications," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6231–6249, Dec. 2017.

[7] *User Equipment (UE) Radio Transmission and Reception; Part 2: Range 2 Standalone (E-UTRA)*, document TS 38.101-2, Version 15.3.0, 3GPP, Sep. 2018.

[8] *IEEE Standard for Information Technology Local And Metropolitan Area Networks Specific Requirements Part 15.3: Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension*, IEEE Stdandard 802.15.3c-2009 (Amendment to IEEE Std 802.15.3-2003) Oct. 2009, pp. 1–200.

[9] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. A. Rahman, R. Funada, F. Kojima, I. Lakkis, H. Harada, and S. Kato, "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.

[10] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proc. 6th Int. ICST Conf. Commun. Netw. China (CHINACOM)*, Aug. 2011, pp. 361–365.

[11] W. Feng, Z. Xiao, D. Jin, and L. Zeng, "Circular-antenna-array-based codebook design and training method for 60 GHz beamforming," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 4140–4145.

[12] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[13] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[14] T. He and Z. Xiao, "Suboptimal beam search algorithm and codebook design for millimeter-wave communications," *Mobile Netw. Appl.*, vol. 20, no. 1, pp. 86–97, Feb. 2015.

[15] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[16] Z. Xiao, H. Dong, L. Bai, P. Xia, and X.-G. Xia, "Enhanced channel estimation and codebook design for millimeter-wave communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9393–9405, Oct. 2018.

[17] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X. G. Xia, "3D beamforming for flexible coverage in millimeter-wave UAV communications," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 837–840, Jun. 2019.

[18] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, Nov. 2017.

[19] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "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.

[20] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.

[21] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[22] J. Wang, Z. Lan, C.-S. Sum, C.-W. Pyo, J. Gao, T. Baykas, A. Rahman, R. Funada, F. Kojima, I. Lakkis, H. Harada, and S. Kato, "Beamforming codebook design and performance evaluation for 60GHz wideband WPANs," in *Proc. IEEE 70th Veh. Technol. Conf. Fall*, Sep. 2009, pp. 1–6.

[23] A. Maltsev, *Channel Models for 60 GHz WLAN Systems*, IEEE Standard 802.11ad, 2010.

[24] J. Mao, Z. Gao, Y. Wu, and M.-S. Alouini, "Over-sampling codebook-based hybrid minimum sum-mean-square-error precoding for millimeter-wave 3D-MIMO," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 938–941, Dec. 2018.

[25] C. A. Balanis, *Antenna Theory and Design*. Hoboken, NJ, USA: Wiley, 2012.

[26] W. Hong, K.-H. Baek, and S. Ko, "Millimeter-wave 5G antennas for smartphones: Overview and experimental demonstration," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6250–6261, Dec. 2017.

[27] V. Raghavan, M.-L. Chi, M. A. Tassoudji, O. H. Koymen, and J. Li, "Antenna placement and performance tradeoffs with hand blockage in millimeter wave systems," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3082–3096, Apr. 2019.

[28] R. Swinbank and R. J. Purser, "Fibonacci grids: A novel approach to global modelling," *Quart. J. Roy. Meteorol. Soc.*, vol. 132, no. 619, pp. 1769–1793, 2006.

[29] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 3779–3784.

[30] S. Zhang and Y. Huang, "Complex quadratic optimization and semidefinite programming," *SIAM J. Optim.*, vol. 16, no. 3, pp. 871–890, 2006.

[31] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[32] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM J. Optim.*, vol. 6, no. 2, pp. 342–361, 1996.

[33] Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg, "Row by row methods for semidefinite programming," *Ind. Eng.*, pp. 1–21, Apr. 2009.

[34] H. Wai, W. Ma, and A. M. So, "Cheap semidefinite relaxation MIMO detection using row-by-row block coordinate descent," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 3256–3259.

[35] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[36] A. M.-C. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Math. Program.*, vol. 110, no. 1, pp. 93–110, 2007.

[37] G. Cui, X. Yu, G. Foglia, Y. Huang, and J. Li, "Quadratic optimization with similarity constraint for unimodular sequence synthesis," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4756–4769, Sep. 2017.

[38] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221–1234, Mar. 2014.

[39] A. T. Kyrillidis and G. N. Karystinos, "Rank-deficient quadratic-form maximization over M-phase alphabet: Polynomial-complexity solvability and algorithmic developments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3856–3859.

[40] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1. 1967, pp. 281–297.

[41] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.

[42] Z. Weixia, G. Chao, D. Guanglong, W. Zhenyu, and G. Ying, "A new codebook design scheme for fast beam searching in millimeter-wave communications," *China Commun.*, vol. 11, no. 6, pp. 12–22, Jun. 2014.

[43] *Code Federal Regulations Title 47 Part 2.1093*, Federal Commun. Commission, Washington, DC, USA, 2019.

[44] *Code Federal Regulations Title 47 Part 1.1310*, Federal Commun. Commission, Washington, DC, USA, 2013.

[45] A. Ahlbom, U. Bergqvist, J. H. Bernhardt, J. P. Cesarini, L. A. Court, M. Grandolfo, M. Hietanen, A. F. McKinlay, M. H. Repacholi, D. H. Sliney, J. A. J. Stolwijk, M. L. Swicord, L. D. Szabo, M. Taki, T. S. Tenforde, H. P. Jammet, and R. Matthes, "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)," *Health Phys.*, vol. 74, no. 4, pp. 494–522, Apr. 1998.

[46] V. Raghavan, L. Akhoondzadeh-Asl, V. Podshivalov, J. Hulten, M. A. Tassoudji, O. H. Koymen, A. Sampath, and J. Li, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3010–3024, Jul. 2019.

[47] A. Alammouri, J. Mo, B. L. Ng, J. C. Zhang, and J. G. Andrews, "Hand grip impact on 5G mmWave mobile devices," *IEEE Access*, vol. 7, pp. 60532–60544, 2019.

SANGHYUN CHANG received the B.S. degree in electrical engineering from Seoul National University, in 2000, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), in 2002 and 2007, respectively.

He was with the California Institute of Technology as a Postdoctoral Scholar from 2007 to 2011, jointly with USC as a Research Associate from 2009 to 2011. Since 2011, he has been with Samsung Electronics Company Ltd. He worked on next generation Wi-Fi standardization, and served as a Vice Chair in IEEE 802.11 TGay from 2012 to 2015. He is currently a Principal Engineer on flagship 5G smartphone product Research and Development. His research interests include wireless communications, ranging, and radar system design.

PENGDA HUANG (M'13) received the Ph.D. degree in electrical engineering from Southern Methodist University.

He is currently with Samsung Research America, Plano, TX, USA. His research interests include signal processing in spectrum sharing networks and source coding.

MANDAR N. KULKARNI (S'13) received the B.Tech. degree in electronics and communications engineering from IIT Guwahati, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from The University of Texas at Austin, in 2015 and 2018, respectively.

He has held internship positions at Nokia Bell Labs, Murray Hill, NJ, USA, in 2016; Nokia Networks, Arlington Heights, IL, USA, in 2014, 2015, and 2017; Technical University, Berlin, Germany, in 2012; and the Indian Institute of Science, Bangalore, India, in 2011. He is currently a Senior Research Engineer with Samsung Research America. He is broadly interested in envisioning how wireless networks should evolve in the future, and try to make suggestions on design choices based on elaborate simulations/experiments and simplified mathematical models to explain the observations.

Dr. Kulkarni received the President of India Gold Medal in 2013, George J. Heuer, Jr., Fellowship at UT Austin in 2014, and Leonard G. Abraham Prize in 2018.

JIANHUA MO (S'12–M'17) received the B.S. and M.S. degrees from Shanghai Jiao Tong University, in 2010 and 2013, respectively, and the Ph.D. degree from The University of Texas at Austin, in 2017, all in electrical engineering.

He is currently a Staff Engineer with Samsung Research America, Plano, TX, USA. His research interests include physical layer security and MIMO communications with low resolution ADCs. He is currently involved in the development of mmWave wireless backhaul transmission and 5G UE beam codebook.

Dr. Mo's awards and honors include the Heinrich Hertz Award for Best Communications Letter in 2013, Stephen O. Rice Prize of 2019, an Exemplary Reviewer of the IEEE Wireless Communications Letters, in 2012 and of the IEEE Communications Letters, in 2015, and Finalist for Qualcomm Innovation Fellowship, 2014.

BOON LOONG NG (S'03–M'08) received the B.E. degree in electrical and electronic and the Ph.D. degree in engineering from the University of Melbourne, Australia, in 2001 and 2007, respectively.

He is currently the Research Director of Samsung Research America—Standards & Mobility Innovation (SMI) Lab, Plano, TX, USA. He had contributed to 3GPP RAN L1/L2 standardizations of LTE, LTE-A, LTE-A Pro and 5G NR technologies from the period of 2008 to 2018. He holds more than 60 USPTO-granted patents on LTE/LTE-A/LTE-A Pro/5G and more than 100 patent applications globally. Since 2018, he has been leading a R&D Team that develop system and algorithm design solutions for commercial 5G and Wi-Fi technologies.
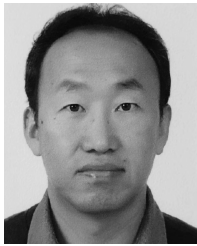
AHMAD ALAMMOURI (S'11) received the B.Sc. degree (Hons.) in electrical engineering from the University of Jordan, Amman, Jordan, in 2014, and the M.Sc. degree in electrical engineering from the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The University of Texas at Austin, TX, USA.

He has held summer internships at Samsung Research America, Richardson, TX, in 2017 and 2018. His research interests include statistical modeling and performance analysis of wireless networks.

JIANZHONG CHARLIE ZHANG (S'00–M'03–SM'09–F'16) received the Ph.D. degree from the University of Wisconsin–Madison, Madison. From 2009 to 2013, he served as the Vice Chairman of the 3GPP RAN1 working group and led development of LTE and LTE-Advanced technologies such as 3D channel modeling, UL-MIMO, CoMP, and carrier aggregation for TD-LTE. He is currently the Vice President and the Head of the Standards and Mobility Innovation Lab, Samsung Research America, where he leads research, prototyping, and standards for 5G and future multimedia networks.

JEONGHEUM LEE (S'01–M'04) received the B.S., M.S., and Ph.D. degrees from Hanyang University, Seoul, South Korea, in 1997, 1999, and 2003, respectively.

Since 2003, he has been with Samsung Electronics Company Ltd., Suwon, South Korea, where he is currently a Principal Research Engineer. From 2003 to 2006, he was engaged in the development of a digital IF/RF transceiver card and crest factor reduction algorithms for code division multiple access (CDMA) base transceiver stations. Since March 2006, he has been with the research and development of smartphone, tablets and other innovative mobile electronic devices. From 2017 to 2018, he participated in 3GPP Radio Access Network Working Group 4 as a Samsung Delegate and led the standardization of 5G millimeter-wave UE RF transmit power requirement. His current research interests include array antenna design, beamforming, 5G mobile communications system design, computational electromagnetics, electromagnetic optimization, microwave and millimeter-wave circuit design, film bulk acoustic wave resonator (FBAR) filter analysis and design, and mobile communication system design.

WON-JOON CHOI received the B.S. degree in electrical engineering from Seoul National University, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, where he conducted intensive research in wireless communication systems. He has several patents issued or pending, and many standards contributions and journal and conference publications.

He joined the Mobile Communication Business Unit, Samsung Electronics Company Ltd., as the Senior Vice President, in March 2016. He is currently the Head of Flagship Product R&D Team, and responsible for Samsung's flagship mobile device development and 5G technologies. He held various leadership positions at Qualcomm, Amicus Wireless, and Atheros Communications, where he had developed wireless SoCs and devices for LTE/LTE-Advanced, CDMA2000, WiMAX, and WiFi. He has more than 20 years of experience in wireless communication technologies, spanning core research, standards, wireless SoCs, and mobile devices.

• • •