Received 5 July 2021; accepted 27 July 2021. Date of publication 5 August 2021; date of current version 11 August 2021.

Digital Object Identifier 10.1109/OJCOMS.2021.3101747

Energy-Efficient MU-Massive-MIMO Hybrid Precoder Design: Low-Resolution Phase Shifters and Digital-to-Analog Converters for 2D Antenna Array Structures

MOBEEN MAHMOOD (Graduate Student Member, IEEE), ASIL KOC (Graduate Student Member, IEEE), AND THO LE-NGOC (Life Fellow, IEEE)

Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0G4, Canada

CORRESPONDING AUTHOR: M. MAHMOOD (e-mail: mobeen.mahmood@mail.mcgill.ca)

This work was supported in part by Huawei Technologies Canada and in part by the Natural Sciences and Engineering Research Council of Canada.

ABSTRACT This paper investigates a multi-user massive multiple-input multiple-output (MU-mMIMO) hybrid precoding (HP) scheme using low-resolution phase shifters (PSs) and digital-to-analog converters (DACs). The proposed HP approach involves two stages: RF beamforming based on the slowly time-varying channel second-order correlation matrix, and baseband MU precoding based on the instantaneous effective baseband channel to mitigate MU-interference by a regularized zero-forcing (RZF) technique. We consider three HP design architectures: (i) HP using full-resolution PSs and DACs, with a baseband transfer block for constant-modulus RF beamformer, (ii) HP using b-bit PSs and full-resolution DACs, with an orthogonal matching pursuit (OMP) based algorithm that can approach the optimal unconstrained RF beamformer, and (iii) HP using b-bit PSs and q-bit DACs, taking into account also DAC quantization noise. Illustrative results show that the proposed HP schemes with low-resolution PSs can approach the sum-rate of full-resolution PSs by using only 2-bit PSs, while offering higher energy efficiency. Furthermore, a study of sum-rate results for various PS and DAC quantization levels reveals that HP can achieve near-optimal performance with only 2-bit PSs and 5-bit DACs. Moreover, a comparison of the different array configurations, namely, uniform linear array (ULA), uniform circular array (UCA), uniform rectangular array (URA), and concentric circular array (CCA), indicates that URA and CCA outperform UCA and ULA in terms of spectral and energy efficiencies.

INDEX TERMS Massive multiple-input multiple-output (mMIMO), hybrid precoding, energy efficiency, low-resolution digital-to-analog converters (DACs), low-resolution phase shifters (PSs), uniform linear array (ULA), uniform circular array (UCA), uniform rectangular array (URA), concentric circular array (CCA).

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) refers to utilizing multiple antennas at base station (BS) for improving data rates through spatial multiplexing in single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO) operation modes [1]. Since user equipment cannot have a large number of antennas due to its limited size, MU-MIMO provides much better advantages than SU-MIMO; however, MU-MIMO is not a scalable technology. To meet the growing demand for increased throughput with limited resources such as bandwidth, massive MIMO (mMIMO) is one of the core technologies for next generation wireless communication systems [2]–[4]. Massive MIMO (also known as large-scale antenna systems, very large MIMO, hyper MIMO) uses a large number of antennas at BS, which significantly improves the spectral efficiency, simplifies the signal processing, and enhance the energy efficiency by directing the beams selectively to the users. One of the key features of mMIMO is the reduction of interference at the BS, commonly known as *precoding*, which

plays an important role in the reliable downlink transmission as it involves the encoding of the signal at the transmitter side. A comparison of various linear precoding techniques is presented in [5], which relates to the design of single-stage fully-digital precoder (FDP) where the number of costly and energy-intensive radio frequency (RF) chains is equal to the number of BS antennas. Hence, the use of single-stage FDP in mMIMO creates two challenging problems: (i) increased expense and energy consumption due to the use of large number of RF chains, and (ii) the use of significant amount of spectral resources and increased channel state information (CSI) overhead. Hybrid precoding (HP) has been explored as a way forward to solve these problems [6].

HP is a two-stage precoder consisting of an analog RF beamforming stage and a digital baseband precoding stage, and with much-reduced system complexity/hardware cost, it can achieve the performance close to FDP which requires full instantaneous CSI. In HP, large-dimensional processing is carried out by phase shifters (PSs) at the transceiver RF front-end, followed by low-dimensional processing at the baseband level. A few RF chains and digital-to-analog converters/analog-to-digital converters (DACs/ADCs) connect the RF and the baseband stages. Compared to FDP, the number of RF chains in HP varies from the number of transmission streams (or equivalently the number of single-antenna users to be served) and the number of antennas at the BS. Furthermore, at the RF-stage, two commonly used methods of processing input paramters are: (i) using fast time-varying CSI [7]–[13], and (ii) using slowly time-varying instantaneous CSI [14]–[22]. Most of the HP designs have assumed the use of full-resolution PSs and DACs to achieve a satisfactory performance close to FDP structures. However, because of the increased complexity and costs, implementing ∞-bit PSs and DACs/ADCs is impractical. The recent studies show that there has been an increased interest in the use of small, power-efficient, and inexpensive devices for hardware efficient transceivers capable of beamforming. Therefore, the use of low-resolution PSs [23]–[31] and DACs/ADCs [32]–[37] allows a more viable HP to be designed.

A. PRIOR WORKS

1) HYBRID PRECODING TECHNIQUES

In [7], a low complexity HP technique is proposed, which gives the sum-rate close to FDP by reducing the number of RF chains but requires full instantaneous CSI. Both the number of RF chains and PSs are reduced in [8], where the HP is designed for multi-carrier mMIMO which offers the same sum-rate as FDP. Refertence [9] presents the idea of hybrid precoder and combiner (HPC), in which matrix decomposition is used to transform an unconstrained high-dimensional digital precoder into a hybrid structure that includes RF and baseband precoders. Similarly, [10] implements an HPC architecture for sub-connected/partially-connected array structures (where each RF chain is connected to only a subset of the antenna elements). To have an energy-efficient HPC, the NP-hard problem is solved by using the two-layer optimization approach (interference alignment and fractional programming). Additionally, [11] describes the design of low-complexity hybrid block diagonalization HPC for SU-mMIMO, which achieves a sum-rate that is comparable to the conventional block diagonalization (BD). Similarly, [12] proposes a low-complexity mmWave MU-MIMO HPC design in which the RF-stage is designed using virtual path selection and the baseband stage is constructed using zero-forcing (ZF). In [13], HP is designed for full-connected and partially-connected structures based on manifold optimization, which gives high performance but requires perfect CSI.

Different HP solutions presented in [7]–[13] provide a comparable sum-rate to FDP but requires the use of large CSI overhead. Among the HP designs based on the slowly time-varying CSI, joint spatial division and multiplexing (JSDM) approach in [14] is used to construct the HP, which requires reduced-dimensional channel state information at the transmitter (CSIT). This work is further extended for large number of users in [15], where the authors split the users into multiple groups based on their identical channel covariance eigenvectors, and then design the two-stage HP by leveraging the near-orthogonality of the different user groups' eigen-spaces. The RF beamformer is designed using an average signal-to-leakage-plus-noise-rate (SLNR) criterion based on dominant eigenvalues in [16], where the baseband precoder is designed using zero-forcing (ZF) to reduce the inter-group interference. The HP is configured for fully-connected (FC) and partially-connected (PC) hybrid structures in the wideband channel in [17], showing a sum-rate performance close to FDP. The RF beamforming and baseband stages are constructed using minimum-mean-squared error (MMSE) and weighted MMSE (WMMSE) approaches in [18], which enhance the sum rate by suppressing inter-group and intra-group interferences. At the RF stage, a phase mode transformation technique is used to design the HP in [19], which transforms the uniform circular array (UCA) into a reduced virtual uniform linear array (ULA). A non-linear HP scheme is subsequently introduced in [20], which takes into account multi-cell mMIMO systems and improves error performance compared to linear HP techniques by using the MMSE approach. Then, in [21], a hybrid linear/Tomlinson-Harashima precoder (HL-THP) architecture is implemented, in which users are divided into groups and, by using CSI, the inner (outer) precoder mitigates inter-group (intra-group) interference. In [22], a mmWave MU-MIMO HP is designed using gradient-projection method and requiring partial CSI. Particularly, the designed HP takes into account the performance loss due to beam-misalignment and thus, achieves performance close to FDP.

2) HP WITH LOW-RESOLUTION PSS

The HP studies [7]–[22] provide a comparable sum-rate to FDP but they involve the use of costly full-resolution

hardware components. In [23], the HP design using finite-resolution PSs is presented, which gives high spectral efficiency for SU-mMIMO. In [24], the authors present the HP architecture with low-resolution PSs for downlink multi-user multiple-input single-output (MU-MISO). A PSs selection technique is suggested in [25] to increase the spectral efficiency and decrease power consumption, but it requires perfect CSI and rich scattering channels. The HPC design is described in [26], which first uses low-resolution PSs to iteratively design the RF beamformer and combiner, and then designs the baseband precoder and combiner on the basis of the reduced-dimensional effective channel. In [27], an analog RF beamformer based on low-resolution PSs is constructed by exploiting the antenna diversity, and achieves a high sum-rate for MU-MISO. Similarly, the RF beamformer using low-resolution PSs is built in [28] with the constraints of discrete phase and constant amplitude for MU-MISO. Later, the authors broaden the research work in [29], where they use low-resolution PSs for MU-MISO to design the dynamic sub-connected HP (each RF chain is dynamically connected to a set of antennas via a switch network), and optimize the sum-rate by using fractional programming (FP). Reference [30] compares the sum-rate performance of SU-mMIMO with few-bits PSs using an HP architecture based on Hadamard RF codebook. Then, a dynamic sub-connected HP design for MIMO orthogonal frequency division multiplexing (MIMO-OFDM) is presented in [31]. By utlizing the switch network to connect each RF chain to a set of antenna and avoid overlapping, block coordination descent (BCD) approach is used to maximize the spectral efficiency while using low-resolution PSs.

3) HP WITH LOW-RESOLUTION DACS/ADCS

The HP techniques in [23]–[31] use low-resolution PSs only and utilizes full-resolution DACs/ADCs. A finite-alphabet MU-mMIMO precoder is introduced in [32] based on alternating direction method of multipliers (ADMM), and shows significant advantages of the reduced-complexity HP structure. A quantized linear precoder is designed in [33], and shows that the sum-rate performance of infinite-resolution DACs can be approached by using finite-resolution DACs. In [34], the authors use additive quantization noise (AQN) model for the analysis of low-resolution DACs and compare the energy efficiency of fully-connected and partially-connected HP with quantized FDP. A HP architecture based on channel inversion and singular value decomposition (SVD), and using only few-bit ADCs is presented in [35], providing a comparable sum-rate to that obtained using full-precision ADCs. In [36], a comparison of the spectral and energy efficiencies of HP with FDP under the hardware constraint of few-bit ADCs is presented, which shows peak performance with ADCs resolution of 5-bits. To reduce the CSI overhead, an angular-based HPC structure using low-resolution DACs/ADCs is introduced in [37], where the RF-stage is designed using slowly time-varying CSI and the baseband-stage is formulated based on MMSE criterion. The HP design in [32]–[37] consider the use of low-resolution DACs/ADCs only while utilizing high-resolution PSs at the RF stage.

4) ANTENNA ARRAY STRUCTURES

In mMIMO, the number of the antenna elements as well as their arrangement at the transmitter side is critical in the generation of narrower beams with reduced sidelobes, as it can offer small interference to the unintended users. The above-mentioned research studies mostly use uniform linear array (ULA) at the BS for the HP design [7]–[11], [14]–[18], [20], [21], [23]–[28], [32]–[36]. However, it is not applicable to deploy a large single-dimensional (1D) ULA at the BS because of: (i) spatial inefficiency, and (ii) restriction to illuminate both azimuth and elevation angles at the same time. The antenna elements can be arranged on a two-dimensional (2D) grid to overcome the aforementioned issues. Different HP designs using uniform rectangular array (URA) [29]–[31], [37]–[41], uniform circular array (UCA) [14], [19] and concentric circular array (CCA) [42] are presented, which shows high spatial, spectral and energy efficiencies. In [43], a comparison of different 2D array structures in mMIMO is presented by using low-resolution DACs only, which shows both URA and CCA can provide high sum-rate.

B. CONTRIBUTIONS AND ORGANIZATION

In this paper, we present the analysis of different 2D array structures (ULA, URA, UCA, and CCA) in mMIMO, where we compare the spatial, spectral and energy efficiencies by designing the HP using low-resolution PSs and quantized DACs. The proposed HP scheme includes two cascaded stages: (i) the *RF beamforming stage* is designed via the eigen-decomposition of mMIMO channel second-order correlation matrix, and (ii) the *baseband MU precoding stage* is constructed via the regularized zero-forcing (RZF) technique [44], [45] to mitigate the MU interference in the reduced-dimension effective MU-channel. The main contributions are summarized below:

- *Energy-Efficient HP Design Using Low-Resolution PSs:* Most of the existing HP designs (e.g., in [7]–[11], [14]–[21], [32]–[37]) assume the use of full-resolution PSs for the implementation of RF-stage. However, the components required for phase shift with high accuracy can be expensive [46]. Therefore, it is reasonable to use cost-effective PSs in HP. One possible way to design the HP with low-resolution PSs is to design the RF beamformer assuming infinite-resolution PSs and then, quantize each PSs value by a finite set [7]. This approach, however, is not suitable for systems with low-resolution phase constraint and requires full CSI. Furthermore, the HP designs which are presented in [23]–[25], [27]–[30] are considered for SU-mMIMO and MU-MISO systems. In this paper, we design the HP using low-resolution PSs for MU-mMIMO by taking

into account the few-bit phase constraint and formulate the optimization problem which is solved using orthogonal matching pursuit (OMP) [47], [48]. The algorithmic precoding solution takes the input as the optimal unconstrained (full-resolution) RF beamformer and approximates the constrained (quantized) RF beamformer by exploiting the dominant eigenvectors at the RF. We present the illustrative results of the proposed HP scheme and show that it is possible to approach the performance limit of the unconstrained RF-stage with as low as 2-bit PSs (i.e., sum-rate degradation $\approx$ 2-3 bits/s/Hz).

- *Comparison of 2-D Antenna Array Structures:* In this work, we compare the spatial, spectral and energy efficiencies of different 2D antenna array structures using low-resolution PSs and DACs. The array structure used in [7]–[11], [14]–[18], [20], [21], [23]–[37] is either ULA or URA. However, in this work we compare four different 2D array structures using low-resolution PSs and DACs. The simulation results show that the sum-rate of both CCA and URA can approach the sum-rate of their FDP counterparts with much-reduced number of RF chains.
- *HP Design Using Low-Resolution PSs and DACs:* As mentioned above, the HP is designed using either only low-resolution PSs [23]–[31] or only low-resolution DACs/ADCs [32]–[37]. To the best of our knowledge, the design of a MU-mMIMO HP using both *low-resolution DACs* and *low-resolution PSs* has not been considered yet. Moreover, we also compare the performance of such a reduced-complexity HP for different 2D array structures, i.e., ULA, URA, UCA, and CCA. The simulation results give useful results about the combination of various low-resolution DACs and PSs for near optimal performance. The proposed HP using 2-bit PSs and 5-bit DACs can achieve almost the same spectral efficiency while offering higher energy efficiency than the HP using full-resolution PSs and DACs. Therefore, the designed HP saves power consumption and cost while having negligible impact on the performance.
- *Design of the Quantized Variable Gain RF Beamformer:* In this work, we design the HP in which the RF beamforming stage is constructed using eigen beamforming based on user's angular location. Eigen beamforming can give sum-rate performance of HP close to FDP [14]–[17], however, the RF beamformer results in having non-constant modulus entities. The quantization of such variable gain RF beamformer is challenging. Also, to the best of our knowledge, the quantization of analog precoder designed using eigen beamforming has not been considered yet. Therefore, we solve this problem to modify our system model to have a constant modulus RF beamformer and introduces a transfer block at the baseband stage by formulating an optimization problem. The resulting RF beamformer requires double-PS structure and follows the constant modulus constraint, i.e., the gain of each entity of RF beamformer can vary between 0 and 2. The resulting HP with constant-modulus gives the same sum-rate performance as the variable gain RF beamformer. The main contributions of this work as compared to the related work in the literature are summarized in Table 1.

The rest of this paper is organized as follows. In Section II, we introduce the system model of the MU-mMIMO HP architecture. In Section III, the HP is designed using low-resolution PSs. In Section IV, we present the HP design using low-resolution PSs and quantized DACs for 2D array structures. Section V gives the illustrative results. Finally, Section VI concludes the paper.

Notation: The following notations are used throughout this paper. Boldface lower-case and upper-case letters denote column vectors and matrices, respectively. $()^T$, $()^H$ and $()^\dagger$ represent the transpose, complex-transpose and Moore-Penrose inverse operations, respectively. $\|.\|_2$ and $\|.\|_F^2$ represent the 2-norm and the Frobenius norm of a vector or matrix, respectively. $\mathbf{I}_k$, $\mathbb{E}[.]$, $\mathbb{Q}(.)$ and tr(.) denote $k\times k$ identity matrix, the expectation operator, quantizer function and the trace operator, respectively. We use $x_k \sim \mathcal{CN}(0,\sigma)$ when x_k is a complex Gaussian random variable with zero-mean and variance σ.

II. SYSTEM AND CHANNEL MODEL

In this section, we introduce the system and channel models of the proposed hybrid MU-mMIMO systems.

A. SYSTEM MODEL

We consider a MU-mMIMO system with three different HP structures shown in Figure 1. At first, the HP architecture is depicted in Figure 1(a), which considers full-resolution hardware components, i.e., ∞-bit DACs and PSs. Then, the HP using b-bit PSs and ∞-bit DACs is shown in Figure 1(b), which offers an energy efficient structure due to less power consumption of finite-bit PSs. Finally, the least power consuming HP model is presented in Figure 1(c), which uses b-bit PSs and q-bit DACs. The BS employs N antenna elements at the transmitter, which are fed by N_{RF} RF chains to serve K single-antenna users, where $K \leq N_{RF} \ll N$. As shown in Figure 1(a), the hybrid precoder $\mathbf{B} = \mathbf{B}_A\mathbf{B}_D$ is divided as the RF beamformer $\mathbf{B}_A \in \mathbb{C}^{N\times N_{RF}}$ and the baseband precoder $\mathbf{B}_D \in \mathbb{C}^{N_{RF}\times K}$. The design of HP helps in reducing the number of RF chains from N to N_{RF}. Then, the received signal at the kth user is expressed as:

$$y_k = \mathbf{h}_k^H\mathbf{B}\mathbf{s} + n_k. \tag{1}$$

The system model shown in Figure 1(c) employs low resolution q-bit DACs between the RF and the baseband precoders to reduce the system complexity/cost. To model the precoded signal at the transmitter, we consider the non-uniform

TABLE 1. A brief comparison of the related literature.

Literature	Contents												
	MU-mMIMO Precoder Design			CSI		DACs/ADCs Resolution		PSs Resolution		Array Structure			
	Hybrid	Digital	Analog	Full	Partial	Infinite	Finite	Infinite	Finite	ULA	URA	UCA	CCA
[7]	✓			✓		✓			✓	✓			
[12], [13]	✓			✓		✓		✓			✓		
[9], [21]	✓			✓		✓		✓		✓			
[11], [22]	✓				✓	✓		✓		✓			
[14], [15], [18], [19]	✓				✓	✓		✓		✓		✓	
[32], [34], [35]	✓			✓			✓	✓		✓	✓		
[33]		✓		✓			✓	✓		✓	✓		
[36], [37]	✓				✓		✓	✓		✓	✓		
[27], [28]			✓	✓		✓			✓	✓			
[23]–[26]	✓			✓		✓			✓	✓			
[29]–[31]	✓			✓		✓			✓		✓		
[38]–[41]	✓				✓	✓		✓			✓		
This work	✓				✓		✓		✓	✓	✓	✓	✓

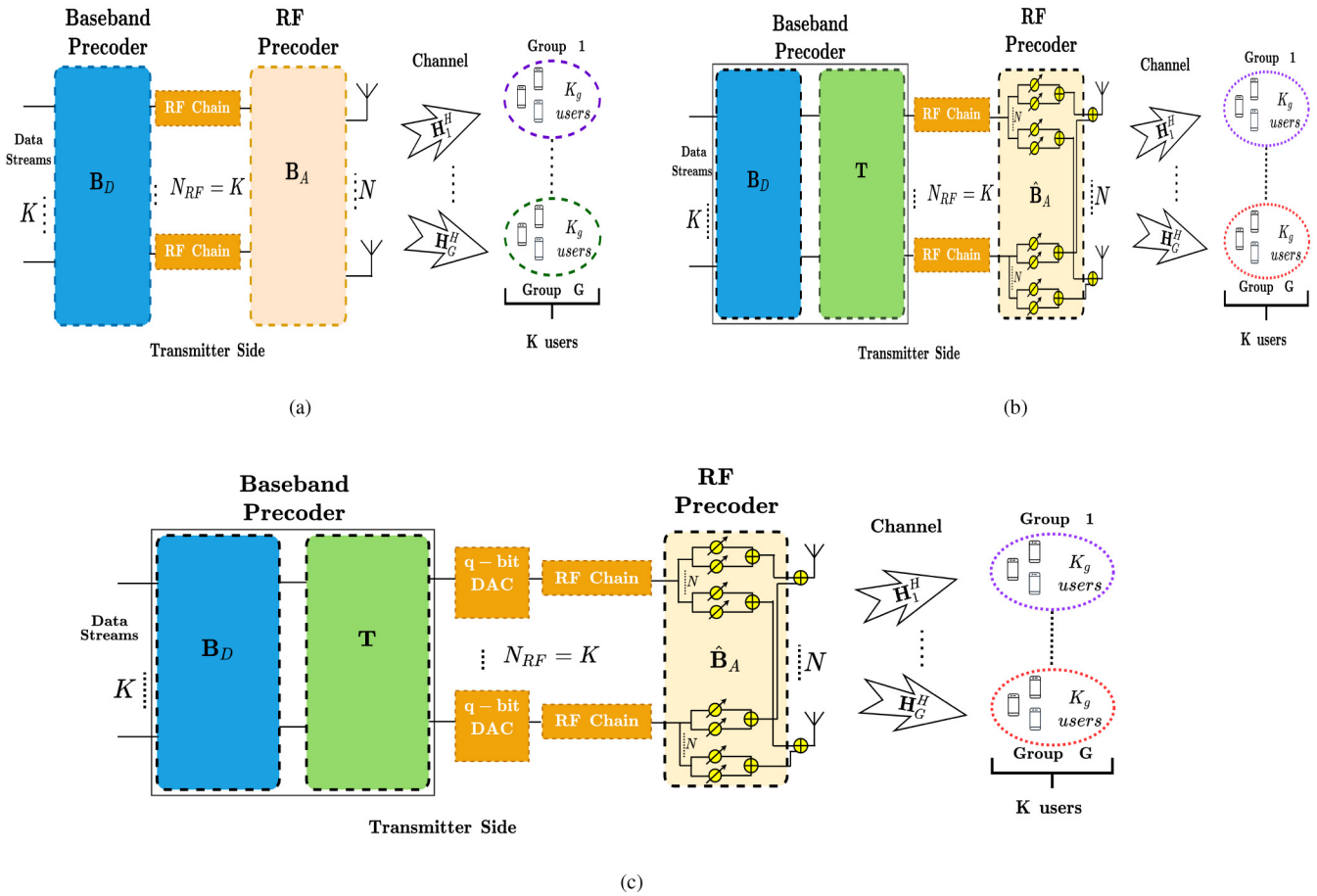


FIGURE 1. MU-mMIMO HP architectures: (a) HP with ∞-bits DACs and PSs (b) HP with *b*-bit PSs and ∞-bit DACs (c) HP with *q*-bit DACs and *b*-bit PSs.

quantizer, and adopt the additive quantization noise (AQN) model as in [49], [50]. Then, the received signal at the *kth* user is given as:

$$\begin{aligned} y_k &= \mathbf{h}_k^H \mathbf{B}_A \mathbb{Q}\{\mathbf{T}\mathbf{B}_D\mathbf{s}\} + n_k, \\ &= \mathbf{h}_k^H \mathbf{B}_A \{\gamma \mathbf{T}\mathbf{B}_D\mathbf{s} + n_q\} + n_k, \\ &= \underbrace{\gamma \mathbf{h}_k^H \mathbf{B}_A \mathbf{T}\mathbf{B}_D\mathbf{s}}_{\text{Desired Signal}} + \underbrace{\mathbf{h}_k^H \mathbf{B}_A n_q}_{\text{Quantization Noise}} + \underbrace{n_k}_{\text{Noise}}, \end{aligned} \tag{2}$$

where $\mathbf{s} \in \mathbb{C}^K$ is the transmitted data signal satisfying the power constraint, i.e., $\mathbb{E}\{\|\mathbf{s}\|_2^2\} \leq P$, where P is the transmit power at the BS, n_k denotes the additive circular symmetric Gaussian noise such that $n_k \sim \mathcal{CN}(0, 1)$, γ is the distortion factor of q-bit DAC, $n_q \sim \mathcal{CN}(0, \mathbf{R}_{n_q})$ is the additive Gaussian quantization noise which is uncorrelated with $\mathbf{s}$, and $\mathbf{T} \in \mathbb{C}^{N_{RF} \times N_{RF}}$ is the transfer block which is introduced at the baseband stage to have a constant-modulus RF beamformer $\hat{\mathbf{B}}_A \in \mathbb{C}^{N \times N_{RF}}$ such that $\hat{\mathbf{B}}_A \mathbf{T} = \mathbf{B}_A$. The covariance matrix of n_q can be written as [37], [51]:

$$\mathbf{R}_{n_q} = \mathbb{E}\left[\mathbf{n}_q \mathbf{n}_q^H\right] = \gamma(1-\gamma)\text{diag}\left(\mathbf{T}\mathbf{B}_D\mathbf{B}_D^H\mathbf{T}^H\right). \tag{3}$$

For $q = 1, 2, \ldots, 5$, the exact values of γ are 0.6366, 0.8825, 0.96546, 0.990503, and 0.997501, respectively, whereas for $q > 5$, the distortion factor γ can be approximated as $\gamma \approx 1 - \frac{\pi\sqrt{3}}{2} 2^{-2q}$ [50].

B. CORRELATION CHANNEL MODEL

Assuming no line-of-sight propagation, the channel of user k is given as:

$$\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{h}_k^{\text{i.i.d}}, \tag{4}$$

where $\mathbf{h}_k^{\text{i.i.d}} \in \mathbb{C}^N \sim \mathcal{CN}(0, \mathbf{I}_N)$. We consider correlated channel coefficients $\mathbf{h}_k$ such as $\mathbf{h}_k \sim (0, \mathbf{R}_k)$, i.e., the channel vector of *kth* user depends on the covariance matrix $\mathbf{R}_k \in \mathbb{C}^{N\times N}$. We let $\mathbf{R}_k = \mathbf{U}_k \Lambda_k \mathbf{U}_k^H$, where $\mathbf{U_k} \in \mathbb{C}^{N\times r}$ is the tall unitary matrix of eigenvectors corresponding to the non-zero eigenvalues and $\Lambda_k \in \mathbb{C}^{r\times r}$ is the diagonal matrix containing r eigenvalues of $\mathbf{R}_k$. Using Karhunen-Loeve decomposition, we can write the channel vector $\mathbf{h}_k$ for user k as:

$$\mathbf{h}_k = \mathbf{U}_k \Lambda_k^{\frac{1}{2}} \mathbf{w}_k, \tag{5}$$

where $\mathbf{w}_k \in \mathbb{C}^r \sim \mathcal{CN}(0, \mathbf{I}_r)$ is the complex coefficient vector. In overall, the channel matrix including all users can be expressed as:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]. \tag{6}$$

The correlation matrix $\mathbf{R}_k = \mathbb{E}[\mathbf{h}_k \mathbf{h}_k^H]$ is obtained by using the one-ring model [14], [52], [53], where a user is located at some distance r_d, and the mean elevation (azimuth) angle is defined as θ (ϕ). For the array structures shown in Figure 2, we assume a uniform distribution of power received from the antennas, and compute the correlation between the channel coefficients of the antennas $1 \le m, n \le N$, which is given as:

$$\mathbf{R}_{m,n} = \frac{1}{2\delta_a\delta_e} \int_{-\delta_a}^{\delta_a} \int_{-\delta_e}^{\delta_e} e^{-j\frac{2\pi}{\lambda}\mathbf{k}^T(\beta+\theta,\mu+\phi)(\mathbf{v}_m-\mathbf{v}_n)}\, d\mu\, d\beta, \tag{7}$$

where λ is the wavelength, δ_a is the angle spread around mean azimuth angle, δ_e is the angle spread around mean elevation angle, $\mathbf{k}(\beta, \mu) = [\sin(\mu)\cos(\beta), \sin(\mu)\sin(\beta)]^T$ is the wave vector for a planar wave, and $\mathbf{v}_m, \mathbf{v}_n \in \mathbb{R}^2$ is the coordinate position of the array element, i.e., $\mathbf{v}_n = [x_n, y_n]^T$. For 1D array structure (i.e., ULA), it is assumed that array elements are placed along y-axis only, which means x_n=0. The correlation matrix $\mathbf{R}$ depends on the antenna array configuration used at the BS as shown in (7). Therefore, the corresponding rank of $\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H$ plays an important role in the spectral and energy efficiencies of the array structure.

III. HYBRID PRECODER DESIGN USING LOW-RESOLUTION PSS

In the eigen beamforming, the hybrid precoder $\mathbf{B} \in \mathbb{C}^{N\times K}$ is constructed by concatenating the RF-precoding stage $\mathbf{B}_A \in \mathbb{C}^{N\times N_{RF}}$ and reduced-dimensional MU baseband precoding stage $\mathbf{B}_D \in \mathbb{C}^{N_{RF}\times K}$. The HP design based on eigen beamforming results in non-constant modulus entities at the RF stage. As discussed earlier, the quantization of this variable gain RF-stage is challenging. To solve this issue, we formulate an optimization problem and introduce a transfer block $\mathbf{T} \in \mathbb{C}^{N_{RF}\times N_{RF}}$ such that $\hat{\mathbf{B}}_A \mathbf{T} = \mathbf{B}_A$, where $\mathbf{B}_A \in \mathbb{C}^{N\times N_{RF}}$ is the variable gain RF beamformer, $\mathbf{T} \in \mathbb{C}^{N_{RF}\times N_{RF}}$ is the transfer block placed at the baseband stage and $\hat{\mathbf{B}}_A \in \mathbb{C}^{N\times N_{RF}}$ is the RF beamformer with constant-modulus entities. The design of the RF-stage together with the transfer block and the baseband stage is explained in the following sections.

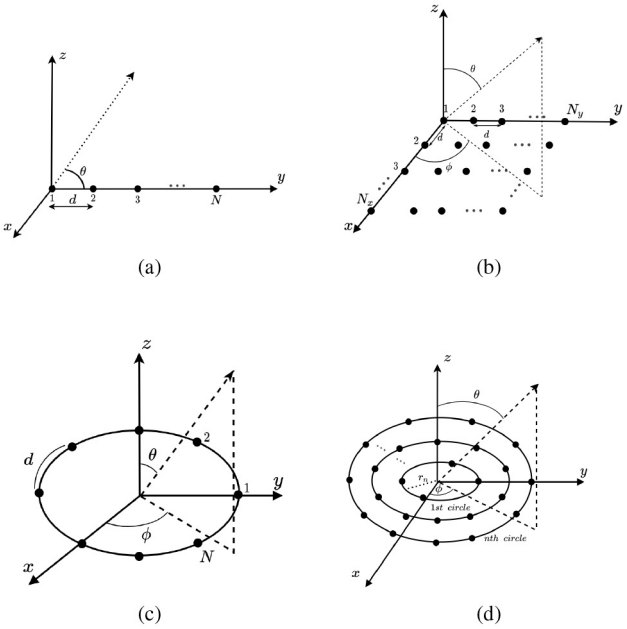


FIGURE 2. 2D Array structures: (a) ULA (b) URA (c) UCA (d) CCA.

A. VARIABLE-GAIN RF BEAMFORMER

We assume K users are clustered into G groups based on their angle-of-departure (AoD) information, where each group contains K_g number of users such that $K = \sum_{g=1}^{G} K_g$. For simplicity, we also assume that the users in the same group have identical covariance matrix $\mathbf{R}_g$, where $g = 1, 2, \ldots, G$. The index $g_k = \sum_{g'=1}^{g-1} K_{g'} + k$ is used to denote the *kth* user in group g. Then, the channel vector for the user k in group g is $\mathbf{h}_{g_k} \sim \mathcal{CN}(0, \mathbf{R}_g)$, where $\mathbf{R}_g = \mathbf{U}_g \Lambda_g \mathbf{U}_g^H$ is the covariance matrix of the group g. Based on the AoD of the users, the correlation matrix can be written as:

$$\mathbf{R}_g = \left[\mathbf{U}_g^+, \mathbf{U}_g^-\right] \left[\Lambda_g^+, \Lambda_g^-\right] \left[\mathbf{U}_g^+, \mathbf{U}_g^-\right]^H, \tag{8}$$

where $\mathbf{U}_g^+ \in \mathbb{C}^{N\times r_g^+}$ is the matrix of eigenvectors corresponding to the dominant eigenvalues for the user group g. Similarly, $\Lambda_g^+ \in \mathbb{C}^{r_g^+\times r_g^+}$ is the diagonal matrix of the dominant eigenvalues. As a result, the data streams per group is bounded by $S_g \in [0, \min\{K_g, r_g^+\}]$. Furthermore, the multiplexing gain limits the maximum number of independent data streams per group that can be transmitted, which is $\min\{K_g, r_g^+\}$. Thus, the total data streams can be written

as $S = \sum_{g=1}^{G} S_g$. The selection of N_{RF_g} is a design parameter and the performance of HP heavily depends on its value such that: (i) $N_{RF_g} \geq S_g$ to ensure that the minimum number of RF chains used per group in HP is not less than the transmitted signals per group, and (ii) the total number of RF chains is determined by the number of groups for K users (i.e., $N_{RF} = \sum_{g=1}^{G} N_{RF_g}$). The RF beamforming matrix $\mathbf{B}_A$ depends on the second-order statistics as it relies on the eigenvalues and eigenvectors extracted from the covariance matrix $\mathbf{R}_g$. If $\mathbf{B}_{A_g} = \mathbf{U}_g^+$ is the RF beamforming matrix of group g, then the complete RF beamforming matrix can be written as:

$$\mathbf{B}_A = \left[\mathbf{B}_{A_1}, \mathbf{B}_{A_2}, \ldots, \mathbf{B}_{A_G}\right]. \tag{9}$$

B. DESIGN OF TRANSFER BLOCK AND CONSTANT-GAIN RF BEAMFORMER

From (9), we can see that $\mathbf{B}_A$ has non-constant modulus entities, which results in a variable-gain RF-stage. To the best of our knowledge, the design of constant-gain analog precoder in mMIMO, which is based on eigen beamforming has not been considered yet. For this challenging issue, we formulate an optimization problem and introduce a transfer block $\mathbf{T} \in \mathbb{C}^{N_{RF} \times N_{RF}}$ as shown in Figure 1(b) such that $\mathbf{B}_A = \hat{\mathbf{B}}_A \mathbf{T}$, where $\mathbf{B}_A \in \mathbb{C}^{N \times N_{RF}}$ is the variable-gain RF beamformer as given in (9), $\mathbf{T} \in \mathbb{C}^{N_{RF} \times N_{RF}}$ is the transfer block placed at the baseband stage and $\hat{\mathbf{B}}_A \in \mathbb{C}^{N \times N_{RF}}$ is the RF beamformer with constant-gain. The transfer block $\mathbf{T}$ and the RF beamformer with constant-modulus $\hat{\mathbf{B}}_A$ is designed as:

$$\begin{aligned} \underset{\mathbf{T}, \hat{\mathbf{B}}_A}{\arg\min} \ & \left\| \mathbf{B}_A - \hat{\mathbf{B}}_A \mathbf{T} \right\|_2^2, \\ \text{s.t.} \ & \hat{\mathbf{B}}_A \in \mathbb{B}_{A_o}, \end{aligned} \tag{10}$$

where $\mathbb{B}_{A_o}$ represents the set of matrices of size $N \times N_{RF}$ satisfying the unit-modulus property for the constrained RF-beamformer $\hat{\mathbf{B}}_A$, which is constructed using double-PS structure as shown in Figure 1(b). The optimization problem is formulated to design the RF beamformer $\hat{\mathbf{B}}_A$ satisfying the modulus constraint and targeting the same performance as $\mathbf{B}_A$, where $\mathbf{B}_A$ is the RF beamformer whose entities does not satisfy the modulus constraint. Since (10) is non-convex because of the unit-modulus constraint, and as proven in [39, Lemma 1], we can dissolve $\hat{\mathbf{B}}_A$ into $\hat{\mathbf{B}}_{A_1} \in \mathbb{C}^{N \times N_{RF}}$ and $\hat{\mathbf{B}}_{A_2} \in \mathbb{C}^{N \times N_{RF}}$ based on double-PS structure, and can rewrite the equivalent optimization problem as:

$$\begin{aligned} \underset{\mathbf{T}, \hat{\mathbf{B}}_{A_1}, \hat{\mathbf{B}}_{A_2}}{\arg\min} \ & \left\| \mathbf{B}_A - \left(\hat{\mathbf{B}}_{A_1} + \hat{\mathbf{B}}_{A_2} \right) \mathbf{T} \right\|_2^2, \\ \text{s.t.} \ & \hat{\mathbf{B}}_{A_1}, \hat{\mathbf{B}}_{A_2} \in \mathbb{B}_{A_o}, \\ & \hat{\mathbf{B}}_A = \hat{\mathbf{B}}_{A_1} + \hat{\mathbf{B}}_{A_2} \end{aligned} \tag{11}$$

Then, the optimal solution for the transfer block $\mathbf{T}$ and constant-modulus RF beamformer $\hat{\mathbf{B}}_A$ is given as:

$$\hat{\mathbf{B}}_{A_1}(a,b) = e^{j\left(\angle \mathbf{B}_A(a,b) + \cos^{-1}\left(\frac{\mathbf{B}_A(a,b)}{2\tau}\right)\right)}, \tag{12a}$$

$$\hat{\mathbf{B}}_{A_2}(a,b) = e^{j\left(\angle \mathbf{B}_A(a,b) - \cos^{-1}\left(\frac{\mathbf{B}_A(a,b)}{2\tau}\right)\right)}, \tag{12b}$$

$$\mathbf{T} = \left(\hat{\mathbf{B}}_{A_1} + \hat{\mathbf{B}}_{A_2} \right)^{\dagger} \mathbf{B}_A, \tag{12c}$$

where $\tau = \frac{1}{2} \max_{a,b} |\mathbf{B}_A(a,b)|$ is the half of the highest modulus element at $\mathbf{B}_A$. Each RF chain is connected to the corresponding antenna element through two PSs, which are summed up together to formulate the RF beamforming gain. This summation allows the gain of each RF beamformer entity to vary between 0 and 2, i.e., by relaxing the unit-modulus constraint, the new constraint of the analog stage is $|\hat{\mathbf{B}}_A(a,b)| \leq 2 \ \forall \ a, b$. By implementing the double-PS structure, we can have PS-only RF beamformer without having any impact on the performance of eigen beamforming. In other words, without $\mathbf{T}$ and double-PS structure at RF stage, we would require variable-gain controllers together with PSs, which makes the implementation of HP using low-resolution PSs challenging.

C. RF BEAMFORMER QUANTIZATION

By the introduction of the transfer block in the HP architecture, each entity of the RF beamformer $\hat{\mathbf{B}}_A$, which is expressed in (12), can be converted to a modulus constraint within 0 and 2 (Figure 1(c)). Thus, the analog precoder $\hat{\mathbf{B}}_A$ can have only PSs for the RF processing. Since the phase of each entry of $\hat{\mathbf{B}}_A$ tends to be highly quantized as well as the use of full-resolution PSs is impractical because of the high cost and power consumption, therefore, we need to investigate the performance of our proposed HP by using low-resolution PSs to be used in a more realistic scenario. For this purpose, we quantize the phase of each entry of $\hat{\mathbf{B}}_A$, i.e., phases of the $(2 \times N_{RF} N)$ entries of $\hat{\mathbf{B}}_A$ are quantized up to b bits of precision. Each (i,j)th entry is quantized to its nearest neighbor based on the closest Euclidean distance. Thus, we can write as:

$$\hat{l} = \underset{l \in \mathbb{L}_o}{\arg\min} \left| \vartheta - \frac{2\pi l}{L} \right|, \tag{13}$$

where ϑ is the unquantized phase of each entity of RF beamformer obtained from (12), $L = 2^b$ and $\mathbb{L}_o$ is the set of all possible quantized phase values, i.e., $\mathbb{L}_o = \{0, \ldots, 2^{b-1}\}$. Then the quantized phase of each entry of $\hat{\mathbf{B}}_A$ can be written as:

$$\hat{\vartheta} = \frac{2\pi \hat{l}}{L}.$$

D. RF BEAMFORMER DESIGN FOR LOW-RESOLUTION PSS

As discussed earlier, the straightforward approach for HP using low-resolution PSs is to design the RF-stage based on full-resolution PSs first, and then quantizing the value of each PSs to a finite set. However, this approach yields a large sum-rate degradation and thus, it is not effective for very low-resolution PSs. To reduce the sum-rate degradation resulting from the constraint of low-resolution PSs, we design the RF beamformer by minimizing the Euclidean distance between

the HP using full-resolution PSs and the HP using low-resolution PSs. The optimization problem can be formulated as:

$$\begin{aligned} \underset{\mathbf{T}^{(q)},\hat{\mathbf{B}}_A^{(q)}}{\arg\min} \quad & \left\| \hat{\mathbf{B}}_A \mathbf{T} - \hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)} \right\|_F, \\ \text{s.t.} \quad & \hat{\mathbf{B}}_A^{(q)} \in \hat{\mathbb{B}}_A^{(q)}, \\ & \left\| \hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)} \right\|_F^2 = N_{RF}, \end{aligned} \tag{14}$$

where $\hat{\mathbb{B}}_A^{(q)}$ represents the set of matrices of size $N \times N_{RF}$ under the following constraints: (i) the gain of each entity of the matrix can vary between 0 and 2, and (ii) the phase of each matrix entity is quantized to enable the use of low-resolution PSs at the RF-stage, $\hat{\mathbf{B}}_A \mathbf{T}$ is the optimal unconstrained precoding matrix, $\hat{\mathbf{B}}_A^{(q)}$ and $\mathbf{T}^{(q)}$ are the RF beamformer and the transfer block designed for low-resolution PSs, respectively. The design problem can be defined as finding the projections of the optimal unconstrained precoder $\hat{\mathbf{B}}_A \mathbf{T}$ onto the set of $\hat{\mathbf{B}}_A^{(q)}$ and $\mathbf{T}$. Each entry of $\hat{\mathbf{B}}_A^{(q)}$ follows modulus constraint and can have a distinct phase value as depicted by the constraint $\hat{\mathbf{B}}_A^{(q)} \in \hat{\mathbb{B}}_A^{(q)}$. Furthermore, $\|\hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)}\|_F^2 = N_{RF}$ represents the normalized transmit power constraint of the RF beamformer. Due to the complex non-convex nature of the feasible set $\hat{\mathbb{B}}_A^{(q)}$, the problem of finding the projections is difficult. However, we can notice that there exits a connection between $\hat{\mathbf{B}}_A$ and $\hat{\mathbf{B}}_A^{(q)}$ by exploiting the structure of the channel $\mathbf{H}$. This can be further explained by the following *remarks*.

1) *Structure of unconstrained RF beamformer*: The design of unconstrained RF beamformer $\hat{\mathbf{B}}_A$ is based on Karhunen-Loeve decomposition of the covariance matrix $\mathbf{R}$. Thus, the eigenvectors (column vectors) for the corresponding dominant eigenvalues form an orthonormal basis.
2) *Relation between* $\hat{\mathbf{B}}_A$ *and* $\hat{\mathbb{B}}_A^{(q)}$: We notice that there exits an association between $\hat{\mathbf{B}}_A$ and $\hat{\mathbb{B}}_A^{(q)}$ as both are based on the dominant eigenvalues extracted from the covariance matrix. In fact, $\hat{\mathbb{B}}_A^{(q)}$ is the set of quantized matrices of $\mathbf{U}^+$, where $\mathbf{U}^+ = [\mathbf{U}_1^+, \mathbf{U}_2^+, \ldots, \mathbf{U}_G^+]$.
3) *Transfer block*: The problem of finding projection of $\hat{\mathbf{B}}_A$ onto $\hat{\mathbf{B}}_A^{(q)}$ alone is challenging. By using the transfer block $\mathbf{T}$, we can find the projection of $\hat{\mathbf{B}}_A \mathbf{T}$ onto the set of $\hat{\mathbf{B}}_A^{(q)}$ and $\mathbf{T}$.
4) *Design of* $\hat{\mathbf{B}}_A^{(q)}$: It must be seen that the feasible quantized RF beamformers in $\hat{\mathbb{B}}_A^{(q)}$ are of size $N \times N_{RF}$. Thus, each column of the constrained RF beamformer $\hat{\mathbf{B}}_A^{(q)}$ can be designed by applying N_{RF} vectors, where each vector follows the constant-modulus and have distinct phase values only. Furthermore, by using the transfer block $\mathbf{T}$, we can form the arbitrary linear combinations of N_{RF} vectors and design the constrained RF beamformer by reducing the Euclidean distance, i.e., $\|\hat{\mathbf{B}}_A \mathbf{T} - \hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)}\|_F$.

Algorithm 1: Constrained RF Beamformer Design via Orthogonal Matching Pursuit

Given: Optimal unconstrained RF beamformer $\hat{\mathbf{B}}_A = [\hat{\mathbf{B}}_{A_1}, \hat{\mathbf{B}}_{A_2}, \ldots, \hat{\mathbf{B}}_{A_G}]$

for $g = 1{:}G$ **do**
- $\hat{\mathbf{B}}_{A_g}^{(q)} = [\,]$
- $\mathbf{B}_{res_g} = \hat{\mathbf{B}}_{A_g} \mathbf{T}_g$
- **for** $j \leq N_{RF_g}$ **do**
  - $\Upsilon_g = (\hat{\mathbf{U}}_g^+)^* \mathbf{B}_{res_g}$
  - Find the index l which maximizes $(\Upsilon_g \Upsilon_g^*)$
  - $\hat{\mathbf{B}}_{A_g}^{(q)} = \left[\hat{\mathbf{B}}_{A_g}^{(q)} | \hat{\mathbf{U}}_g^{+(l)}\right]$
  - Compute $\mathbf{T}_g^{(q)}$ by using least squares $\mathbf{T}_g^{(q)} = \left(\hat{\mathbf{B}}_{A_g}^{(q)*} \hat{\mathbf{B}}_{A_g}^{(q)}\right)^{-1} \hat{\mathbf{B}}_{A_g}^{(q)} \hat{\mathbf{B}}_{A_g}$
  - $\mathbf{B}_{res_g} = \frac{\hat{\mathbf{B}}_{A_g} \mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)} \mathbf{T}_g^{(q)}}{\left\| \hat{\mathbf{B}}_{A_g} \mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)} \mathbf{T}_g^{(q)} \right\|_F}$
- **end**
- $\tilde{\mathbf{T}}_g^{(q)} = \frac{N_{RF_g}}{\left\| \hat{\mathbf{B}}_{A_g}^{(q)} \mathbf{T}_g^{(q)} \right\|_F} \mathbf{T}_g^{(q)}$

end

$\tilde{\mathbf{T}}^{(q)} = [\tilde{\mathbf{T}}_1^{(q)}, \tilde{\mathbf{T}}_2^{(q)}, \ldots, \tilde{\mathbf{T}}_G^{(q)}]$

$\hat{\mathbf{B}}_A^{(q)} = [\hat{\mathbf{B}}_{A_1}^{(q)}, \hat{\mathbf{B}}_{A_2}^{(q)}, \ldots, \hat{\mathbf{B}}_{A_G}^{(q)}]$

return $\tilde{\mathbf{T}}^{(q)}$, $\hat{\mathbf{B}}_A^{(q)}$

Based on remark 2, the optimal constrained RF beamformer can be found by replacing $\hat{\mathbb{B}}_A^{(q)}$ with $\hat{\mathbb{U}}^+$, where $\hat{\mathbb{U}}^+$ represents the set of matrices of quantized eigenvectors corresponding to the dominant eigenvalues, i.e., $\hat{\mathbf{U}}^+$. Subsequently, we can rewrite the optimization problem as follows:

$$\begin{aligned} \underset{\mathbf{T}^{(q)},\hat{\mathbf{B}}_A^{(q)}}{\arg\min} \quad & \left\| \hat{\mathbf{B}}_A \mathbf{T} - \hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)} \right\|_F, \\ \text{s.t.} \quad & \hat{\mathbf{B}}_A^{(q)} \in \hat{\mathbb{U}}^+, \\ & \left\| \hat{\mathbf{B}}_A^{(q)} \mathbf{T}^{(q)} \right\|_F^2 = N_{RF}. \end{aligned} \tag{15}$$

We can append the constraint $\hat{\mathbf{B}}_A^{(q)}$ into the optimization objective and formulate the equivalent problem as [48]:

$$\begin{aligned} \underset{\mathbf{T}^{(q)}}{\arg\min} \quad & \left\| \hat{\mathbf{B}}_A \mathbf{T} - \hat{\mathbf{U}}^+ \mathbf{T}^{(q)} \right\|_F, \\ \text{s.t.} \quad & \left\| \text{diag}\left(\mathbf{T}^{(q)} \mathbf{T}^{(q)*} \right) \right\|_0 = N_{RF}, \end{aligned} \tag{16}$$

where the constraint $\|\text{diag}(\mathbf{T}^{(q)}\mathbf{T}^{(q)*})\|_0 = N_{RF}$ states that the matrix $\mathbf{T}^{(q)}$ acts as an auxilliary variable and can only have a maximum of N_{RF} non-zero rows. As a result, only N_{RF} columns of $\hat{\mathbf{U}}^{+}$ are selected and thus, the transfer block $\mathbf{T}^{(q)}$ can be designed by N_{RF} non-zero rows. The constrained RF beamformer is formulated by projecting the N_{RF} columns of $\hat{\mathbf{U}}^{+}$ onto $\hat{\mathbf{B}}_A$. By employing the orthogonal matching pursuit (OMP) technique proposed in [47], [48], we present an algorithmic solution to design the constrained RF beamformer for low-resolution PSs. The pseudo-code is given in Algorithm, which can be summarized as: (i) the algorithm starts by initializing the constrained precoder $\hat{\mathbf{B}}_{A_g}^{(q)}$ as an empty matrix, (ii) finds the vector along which the unconstrained precoder has the maximum projection, (ii) append the selected column vector to the empty matrix $\hat{\mathbf{B}}_{A_g}^{(q)}$, (iv) using least squares, find the transfer block $\mathbf{T}_g^{(q)}$, (v) remove the contribution of the selected vector, and (vi) finding the column along which the residual precoding matrix $\mathbf{B}_{res_g}$ has the largest projection. The algorithm continues until we find N_{RF_g} precoding vectors. As a result, we get an $N \times N_{RF_g}$ RF precoding matrix and $N_{RF_g} \times N_{RF_g}$ transfer matrix. Finally, from Lemma 1, step 11 ensures that the transmit power constraint is satisfied. The process is repeated for all G user groups. Combining the RF and transfer matrix of all groups, we get a large $N \times N_{RF}$ RF preoding matrix and $N_{RF} \times N_{RF}$ transfer matrix which minimizes $\|\hat{\mathbf{B}}_A\mathbf{T} - \hat{\mathbf{B}}_A^{(q)}\mathbf{T}^{(q)}\|_F^2$.

Lemma 1: If the Euclidean distance before normalization is $\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\|_F \leq \zeta$, then the Euclidean distance after normalization is $\|\hat{\mathbf{B}}_{A,g}\mathbf{T}_g - \hat{\mathbf{B}}_{A,g}^{(q)}\tilde{\mathbf{T}}_g^{(q)}\|_F \leq 2\zeta$.

Proof: See the Appendix. ■

E. BASEBAND MU PRECODER DESIGN

After the design of RF beamformer $\hat{\mathbf{B}}_A^{(q)}$ for b-bit PSs, the baseband MU precoder $\mathbf{B}_D$ can be determined by using the joint group processing (JGP) [14]. The overall reduced dimensional effective channel matrix is expressed as $\boldsymbol{\mathcal{H}} = \tilde{\mathbf{T}}^{(q)}\hat{\mathbf{B}}_A^{(q)H}\mathbf{H}$. Using (6) and (9), $\boldsymbol{\mathcal{H}}^H$ can be written as:

$$\boldsymbol{\mathcal{H}}^H = \begin{bmatrix} \mathbf{H}_1^H\hat{\mathbf{B}}_{A_1}^{(q)}\tilde{\mathbf{T}}_1^{(q)} & \mathbf{H}_1^H\hat{\mathbf{B}}_{A_2}^{(q)}\tilde{\mathbf{T}}_2^{(q)} & \dots & \mathbf{H}_1^H\hat{\mathbf{B}}_{A_G}^{(q)}\tilde{\mathbf{T}}_G^{(q)} \\ \mathbf{H}_2^H\hat{\mathbf{B}}_{A_1}^{(q)}\tilde{\mathbf{T}}_1^{(q)} & \mathbf{H}_2^H\hat{\mathbf{B}}_{A_2}^{(q)}\tilde{\mathbf{T}}_2^{(q)} & \dots & \mathbf{H}_2^H\hat{\mathbf{B}}_{A_G}^{(q)}\tilde{\mathbf{T}}_G^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_G^H\hat{\mathbf{B}}_{A_1}^{(q)}\tilde{\mathbf{T}}_1^{(q)} & \mathbf{H}_G^H\hat{\mathbf{B}}_{A_2}^{(q)}\tilde{\mathbf{T}}_2^{(q)} & \dots & \mathbf{H}_G^H\hat{\mathbf{B}}_{A_G}^{(q)}\tilde{\mathbf{T}}_G^{(q)} \end{bmatrix}, \tag{17}$$

where the diagonal matrices $\mathbf{H}_g^H\hat{\mathbf{B}}_{A_g}^{(q)}\tilde{\mathbf{T}}_g \in \mathbb{C}^{K_g \times N_{RF_g}}$ are the effective channel matrix for group g and the off-diagonal matrices $\mathbf{H}_g^H\hat{\mathbf{B}}_{A_{\hat{g}}}^{(q)}\tilde{\mathbf{T}}_{\hat{g}} \in \mathbb{C}^{K_g \times N_{RF_{\hat{g}}}}$ represent the effective interference channel matrix between groups g and $\hat{g}$, $\forall \hat{g} \neq g$. By applying the well-known RZF technique [44], [45], $\mathbf{B}_D$ can be defined as:

$$\mathbf{B}_D = \eta\mathbf{X}\boldsymbol{\mathcal{H}}^H, \tag{18}$$

where $\boldsymbol{\mathcal{H}}^H$ is the reduced-dimension effective channel as given in (17), $\mathbf{X} = [\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}} + \alpha N_{RF}\mathbf{I}_{N_{RF}}]^{-1}$, $\mathbf{I}_{N_{RF}} \in \mathbb{C}^{N_{RF} \times N_{RF}}$, α is the regularization parameter and η is the normalization factor used to satisfy the power constraint, which can be written as:

$$\eta = \sqrt{\frac{S}{\text{tr}\{\boldsymbol{\mathcal{H}}^H\mathbf{X}^H\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\boldsymbol{\mathcal{H}}\}}}. \tag{19}$$

Therefore, the signal-to-interference-plus-noise ratio (SINR) at the user k in group g can be computed as follows:

$$\text{SINR}_{g_k} = \frac{\frac{P}{S}\left|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_{g_k}\right|^2}{1 + \frac{P}{S}\sum_{i \neq g_k}\left|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_i\right|^2}, \tag{20}$$

where $\frac{P}{S}\sum_{i \neq g_k}|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_i|^2$ accounts for MU interference. Hence, the corresponding sum-rate performance can be computed as:

$$R_{\text{sum}} = \sum_{g=1}^{G}\sum_{g_k=1}^{K_g}\mathbb{E}\big[\log_2\big(1 + \text{SINR}_{g_k}\big)\big]. \tag{21}$$

IV. HYBRID PRECODER DESIGN USING LOW-RESOLUTION PSS AND DACS

In this section, we design the HP using low-resolution hardware components (i.e., q-bit DACs and b-bit PSs). The HP structure is shown in Figure 1(c). Based on the AoD information of users, we design the RF beamformer and the baseband MU precoder along with a transfer block such that quantization error is minimized.[1]

A. RF BEAMFORMER AND TRANSFER BLOCK DESIGN

Following the same user distribution as discussed in Section III-A, K users are clustered into G groups. Based on the AoD of the users, the correlation matrix of group g is given in (8), which consists of the dominant eigenvectors $\mathbf{U}_g^{+} \in \mathbb{C}^{N \times r_g^{+}}$ corresponding to the dominant eigenvalues $\Lambda_g^{+} \in \mathbb{C}^{r_g^{+} \times r_g^{+}}$. Moreover, $\min\{K_g, r_g^{+}\}$ limits the maximum number of the independent data streams that can be transmitted. By using Karhunen-Loeve decomposition, we can write the user channel vector $\mathbf{h}_{g_k}$ and the overall channel matrix $\mathbf{H}$ as given in (5) and (6), respectively. If $\mathbf{B}_{A_g} = \mathbf{U}_g^{+}$ is the RF beamforming matrix of group g having non-constant modulus, then the complete RF beamforming matrix can be written as given in (9). As discussed in Section III-B, the variable-gain RF beamforming matrix $\mathbf{B}_{A_g}$ can be converted to a constant-gain beamforming matrix $\hat{\mathbf{B}}_{A_g}$ by using a transfer block $\mathbf{T}_g$ at the baseband. Using this approach, we can rewrite (9) as:

$$\mathbf{B}_A = \Big[\hat{\mathbf{B}}_{A_1}\mathbf{T}_1, \hat{\mathbf{B}}_{A_2}\mathbf{T}_2, \dots, \hat{\mathbf{B}}_{A_G}\mathbf{T}_G\Big]. \tag{22}$$

1. Part of the work done in this Section IV has been presented in a conference paper [43].

$\hat{\mathbf{B}}_{A_g}$ and $\mathbf{T}_g$ can be found using (12), respectively. However, since the RF beamformer employs b-bit PSs for the phase quantization, we can rewrite (22) as:

$$\mathbf{B}_A^{(q)} = \left[\hat{\mathbf{B}}_{A_1}^{(q)}\mathbf{T}_1^{(q)}, \hat{\mathbf{B}}_{A_2}^{(q)}\mathbf{T}_2^{(q)}, \ldots, \hat{\mathbf{B}}_{A_G}^{(q)}\mathbf{T}_G^{(q)}\right]. \quad (23)$$

By using the solution given in Algorithm, we can find $\hat{\mathbf{B}}_{A_g}^{(q)}$ and $\mathbf{T}_g^{(q)}$ for HP using b-bit PSs at the RF-stage. The overall RF beamforming matrix and the transfer matrix for low-resolution PSs can be given as:

$$\hat{\mathbf{B}}_A^{(q)} = \left[\hat{\mathbf{B}}_{A_1}^{(q)}, \hat{\mathbf{B}}_{A_2}^{(q)}, \ldots, \hat{\mathbf{B}}_{A_G}^{(q)}\right]. \quad (24)$$

$$\mathbf{T}^{(q)} = \left[\mathbf{T}_1^{(q)}, \mathbf{T}_2^{(q)}, \ldots, \mathbf{T}_G^{(q)}\right]. \quad (25)$$

B. BASEBAND MU PRECODER DESIGN WITH FINITE-RESOLUTION DACS

In this section, the design of the baseband MU precoder $\hat{\mathbf{B}}_D^{(q)}$ using q-bit DACs is presented. After designing RF beamformer $\hat{\mathbf{B}}_A^{(q)}$ and the transfer block $\mathbf{T}^{(q)}$ for b-bit PSs as given in (24) and (25), respectively, the quantized baseband precoder $\hat{\mathbf{B}}_A^{(q)}$ can be designed using RZF as given in (18). The use of q-bit DACs introduces the quantization noise. Therefore, the signal to quantization, interference and noise ratio (SQINR) at the user k in group g can be computed as given in (26), as shown at the bottom of the page, where $\frac{P}{S}\sum_{i\neq g_k}|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_i|^2$ accounts for the interference experienced by users in group g from the users in group $\hat{g}\neq g$ and $\frac{P}{S}|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{R}_{n_q}(\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_{g_k})^H|^2$ is the quantization noise when using q-bit DACs. Hence, the corresponding sum-rate performance of HP with b-bit PSs and q-bit DACs can be computed as:

$$R_{\text{sum}} = \sum_{g=1}^{G}\sum_{g_k=1}^{K_g}\mathbb{E}\left[\log_2\left(1+\text{SQINR}_{g_k}\right)\right]. \quad (27)$$

V. ILLUSTRATIVE RESULTS

In this section, the Monte-Carlo simulation results are presented based on the hybrid precoding for various array structures. For the presented results, we equip the BS with 100 isotropic antennas, which are arranged in 1×100 ULA, which requires 50λ space in one direction with half wavelength (i.e., $\frac{\lambda}{2}$) spacing between antenna elements. On the other hand, 10×10 URA configuration requires just 5λ space in horizontal and vertical directions, which is equivalent to a space of $25\lambda^2$. For UCA, we assume the antennas are equally distributed on a circle of radius $\lambda\Delta$ where λ is the wavelength and $\Delta = \frac{1}{2\sqrt{(1-\cos(\frac{2\pi}{N}))^2+\sin(\frac{2\pi}{N})^2}}$. Thus, UCA with 100 antennas has a radius of about 7.96λ (i.e., diameter of $15.92\ \lambda$) and requires an array size of approximately $199\lambda^2$. In CCA, the antennas are distributed on multiple circular rings, where the radius of each *nth* ring is λR_n, while $R_n = nL$ [54]. Following this arrangement, the corresponding number of antennas on *nth* ring is $N_n = \frac{2\pi R_n}{d}$, where d is the antenna spacing which is selected as $\frac{\lambda}{2}$ for all array structures and L is the inter-ring spacing which is selected as 0.55λ. Thus, 100 antennas can be arranged in just 5 rings in CCA and the maximum ring radius in CCA is about 2.75λ, which is more than *8 times* smaller in size than UCA having the same number of antennas. In other words, this is equivalent to a space of approximately $24\lambda^2$. In terms of the spatial efficiency, considering the carrier frequency as 6 GHz, then the array sizes of ULA, URA, UCA, and CCA are respectively 2.5 m, 0.0625 m^2, 0.49 m^2, and 0.0592 m^2. Thus, CCA is the highest spatially efficient structure, whereas ULA offers the least spatial efficiency. For all the illustrative results provided hereafter, we compare the performance of different array structures, when $N = 100$. We consider three different user groups ($G = 3$), which are located around the BS at mean elevation angle $\theta = 73°$, mean azimuth locations $(\phi_1, \phi_2, \phi_3) = (30°, 90°, 150°)$ with $\delta_a = 15°$ and $\delta_e = 12.5°$. The signal-to-noise-ratio (SNR) is defined as $\frac{P}{\sigma^2 K}$ and the regularization parameter $\alpha = \frac{\sigma^2}{P}$ [45]. Based on the AoD information of the user groups, the HP is designed by concatenating RF beamformer and the MU baseband precoder. The rank of the covariance matrix $\mathbf{R}$ (please see (7)) can vary for different array structures, which can impact the performance of the proposed HP schemes. A detailed discussion about the rank of ULA, URA, UCA, and CCA for a varying number of antennas is provided in [42]. Since the RF beamformer can be designed for constant and variable-gain entities, therefore, we first compare the performance of the HP for the two cases: (i) HP with variable-gain RF beamformer, and (ii) HP with transfer block and constant-gain RF beamformer. To compare the performance of these two HP schemes, we assume full-resolution hardware components (i.e., DACs and PSs).

In Figure 3(a), the sum-rate of proposed HP scheme with variable-gain RF beamformer is compared to FDP as well as different HP techniques (i.e., phase extraction alternative minimization (PE-AltMin) method in [13] and iterative

$$\text{SQINR}_{g_k} = \frac{\frac{P}{S}\left|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_{g_k}\right|^2}{1+\frac{P}{S}\sum_{i\neq g_k}\left|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{X}\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_i\right|^2+\frac{P}{S}\left|\mathbf{h}_{g_k}^H\hat{\mathbf{B}}_A^{(q)}\tilde{\mathbf{T}}^{(q)}\mathbf{R}_{n_q}\left(\tilde{\mathbf{T}}^{(q)H}\hat{\mathbf{B}}_A^{(q)H}\mathbf{h}_{g_k}\right)^H\right|^2} \quad (26)$$

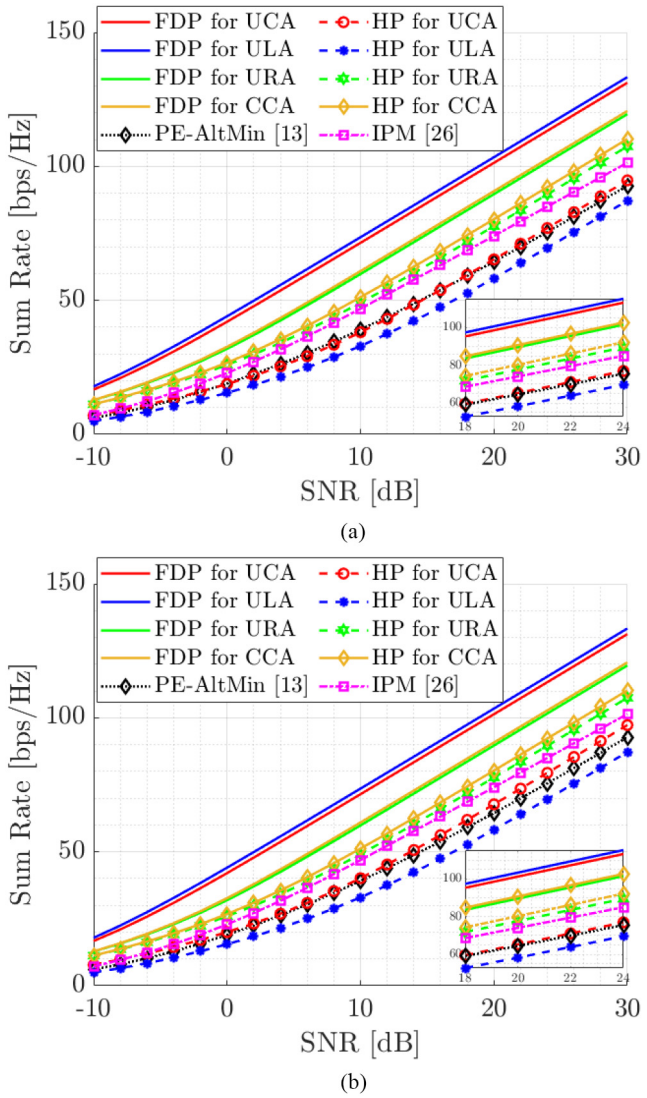


FIGURE 3. Sum-rate comparison of ULA, UCA, URA and CCA using full-resolution PSs and DACs: (a) variable-gain RF beamformer (b) constant-modulus RF beamformer (with transfer block).

phase matching (IPM) HP technique in [26]) for different array structures, whereas in Figure 3(b), the sum-rate of HP with transfer block and constant-gain RF beamformer is presented. It can be seen that both HP schemes (i.e., two-stage HP as in Fig. 3(a) and three-stage HP as in Fig. 3(b)) yield similar performance for ULA, URA, UCA, and CCA. Furthermore, FDP achieves higher sum-rate at the expense of large number of RF chains, i.e., $N_{RF} = N = 100$, whereas for both proposed HP schemes, we can approach the sum-rate of FDP with relatively small number of RF chains, i.e., $N_{RF} = 9$. Also, both CCA and URA can give high sum-rate when compared to ULA and UCA as well as the HP solutions presented in [13] and [26]. In the following analysis, we provide an extensive comparison of spectral and energy efficiencies of the array structures for the case of: (i) HP with b-bit PSs and ∞-bit DACs, and (ii) HP with b-bit PSs and q-bit DACs.

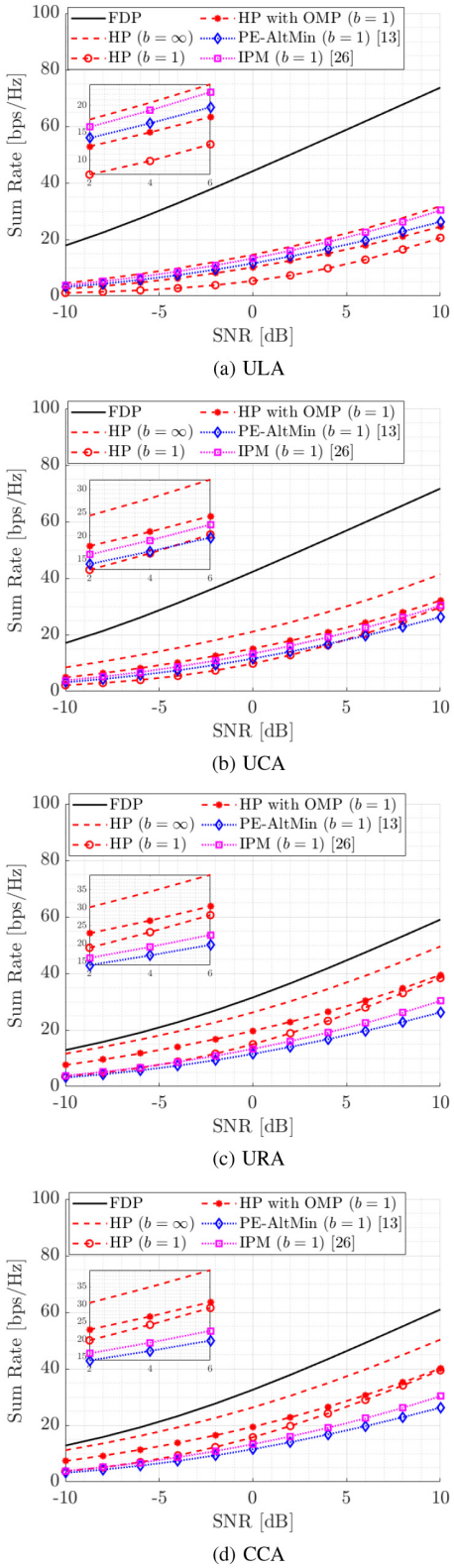
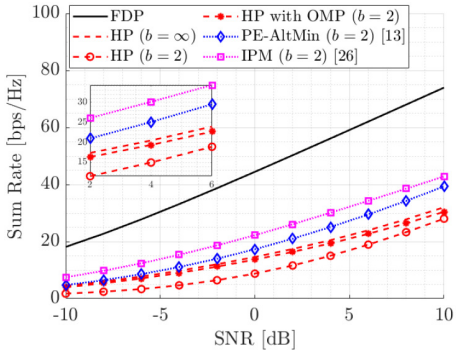


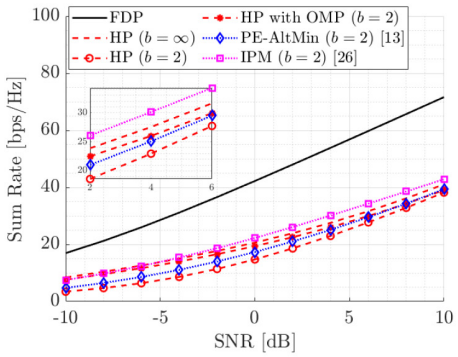
FIGURE 4. Sum-rate comparison with 1-bit PSs: (a) ULA (b) UCA (c) URA (d) CCA.

A. HP WITH B-BIT PSS AND ∞-BIT DACS

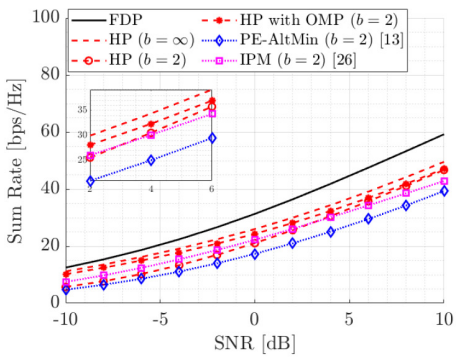
In this section, we analyze the performance of ULA, URA, UCA, and CCA by evaluating the proposed HP schemes when using few-bit PSs only. Figure 4 depicts the sum-rate



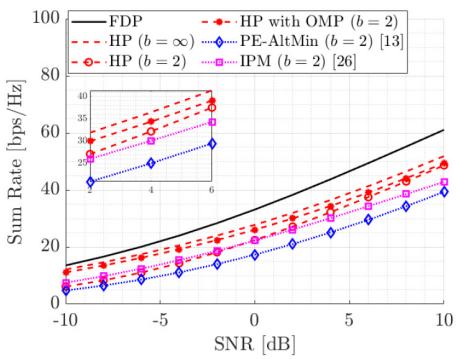
(a) ULA



(b) UCA



(c) URA



(d) CCA

FIGURE 5. Sum-rate comparison with 2-bit PSs: (a) ULA (b) UCA (c) URA (d) CCA.

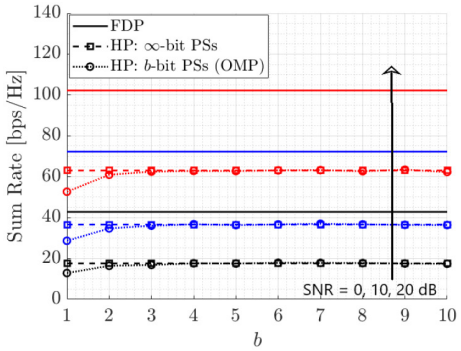
of different array structures for the following cases: (i) FDP, (ii) HP using ∞-bit PSs, (iii) HP using 1-bit PSs (phase quantization of (ii)), (iv) HP using 1-bit PSs (OMP), (v) HP using PE-AltMin [13], and (vi) HP using IPM [26]. By using the same simulation parameters as mentioned earlier, it can be seen that the proposed OMP based HP design for 1-bit PSs can provide higher sum-rates than the HP designed with simple phase quantization as well as different HP solutions (i.e., [13], [26]) for URA and CCA array structures. Also, the sum-rate degradation is reduced to around (3-4) [bps/Hz] at low SNR, whereas at high SNR, the degradation is $\approx (8-10)$ [bps/Hz] due to high quantization error. However, by using only 2-bit PSs, we can approach the sum-rate of ∞-bit PSs as shown in Figure 5. In this case, the sum-rate degradation is reduced to $\approx (2-3)$ [bps/Hz] at all SNR values. From Figures 4 and 5, we can also see that URA and CCA can give higher sum-rate than ULA and UCA for both 1-bit and 2-bit PSs.

Figure 6 plot the sum-rate vs. b, where b is the number of the bits of the PSs, and compares the spectral efficiency of ULA, URA, UCA, and CCA. The results are presented at SNR of 0, 10 and 20 dB. Figure 6 indicates that the proposed HP scheme with 2-bit PSs can approach the performance of HP with ∞-bit PSs. For a given SNR, the sum-rate performance gap between the FDP and HP with ∞-bit PSs in the case of URA and CCA is much smaller than in the case of ULA and UCA. Furthermore, for all array structures, the sum-rate gap increases with increasing SNR. Similarly, for the same SNR values, we investigate the sum-rate vs. N_{RF_g} of FDP, HP's with ∞-bit and 1-bit PS in different array structures as plotted in Figure 7. The results show that both URA and CCA can give high sum-rate with a slight increase in N_{RF_g}. The sum-rate increases by approximately 20 % for ULA and UCA, and almost 10 % for URA and CCA by increasing N_{RF_g} from 3 to 4. Also, in Figure 8, we plot the sum-rate vs. N_{RF_g} of FDP, HP's with ∞-bit and 2-bit PS. It can be seen that by increasing N_{RF_g} from 3 to 4, the proposed HP scheme using either URA or CCA can approach the performance of FDP with ∞-bit PSs, and offers a slightly lower sum-rate with 2-bit PSs. On the other hand, the proposed HP scheme using ULA and UCA requires more RF chains to approach the performance of the FDP.

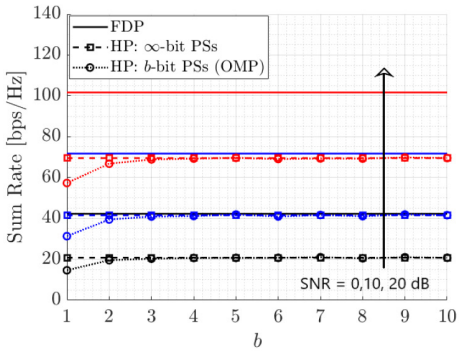
The use of low-resolution PSs with the proposed HP scheme not only offers high spectral efficiency but can also provide high energy efficiency compared to FDP and HP using full-resolution PSs. The energy efficiency is defined as the ratio of the sum-rate R_{sum} and the total power consumption P_{total}, i.e.,

$$\epsilon = \frac{R_{sum}}{P_{total}} = \frac{R_{sum}}{P_T + N_{RF}P_{RF} + N_{PS}P_{PS_b}}, \tag{28}$$

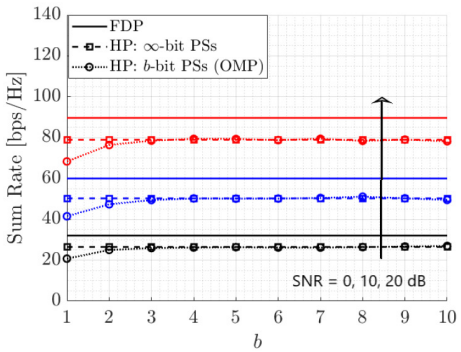
where P_T represents the total transmission power, N_{RF} is the total number of RF chains, P_{RF} is the power consumption of each RF chain, and P_{PS_b} is the power consumption of b-bit PSs. As in [39], we use $P_T = 1$ W and $P_{RF} = 250$ mW. The use of transfer block yields a constant-gain RF beamformer but also doubles the number of PSs per each RF chain due to double-PS structure (as discussed in Section III-B). Thus, $N_{PS} = 2 \times N_{RF} \times N$. We use 5 mW and 10 mW as the power consumption values for 1-bit and 2-bit PSs,



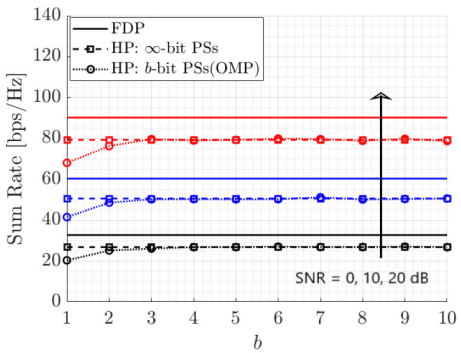
(a) ULA



(b) UCA

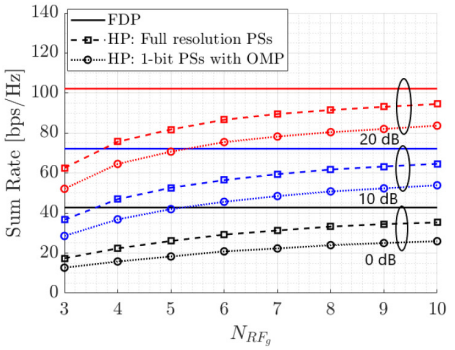


(c) URA

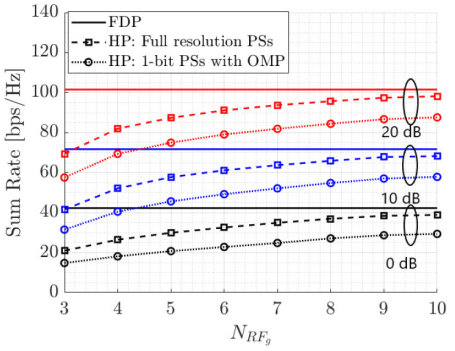


(d) CCA

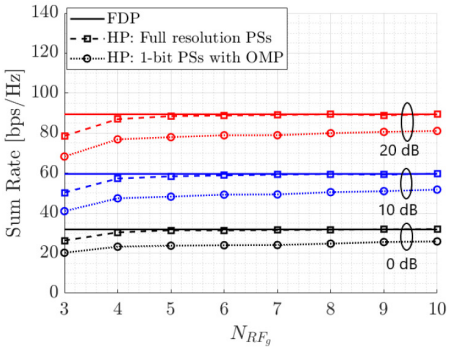
FIGURE 6. Sum-rate vs b at SNR = (0,10,20) dB: (a) ULA (b) UCA (c) URA (d) CCA.



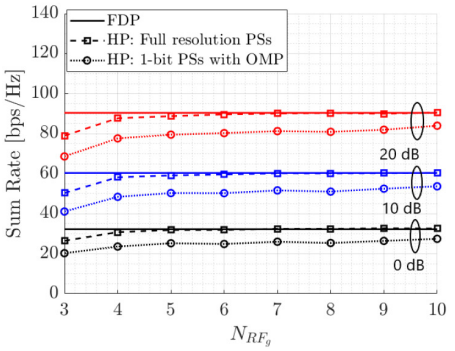
(a) ULA



(b) UCA



(c) URA



(d) CCA

FIGURE 7. Sum-rate vs N_{RF_g} with 1-bit PSs at SNR = (0,10,20) dB: (a) ULA (b) UCA (c) URA (d) CCA.

i.e., ($P_{PS_1} = 5$ mW, $P_{PS_2} = 10$ mW) [55], [56]. Figure 9 analyzes the energy efficiency of ULA, UCA, URA, and CCA vs. transmit power, which varies from -30 dB to 30 dB and compares ϵ at SNR = 15 dB for four cases which are: (i) FDP, (ii) HP using ∞-bit PSs, (iii) HP using 1-bit PSs, and (iv) HP using 2-bit PSs. For the full-resolution PSs, we assume 5-bit PSs be a good choice to replicate the performance of ∞-bit PSs.

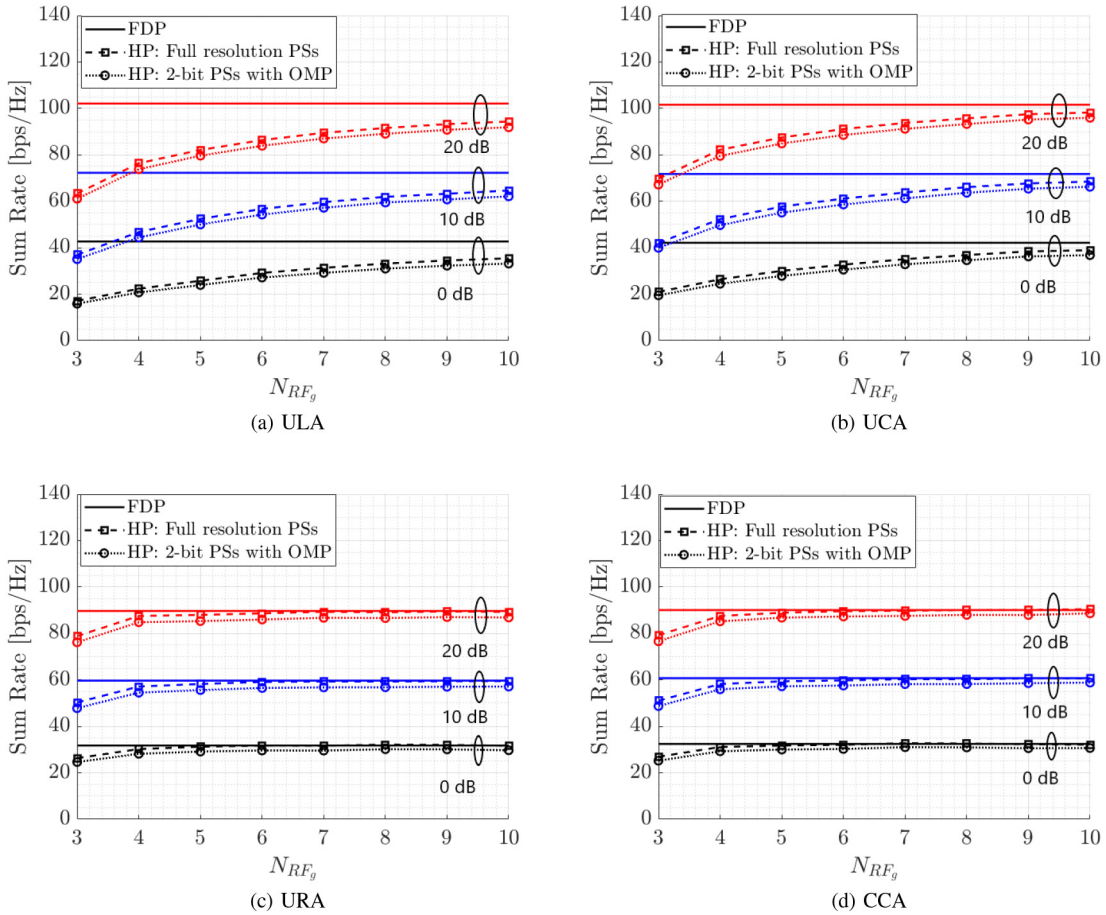


FIGURE 8. Sum-rate vs N_{RF_g} with 2-bit PSs at SNR = (0,10,20) dB: (a) ULA (b) UCA (c) URA (d) CCA.

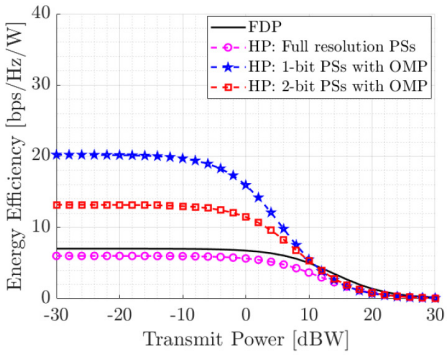
Figure 9 shows that both 1-bit and 2-bit PSs can offer high energy efficiency compared to FDP and HP using ∞-bit PSs. However, the energy efficiency decreases sharply beyond 0 dB for all array structures because of high transmit power. Compared to ULA and UCA, both URA and CCA can offer higher energy efficiency when using only 1 or 2-bit PSs at the RF-stage.

B. HP WITH b-BIT PSs AND q-BIT DACS

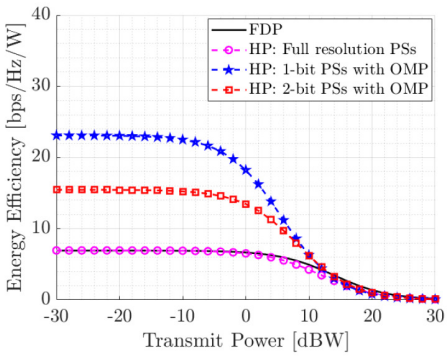
In this section, we examine the spectral efficiency of the proposed HP scheme with low-resolution DACs and PSs, i.e., the combined effect of q-bit DACs and b-bit PSs on the sum-rate performance for ULA, URA, UCA, and CCA. To better understand the effect of low-resolution PSs and DACs on the sum-rate degradation R_{deg}, we first plot R_{deg} due to each individual hardware component (i.e., DACs or PSs). Figure 10 shows various R_{deg} plots vs. b for different values of q. It can be seen that 1-bit DAC can give large sum-rate degradation ($\approx 65\%$). The large quantization noise introduced by the use of only 1-bit DAC impede the performance improvement even by increasing PSs resolution (i.e., b). However, by increasing DAC resolution, we can see that R_{deg} decreases significantly, e.g., R_{deg} reduces to $\approx 25\%$ with 5-bit DACs and 1-bit PSs. In other words, we can approach the performance of HP with ∞-bit resolution with just 5-bit DACs, offering significant advantages in power consumption and cost.

Figure 11 plots R_{deg} vs. q for different PS resolutions. Similar to the case of 1-bit DACs, HP using 1-bit PSs can give high R_{deg} ($\approx 70\%$). However, this degradation can be decreased significantly by increasing the resolution of DACs. Thus, HP with 1-bit PSs can provide higher sum-rate than HP with 1-bit DACs. Also, for $b \geq 2$, we can see that R_{deg} is almost the same for all resolution levels, i.e., $b = 2, 3, 4, 5$, which indicates that the performance of HP depends more on DAC resolution than on PS resolution. Furthermore, this analysis gives us the least numbers of bits of DACs and PSs required to achieve a satisfactory performance close to FDP and HP using ∞-bit components.

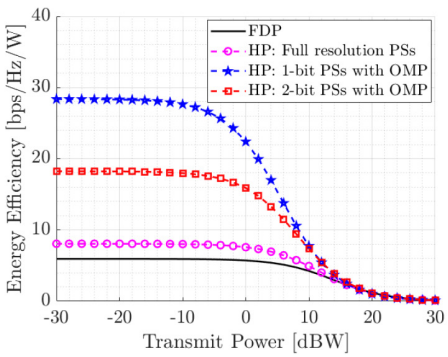
Figure 12 plots the sum-rate vs. q and b for ULA, UCA, URA, and CCA at SNR = 10 dB and b and q varying from 1 to 10. The sum-rate performance gap between the FDP and proposed HP using ULA and UCA is larger than the case using URA and CCA. Moreover, the analysis of the results obtained in Figure 10 and Figure 11 gives some following useful notes: (i) the use of low-resolution DACs significantly degrades the sum-rate performance as compared to the use of low-resolution PSs (e.g., $R_{(q=1,b=10)} \ll R_{(q=10,b=1)}$), (ii)



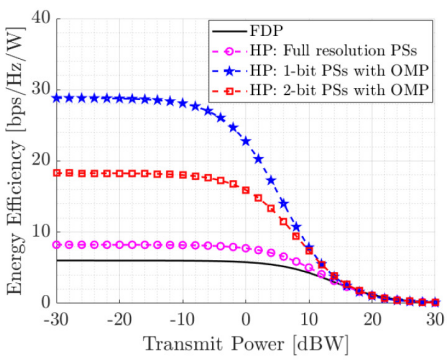
(a) ULA



(b) UCA

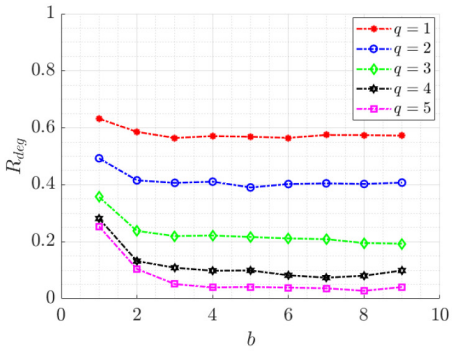


(c) URA

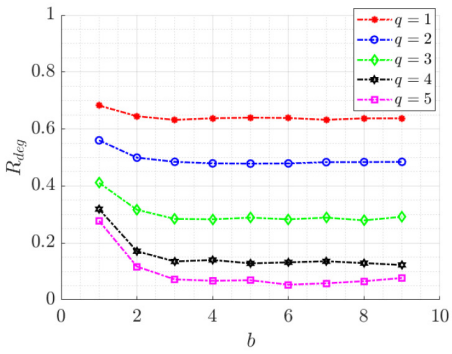


(d) CCA

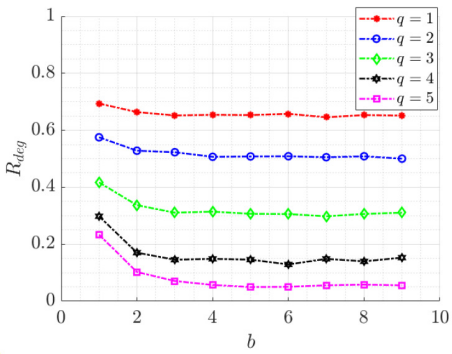
FIGURE 9. Energy efficiency comparison vs. transmit power at SNR= 15 dB: (a) ULA (b) UCA (c) URA (d) CCA.



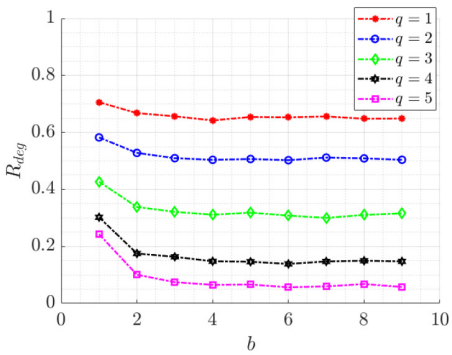
(a) ULA



(b) UCA



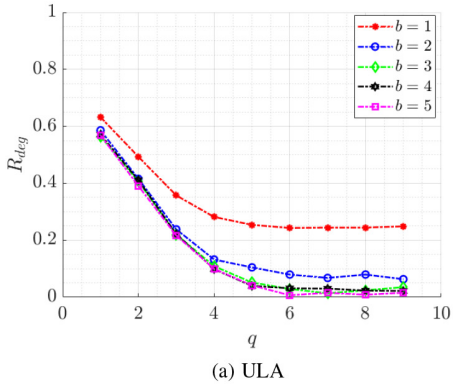
(c) URA



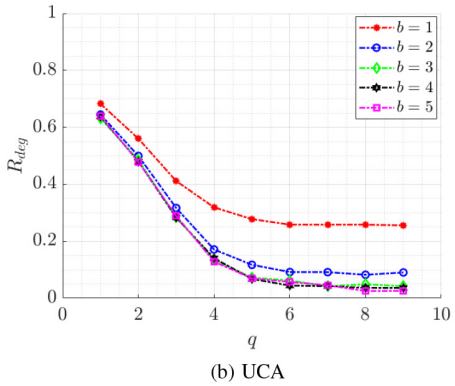
(d) CCA

FIGURE 10. Sum-rate degradation vs. b for different DAC resolutions at SNR = 10 dB: (a) ULA (b) UCA (c) URA (d) CCA.

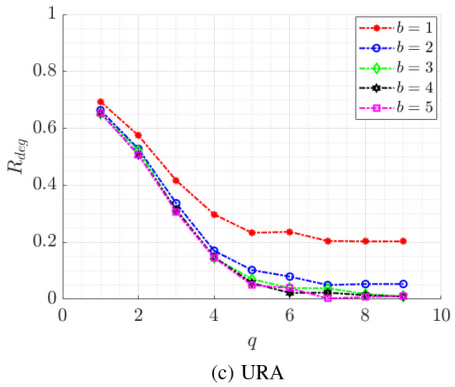
single-layered array structures (i.e., ULA and UCA) give low sum-rate, whereas, the multi-layered array structures (i.e., URA and CCA) can approach the sum-rate of FDP, and (iii) the optimal sum-rate can be obtained by using different combinations of q and b and thus, offers the choice to use which combination of low-resolution hardware components



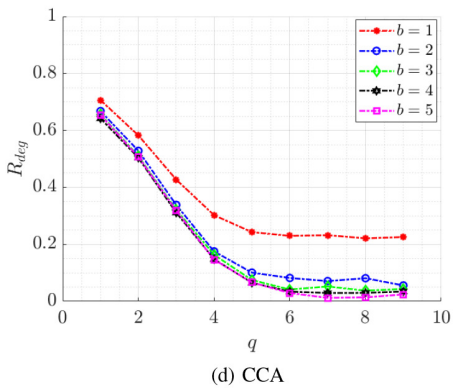
(a) ULA



(b) UCA

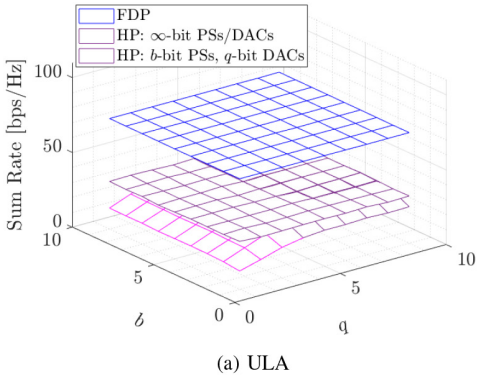


(c) URA

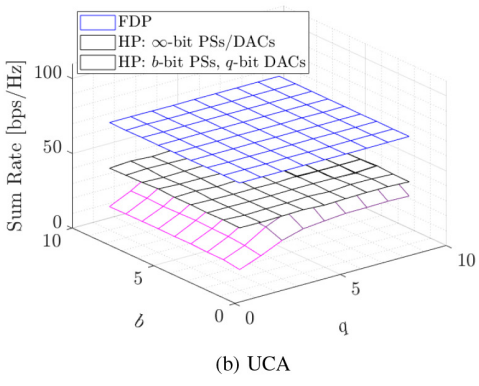


(d) CCA

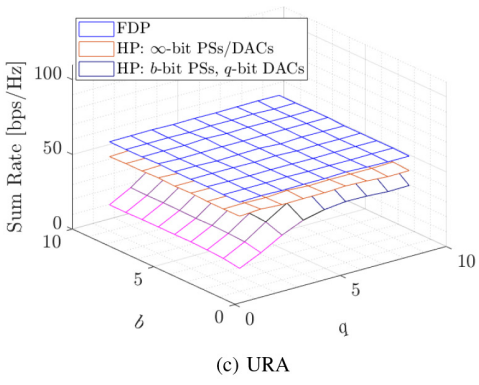
FIGURE 11. Sum-rate degradation vs. q for different PS resolutions at SNR = 10 dB: (a) ULA (b) UCA (c) URA (d) CCA.



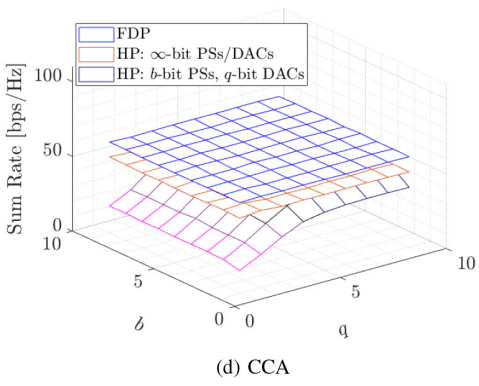
(a) ULA



(b) UCA



(c) URA



(d) CCA

FIGURE 12. Sum-rate vs. q and b at SNR = 10 dB: (a) ULA (b) UCA (c) URA (d) CCA.

for increased sum-rate. This can also be better understood by the equal-sum-rate contour plots shown in Figure 13. It can be observed that we can get reasonably high sum-rate by using the proposed HP with $(q \geq 5, b \geq 2)$, which is, in particular, close to that of FDP for both CCA and URA. Table 2 provides the sum-rate performance comparison for different quantization levels of DACs and PSs in mMIMO.

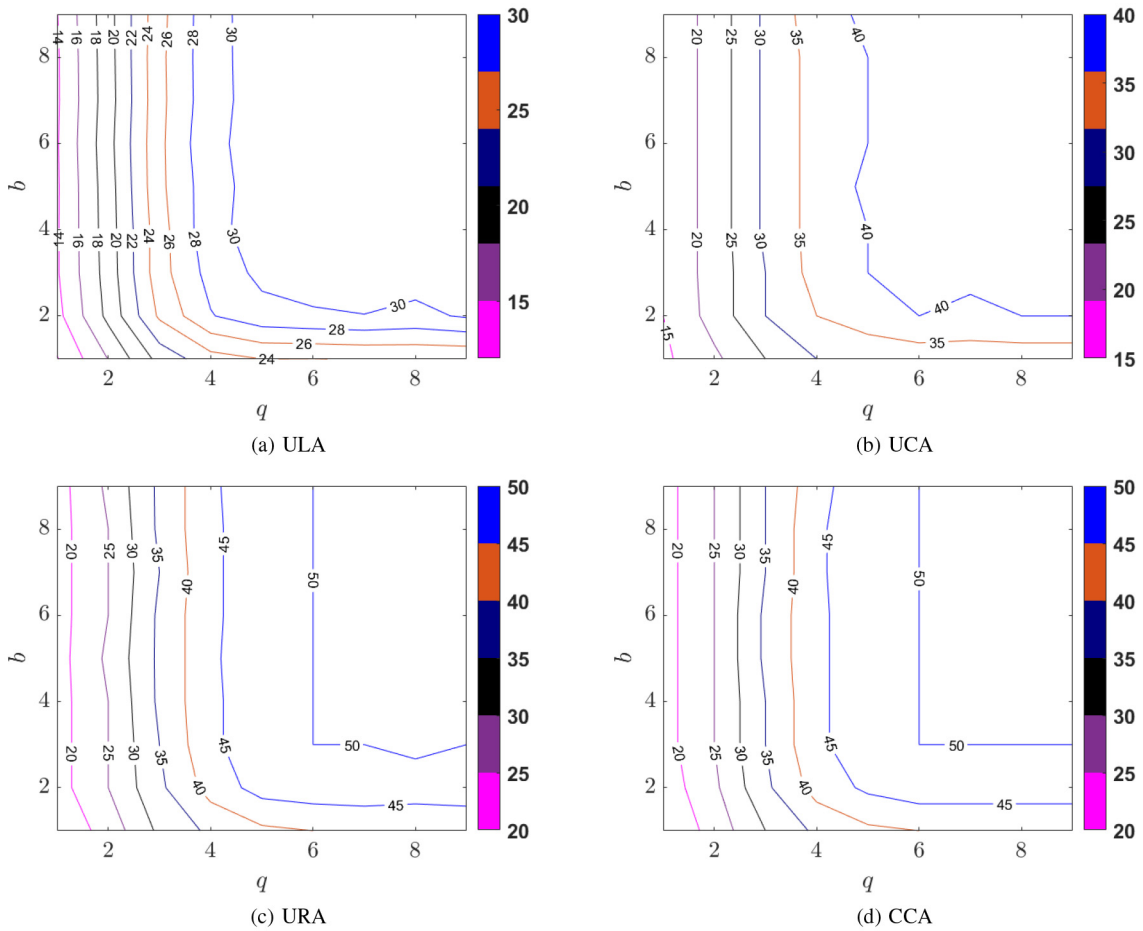


FIGURE 13. Equal-sum-rate contour plots at SNR = 10 dB: (a) ULA (b) UCA (c) URA (d) CCA.

TABLE 2. Performance comparison for different resolutions of DACs and PSs.

PSs \ DACs	$\infty-$ bit DACs	$q-$ bit DACs
$\infty-$ bit PSs	High sum-rate close to FDP	High sum-rate for $q \geq 5$
$b-$ bit PSs	High sum-rate for $b \geq 2$	High sum-rate for $q \geq 5, b \geq 2$

TABLE 3. Performance comparison of ULA, UCA, URA and CCA using low-resolution PSs and DACs.

Comparison paramter	ULA	UCA	URA	CCA
Spatial efficiency	Low	Low	High	High
Sum-rate using b-bit PSs (see Figure 4 - Figure 8)	Low	High	High	High
Energy efficiency using b-bit PSs (see Figure 9)	Low	High	High	High
Sum-rate using q-bit DACs and b-bit PSs (see Figure 10 - Figure 11)	Low	Low	High	High

Finally, Table 3 gives the comprehensive comparison of spatial, spectral and energy efficiencies of ULA, UCA, URA, and CCA when using few-bit DACs and PSs.

VI. CONCLUSION

In this paper, we have presented the MU-mMIMO hybrid precoder (HP) design for low-resolution PSs and/or DACs, and investigated its performance using different 2D antenna array structures. In the HP design, the RF precoder has been developed via the slow time-varying AoD information, whereas the reduced-size effective CSI is utilized at the baseband precoder. Considering the hardware complexity, we have considered the following cases in the HP design: (i) low-resolution PSs only, and (ii) low-resolution PSs and DACs. An optimization problem has been formulated to provide an RF beamformer with uniform gain entities by introducing a transfer block at the baseband stage. Furthermore, we have proposed an algorithmic solution using orthogonal matching pursuit (OMP) to reduce the Euclidean distance between the HP with full-resolution PSs and the HP with low-resolution PSs. Based on OMP, we have first designed HP using b-bit PSs and ∞-bit DACs, then constructed the HP using b-bit PSs and q-bit DACs to significantly reduce hardware complexity and costs in MU-mMIMO HP systems. The spectral and energy efficiencies of ULA, UCA, URA, and CCA have been compared using Monte Carlo simulations. Illustrative results indicate that the

proposed HP using both URA and CCA can provide high sum-rates, which are comparable to their FDP counterparts. Also, the use of low-resolution PSs and DACs can provide high energy efficiency. Finally, we have presented different combinations of PS and DAC quantization levels to achieve near-optimal sum-rate. It has been shown that HP using only 2-bit PSs and 5-bit DACs can provide high performance close to FDP.

APPENDIX

A. PROOF OF LEMMA 1:

Let the normalization factor $\frac{N_{RF_g}}{\left\|\hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F}$ is equal to $\frac{1}{\mu}$. Then, we can write:

$$\left\|\hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F = \mu N_{RF_g} = \mu\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g\right\|_F. \tag{29}$$

Using norm inequality, we have:

$$\begin{aligned}\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F &\geq \left|\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g\right\|_F - \left\|\hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F\right|,\\ &= |1-\mu|\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g\right\|_F, \end{aligned}\tag{30}$$

which is equivalent to $\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g\right\|_F \leq \frac{1}{|\mu-1|}\zeta$. When $\mu \neq 1$, we have $\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F \neq 0$. Then,

$$\begin{aligned}\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\tilde{\mathbf{T}}_g^{(q)}\right\|_F &= \left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)} + \left(1-\frac{1}{\mu}\hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right)\right\|_F, &(31)\\ &\leq \left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\mathbf{T}_g^{(q)}\right\|_F + \left|1-\frac{1}{\mu}\right|\left\|\hat{\mathbf{B}}_{A_g}^{(q)}\tilde{\mathbf{T}}_g^{(q)}\right\|_F, &(32)\\ &\leq \zeta + |\mu-1|\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g\right\|_F\\ &\leq \zeta + \frac{|\mu-1|}{|\mu-1|}\zeta,\\ &= 2\zeta. &(33)\end{aligned}$$

From (33), we can write as:

$$\left\|\hat{\mathbf{B}}_{A_g}\mathbf{T}_g - \hat{\mathbf{B}}_{A_g}^{(q)}\tilde{\mathbf{T}}_g^{(q)}\right\|_F = 2\zeta. \tag{34}$$

REFERENCES

[1] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Techn. J.*, vol. 20, pp. 11–22, Mar. 2015.

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

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[5] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018.

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

[7] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[8] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog–digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[9] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[10] S. He, C. Qi, Y. Wu, and Y. Huang, "Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems," *IEEE Access*, vol. 4, pp. 9895–9905, 2016.

[11] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[12] A. Li and C. Masouros, "Hybrid analog-digital millimeter-wave MU-MIMO transmission with virtual path selection," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 438–441, Feb. 2017.

[13] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[14] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[15] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[16] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, Jun. 2015.

[17] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[18] Y. Jeon, C. Song, S.-R. Lee, S. Maeng, J. Jung, and I. Lee, "New beamforming designs for joint spatial division and multiplexing in large-scale MISO multi-user systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3029–3041, May 2017.

[19] A. Koc, A. Masmoudi, and T. Le-Ngoc, "Hybrid beamforming for uniform circular arrays in multi-user massive MIMO systems," in *Proc. 32nd IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Edmonton, AB, Canada, May 2019, pp. 1–4.

[20] R. Mai and T. Le-Ngoc, "Nonlinear hybrid precoding for coordinated multi-cell massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2459–2471, Mar. 2019.

[21] S. Zarei, W. Gerstacker, and R. Schober, "Low-complexity hybrid linear/Tomlinson-Harashima precoding for downlink large-scale MU-MIMO systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Washington, DC, USA, Dec. 2016, pp. 1–7.

[22] C. Pradhan, A. Li, L. Zhuo, Y. Li, and B. Vucetic, "Hybrid-precoding for mmWave multi-user communications in the presence of beam-misalignment," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6083–6099, Sep. 2020.

[23] F. Sohrabi and W. Yu, "Hybrid beamforming with finite-resolution phase shifters for large-scale MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun./Jul. 2015, pp. 136–140.

[24] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[25] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.

[26] Z. Wang, M. Li, Q. Liu, and A. L. Swindlehurst, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

[27] H. Li, Q. Liu, Z. Wang, and M. Li, "Transmit antenna selection and analog beamforming with low-resolution phase shifters in mmWave MISO systems," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1878–1881, Sep. 2018.
[28] H. Li, Q. Liu, Z. Wang, and M. Li, "Joint antenna selection and analog precoder design with low-resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 967–971, Jan. 2019.
[29] H. Li, M. Li, and Q. Liu, "Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 602–614, Jan. 2020.
[30] H. Seleem, A. I. Sulyman, and A. Alsanie, "Hybrid precoding-beamforming design with Hadamard RF codebook for mmWave large-scale MIMO systems," *IEEE Access*, vol. 5, pp. 6813–6823, 2017.
[31] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution PSs for wideband mmWave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2168–2181, Sep. 2020.
[32] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Finite-alphabet precoding for massive MU-MIMO with low-resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4706–4720, Jul. 2018.
[33] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.
[34] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.
[35] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.
[36] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 484–498, Jun. 2018.
[37] A. Koc and T. Le-Ngoc, "Hybrid millimeter-wave massive MIMO systems with low CSI overhead and Few-Bit DACs/ADCs," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Victoria, BC, Canada, Nov./Dec. 2020, pp. 1–5.
[38] A. Koc, A. Masmoudi, and T. Le-Ngoc, "Angular-based 3D hybrid precoding for URA in multi-user massive MIMO systems," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Honolulu, HI, USA, Sep. 2019, pp. 1–5.
[39] A. Koc, A. Masmoudi, and T. Le-Ngoc, "3D angular-based hybrid precoding and user grouping for uniform rectangular arrays in massive MU-MIMO systems," *IEEE Access*, vol. 8, pp. 84689–84712, 2020.
[40] A. Koc, A. Masmoudi, and T. Le-Ngoc, "3D angular-based hybrid precoding for multi-cell MU-massive-MIMO systems in C-RAN architecture," in *Proc. IEEE 31st Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, London, U.K., Aug./Sep. 2020, pp. 1–6.
[41] W. Zheng, A. Koc, and T. Le-Ngoc, "Sub-connected hybrid precoding architectures in massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
[42] M. Mahmood, A. Koc, and T. Le-Ngoc, "2D antenna array structures for hybrid massive MIMO precoding," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
[43] M. Mahmood, A. Koc, and T. Le-Ngoc, "Massive-MIMO hybrid precoder design using few-bit DACs for 2D antenna array structures," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, Canada, Jun. 2021.
[44] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[45] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[46] J. D. Krieger, C.-P. Yeang, and G. W. Wornell, "Dense delta-sigma phased arrays," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1825–1837, Apr. 2013.
[47] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[48] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[49] Y. Xiong, "Achievable rates for massive MIMO relaying systems with variable-bit ADCs/DACs," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 991–994, May 2020.
[50] C. Kong, C. Zhong, S. Jin, S. Yang, H. Lin, and Z. Zhang, "Full-duplex massive MIMO relaying systems with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5033–5047, Aug. 2017.
[51] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.
[52] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
[53] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.
[54] R. L. Haupt, "Optimized element spacing for low sidelobe concentric ring arrays," *IEEE Trans. Antennas Propag.*, vol. 56, no. 1, pp. 266–268, Jan. 2008.
[55] J.-C. Chen, "Energy-efficient analog combiner design using low-resolution phase shifters and antenna selection for mmWave D2D communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13979–13984, Nov. 2020.
[56] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

MOBEEN MAHMOOD (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2013, and the M.Sc. degree (Hons.) in electrical engineering from the American University of Sharjah (AUS), Sharjah, UAE, in 2019. He is currently pursuing the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada.

He served as the Graduate Teaching Assistant and a Graduate Research Assistant with AUS from 2017 to 2019. From 2014 to 2017, he was with China Mobile Pakistan, Islamabad, Pakistan. His main research interests include massive MIMO, hybrid precoding and array processing. He is the recipient of AUS Graduate Teaching Assistantship, AUS Graduate Research Assistantship, McGill Graduate Excellence Fellowship, McGill Engineering Class of 1936 Fellowship, and J. W. McConnell Memorial Fellowship as part of McGill Engineering Doctoral Award.

ASIL KOC (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in electronics and communication engineering and the M.Sc. degree (Hons.) in telecommunication engineering from Istanbul Technical University, Istanbul, Turkey, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with McGill University, Montreal, QC, Canada.

From 2015 to 2017, he was a Research and a Teaching Assistant with the Electronics and Communication Engineering Department, Istanbul Technical University. Since 2017, he has been a Teaching Assistant with the Electrical and Computer Engineering Department, McGill University. His research interests include, but not limited to wireless communications, massive MIMO, full-duplex, spatial modulation, energy harvesting, and cooperative networks. He was a recipient of the Erasmus Scholarship Award funded by the European Union, the McGill Engineering Doctoral Award, the STARaCom Collaborative Grant funded by the FRQNT, and the Graduate Research Enhancement and Travel Award funded by McGill University.

THO LE-NGOC (Life Fellow, IEEE) received the B.Eng. degree in electrical engineering in 1976, the M.Eng. degree in microprocessor applications from McGill University, Montreal, in 1978, and the Ph.D. degree in digital communications from the University of Ottawa, Canada, in 1983.

From 1977 to 1982, he was with Spar Aerospace Ltd., Sainte-Anne-de-Bellevue, QC, Canada, involved in the development and design of satellite communications systems. From 1982 to 1985, he was with SRTelecom, Inc., SaintLaurent, QC, Canada, where he developed the new point-tomultipointDATDMA/TDM Subscriber Radio System SR500. From 1985 to 2000, he was a Professor with the Department of Electrical and Computer Engineering, Concordia University, Montreal. Since 2000, he has been with the Department of Electrical and Computer Engineering, McGill University. His research interest includes broadband digital communications. He was a recipient of the 2004 Canadian Award in Telecommunications Research and the IEEE Canada Fessenden Award in 2005. He is a Distinguished James McGill Professor, and a Fellow of the Engineering Institute of Canada, the Canadian Academy of Engineering, and the Royal Society of Canada.