IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 28, 2019, accepted September 19, 2019, date of publication September 27, 2019, date of current version October 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944061

Hybrid Precoding for Beamspace MIMO Systems With Sub-Connected Switches: A Machine Learning Approach

TING DING[1,2], YONGJUN ZHAO[1], LIXIN LI[3], (Member, IEEE), DEXIU HU[1], AND LEI ZHANG[4], (Senior Member, IEEE)

[1]National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450001, China
[2]Henan High-speed Railway Operation and Maintenance Engineering Research Center, Zhengzhou 450000, China
[3]School of Electronics and Information, Northwestern Polytechnical University, Xi' an 710129, China
[4]School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

Corresponding author: Ting Ding (dingting_ndsc@foxmail.com)

This work was supported by the National Natural Science Foundation of China under Grant NSFC61703433.

ABSTRACT By employing lens antenna arrays, the number of radio frequency (RF) chains in millimeter-wave (mmWave) communications can be significantly reduced. However, most existing studies consider the phase shifters (PSs) as the main components of the analog beamformer, which may result in a significant loss of energy efficiency (EE). In this paper, we propose a switch selecting network to solve this issue, where the analog part of the beamspace MIMO system is realized by a sub-connected switch selecting network rather than the PS network. Based on the proposed architecture and inspired by the cross-entropy (CE) optimization developed in machine learning, an optimal hybrid cross-entropy (HCE)-based hybrid precoding scheme is designed to maximize the achievable sum rate, where the probability distribution of the hybrid precoder is updated by minimizing CE with unadjusted probabilities and smoothing constant. Simulation results show that the proposed HCE-based hybrid precoding can not only effectively achieve the satisfied sum-rate, but also outperform the PSs schemes concerning energy efficiency.

INDEX TERMS mmWave Massive MIMO, hybrid precoding, beamspace, machine learning, cross-entropy, lens array.

I. INTRODUCTION

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) can achieve orders of magnitude increase in spectral efficiency, which makes it a promising technique for 5G-and-Beyond-wireless communication systems [1]–[3]. In conventional MIMO systems, each antenna requires one radio-frequency (RF) chain to realize the full freedom in terms of multiplexing gain. However, the employment of a large number of antennas in mmWave massive MIMO systems leads to a large number of RF chains, which will result in unaffordable hardware cost and power consumption [4].

To reduce hardware cost and achieve high array gain, the hybrid analog-digital precoding has been proposed [5] [6], where the typical analog beamformer is composed by a network of phase shifters (PSs) [7], [8]. In most of existing hybrid precoding architectures, each RF chain is connected to all antennas with PSs, which requires a complex PS network at link ends [9]. To reduce the number of RF chains in mmWave massive MIMO systems, one approach of hybrid precoding was provided, which decompose the fully digital precoder into a large-dimension analog precoder and a small-dimension digital precoder [10], [11].

Another approach of beamspace MIMO has also been proposed by utilizing discrete lens antenna arrays to transform the conventional spatial channel into the beamspace channel [12], [13]. Since the beamspace channel is sparse, a small number of more powerful beams is selected to reduce the dimension of the MIMO system [14]–[16]. Several beam selection algorithms have been discussed in previous works. By employing uniform linear arrays (ULA), a simple scheme for the beam selection with maximum magnitude (MM) is proposed in [13]. In [16], X. Gao et al. proposed

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

an interference-ware beam selection (IA-BS) to address the problem of inter-users interference when the same beam is assigned to different users. However, the existing architectures still require energy-intensive analog PS network.

In order to solve this issue, the switches network instead of the PSs network [17] can be applied in the analog network, as the energy consumption of the switch is much less than that of the PS. However, the optimal solution for the architecture is still an open problem.

In addition, machine learning has great advantages in structured information and massive data, and also have a wide range of applications and developments in many fields, including 5G wireless physical-layer [18]–[20]. In [18], several potential applications based on deep learning is reviewed. In [19], [20], [22], [23], machine learning inspired hybrid precoding scheme for massive MIMO is presented, and their effectiveness has also been confirmed.

In this paper, an energy-saving switch selecting network with lens array architecture is proposed to solve the energy consumption problem, where the analog part of the proposed architecture is realized by a sub-connected switch network with lens array. By incorporating an unadjusted probability and smoothing constant to the conventional cross-entropy (CE) method in machine learning [21], a hybrid CE (HCE) based precoding scheme for the proposed architecture is further designed to maximize the achievable sum-rate. Specifically, the analog precoder samples are generated according to the probability distributions, and the corresponding digital precoder samples are optimized by the classical wiener filter (WF) digital precoder scheme. Then, these hybrid precoders are weighted according to their achievable sum-rate. Then, we adopt the unadjusted probabilities and smoothing constant for updating the probability distribution of hybrid precoder by minimizing CE. Repeating such procedures, the optimal hybrid precoder with sufficient distributions is finally obtained.

The contributions in this paper are summarized as follows:

- We propose a novel sub-connected switch selecting network architecture with lens array, which is more feasible than the full digital precoding architecture.
- Based on the proposed architecture, an optimal hybrid cross-entropy (HCE)-based hybrid precoding scheme is proposed, which is close to optimal performance with a low-cost hybrid precoding architecture.
- By comparing four beamspace MIMO schemes, the proposed algorithm is verified to achieve satisfactory sum-rate performance and much higher energy efficiency (EE) than PS network schemes.

Symbol descriptions: Symbol descriptions used in this paper are defined in Table 1.

II. SYSTEM MODEL OF MMWAVE MASSIVE MIMO

The mmWave massive MIMO system is considered in the paper, where the base station (BS) is equipped with N transmit antennas and N_{RF} RF chains to serve K single-antenna users simultaneously.

TABLE 1. List of notations.

B	Matrix B
a	Vector a
$[B]_{m,n}$	(m,n)-th element of the matrix B
B^T	Transpose of the matrix B
B^*	Conjugate of the matrix B
B^{-1}	Inverse of the matrix B
B^H	Conjugate transpose of the matrix B
$tr\,(B)$	Trace of the matrix B
$abs\,(B)$	Take the absolute value of the matrix B
$\|B\|_F$	Frobenius norm of the matrix B
$\|B\|$	Determination of the matrix B
I_{N_r}	Identity matrix of $N_r \times N_r$
$\mathcal{CN}\,(\mu,\delta)$	Complex circular Gaussian distribution with mean μ and covariance δ^2
$\mathbf{1}_{M\times N}$	All one matrix of $M \times N$

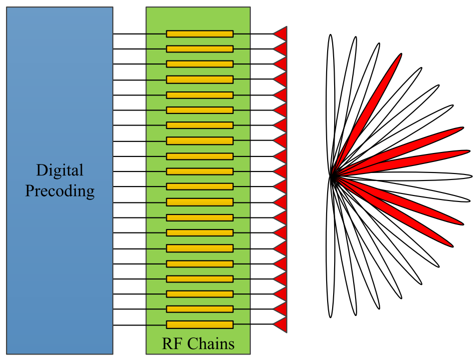

FIGURE 1. Traditional MIMO in the spatial domain.

A. TRADITIONAL MIMO IN THE SPATIAL DOMAIN

The traditional MIMO in the spatial domain is shown in Fig. 1, in the downlink, the $K \times 1$ received signal vector $\mathbf{y}$ for the K users can be presented as

$$\mathbf{y} = \mathbf{H}^H\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]$ is the channel matrix containing all users and $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ for $k = 1, 2, \cdots, K$. $\mathbf{P} \in \mathbb{C}^{N\times K}$ is the precoding matrix with the total power constraint, i.e., $E\left\{\|\mathbf{P}\|_F^2\right\} = P_T$, where P_T is the transmitted power. $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T$ is the $K \times 1$ vector of independent symbols for different users. And $\mathbf{n}$ of size $K \times 1$ is the vector of additive white Gaussian noise (AWGN) with zero-mean and variance σ^2.

B. BEAMSPACE MIMO

The conventional channel in the spatial domain can be converted to the beamspace channel by employing a carefully designed discrete lens array (DLA) [13], as shown in Fig. 2. The lens antenna array contains the array steering vectors

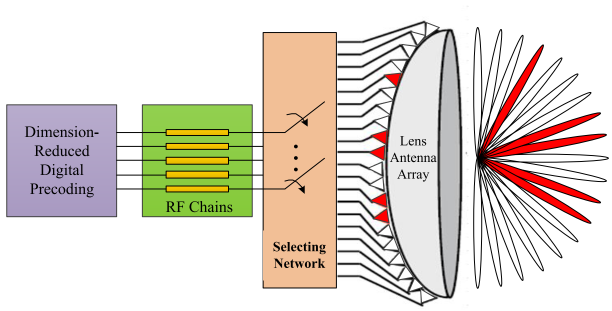

FIGURE 2. **Beamspace MIMO.**

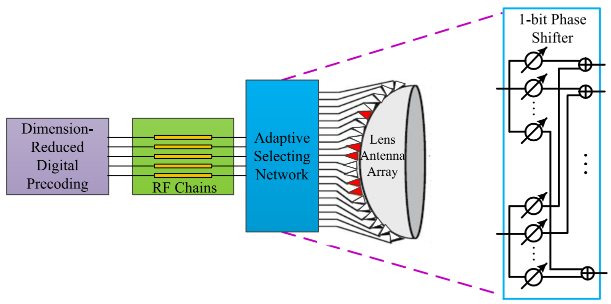

FIGURE 3. **The hybrid precoding architectures with full-connected phase shifters.**

of N orthogonal directions (beams) covering the entire angle space as [13],

$$\mathbf{U} = \left[\alpha\left(\bar{\varphi}_1\right), \alpha\left(\bar{\varphi}_2\right), \cdots, \alpha\left(\bar{\varphi}_N\right)\right]^H, \tag{2}$$

where $\bar{\varphi}_n = \frac{1}{N}\left(n - \frac{N+1}{2}\right)$, for $n = 1, 2, \cdots, N$, represents the normalized spatial direction. $\alpha(\varphi)$ is the array steering vector. For the typical ULA with N antennas, we have

$$\alpha(\varphi) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\varphi m}\right]_{m\in\tau(N)}, \tag{3}$$

where $\tau(N) = \{l - (N-1)/2, l = 0, 1, \cdots, N-1\}$ is a symmetric set of indices centered around zero. The spatial direction is defined as $\varphi \triangleq \frac{d}{\lambda} sin\theta$, where θ is the physical direction, λ is the antenna spacing which usually satisfies $d = \lambda/2$ at mmWave frequencies.

By employing the DLA to include negligible performance loss [14]–[16], the beamspace MIMO can transform the conventional spatial channel to the beamspace channel to capture the channel sparsity at mmWave frequencies. Fig. 3 is a finite-resolution PS-based hybrid precoding architecture, where each RF chain is connected to all antennas via PSs. It can achieve the full array gains and the near-optimal performance. However, it usually requires a large number of phase shifters and suffers from the high cost of hardware and consumption of energy.

The system model of beamspace MIMO can be represented by

$$\tilde{\mathbf{y}} = \mathbf{H}^H\mathbf{U}^H\mathbf{P}\mathbf{s} + \mathbf{n} = \tilde{\mathbf{H}}^H\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{4}$$

where $\tilde{\mathbf{y}}$ is the received signal vector in the beamspace, and the beamspace channel $\tilde{\mathbf{H}}$ is defined as $\tilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_K\right]$, where $\tilde{\mathbf{h}}_k$ is the beamspace channel of the k-th user, then

$$\tilde{\mathbf{H}} = \mathbf{U}\mathbf{H} = \left[\mathbf{U}\mathbf{h}_1, \mathbf{U}\mathbf{h}_2, \cdots, \mathbf{U}\mathbf{h}_K\right]. \tag{5}$$

In (5), the columns of $\tilde{\mathbf{H}}\left(\tilde{\mathbf{h}}_k\right)$ correspond to N orthogonal beams whose spatial directions are $\tilde{\varphi}_1, \tilde{\varphi}_2, \cdots, \tilde{\varphi}_N$, respectively. In mmWave propagation environments, the number of predominant scatters is quite limited. To reflect the sparse scatterers of the mmWave channel, we adopt the Saleh-Valenzuela channel model [13], [14], [16] in the paper, which is widely used as (6), $\mathbf{h}_k$ of the kth user can be expressed by

$$\mathbf{h}_k = \sqrt{\frac{N}{L+1}}\sum_{l=1}^{L}\beta_k^{(l)}\mathbf{a}\left(\varphi_k^{(l)}, \theta_k^{(l)}\right) = \sqrt{\frac{N}{L+1}}\sum_{l=0}^{L}\mathbf{c}_i, \tag{6}$$

where $\varphi_k^{(l)}$ and $\theta_k^{(l)}$ denote the azimuth (AOD) and angle of arrival (AOA) of the path of kth user, respectively. $\mathbf{a}\left(\varphi_k^{(l)}, \theta_k^{(l)}\right)$ represents the $N \times 1$ array steering vector. $\mathbf{c}_0 = \beta_k^{(0)}\mathbf{a}\left(\varphi_k^{(0)}\right)$ is the line-of-sight (LoS) component of the kth user with $\beta_k^{(0)}$ presenting the complex gain and $\varphi_k^{(0)}$ denoting the spatial direction, $\mathbf{c}_i = \beta_k^{(l)}\mathbf{a}\left(\varphi_k^{(l)}\right)$ for $1 \leq l \leq L$ is the lth non-line-of-sight (NLoS) component of the lth user, and L is the total number of NLoS components.

C. BEAMS SELECTION

The number of more powerful elements of each beamspace channel vector $\tilde{\mathbf{h}}_k$ is much smaller than N, that is because of the limitation of the number of dominant scatters [16]. The most important property of $\tilde{\mathbf{H}}$ is that it has a sparse structure representing the directions of the different users. It can be exploited that the number of RF chains is reduced without obvious performance loss from beam selection. Considering there are K single antenna users to be served, the received signal vector for the users in the downlink can be expressed by

$$\tilde{\mathbf{y}} \approx \tilde{\mathbf{H}}_r^H\mathbf{P}_r\mathbf{s} + \mathbf{n}, \tag{7}$$

where $\tilde{\mathbf{H}}_r$ is the channel matrix with reduced dimension, which is caused by the sparseness of the beam spatial channel matrix due to the limited scattering, and the switching network can reduce the MIMO dimension to $\tilde{\mathbf{H}}_r$ by selecting a small number of main beams. $\mathbf{P}_r$ is the reduced-dimension precoder.

From the perspective of hardware complexity, the dimension of $\mathbf{P}_r$ in (7) is much smaller than the full digital precoder matrix $\mathbf{P}$ in (1), so we can also observe that beamspace MIMO can obviously reduce the number of RF chains, the power consumption and complexity of the system as shown in Fig. 2.

III. HYBRID PRECODING DESIGN

In this section, we first describe the proposed sub-connected switch network with lens array architecture, and then an

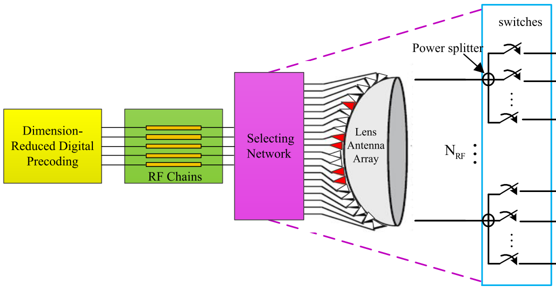

FIGURE 4. Proposed sub-connected switches based precoding architecture with lens array.

optimal HCE-based hybrid precoding scheme is designed to maximize the achievable sum-rate.

A. PROPOSED HYBRID PRECODING ARCHITECTURE

We propose a sub-connected switch based-hybrid precoding architecture with lens array for beamspace MIMO systems to overcome the high energy consumption issue. As shown in Fig. 4, the BS employs an N elements lens array and N_{RF} RF chains to serve K single-antenna users. The received signal $\tilde{\mathbf{y}} \in \mathbb{C}^{K\times 1}$ at users can be expressed by

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}^H \tilde{\mathbf{F}}_{RF} \tilde{\mathbf{F}}_{BB} \mathbf{s} + \mathbf{n}, \tag{8}$$

where $\tilde{\mathbf{F}}_{RF} = \left[\tilde{\mathbf{f}}_{RF}^{(1)}, \tilde{\mathbf{f}}_{RF}^{(2)}, \cdots, \tilde{\mathbf{f}}_{RF}^{(N_{RF})}\right] \in \mathbb{C}^{N\times N_{RF}}$ is the beamspace analog beamformer. $\tilde{\mathbf{F}}_{BB} = \left[\tilde{\mathbf{f}}_{BB}^{(1)}, \tilde{\mathbf{f}}_{BB}^{(2)}, \cdots, \tilde{\mathbf{f}}_{BB}^{(K)}\right] \in \mathbb{C}^{N_{RF}\times K}$ is the beamspace digital precoder.

Specifically, the $\tilde{\mathbf{F}}_{RF}$ is achieved via PSs and switches, $\tilde{\mathbf{F}}_{RF}$ should fulfill the following constraints.

In Fig. 4, the analog precoder matrix $\tilde{\mathbf{F}}_{RF}$ should be a block diagonal matrix instead of a full matrix as

$$\tilde{\mathbf{F}}_{RF} = \begin{bmatrix} \mathbf{f}_1^{RF} & 0 & \cdots & 0 \\ 0 & \mathbf{f}_2^{RF} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{f}_{N_{RF}}^{RF} \end{bmatrix}, \tag{9}$$

where $\mathbf{f}_n^{RF}$ is the $M\times 1$ analog precoder on the nth sub antenna array with $M = N/N_{RF}$. Since only the switches are used, all the N elements of $\tilde{\mathbf{F}}_{RF}$ should belong to

$$\frac{1}{\sqrt{N}}\{0,1\}. \tag{10}$$

Due to the special hardware constraints of (9) and (10), it is difficult to extend conventional beamspace MIMO precoding schemes [13], [16] to the hybrid precoding algorithm based on the proposed architecture. As a result, we need to conceive an efficient precoding scheme with near-optimal performance, which is discussed in the following section.

IV. THE PROPOSED HYBRID ALGORITHM

A. PROPOSED HYBRID PRECODING SCHEME

Our aim is to design the analog beamformer $\tilde{\mathbf{F}}_{RF}$ and the digital precoder $\tilde{\mathbf{F}}_{BB}$ to maximize the achievable sum-rate R_{sum}, which can be presented as

$$\begin{aligned} \left(\tilde{\mathbf{F}}_{RF}^{opt}, \tilde{\mathbf{F}}_{BB}^{opt}\right) = \ & \underset{\tilde{\mathbf{F}}_{RF},\, \tilde{\mathbf{F}}_{BB}}{argmax} \ R_{sum}, \\ & s.t. \ \tilde{\mathbf{F}}_{RF} \in \Gamma, \\ & \left\|\tilde{\mathbf{F}}_{RF}\tilde{\mathbf{F}}_{BB}\right\|_F^2 = P_T, \end{aligned} \tag{11}$$

where Γ represents the set with all possible analog beamformers satisfying the constraints (10) and (11), $\tilde{\mathbf{f}}_{BB}^{(k)}$ is the kth column of $\tilde{\mathbf{F}}_{BB}$, and the sum-rate R_{sum} for a given channel realization can be presented by

$$R_{sum} = \sum_{k=1}^{K} log_2\left(1+\gamma_k\right), \tag{12}$$

where γ_k refers to the signal-to-interference-plus-noise ratio (SINR) of the kth user as

$$\gamma_k = \frac{\left|\tilde{\mathbf{h}}_k^H \tilde{\mathbf{F}}_{RF} \tilde{\mathbf{f}}_{BB}^{(k)}\right|^2}{\sum_{k'\neq k}^{K}\left|\tilde{\mathbf{h}}_k^H \tilde{\mathbf{F}}_{RF} \tilde{\mathbf{f}}_{BB}^{(k')}\right|^2 + \sigma^2}. \tag{13}$$

To solve the problem, we first decouple the joint design of $\tilde{\mathbf{F}}_{RF}$ and $\tilde{\mathbf{F}}_{BB}$. Given $\tilde{\mathbf{F}}_{RF}$, the constraints become convex, and $\tilde{\mathbf{F}}_{BB}$ can be obtained by classical schemes based on the effective channel. Since the number of possible $\tilde{\mathbf{F}}_{BB}$ is finite, the optimization problem actually can be solved by trying all possible $\tilde{\mathbf{F}}_{RF}$. Unfortunately, this will involve unaffordable complexity because of the large number of N. So we need to search a more intelligent scheme to reduce the complexity.

We first introduce the conventional CE method, which is a probabilistic model-based algorithm to solve the combining problem by an iterative procedure [21]. During the ith iteration, firstly, generate a random sample from a pre-specified probability distribution function. The CE method generates S possible precoder $\left\{\tilde{\mathbf{F}}_{RF}\right\}_{s=1}^{S}$ based on the probability distribution $\Xi\left(\Gamma;\mathbf{p}^{(i)}\right)$ as the candidates, where $\Xi\left(\Gamma;\mathbf{p}^{(i)}\right)$ represents the probability of $\tilde{\mathbf{F}}_{RF}^{(s)}$ producing, and $\mathbf{p}^{(i)}$ denotes the probability parameter of the ith iteration. Then, use the sample to modify the parameters of the probability distribution in order to produce a "elitist" samples in the next iteration. Specifically, compute the objective value R_{sum} of all candidates, and select $S_{elitist}$ best candidates as elitists. Finally, update the probability distribution by minimizing the CE according to the elitists preservation. We can achieve the optimal probability distribution by repeating such procedures. However, the conventional CE requires a very large samples for finding high-quality solutions and the search parameters are generally difficult to adjust.

To overcome the disadvantages and aid convergence and reduce the sample size, we recommend the use of an unadjusted probability and smoothing constant in the conventional CE, which is labeled the HCE-based hybrid precoding algorithm. Firstly, we formulate the non-zero elements in $\tilde{\mathbf{F}}_{RF}$

as $N \times 1$ vector

$$\mathbf{f} = \left[\left(\mathbf{f}_1^{RF}\right)^T, \left(\mathbf{f}_2^{RF}\right)^T, \cdots, \left(\mathbf{f}_{N_{RF}}^{RF}\right)^T\right]^T, \tag{14}$$

and set the probability parameter $\mathbf{p} = [p_1, p_2, \cdots, p_N]^T$, $0 \leq p_n \leq 1$. Each element of $\mathbf{f}$ is modeled as an independent Bernoulli random variable with a probability mass function

$$\begin{aligned} &Pr\left\{f_i = \frac{1}{\sqrt{N}}\right\} = p_i, \\ &Pr\{f_i = 0\} = 1 - p_i, \\ &\quad for\ i = 1, 2, \cdots, N. \end{aligned} \tag{15}$$

Thus, a family of Bernoulli pdfs associated with $\mathbf{f}$ can be written as

$$\Xi(\mathbf{f}; \mathbf{p}) = \prod_{n=1}^{N} p_n^{\sqrt{N}\mathbf{f}_n} (1 - p_n)^{1-\sqrt{N}\mathbf{f}_n}. \tag{16}$$

That is, first generate a random sample $\{\mathbf{f}^s\}_{s=1}^S$ from a probability distribution $\mathbf{p}^{(i)}$, and reformulate them in terms of matrix that pertain to Γ.

During the ith iteration, as all the N non-zero elements of $\mathbf{f}$ belong to $\frac{1}{\sqrt{N}}\{0, 1\}$.

After the initializing the probability distribution, we first generate the beamspace channel $\tilde{\mathbf{H}}$ in step 1.

In step 2, we then generate S possible precoder $\left\{\tilde{\mathbf{F}}_{RF}\right\}_{s=1}^S$ based on the probability distribution $\Xi\left(\mathbf{f}; \mathbf{p}^{(i)}\right)$ as samples.

In step 3, according to the effective channel $\tilde{\mathbf{H}}_{eq}^s = \tilde{\mathbf{H}}\tilde{\mathbf{F}}_{RF}^s$ for $1 \leq s \leq S$, we calculate the corresponding digital precoder $\tilde{\mathbf{F}}_{BB}^s$.

There are three major kinds of linear precoder schemes: the matched filter (MF), zero-forcing (ZF), and wiener filter (WF). In this paper, we adopt the classical WF digital precoder scheme with the near-optimal performance as an example, and $\tilde{\mathbf{F}}_{BB}^s$ can be calculated as

$$\mathbf{G}^s = \left(\tilde{\mathbf{H}}_{eq}^s \left(\tilde{\mathbf{H}}_{eq}^s\right)^H + \zeta \mathbf{I}\right)^{-1} \tilde{\mathbf{H}}_{eq}^s, \tag{17}$$

$$\zeta = \sigma^2 K / P_T, \tag{18}$$

$$\tilde{\mathbf{F}}_{BB}^s = \beta^s \mathbf{G}^s, \tag{19}$$

$$\beta^s = \sqrt{\frac{P_T}{tr\left(\mathbf{G}^s \Lambda (\mathbf{G}^s)^H\right)}}, \tag{20}$$

where $\Lambda = E[\mathbf{ss}]^H$ denotes the diagonal correlation matrix of $\mathbf{s}$.

In step 4 to 5, we calculate the $\left\{R_{sum}\left(\tilde{\mathbf{F}}_{RF}^s\right)\right\}_{s=1}^S$ and order the $\left\{R_{sum}\left(\tilde{\mathbf{F}}_{RF}^s\right)\right\}_{s=1}^S$ in a descend sequence. The sum-rate can be obtained by substituting $\tilde{\mathbf{F}}_{RF}^s$ and $\tilde{\mathbf{F}}_{BB}^s$ into (12) (13). We sort the calculated sum-rate in a descending sequence, and get the descending series $R_{sum}\left(\tilde{\mathbf{F}}_{RF}^{[1]}\right) \geq R_{sum}\left(\tilde{\mathbf{F}}_{RF}^{[2]}\right) \geq \cdots \geq R_{sum}\left(\tilde{\mathbf{F}}_{RF}^{[S]}\right)$.

In step 6, updating $\mathbf{P}^{(i+1)}$ by minimizing CE. In the conventional CE algorithm, the probability distribution is updated by minimizing CE according to the elitists' preservation, which is equivalent to solving

$$\mathbf{P}^{(i+1)} = \underset{\mathbf{P}^{(i)}}{arg\ max} \frac{1}{\lceil \rho N \rceil} \sum_{s=1}^{\lceil \rho N \rceil} ln\Xi\left(\tilde{\mathbf{F}}_{RF}^{[s]}; \mathbf{p}^{(i)}\right), \tag{21}$$

where ρ is the cutoff point for high-quality observations.

In the paper, unadjusted probabilities for the current iteration are adopted to avoid the search parameters optimum, which is denoted by

$$\nu_j = \frac{\sum_{s=1}^{\lceil \rho N \rceil} \sqrt{N} f_j^{(s)}}{\lceil \rho N \rceil}, \tag{22}$$

where $f_j^{(s)} \in \frac{1}{\sqrt{N}}\{0, 1\}, j = 1, 2, \cdots, N$.

Instead of directly updating parameter $\mathbf{P}^{(i+1)}$ to $\mathbf{P}^{(i)}$, a smoothing process is adopted to avoid the local optimum, which is denoted by

$$p_j^{(i+1)} = \alpha \nu_j + (1 - \alpha) p_j^{(i)}, \quad j = 1, 2, \cdots, N, \tag{23}$$

where α is smoothing constant for updating p. When $\alpha = 1$ the smoothing process degenerates to the original updating formulation (21).

In step 7 to 10, we update the best candidate and reset the number of iterations without improvement. Also, we update the global number of iterations.

Finally, the optimal analog beamformer and digital precoder will be calculated as $\tilde{\mathbf{F}}_{RF}^{(*)}$ and $\tilde{\mathbf{F}}_{BB}^{(*)}$ by repeating such procedure respectively until the maximum number of iterations I is met. The proposed HCE-based hybrid precoding scheme is summarized in **Algorithm 1**.

B. COMPUTATINAL COMPLEXITY ANALYSIS

In the subsection, the computational complexity of the proposed HCE-based hybrid precoding scheme is explained. In beamspace MIMO, the equivalent full-dimension precoder can be obtained by replacing $\mathbf{H}$ with $\tilde{\mathbf{H}}$. The full dimensional precoders require $\mathcal{O}(N)$ MIMO precoding for determining the $N \times K$ precoder matrix. However, $\mathcal{O}(N_{RF})$ MIMO precoding is needed for the reduced-dimensional beamspace MIMO determined by the $N_{RF} \times K$ $\tilde{\mathbf{H}}_r$. This reduces the hardware complexity from $\mathcal{O}(N)$ to $\mathcal{O}(N_{RF})$.

From **Algorithm 1**, during the ith iteration, we can observe that the complexity of the HCE-based hybrid precoding scheme only involves steps 3, 4, 5, and 6. In step 3, according to the effective channel $\tilde{\mathbf{H}}_{eq}^s = \tilde{\mathbf{H}}\tilde{\mathbf{F}}_{RF}^s$ for $1 \leq s \leq S$, we calculate the corresponding digital precoder $\tilde{\mathbf{F}}_{BB}^s$. So the complexity of the step is $\mathcal{O}\left(SN_{RF}K^2\right)$. In step 4 to 5, the achievable sum rate is calculated, we adopt the classical WF digital precoder scheme, the SINR_k of the kth user for the sth candidate, so the complexity of this step is $\mathcal{O}(S)$. And we order the sum-rate in a descend order, which is quite simple with the complexity $\mathcal{O}(S)$. Finally, in step 6, the probability $\mathbf{p}^{(i+1)}$ is updated by (22) and (23) with the complexity $\mathcal{O}(N_{RF}S)$.

Algorithm 1 HCE-Based Hybrid Precoding Scheme

Input: Channel matrix $\mathbf{H}$, Number of iterations I, Number of iterations without improvement I', Number of sample S, cutoff point for high-quality observations ρ, smoothing constant α.

Initialization: $i = 1$; $\mathbf{p}^{(0)} = \frac{1}{2} \times 1_{N\times 1}$.

While $\left(i' < I'\right)$ and $(i < I)$

1: Beamspace channel: $\tilde{\mathbf{H}} = \left[\tilde{h}_1, \tilde{h}_2, \cdots, \tilde{h}_K\right] = \mathbf{UH}$;

2: Randomly generate S candidate analog beamformers $\left\{\tilde{\mathbf{F}}^s_{RF}\right\}_{s=1}^{S}$ based on $\Xi\left(\Gamma; \mathbf{p}^{(i)}\right)$;

3: Compute S corresponding digital precoder $\left\{\tilde{\mathbf{F}}^s_{BB}\right\}_{s=1}^{S}$ based on (17), (18), (19) and (20);

4: Calculate the achievable sum-rate $\left\{R_{sum}\left(\tilde{\mathbf{F}}^s_{RF}\right)\right\}_{s=1}^{S}$;

5: Order the $\left\{R_{sum}\left(\tilde{\mathbf{F}}^s_{RF}\right)\right\}_{s=1}^{S}$ in a descend sort as $R_{sum}\left(\tilde{\mathbf{F}}^{[1]}_{RF}\right) \geq R_{sum}\left(\tilde{\mathbf{F}}^{[2]}_{RF}\right) \geq \cdots \geq R_{sum}\left(\tilde{\mathbf{F}}^{[S]}_{RF}\right)$;

6: Calculate v_j according to (22), and update $\mathbf{p}^{(i+1)}$ by (23);

7: **if** $\left(R_{sum}\left(\tilde{\mathbf{F}}^{[*]}_{RF}\right) \geq R_{sum}\left(\tilde{\mathbf{F}}^{[1]}_{RF}\right)\right)$

then $i' = i' + 1$

else

$\tilde{\mathbf{F}}^{[*]}_{RF} = \tilde{\mathbf{F}}^{[1]}_{RF}$, $i' = 0$, $i = i + 1$;

8: **End for**

9: Sum rate:

$$R_{sum} = \sum_{k=1}^{K} log_2\left(1 + \gamma_k\right),$$

$$\gamma_k = \frac{\left|\tilde{\mathbf{h}}_k^H \tilde{\mathbf{F}}_{RF} \tilde{\mathbf{f}}_{BB}^{(k)}\right|^2}{\sum_{k' \neq k}^{K} \left|\tilde{\mathbf{h}}_k^H \tilde{\mathbf{F}}_{RF} \tilde{\mathbf{f}}_{BB}^{(k')}\right|^2 + \sigma^2},$$

Output: Analog precoder $\mathbf{F}^{(*)}_{RF}$; Digital precoder $\mathbf{F}^{(*)}_{BB}$; Sum rate R_{sum}.

Therefore, after I iterations, the total complexity of the proposed HCE-based precoding scheme is $\mathcal{O}\left(SN_{RF}K^2 I\right)$, and the complexity of the proposed HCE-based hybrid precoding scheme is acceptable.

V. ENERGY EFFICIENCY

The energy efficiency (EE) is the ratio between the achievable sum-rate R_{sum} and the entire power consumption [16], which can be defined as,

$$EE = \frac{R_{sum}}{P_T + P_{BB} + P_H} (bps/Hz/W), \tag{24}$$

where P_{BB} is the power consumption of baseband, P_H is the energy consumption of hardware architecture.

The traditional MIMO in the spatial domain is shown in Fig. 1, where each RF chain is connected with one antenna. In this architecture, the energy consumption of hardware only involves RF chains. So we have

$$P_H = NP_{RF}, \tag{25}$$

where P_{RF} is the energy consumption of the RF chains. The number of required RF chains is equal to the number of antennas, i.e., $N_{RF} = N$, which is usually large for mm-wave massive MIMO systems, e.g., $N_{RF} = N = 256$.

The beamspace MIMO is shown in Fig. 2, where the selecting network is composed of switches and the number of switches is equal to the number of RF chains. The beamspace MIMO can significantly reduce the number of required RF chains without obvious performance loss [16]. In this architecture we have

$$P_H = N_{RF}P_{RF} + N_{RF}P_{SW}, \tag{26}$$

where P_{SW} is the energy consumption of switch, which is much lower than P_{PS}.

The hybrid precoding architectures with fully-connected phase shifters is shown in Fig. 3, where each RF chain is connected with all antennas of BS via phase shifters. In this architecture, we have

$$P_H = N_{RF}P_{RF} + N_{RF}NP_{PS} + NP_{COM}, \tag{27}$$

where P_{PS}, P_{COM}, P_{RF} is the energy consumption of PS, power combiner, RF chain, respectively.

The proposed sub-connected switches based precoding architecture with lens array, as shown in Fig. 4, where each RF chain in connection with one subarray. The architecture employs an amount (N) of energy-efficient switches. In this architecture we have

$$P_H = N_{RF}P_{RF} + N_{RF}NP_{SW} + N_{RF}P_{COM} + N_{RF}P_{SP}, \tag{28}$$

where P_{SP} is the energy consumption of power splitter.

VI. SIMULATION RESULTS

The simulation parameters are described as follows: the BS is equipped with ULA of $N = 256$ antennas and $N_{RF} = 16$ RF chains to serve $K = 16$ users simultaneously.

The mmWave MIMO channel is generated according to (6) with the following steps:

- One LoS component with $L = 2$ NLoS components;
- $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$, $\beta_k^{(l)} \sim \mathcal{CN}\left(0, 10^{-1}\right)$ for $l = 1, 2$;
- $\varphi_k^{(0)}$ and $\varphi_k^{(l)}$ follow the i.i.d. uniform distribution within $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

The rest of the simulation parameters is summarized as in Table 2. The sum-rate performance against the conventional two-stage hybrid precoding scheme is designed for PS-based architecture [11], and the conventional antenna selection (AS)-based hybrid precoding with switches in [17]. For the proposed HCE-based hybrid precoding scheme, we set the smoothing parameter $\alpha = 0.8$, and the cutoff point for high-quality observations $\rho = 0.2$, number of samples $S = 100$, number of iterations $I = 20$.

Fig. 5 shows that the proposed HCE algorithm outperforms the traditional CE algorithm. The proposed algorithm is superior to the traditional CE-based algorithm since the HCE algorithm only involves one additional step (i.e., step 6 in **Algorithm 1**) with low complexity. Moreover, we can observe from Fig. 5 that the proposed HCE-based hybrid

TABLE 2. **List of parameters.**

Parameters	Values
Wavelength (λ)	1
Antenna Spacing(d)	$\lambda/2$
Number of path(L)	3
Transmitted power (P_T)	30mW
Power consumption of baseband (P_{BB})	200mW
Energy consumed of each RF chain (P_{RF})	250mW
Energy consumed of phase shifter (P_{PS})	40mW
Energy consumed of power combiner (P_{COM})	20mW
Energy consumed of switch (P_{SW})	5mW
Energy consumed of power splitter (P_{SP})	5mW
Signal-to-noise ratio SNR	10dB

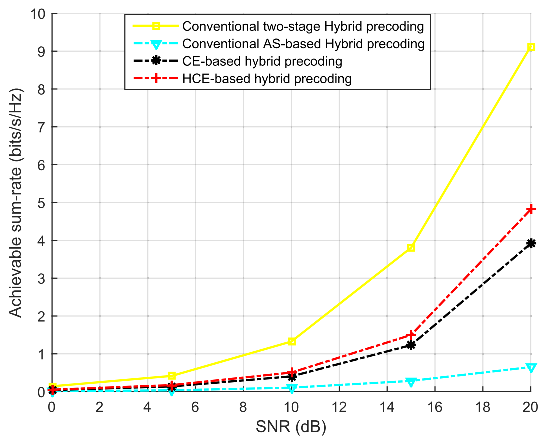

FIGURE 5. **Achievable sum-rate comparison.**

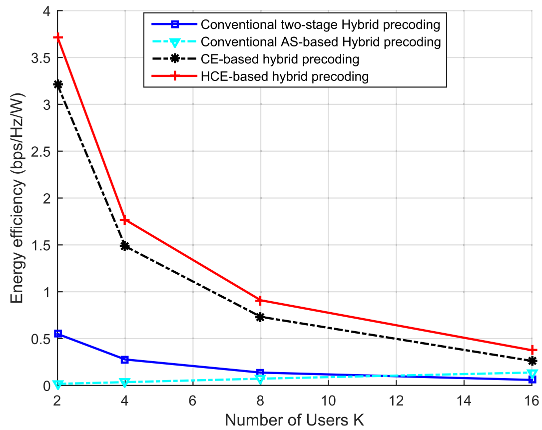

FIGURE 6. **Energy efficiency comparison.**

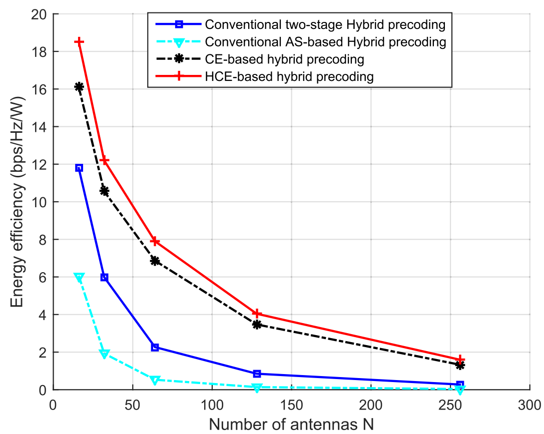

FIGURE 7. **Energy efficiency comparison as $K = 4$.**

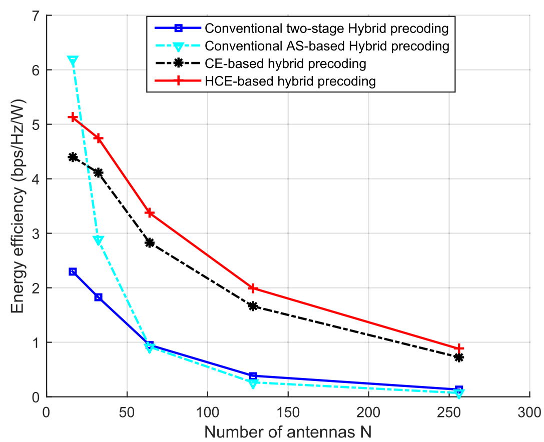

FIGURE 8. **Energy efficiency comparison as $K = 8$.**

precoding achieve much higher sum-rate than the conventional AS-based hybrid precoding, as it can get the potential array gains. Compared with the near-optimal PS-based architecture, We can observe that the proposed switch network based-architecture suffers from some loss of array gains. The performance loss is unavoidable because system architecture is not perfect, and part of the system array gain is lost. The proposed analog precoding architecture is more feasible than the full digital precoding architecture.

Fig. 6 shows the EE comparison when $N = 256$ is fixed and $N_{RF} = K$ varies from 2 to 16. The energy efficiency (EE) is defined as (24), the energy consumption is defined in section V. The parameters are summarized in Table 2. As the number of users increases, the EE of the hybrid precoding architecture based on the lens array is enormous. The energy consumption of the full digital precoding architecture is increasing as the number of users increases. This is because of the huge energy consumption of the numerous RF chains in the fully digital precoding architecture. In Fig. 6, it can be observed that the proposed HCE-based hybrid precoding can achieve much higher EE than the others, especially when K is not very large. Fig. 6 also shows that the EE of the two-stage hybrid precoding scheme with PS-based architecture is even lower than that of the conventional AS-based hybrid precoding with switches. This is due to that as the number of users K increases, the number of PSs in PS-based architecture increases markedly. Therefore, the energy consumption of PS will be large, even higher than that of switches and RF chains. Fig. 7 and Fig. 8 shows the EE comparisons of the four schemes, with $K = 4$ and $K = 8$ respectively and N varies from 16 to 256. From Fig. 7, it can be observed that the highest EE can always be achieved by the HCE-hybrid precoding scheme, which is based on the switch selection network architecture. Fig. 8 shows that as N grows, the EE of the HCE-based gets higher until $N = 16$ compared with the other three algorithms. We can also observe that the EE of the HCE-based gets higher until $N = 32$ compared with the CE-based hybrid precoding scheme.

VII. CONCLUSION

In this paper, we proposed a switch selecting network architecture for beamspace MIMO system, where the analog precoding is realized by a sub-connected switch network. The lens array-based hybrid precoding architecture can be equipped with 5G BS. Based on the proposed architecture, we design an HCE-based hybrid precoding algorithm by adding unadjusted probabilities and smoothing constant to the conventional CE method in machine learning to maximize the achievable sum-rate. Simulation results showed that the HCE-based hybrid precoding algorithm could achieve good performance in terms of both sum-rate and EE with low hardware complexity and low computational complexity. Furthermore, 5G BS can be a key enabler for mmWave communications. We expect that machine learning will also play a major role in the digital signal processing carried out in the baseband processor of the BS. Another promising direction is to apply deep learning [25] to enhance hybrid precoding techniques for mmWave massive MIMO.

REFERENCES

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

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

[6] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[7] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1445–1459, Mar. 2018.

[8] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[9] Y. Xing, Y. Chen, and L. Yang, "MMSE-based wideband hybrid precoding for massive MIMO systems," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2016, pp. 19–21.

[10] 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.

[11] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[12] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[13] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3679–3684.

[14] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[15] R. Qian, M. Sellathurai, and X. Fang, "Antenna selection for multi-user MIMO at millimeter-wave spectrum with lens antenna arrays," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–5.

[16] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[17] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 90–97.

[18] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of Wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, to be published.

[19] H. Ren, L. Li, W. Xu, W. Chen, and Z. Han, "Machine learning-based hybrid Precoding with robust error for UAV mmWave massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[20] L. Li, H. Ren, X. Li, W. Chen, and Z. Han, "Machine learning-based spectrum efficiency hybrid precoding with lens array and low-resolution ADCs," *IEEE Access*, vol. 7, pp. 117986–117996, 2019.

[21] R. Rubinstein and D. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer, 2013.

[22] X. Gao, L. Dai, Y. Sun, S. Han, and I. Chih-Lin, "Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[23] M. Tian, J. Zhang, Y. Zhao, J. Yang, and G. Gui, "Switch and inverter based hybrid precoding algorithm for MmWave massive MIMO system: Analysis on sum-rate and energy-efficiency," *IEEE Access*, vol. 7, pp. 49448–49455, 2019.

[24] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2018.

[25] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

TING DING received the B.S. and M.S. degrees from Northwestern Polytechnical University, in 2005 and 2008, respectively. She is currently pursuing the Ph.D. degree with National Digital Switching System Engineering and Technological Research Center. Her current research interests include mmWave Massive MIMO, hybrid precoding, and channel estimation.

YONGJUN ZHAO received the M.S. and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1993 and 2008, respectively. He is currently a Professor with the Department of Information and Communication Engineering, National Digital Switching System Engineering Technology Center, Zhengzhou, China. His current research interests include array signal processing and mmWave Massive MMO.

LIXIN LI (M'12) received the B.Sc. and M.Sc. degrees in communication engineering and the Ph.D. degree in control theory and its applications from Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2008, respectively, where he was a Postdoctoral Fellow, from 2008 to 2010. In 2017, he was a Visiting Scholar with the University of Houston, USA. He is currently an Associate Professor with the School of Electronics and Information, NPU. He has authored or coauthored more than 80 technical articles in journals and international conferences, and he holds ten patents. His current research interests include wireless communications, game theory, and machine learning. He received the 2016 NPU Outstanding Young Teacher Award, which is the highest research and education honors for young faculties in NPU. He has reviewed articles for many international journals.

DEXIU HU received the Ph.D. degree from Tsinghua University, Beijing, China, in 2017, and the M.E. degree from Zhengzhou Information Technology Institute, Zhengzhou, China, in 2010. He is currently an Assistant Professor with the Department of Information and Communication Engineering, National Digital Switching System Engineering Technology Center, Zhengzhou. His current research interests include DOA estimation, passive location, and satellite communication.

LEI ZHANG received the Ph.D. degree from the University of Sheffield, U.K. He is currently a Lecturer with the University of Glasgow, U.K. He also holds a visiting position in 5GIC at the University of Surrey. He has 16 US/UK/EU/Chinese granted patents on wireless communications. His research interests include the communications and array signal processing, including radio access network slicing (RAN slicing), wireless blockchain systems, new air interface design, the Internet of Things (IoT), multiantenna signal processing, and massive MIMO systems.