IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 21, 2021, accepted April 29, 2021, date of publication May 10, 2021, date of current version May 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3078434

Hierarchical Energy Efficient Hybrid Precoding for Configurable Sub-Connected MIMO Systems

XIANG LI, YANG HUANG, WEI HENG, AND JING WU
National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
Corresponding author: Wei Heng (wheng@seu.edu.cn)

This work was supported by the National Natural Science Foundation of China (NSFC) under Project 61771132.

ABSTRACT This paper proposed a configurable sub-connected architecture with a framework that dynamically activates the near-optimal subset of antennas and RF chains to implement energy-efficient hybrid precoding in millimeter wave multiple-input multiple-output system. Since the exhaust search is computational intractable, we propose a two-stage hybrid precoding algorithm, where the digital precoder is designed to eliminate the inter-user interference by zero-forcing rule. Specifically, in the first stage, we introduce an extended cross-entropy algorithm that adaptively updates the probability distribution of potential states in the analog precoder matrix, which can generate a solution that is close to the optimal with a sufficiently high probability. In the second stage, a QR-based subset selection algorithm is proposed to pick the near-optimal subset of the RF chains to further cut down on the energy cost. Simulation results show that proposed extend-CE algorithm gets favorable performance in terms of energy efficiency, and QR-based RF chain selection can achieve a near-optimal performance. Other hybrid precoding algorithms can also be incorporated into the proposed RF chain selection algorithm.

INDEX TERMS Cross entropy, energy efficiency, hybrid precoding, mmWave, MU-MIMO.

I. INTRODUCTION

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) has been regarded as a key technology to meet the increasing traffic and critical energy efficiency demand in cellular network [1], [2]. The sub-6 GHz microwave frequency spectrum is very crowd for mobile communication [3]. MmWave frequency band is considered as a promising spectrum resource to meet the demand of 5G wireless communication and beyond, which provides larger capacity, higher data rate, and more reliability [4]. It also faces challenges like severe path loss, unconventional fading channel, hardware complexity and high energy consumption. In the large-scale MIMO transmission scenario, the array gain is reaped to combat severe path loss by massive antennas at an order of a hundred or more [5].

Base station (BS) equipped with a large number of antennas is able to serve multiple users simultaneously. Precoding schemes can be categorized into three types, namely, fully-digital structure, fully-analog structure, and hybrid structure. In the fully-digital system, it requires a dedicated radio frequency (RF) chain consisting of digital-to-analog converter (DAC), mixer, filter, power amplifier to serve each antenna [6]. This leads to unacceptable hardware cost as well as high power consumption when plenty of transmit antennas are adopted in the MIMO scenario. Hybrid precoding involves a low-dimension digital precoder but preserves the high-dimension analog array at the same time [7]–[9]. It provides more concurrent data streams than the fully-analog scheme, and avoid a large number of RF chains when compared with fully-digital structure. Two typical precoding architectures are generally used to realize the hybrid precoding, i.e., fully-connected (FC) and sub-connected (SC) structure. Numerical results show that SC structure achieves comparable spectral efficiency performance with its counterpart but saves a lot of hardware complexity and power consumption [10].

In the hybrid process, analog beamformer is designed to harvest the array gain to improve the spectral efficiency while digital baseband precoder is aimed at eliminating the inter-chain interference [11]. Analog precoding design has been studied in many literatures. Utilizing the spatial sparsity,

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

hybrid precoding based on orthogonal matching pursuit (OMP) is proposed in [6], however, the pre-defined codebook limits the performance of analog beamformer. By factorizing the equivalent channel matrix, near-optimal hybrid precoding algorithms for both FC and SC structure are proposed based on alternating minimization and then extended to mmWave multi-carrier system in [12]. In [13], a deep neural network (DNN) is introduced as a framework to train the precoder and combiner which achieves substantial performance than conventional codebook-based block diagonalization (BD). Hybrid precoding in terahertz communications is investigated in [14], where ultra-massive antennas are deployed to overcome the huge propagation loss. Then a dynamic subarray hybrid precoding scheme is proposed to balance the spectral efficiency and power consumption.

Matrix theory demonstrates that linear digital processing schemes, such as match filter (MF) [15], zero forcing (ZF) [16] or BD [17], can help promote the system throughput and simplify the precoding and combining design for multiuser communication. ZF precoding, which cancels the inter-user interference through simple matrix inversion, is a practical scheme for implementation. Due to the significantly simplified receiver design, the mobile stations (MSs) no longer require any combining process [18]. When the MSs equip extra antennas to promote their array gain, a hybrid precoding and combining scheme should be designed to cope with the cross-stream interference as well as inter-user interference. BD can be viewed as a generalization of the ZF to deal with this situation which uses singular value decomposition (SVD) as the matrix tool to eliminate the inter-chain interference, and provides a close-form solution to calculate the digital precoding and combining matrix [19]. The traditional BD achieves a sub-optimal capacity performance by selecting from DFT set as the RF combiner, and using equal gain transmission (EGT) as the RF precoder design criterion. The spectral and energy efficiency comparison of MU-MIMO systems with different linear processing schemes can be found in [20].

Energy efficiency is another issue when designing the hybrid precoding system. User can scale down its transmitted power proportional to the number of antennas at the BS [21], but equipping massive antennas results in a higher power consumption. It requires a tradeoff between achieving higher data rates and power consumption [22]–[24]. Owing to the increase in both ecological and economic concerns, EE optimization as a fractional programming problem for wireless communications has received extensive interest recent years [25]–[27]. Although both academia and industry have already focused on the EE of cellular networks [28], existing architectures cannot face the increasing complexity of future networks towards new frequency bands, new radio, new service and applications. For this reason, innovative architectures are required to address the crucial demanding green specifications and other considerations in next-generation communications. The use of a SC structure can naturally reduce the power consumption than the FC structure [29]. The power consumption can be reduced by implementing low resolution quantization for both precoder and combiner [30]. With 1-bit sampling resolution, optimizing the number of RF chains also shows better energy efficiency than the conventional hybrid beamforming architecture [31]. The power consumption model investigated in [32] has been widely used for analyzing the energy efficiency optimization problem. Based on this model, many energy-efficient hybrid precoding schemes have been proposed in mmWave MIMO cellular communications [33]–[43]. However, the major drawback of these works is that they adopt a fixed architecture. A better energy efficiency performance can be achieved if the system has a dynamic architecture in accordance with variable clients or channel conditions [44].

In this paper, we consider the downlink transmission of the sub-connected MU-MIMO system. As shown in Fig. 1, a dynamic RF precoding structure is adopted with configurable connections. The nature of "dynamic" lies in the fact that the RF chains and antennas can be closed to save energy consumption which enables this structure to activate the near-optimal subset of antennas and RF chains to implement energy-efficient hybrid precoding. The hybrid precoding algorithm is designed hierarchically. Inspired by cross-entropy (CE) [35], we propose an analog precoding algorithm where the candidate precoding matrices are generated according to equiprobable distribution, and then the elite candidates are selected to update the distribution in each iteration. After several iterations, the distribution will finally converge to a stable state, and the obtained precoder can be sufficiently near the optimal. CE has shown potential sum-rate performance in recent studies [36], but its searching space is significantly reduced than the exhaust search. The baseband channel matrix should have a sufficient rank to transmit enough streams to support all the clients, while at the same time, the number of active RF chains should be as less as possible to save energy consumption. This problem can be derived as identifying the most representative columns of the equivalent analog channel matrix. In [45], author provides a rule to increase the ratio of determinants by permutating the columns. Base on this rule, we develop an iterative algorithm to systematically pick the subset that represents the entire channel well. The contribution of this paper can be summarized as follows:

- A configurable sub-connected hybrid precoding structure is proposed. By adjusting the states of the network, this dynamic architecture can achieve a balance between spectral efficiency and power consumption.
- We develop a framework utilizing the extend-CE minimization to generate the analog precoding matrix in an iterative way. Energy efficiency is optimized by calculating the antenna activity and phase of each shifter at the same time.
- Based on the QR decomposition, we formulate how to discard the unnecessary columns in the equivalent analog channel matrix, and accordingly turn off the RF chain to further save the power consumption.

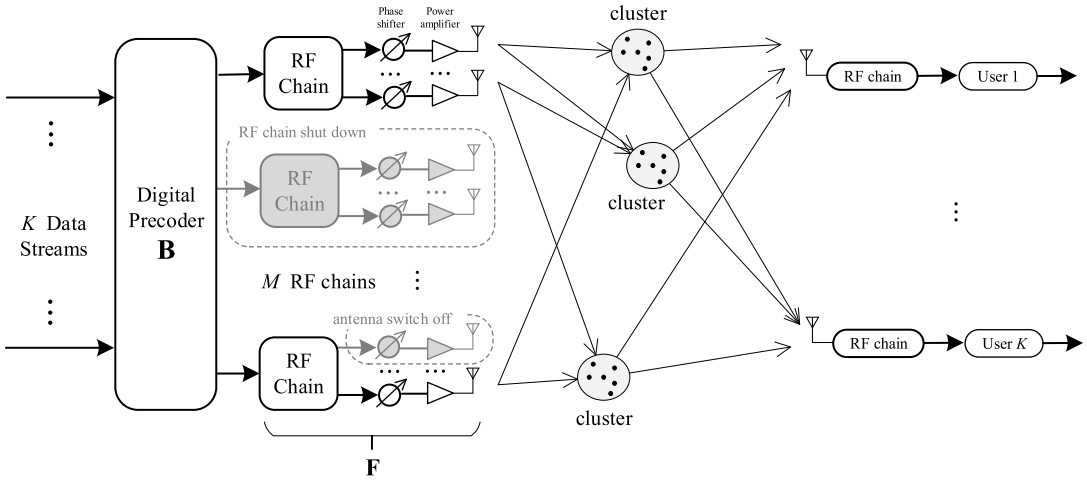

FIGURE 1. System model for an mmWave A/D hybrid MIMO system with configurable sub-connect structure.

To the best of our knowledge, there is no prior work on energy efficiency optimization considering this dynamic architecture, which is still an open problem and the focus of our work. The rest of this paper is organized as follows. Section II presents the system model of downlink mmWave MU-MIMO system. Energy efficient hybrid precoding is studied in Section III. Section IV provides the simulation results and performance comparison. Finally, the conclusion is drawn in Section V.

Notation: Upper- and lower-case boldface denotes matrices and vectors, respectively. $\mathbb{E}[\cdot]$ denotes the expectation and $\mathbb{C}$ denotes the ensemble of complex numbers. $\mathbf{X}^{\mathrm{H}}$, and $\mathbf{X}^{-1}$ represent the conjugate transpose and inversion. $||\cdot||_{\mathrm{F}}$ and $\det(\cdot)$ denote the Frobenius norm and determinant of matrix $\mathbf{X}$. Finally, $\mathbf{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{0}$ is the all-zero matrix.

II. SYSTEM MODEL

A. A/D HYBRID MIMO SYSTEM MODEL

We consider the downlink transmission of the MU-MIMO system. BS is equipped with N antennas and M RF chains, and K single-antenna MSs are active simultaneously, so there are totally K data streams to be scheduled at the BS. To guarantee the effectiveness of communications, and maximize the system throughput with limited hardware resources, the number of RF chains should be constrained by $K \leq M \leq N$.

At the BS, the transmitted symbol vector $\mathbf{s}$ is addressed by a digital baseband precoder $\mathbf{B}$ of $M \times K$ and then an $N \times M$ RF precoder $\mathbf{F}$. Baseband precoder allows both amplitude and phase modification on the signal, while only phase shift can be realized by RF precoder via analog circuit. We assume each entry of $\mathbf{F}$ has unit amplitude and $\mathbf{B}$ is designed to meet the transmitted power limitation, i.e., $||FB||_{\mathrm{F}}^2 = K$. We assume a non-frequency selective fading channel, the signal arrived at the k^{th} MS is

$$y_k = \mathbf{h}_k \mathbf{F}\mathbf{B}\mathbf{s} + n_k, \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_K]^{\mathrm{T}} \in \mathrm{C}^{K\times 1}$ is the vector contains the signal of all the MSs. The signal vector satisfies $\mathbb{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = P/K\mathbf{I}_K$, where P is the average transmit power. $\mathbf{h}k \in \mathbb{C}^{1\times N}$ is the channel matrix of the k^{th} MS, and n_k is the i.i.d additive noise vector obeying $\mathcal{CN}(0, \sigma^2)$. To simplify the analysis, the equivalent analog channel coefficient of k^{th} MS is defined as

$$\tilde{\mathbf{h}}_k = \mathbf{h}_k \mathbf{F}. \tag{2}$$

Then the signal at the k^{th} MS can be written as

$$\hat{s}_k = \tilde{\mathbf{h}}_k \mathbf{b}_k s_k + \underbrace{\sum_{i=1, i\neq k}^{K} \tilde{\mathbf{h}}_k \mathbf{b}_i s_i}_{\text{interference}} + \underbrace{n_k}_{\text{noise}} \tag{3}$$

in which $\mathbf{b}_k$ is the digital baseband precoding matrix corresponding to the k^{th} MS, i.e., the k^{th} columns of $\mathbf{B}$. Assuming the modulated symbols are Gaussian, the sum spectral efficiency can be expressed as

$$R = \sum_{k=1}^{K} \log_2 (|1 + \mathrm{SINR}_k|), \tag{4}$$

where SINR_k is the signal-to-interference-plus-noise ratio of the k^{th} MS as

$$\mathrm{SINR}_k = \frac{\tilde{\mathbf{h}}_k \mathbf{b}_k s_k}{\sum_{i=1, i\neq k}^{K} \tilde{\mathbf{h}}_k \mathbf{b}_i s_i + n_k}. \tag{5}$$

B. mmWave CHANNEL MODEL

MmWave channel no longer obeys the conventional Rayleigh fading, and it has different propagation characteristics when compared with low-frequency channels [37]. Clustered mmWave channel model characterizes the limited scattering feature of the mmWave channel. The normalized mmWave downlink channel for the k^{th} MS is modelled as the sum of all the propagation paths which are scattered in N_{C} clusters

where each cluster involves total N_{P} paths. It can be expresses as

$$\mathbf{h}_k = \sqrt{\frac{1}{N_{\mathrm{C}} N_{\mathrm{P}}}} \sum_{c=1}^{N_{\mathrm{C}}} \sum_{p=1}^{N_{\mathrm{P}}} g_{c,p}^{k} \mathbf{a}_{\mathrm{BS}} \left(\varphi_{c,p}^{k}\right) \tag{6}$$

in which $gk\ c, p$ corresponds to the complex channel gain of the p^{th} path of the c^{th} cluster. $\varphi k\ c, p$ is the azimuth angle of the departure (AoD). $\mathbf{a}_{\mathrm{BS}}(\varphi kc, p)$ is the transmit array response vector while the elevation dimension is ignored. In each cluster, we assume $\varphi kc, p$ is distributed within a specified range which is generated by truncated Laplacian distribution.

Without loss of generality, we chose the uniform linear array (ULA) to model the BS and MSs array in this study. For a N antenna array, the response vector can be given as

$$\mathbf{a}_{\mathrm{BS}} \left(\varphi_{c,p}^{k}\right) = \frac{1}{\sqrt{N}} \left[1,\ e^{j2\pi d \sin(\varphi_{c,p}^{k})/\lambda},\ \ldots,\ e^{j2\pi (N-1) d \sin(\varphi_{c,p}^{k})/\lambda}\right]^{\mathrm{T}}, \tag{7}$$

where λ is the wavelength of the carrier frequency and d is the distance between the adjacent antennas. Assuming a half wavelength antenna spacing in this study, the array response can be calculated from (7). Other antenna array patterns can also be adopted, and proposed scheme can be directly applied to arbitrary antenna arrays. Due to the sparse nature of the mmWave, the channel coefficients can be effectively estimated via compressing sensing algorithms [38], [39], and we assume that BS has full knowledge of $\mathbf{h}_k$.

C. HYBRID PRECODER

The broadly discussed FC hybrid precoder, where each RF chain is connected to all antennas through variable-degree phase shifters (PSs) and RF adders, is able to approach the performance of the unconstrained maximization [40]. Hybrid precoding adds constraints to the RF precoder where only B-bit quantized shift can be applied to each element, i.e.,

$$f^{(p,q)} \in \left\{1,\ e^{j\frac{2\pi}{2^B}},\ \ldots,\ e^{j\frac{2\pi \times (2^B-1)}{2^B}}\right\}, \tag{8}$$

where $f^{(p,q)}$ is the $(p, q)^{\mathrm{th}}$ entry of $\mathbf{F}$. It has been considered in many studies to achieve full precoding gain [9]–[20], [29]–[41]. Most of these algorithms decompose the problem into analog and digital precoder design separately. Specifically, the analog precoder $\mathbf{F}$ is first generated or selected from a predefined codebook to maximize the sum rate according to the channel matrix $\mathbf{h}_k$. After that, the digital precoder $\mathbf{B}$ is designed to eliminate the inter-chain interference.

To reduce the hardware complexity, one feasible solution is to replace the FC RF network with SC structure [41] where each RF chain is only connected to Q variable-degree PSs where $Q = N/M$. The format of this RF precoding matrix can be written in a block-diagonal form as

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{f}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{f}_M \end{bmatrix}, \tag{9}$$

where $\mathbf{f}_m$ denotes the analog precoding vector of the m^{th} RF chain with size $Q \times 1$ and each non-zero element should be in the set of (8). A further simplification is to replace the PSs by only one inverter and RF switches [33], [41]. However, due to the elementary on-off connection, this switch-based structure fails to realize the full array gain.

In typical design procedure, the optimization on precoder should be an essential method to achieve maximal energy efficiency or minimal distance to optimal precoder. Finding the energy efficient precoding matrix are non-coherent combining optimization problems [42], and the large-scale precoding matrix makes it almost intractable to find the global optimum through exhaust search while maintaining the constraints imposed on the precoder [6]. Even in the fully-digital MU-MIMO systems, it still requires enormous efforts to find a local optimum of energy efficiency [43].

III. ENERGY-EFFICIENT HYBRID PRECODING

With a large number of antennas deployed in the considered scenario, array gain should be cultivated through appropriate designed hybrid precoding. Two conditions should be met at the same time: one is that the equivalent channel should be well-conditioned enough to support reliable transmission of totally K streams; another is that after closing nonessential RF chains and antennas, energy efficiency should be as larger as possible.

We formulate how to design the hybrid precoder in this section. The parameters of power consumption model are defined in subsection A. Then, the overall framework of proposed algorithm and the procedure of generating the RF precoder are provided in subsection B. Subsection C gives the active RF chain selection algorithm to calculate the baseband precoder $\mathbf{B}$ from the equivalent baseband channel matrix, which is a necessary step in the algorithm 1 framework.

A. ENERGY EFFICIENCY

In terms of achievable sum rate R and consumed power P, the energy efficient is defined as the ratio

$$E(\mathbf{F}, \mathbf{B}) \triangleq \frac{R(\mathbf{F}, \mathbf{B})}{P(\mathbf{F}, \mathbf{B})}\ (\text{bits/Hz/J}), \tag{10}$$

where $R(\cdot)$ represents the information rate in bits/s/Hz and $P(\cdot)$ is the required power in watt. We adopt the power consumption model in [33], where the power consumption can be expressed as

$$P = N_{\mathrm{BS}} \times (P_{\mathrm{PA}} + P_{\mathrm{PS}}) + N_{\mathrm{S}} \times (P_{\mathrm{RFC}} + P_{\mathrm{DAC}}) + P_{\mathrm{BB}} \tag{11}$$

in which N_{BS} and N_{S} are the number of active antennas and RF chains. P_{PA}, P_{PS}, P_{RFC} and P_{DAC} are the power of

the power amplifier, PS, RF chain, and DAC respectively. P_{BB} denotes the power of the digital signal processing. It can be observed from (11) that SC structure requires less power on P_{PS} than the FC structure which contains $N \times M$ PSs and power of RF adders is omitted. The values of each parameter employed in this work are as follows: $P_{\mathrm{PA}} = 20\mathrm{mW}$, $P_{\mathrm{PS}} = 30\mathrm{mW}$, $P_{\mathrm{RFC}} = 30\mathrm{mW}$, $P_{\mathrm{DAC}} = 200\mathrm{mW}$, and $P_{\mathrm{BB}} = 4$ [41].

B. CE-BASED ANALOG PRECODING

In the first stage, we optimize the energy efficiency by turning off nonessential antennas (and corresponding PSs and amplifiers) which contribute little to the total energy efficiency. To better explain the antenna selection mechanism, we introduce an extra state, i.e., 0, that indicates the off switch of the antenna, while the non-zero values still determine the phase of the corresponding active antenna. Clearly, increasing the number of active antennas, we might have a higher information rate but there is also a higher power consumption. The possible value of the entries of $\mathbf{F}$ can be extended as

$$f^{(p,q)} \in \left\{0,\ 1,\ e^{j\frac{2\pi}{2^B}}, \ldots, e^{j\frac{2\pi\times(2^B-1)}{2^B}}\right\}, \tag{12}$$

Based on the SC structure, the purpose of energy efficient hybrid precoding is to maximize energy efficiency with respect to $\mathbf{F}$ and $\mathbf{B}$ as

$$\begin{aligned}\{\mathbf{F},\mathbf{B}\} = \underset{\mathbf{F},\mathbf{B}}{\arg\max}\ E(\mathbf{F},\mathbf{B}) = \underset{\mathbf{F},\mathbf{B}}{\arg\max}\ \frac{\sum_{k=1}^{K}\log_2(1+\gamma_k)}{P(\mathbf{F},\mathbf{B})}\\ \text{subject to } \mathbf{F}\in\mathbb{F},\\ \|\mathbf{FB}\|_{\mathrm{F}}^2 = K.\end{aligned} \tag{13}$$

in which $\mathbb{F}$ is the set of all the possible analog precoding matrix satisfying (9) and (12). We define the equivalent baseband channel as

$$\mathbf{H}_{\mathrm{eq}} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_K^{\mathrm{T}}]^{\mathrm{T}}\mathbf{F}. \tag{14}$$

For ZF digital precoder, the SINR of the k^{th} user γ_k can be calculated as

$$\gamma_k = \frac{\left\|\mathbf{h}_k^{\mathrm{T}}\mathbf{F}\mathbf{b}_k\right\|_{\mathrm{F}}^2}{\sigma_k^2 + \sum_{i\neq k}\left\|\mathbf{h}_k^{\mathrm{T}}\mathbf{F}\mathbf{b}_i\right\|_{\mathrm{F}}^2}, \tag{15}$$

where $\mathbf{b}_k$ is the k^{th} column of the digital baseband precoder $\mathbf{B}$ and the detail procedure of designing $\mathbf{B}$ is provided in next subsection. The core of hybrid-ZF precoding is to design $\mathbf{F}$ under given constraints. One solution is the exhaustion search: $\mathbf{F}$ has up to N non-zero elements and each element have $2^B + 1$ potential values, so there are $(2^B + 1)^N$ possibilities. However, in the massive MIMO configuration, N is usually very large, so finding the global optimum while maintaining the constraints imposed on the RF precoder is often computational intractable. In the proposed algorithm, we resort to iteratively generating near-optimal tests and optimizing the distribution by minimizing the cost function in a close form. Specific steps are detailed as follows:

Step I, initializing the equal-probable matrix $\mathbf{P}^1$ whose element $p1$ l,n indicates the possibility of the l^{th} phase of the n^{th} non-zero element in $\mathbf{F}$ where $1 \leq l \leq 2^B + 1$ and $1 \leq n \leq N$.

Step II, according to $\mathbf{P}^m$, where m represents the iteration index, randomly generating S individuals $\mathbf{F}^s$, where $1 \leq s \leq S$, and computing the corresponding $\mathbf{B}^s$ whose detail procedure is given in next subsection.

Step III, calculating the energy efficiency E^s by (10) before sorting $\{E^s\}$ in a decent order as $\{E_1, E_2, \ldots, E_S\}$. Then we pick the T largest as the elite-batch.

Step IV, the extend-CE using with weighted log-probability distribution is

$$\mathcal{E}\left(\mathbf{P}^m\right) = \sum_{t=1}^{T} w_t \sum_{n=1}^{N}\sum_{l=1}^{2^B+1}\delta_{t,n,l}^m \ln p_{n,l}^m, \tag{16}$$

where $\delta_{t,n,l}$ is the binary activity indicator. Only when the l^{th} phase candidate of the n^{th} antenna is selected at t^{th} test, $\delta_{t,n,l}$ is active. w_t is the weight of the elite-batch defined by

$$w_t = \frac{|R_t - R_T|}{\sum_{t=1}^{T}|R_t - R_T|} \quad (1 \leq t \leq T). \tag{17}$$

Step V, updating the probability matrix following

$$\mathbf{P}^{m+1} = \underset{\mathbf{P}^m}{\arg\min}\,\mathcal{E}\left(\mathbf{P}^m\right). \tag{18}$$

The problem can be formulated as

$$\begin{aligned}&\text{minimize}\ \ \mathcal{E}\left(\mathbf{P}^m\right)\\ &\text{subject to}\ \sum_{l=1}^{2^B+1} p_{n,l}^m = 1, \quad \text{for } n = 1, \ldots, N.\end{aligned} \tag{19}$$

We introduce multipliers ε_n where $n = 1, \ldots, N$, the Lagrange function can be constructed as

$$\begin{aligned}\mathcal{L}\left(\mathbf{P}^m, \varepsilon_1, \ldots, \varepsilon_N\right) = &\sum_{t=1}^{T} w_t \sum_{n=1}^{N}\sum_{l=1}^{2^B+1}\delta_{t,n,l}^m \ln p_{n,l}^m\\ &- \sum_{r=1}^{N}\varepsilon_r\left(\sum_{l=1}^{2^B+1} p_{l,n} - 1\right),\end{aligned} \tag{20}$$

and solving totally $N \times (2^B + 1)$ equations

$$\begin{aligned}&\nabla_{p_{n,l}^m,\,\varepsilon_r}\mathcal{L}\left(\mathbf{P}^m, \varepsilon_1, \ldots, \varepsilon_N\right) = 0,\\ &\text{where } 1 \leq l \leq 2^B + 1,\ 1 \leq n \leq N,\ 1 \leq r \leq N.\end{aligned} \tag{21}$$

Finally, we have

$$\begin{aligned}p_{n,l}^{m+1} = \frac{\sum_{t=1}^{T} w_t \delta_{e,n,l}^m}{\sum_{t=1}^{T}\sum_{l=1}^{2^B+1} w_t \delta_{t,n,l}^m},\\ \text{for } 1 \leq l \leq 2^B + 1, \quad 1 \leq n \leq N.\end{aligned} \tag{22}$$

Step VI, let $m \leftarrow m + 1$ and restart the loop from step II until $w_t = 0$ for all t or reaches the maximum number of iterations. Finally, we output $\mathbf{F}^{opt}$ and $\mathbf{B}^{opt}$ which have the largest E^{opt}.

The complete process of proposed algorithm is listed in Algorithm 1. In the proposed algorithm, w_t in (17) creates the unfairness for each elite to accelerate the convergency and provides a simple ending criterion in Step VI.

Algorithm 1 Proposed Hybrid Precoding Algorithm

Input: $\mathbf{H}$, σ^2
1: **Loop**:
2: **for** s **in** $1 \ldots S$ **do**
3: Generate $\mathbf{F}^s$ according to $\mathbf{P}^m$;
4: Calculate $\mathbf{B}^s$ by (26);
5: Calculate E^s by (10);
6: **end for**
7: **for** t **in** $1 \ldots T$ **do**
8: Calculate w_t by (17);
9: **end for**
10: Check convergency;
11: **for** n **in** $1 \ldots N$ **do**
12: **for** l **in** $1 \ldots 2^B + 1$ **do**
13: Update $pm{+}1$ l, n by (22);
14: **end for**
15: **end for**
16: $m \leftarrow m + 1$;
17: **end loop**
18: **return** $\mathbf{F}^{\text{opt}}$, $\mathbf{B}^{\text{opt}}$ and E^{opt}.

C. ACTIVE RF CHAIN SELECTION

In this subsection, we show how to calculate the baseband precoder $\mathbf{B}$ through active RF chain selection. From (14), we have an equivalent baseband channel matrix $\mathbf{H}_{\text{eq}}$ of $K \times M$.

Algorithm 2 Near-Optimal RF Chain Selection

Input: $\mathbf{H}_{\text{eq}}$
1: Initialize $\Pi = \mathbf{I}_N$, $\mathbf{H} = \mathbf{H}_{\text{eq}}\Pi_{i,j+K}$
2: Decomposition: $\mathbf{H} = \mathbf{QR}$ where $\mathbf{R} = [\mathbf{R}_K\, \mathbf{C}]$
3: **while**: exist i, j such that $|\mathbf{R}\text{-}1\ K\mathbf{C}|_{i,j} > 1$
4: Update $\mathbf{H}$: $\mathbf{H} = \mathbf{H}\Pi_{i,j+K}$
5: Decomposition: $\mathbf{H} = \mathbf{QR}$ where $\mathbf{R} = [\mathbf{R}_K\, \mathbf{C}]$
6: **end while**
7: Calculate $\mathbf{B}$ by (26) where $\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2]$;
8: **return** $\mathbf{B}$.

The problem of RF chain selection can be expressed as finding a permutation matrix $\mathbf{\Pi}$ such that

$$\mathbf{H}_{\text{eq}}\mathbf{\Pi} = [\mathbf{H}_1\ \mathbf{H}_2] \tag{23}$$

where $\mathbf{H}_1$ contains the K most linearly independent columns in $\mathbf{H}_{\text{eq}}$ (condition 1), or equivalently the best linear combination of $\mathbf{H}_1$ is close to $\mathbf{H}_2$ such that $\min_z \|\mathbf{H}_1 z - \mathbf{H}_2\|_2$ is sufficiently small (condition 2). Identifying the most important columns in a matrix has been studied in solving rank-deficient least square (LS) problems [47], information retrieval [48], genetics [49], and wireless communication [50].

The concept of rank-deficiency is important in subset selection where the singular value is considered as the metric to check if a matrix is or close to rank-deficient. Our near-optimal RF chain selection is based on QR decomposition [47] as

$$\mathbf{H}_{\text{eq}} = \mathbf{QR} = \mathbf{Q}\left[\mathbf{R}_K\ \mathbf{C}\right], \tag{24}$$

where $\mathbf{R}_K \in \mathbb{C}^{K\times K}$ and $\mathbf{C} \in \mathbb{C}^{K\times(M-K)}$. To gauge the linear independence of a set of columns and find the most representative columns. We pivot the columns of $\mathbf{R}$ with the goal of increasing the singular value of $\mathbf{R}_K$. Based on the fact that $|\det(\mathbf{H}_{\text{eq}})| = |\det(\mathbf{R})| = |\det(\mathbf{R}_K)| \times |\det(\mathbf{C})|$, the permutation moves the selected columns towards satisfying the conditions 1 or 2 by increasing the singular values of $\mathbf{R}_K$ as large as possible.

In [46], a useful conclusion is drawn to detect the increase of the singular value before really performing a permutation. Define $\tilde{\mathrm{R}} = \mathbf{R}\mathbf{\Pi}_{i,j}$, the ratio of two determinants can be expressed as

$$\frac{\left|\det\left(\tilde{\mathbf{R}}\right)\right|}{|\det(\mathbf{R})|} = \left|\mathbf{R}_K^{-1}\mathbf{C}\right|_{i,j}. \tag{25}$$

We only permutate the i^{th} and j^{th} column of $\mathbf{R}$ if (25) is greater than threshold. In our algorithm, the loop runs as long as the determinant increases in magnitude, i.e., $|\det(\tilde{\mathrm{R}})| \,/\, |\det(\mathbf{R})| > 1$. Finally, the baseband precoder can be calculated by

$$\mathbf{B} = \frac{\sqrt{K}\,(\mathbf{H}_1)^{\mathrm{H}}\left(\mathbf{H}_1\,(\mathbf{H}_1)^{\mathrm{H}}\right)^{-1}}{\left\|\mathbf{F}\,(\mathbf{H}_1)^{\mathrm{H}}\left(\mathbf{H}_1\,(\mathbf{H}_1)^{\mathrm{H}}\right)^{-1}\right\|_{\mathrm{F}}^2}. \tag{26}$$

The procedure of near-optimal RF chain selection is provided in Algorithm 2. Note that proposed algorithms do not rely on any precondition of the channel matrix. It can be adopted in the mmWave sparse scattering channel as well as the typical i.i.d Rayleigh channel as long as the channel matrix is available.

IV. SIMULATION RESULTS

In this section, the performance of our proposed precoding algorithm is evaluated in the downlink mmWave MIMO systems. The number of antennas at the BS are 64, and 16 RF chains are employed to serve 8 users. The SNR is defined as $P \,/\, \sigma^2$. We model the mmWave propagation channel with $N_{\text{C}} = 8$ clusters and each cluster involves $N_{\text{P}} = 10$ paths. The angle spread of θ^k and φ^k are both equal to 7.5°, and each path factor follows a complex Gaussian distribution with zero mean and unit variance. Other parameters of proposed algorithm are set as follows: $S = 500$, $E = 0.15S$, and the maximum number of iterations is 20.

In Fig. 2, we compare the achieved energy efficiency against SNR of proposed algorithm with state-of-art solutions where $B = 2$. Dinkelbach Method (DM) [51] is an efficient algorithm which replaces the fractional cost function of (10) by iteratively solving a sequence of difference-based problems. Iterative optimization (IO) [43] is performed to maximize the spectral efficiency after exhaust searching (ES) the most energy efficient antenna pattern. Both full-rank and QR-based RF chain selection are presented for comparison. As observed from the figure, fully-digital approach provides the lower bound for all algorithms, because it consumes much more power than others. For all the schemes, QR-based

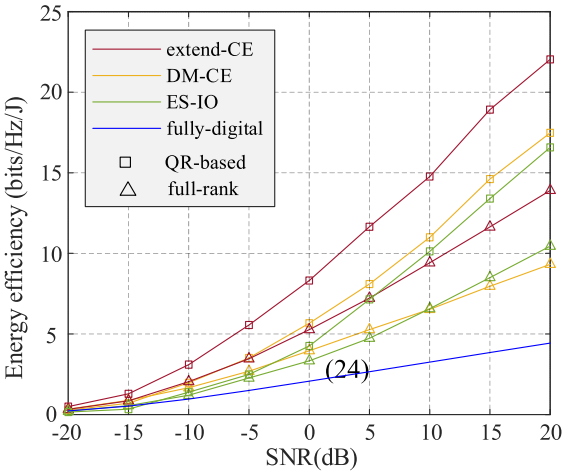

FIGURE 2. Energy efficiency comparison of different schemes.

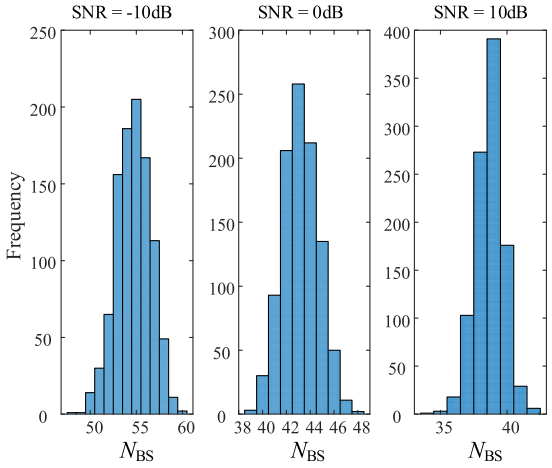

FIGURE 4. PMF plots of the number of active antennas at different SNR.

RF chain selection shows superior performance than the full-rank matrix inversion. DM-CE and ES-IO have comparable performance and proposed extend-CE outperforms both over a wide range of SNR.

Fig. 3 plots the achievable energy efficiency cumulative distribution function (CDF) curve of proposed algorithm with different phase resolution when SNR = 20 dB. Simulations start from similar initial population. The multi-phase superiority helps proposed algorithm promote performance faster than the binary phase condition after the first 15 loops. As the generation evolves, the range of achievable energy efficiency continues narrowing and the curve becomes steeper. After 20 loops, both the algorithms achieve their convergency because the weight w_t decreases to zero, i.e., $R_t = R_T$ for $1 \leq t \leq T$, and the probability evolution terminated. Finally, proposed algorithm with $B = 2$ advances the curve where $B = 1$ about 0.2 bits/Hz/J.

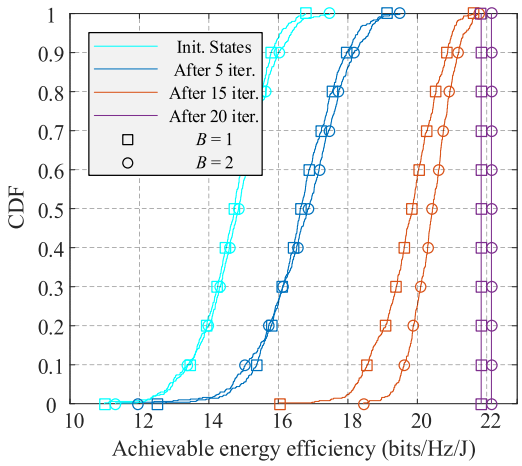

FIGURE 3. CDF of the achieved energy efficiency.

In Fig. 4, the comparison is made to the proposed algorithm with different SNR, in terms of the probability mass function (PMF) for the number of active antennas. The PMF plots indicate the histogram that for how many realizations a particular value of the variable defined on the x-axis is achieved. In this figure, we show the PMF of the distribution of proposed extend-CE method over the number of antennas, i.e., NBS. For example, at SNR = 0dB, proposed algorithm chooses NBS = 43 for 260 different realizations. We also observe that at a lower SNR, it prefers to activate more antennas to maximize energy efficiency, while at a higher SNR, it prefers to turn off nonessential antennas to save power consumption. Because calculating the object function in Eq. (13) is non-convex and hard to tackle directly [23], we can relax the constrain by finding the optimal number of active antennas instead of calculating the optimal precoder. The figure gives an idea on how close the proposed algorithm is near the optimal in terms of the number of active antennas. Because CE is a global random search procedure, it can find an optimal solution with a probability arbitrarily close to 1 [52]. It is sure that our results fall around the optimal number of antennas. The probability distribution is more concentrated and the results are more reliable at a high SNR.

In Fig. 5, we compare the performance of proposed algorithm with different RF antennas as a function of quantizing resolution of the variable-degree PS. The parameters are set as follows: $K = 8$, $M = 16$ and SNR = 10 dB. It can be observed that doubling the number of RF antennas promotes the energy efficiency by about 3 bits/Hz/J. Increasing the phase candidates can also improve the performance. When $N = 32$, the promotion is not significant, but with more than 128 antennas, increasing the resolution will remarkably improves the energy efficiency. However, as the number of potential states increases, the size of population becomes the major limitation. Comparing the curves where $N = 256$, potential improvement can be achieved when we tenfold the population. System with $N = 256$ and $S = 5000$ promotes about 3 bits/Hz/J than $S = 500$, almost the same performance of doubling the number of RF antennas.

In Fig.6, we compare the achievable energy efficiency performance of different RF chain selection schemes,

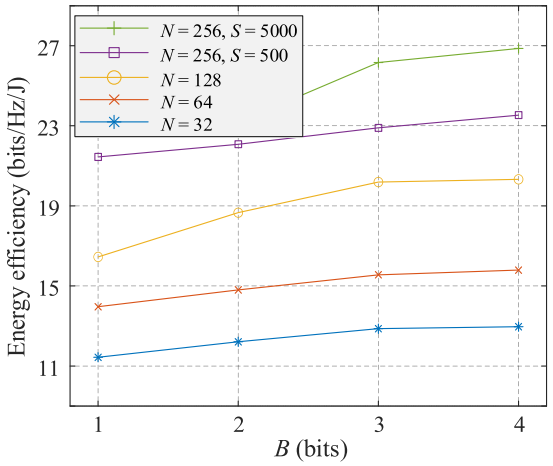

FIGURE 5. Energy efficiency comparison of different phase resolution.

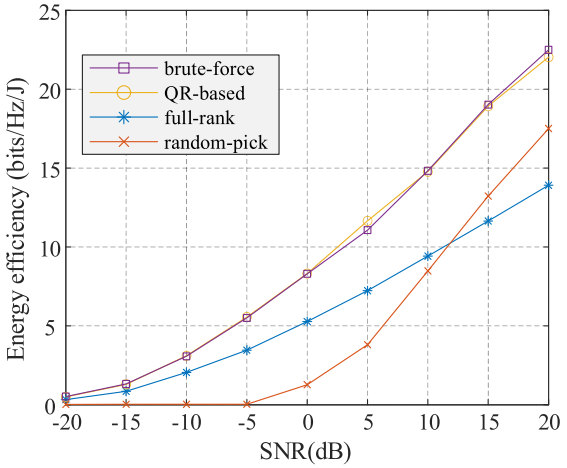

FIGURE 6. Comparison of achieved energy efficiency with different RF chain selection schemes.

namely, full-rank which keeps all the RF chains at service, random-pick which randomly picks K RF chains and turn off others, proposed QR-based RF chain selection scheme, and brute-force which searches over all the possible combination. 64 antennas and 16 RF chains are employed to serve 8 users in the simulation. Because the full-rank scheme consumes the most energy, its energy efficiency is way from the optimal especially at a high SNR. From the figure, we can see that proposed QR-based RF chain selection scheme achieves similar performance with the brute-force, and more than 7 dB advances the random-pick over a wide range of SNR although they have identical baseband power consumption. Because our method is not deterministic, we measure the convergency performance of above methods in terms of their running-time on my laptop (with AMD 4800u processor). Averaging 10^5 realizations at 10dB, the time consumption of brute-force, QR-based, full-rank and random-pick are 7.13×10^{-2}s (using *parfor*), 3.20×10^{-3}s, 2.97×10^{-5}s and 2.81×10^{-5}s, respectively. From above results, we observe that full-rank and random-pick show comparable running time. QR-based subset selection is about 10^2 time consuming that random-pick on average, but still far less than the brute-force.

V. CONCLUSION

In this paper, a sub-connected hybrid precoding structure with configurable digital and analog connections has been proposed for the mmWave MU-MIMO downlink transmission. This structure adopts a dynamic connection to achieve a balance between spectral efficiency and power consumption. The design of the precoder is formulated as a non-convex optimization problem, and we solve the problem in two stages. In the first stage, we adopt the extend-CE with weighted log-probability distribution as the metric to solve the CE minimization problem, and a close-form solution is derived to update the probability matrix in each iteration. In the second stage, a QR-based subset selection algorithm is proposed to pick the active RF chains to further cut down on the power consumption, thus a near-optimal baseband precoding matrix is obtained. The performance of proposed framework is examined with different configurations, and simulation results indicate that proposed algorithm is able to match the channel to improve the system's energy efficiency. Further study will be explored to minimize the population size and achieve faster convergency in the future work.

REFERENCES

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

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[4] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[5] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.

[6] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

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

[8] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[9] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[10] X. Song, T. Kuhne, and G. Caire, "Fully-connected vs. sub-connected hybrid precoding architectures for mmWave MU-MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[11] R. W. Heath, Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
[12] 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.
[13] 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.
[14] L. Yan, C. Han, and J. Yuan, "A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2041–2056, Sep. 2020.
[15] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, 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.
[16] 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.
[17] 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.
[18] M. Hanif, H. Yang, G. Boudreau, E. Sich, and H. Seyedmehdi, "Low-complexity hybrid precoding for multi-user massive MIMO systems: A hybrid EGT/ZF approach," *IET Commun.*, vol. 11, no. 5, pp. 765–771, Mar. 2017.
[19] A. R. Flores, R. C. de Lamare, and B. Clerckx, "Linear precoding and stream combining for rate splitting in multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 890–894, Apr. 2020.
[20] H. Quoc Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
[21] 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.
[22] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
[23] Y. Huang, S. He, J. Wang, and J. Zhu, "Spectral and energy efficiency tradeoff for massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6991–7002, Aug. 2018.
[24] A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Towards optimal energy efficiency in cell-free massive MIMO systems," *IEEE Trans. Green Commun. Netw.*, early access, Feb. 12, 2021, doi: 10.1109/TGCN.2021.3059206. [Online]. Available: https://arXiv:2005.07459
[25] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[26] H. Vaezy, M. J. Omidi, M. M. Naghsh, and H. Yanikomeroglu, "Energy efficient transceiver design in MIMO interference channels: The selfish, unselfish, worst-case, and robust methods," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5377–5389, Aug. 2019.
[27] L. You, J. Xiong, X. Yi, J. Wang, W. Wang, and X. Gao, "Energy efficiency optimization for downlink massive MIMO with statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2684–2698, Apr. 2020.
[28] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.
[29] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog–digital transceivers for single- and multi-carrier large antenna array systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.
[30] A. Kaushik, E. Vlachos, and J. Thompson, "Energy efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[31] E. Vlachos, A. Kaushik, and J. Thompson, "Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–5.
[32] J.-C. Guo, Q.-Y. Yu, W.-X. Meng, and W. Xiang, "Energy-efficient hybrid precoder with adaptive overlapped subarrays for large-array mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1484–1502, Mar. 2020.
[33] 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)*, May 2017, pp. 1–6.
[34] M. Tian, J. Zhang, Y. Zhao, L. Yuan, 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, Apr. 2019.
[35] D. P. Kroese, S. Porotsky, and R. Y. Rubinstein, "The cross-entropy method for continuous multi-extremal optimization," *Methodol. Comput. Appl. Probab.*, vol. 8, no. 3, pp. 383–407, Sep. 2006.
[36] Y. Zhang, X. Dong, and Z. Zhang, "Machine learning-based hybrid precoding with low-resolution analog phase shifters," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 186–190, Jan. 2021.
[37] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart., 2018.
[38] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[39] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[40] X. Bao, W. Feng, J. Zheng, and J. Li, "Deep CNN and equivalent channel based hybrid precoding for mmWave massive MIMO systems," *IEEE Access*, vol. 8, pp. 19327–19335, 2020.
[41] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
[42] R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. New York, NY, USA: Springer, 2004.
[43] C. Hu, J. Liu, X. Liao, Y. Liu, and J. Wang, "A novel equivalent baseband channel of hybrid beamforming in massive multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 764–767, Apr. 2018.
[44] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.
[45] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing QR factorization," *SIAM J. Sci. Comput.*, vol. 17, no. 4, pp. 848–869, Jul. 1996.
[46] M. Broadbent, M. Brown, and K. Penner, "Subset selection algorithms: Randomized vs. Deterministic," *SIAM Undergraduate Res. Online*, vol. 3, pp. 50–71, Oct. 2010.
[47] A. Cortinovis and D. Kressner, "Low-rank approximation in the frobenius norm by column and row subset selection," *SIAM J. Matrix Anal. Appl.*, vol. 41, no. 4, pp. 1651–1673, Nov. 2020.
[48] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. E. Rouayheb, "Private information retrieval schemes for coded data with arbitrary collusion patterns," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1908–1912.
[49] P. Krömer, J. Platoá, J. Nowaková, and V. Snáel, "Optimal column subset selection for image classification by genetic algorithms," *Ann. Oper. Res.*, vol. 265, no. 2, pp. 205–222, Jun. 2018.
[50] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Transmit antenna subset selection for single and multiuser spatial modulation systems operating in frequency selective channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6156–6169, Jul. 2018.
[51] A. Kaushik, J. Thompson, E. Vlachos, C. Tsinos, and S. Chatzinotas, "Dynamic RF chain selection for energy efficient and low complexity hybrid beamforming in millimeter wave MIMO systems," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 886–900, Dec. 2019.
[52] A. Costa, O. D. Jones, and D. Kroese, "Convergence properties of the cross-entropy method for discrete optimization," *Operations Res. Lett.*, vol. 35, no. 5, pp. 573–580, Sep. 2007.

XIANG LI was born in 1990. He received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests include wireless transmission, wireless networks, and signal processing and its applications.

YANG HUANG was born in 1997. He received the B.S. degree in information engineering from Southeast University, Nanjing, China, where he is currently pursuing the M.S. degree with the National Mobile Communications Research Laboratory. His research interests include cloud computing and big data, machine learning, and wireless signal processing.

WEI HENG was born in 1965. He received the B.S. degree in information engineering from the Huazhong University of Science and Technology, in 1987, and the M.S. degree in signal and information processing and the Ph.D. degree in communications and information system from Southeast University, in 1990 and 1997, respectively. From 1997 to 2002, he was a Research Scholar with The Catholic University of America. Since 2002, he has been a Professor with the National Mobile Communications Research Laboratory, Southeast University. His research interests include wireless communication theory, cognitive radio, and signal processing.

JING WU received the B.S. degree in electronic science and technology, and the M.S. degree in communication and information system from the Nanjing University of Posts and Telecommunications, China, in 2013 and 2016, respectively. She is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering, Southeast University, China. Her research interest includes statistical signal processing and its applications in wireless communications and cyber-physical systems.