IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INTELLIGENT AND COGNITIVE TECHNIQUES FOR INTERNET OF THINGS

Received April 1, 2019, accepted April 29, 2019, date of publication May 15, 2019, date of current version May 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916883

Machine Learning Inspired Hybrid Precoding for Wideband Millimeter-Wave Massive MIMO Systems

TALHA MIR[1], MUHAMMED ZAIN SIDDIQI[1], USAMA MIR[2], RICHARD MACKENZIE[3], AND MO HAO[4]

[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Department of Computer Sciences and IT, Saudi Electronic University, Dammam 32256, Saudi Arabia
[3]British Telecom Technology, Adastral Park, Ipswich IP5 3RF, U.K.
[4]Tsinghua SEM Advanced ICT Lab, Tsinghua University, Beijing 100084, China

Corresponding author: Talha Mir (bah15@mails.tsinghua.edu.cn)

This work was supported by the National Science and Technology Major Project of China under Grant 2018ZX03001004-003.

ABSTRACT Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) has already been considered as a promising solution to meet the requirement of the higher data rate for the future Internet of Things (IoTs). Hybrid precoding is an effective solution for the mmWave massive MIMO systems to significantly decrease the number of radio frequency (RF) chains without an apparent sum-rate loss. However, the current literature on hybrid precoding considers either the high-resolution (HR) phase shifters (PSs) with high power consumption or the impractical narrowband mmWave channel model. To this end, in this paper, we investigate an energy-efficient hybrid precoding scheme using one-bit PSs for practical frequency-selective wideband mmWave massive MIMO systems. Specifically, we provide the energy consumption analysis to reveal the fact that the energy consumed by the one-bit PSs is much lower than that by the HR-PSs, and the array gain loss incurred by using one-bit PSs is minimal. Moreover, motivated by the cross-entropy optimization (CEO) algorithm evolved for machine learning, we propose the CEO-based hybrid precoding scheme to maximize the achievable sum-rate of the considered system. In the CEO-based hybrid precoding, we update the probability distributions of the elements in the hybrid precoder to minimize the cross-entropy between the two probability distributions so that we can generate the final solution close to the optimal one. Furthermore, we extend the proposed CEO-based hybrid precoding scheme from the case with one-bit PSs to the general case with HR-PSs to show that our solution can also be applied in other scenarios. The performance evaluation demonstrates that our proposed scheme can obtain near-optimal sum-rate and considerably higher energy efficiency than some existing solutions.

INDEX TERMS Millimeter-wave, massive MIMO, hybrid precoding, energy efficiency, machine learning.

I. INTRODUCTION

Internet of things (IoT) is comprised of billions of different devices and can enable various advanced technologies such as ecological protection and smart homes [1]. In some applications, IoT also imposes a demanding requirement in data rates. Fortunately, millimeter-wave (mmWave) communication can be considered as an effective solution to satisfy the high data rate requirement for IoT, since it enjoys rich and unexplored spectrum resources [2]. However, high-frequency mmWave signals at 30-300 GHz suffer from severe propagation loss due to their short wavelengths. To overcome this problem, a large number of antennas can be used to achieve a high antenna array gain, which is attributed to the short wavelengths of mmWave signals [3]. However, using a massive antenna array results in a new challenge, i.e., in the existing sub-6 GHz multiple input multiple output (MIMO) systems, a dedicated radio frequency (RF) chain is usually required for each antenna to realize the fully-digital precoding [4]. However, this fully-digital precoding is not affordable in mmWave massive MIMO systems due to a large number of required RF chains (i.e., 256 or 1024) [5]. Analog precoding with a single

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

RF chain is a suitable solution to this problem [6], but it can only support the single-stream data transmission and does not allow for spatial multiplexing [7].

To realize spatial multiplexing and to line up with the sum-rate of fully-digital precoding, hybrid (digital and analog) precoding has been recently proposed [8]. According to the basic concept of hybrid precoding, the large-size fully-digital precoder is divided into a large-size analog precoder realized by analog PSs and a small-size digital precoder requiring a small number of RF chains [8]. Due to the low-rank property of mmWave channels, a small-size digital precoder is adequate to realize the required spatial multiplexing. In this way, hybrid precoding is capable of inheriting the near-optimal sum-rate of the fully-digital precoding but significantly reduce the number of required RF chains [9].

A. PRIOR WORKS

Very recently, machine learning (ML) has been proposed to integrate with MIMO systems. For example, the support vector machines (SVM) for analog beam selection to achieve the near-optimal sum-rate with low complexity was proposed in [10]. Furthermore, [11] presented a real-time MIMO CSI feedback architecture, which uses the long-short-term memory (LSTM) to design the CSI codebook to reduce the feedback overhead of MIMO systems. Moreover, [8] discussed to utilize the matrix factorization to formulate the hybrid precoding problem and employed the orthogonal matching pursuit (OMP) algorithm to achieve the optimal solution. In [12] Yu *et al.* used the same formulated problem as in [8] and proposed an alternating minimization algorithm to obtain the near-optimal hybrid precoder. The proposed solutions in [8] and [12] can obtain a sum-rate close to fully-digital precoding, but the channel model used in these works is a narrowband mmWave channel. However, the actual mmWave systems are likely to be wideband due to the availability of substantial bandwidth at mmWave frequencies (e.g., 2 GHz) [13].

In conventional fully-digital precoding, the extension from the narrowband to the wideband channel is straightforward by utilizing the orthogonal frequency division multiplexing (OFDM). It converts the frequency-selective wideband channel to a series of frequency-flat subchannels. However, it cannot work for hybrid precoding, where the subcarrier-dependent digital precoder is designed in the frequency domain, while the subcarrier-independent analog precoder realized by the PSs is designed in the time domain which is common across all the subcarriers. This unique characteristic of frequency-selective hybrid precoding differentiates it from the traditional fully-digital precoding and makes it more challenging in the wideband mmWave channels.

Some recent works have considered hybrid precoding for practical frequency-selective wideband mmWave channels. For example, the hybrid precoding design for a mmWave MIMO-OFDM system was presented in [7] where HR-PSs with high energy consumption were considered. Then, the wideband mmWave hybrid precoding problem for maximizing the spectral efficiency with limited feedback is considered in [14]. The authors first developed a hybrid codebook for wideband mmWave systems, and then proposed the Gram-Schmidt orthogonalization based hybrid precoding algorithm for the given codebook. Likewise, a closed-form wideband hybrid precoding solution was proposed in [15] but again the HR-PSs with high energy consumption were considered.

To address the high energy consumption problem, the usage of low-resolution PSs instead of HR-PSs appears to be a feasible solution, as the power consumed by low-resolution PSs is considerably lower than that consumed by HR-PSs. To this end, Gao *et al.* [16] and [17] have proposed to use one-bit PSs for hybrid precoding, where the analog precoder was realized by energy-efficient one-bit PSs to reduce the power consumption considerably. However, similar to [8] and [12], these works also considered the narrowband mmWave channel model.

To sum up, most of the existing research works on hybrid precoding either consider the high-resolution PSs with high energy consumption or consider the frequency-flat narrowband mmWave channel model. To the best of authors' knowledge, the hybrid precoding with one-bit PSs over practical frequency-selective wideband mmWave channels has not been investigated in the literature.

B. CONTRIBUTIONS

Against the above background, this paper investigates energy-efficient hybrid precoding with one-bit PSs for frequency-selective wideband mmWave massive MIMO systems. In this considered system, one-bit PSs instead of HR-PSs are used to realize the analog precoder. Then, we formulate the sum-rate maximization problem under the constraints of constant-modulus and total power. In our considered system, the frequency-dependent digital precoder is designed in the frequency domain, while the analog precoder is frequency-independent and common across all the subcarriers.

To solve this problem, we utilize the idea of cross-entropy optimization (CEO), which was developed initially for machine learning applications, to search the optimal analog precoder intelligently. The cross-entropy (CE) algorithm is a general optimization method that has been applied successfully to many NP-hard combinatorial optimization problems [18]. It was firstly designed for rare event simulation by Rubinstein in 1997 and then was formulated as an optimization tool for both continuous and discrete optimization problems [19], [20]. Recently, it was found that the CEO has a strong connection with neural computation and reinforcement learning [21]. In our considered problem, based on the probability distributions of the elements in the hybrid precoder, several candidate hybrid precoders are generated randomly at first. Then, based on the achievable sum-rates of these candidate hybrid precoders, the proposed CEO-based hybrid precoding scheme improves the probability distributions of the elements in the hybrid precoder by reducing the

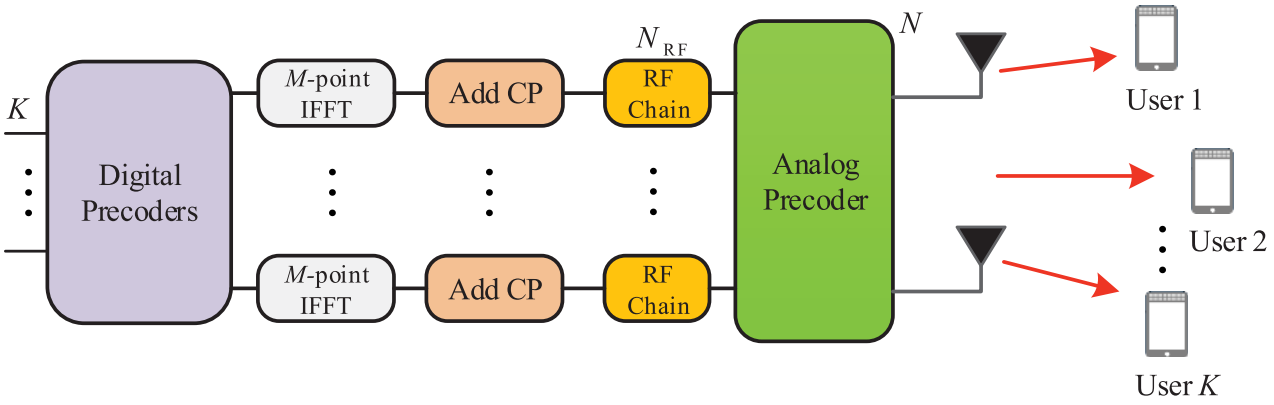

FIGURE 1. Hybrid precoding for frequency-selective wideband mmWave massive MIMO-OFDM system.

cross-entropy between the current probability distributions of the hybrid precoders and the experimental distributions of the elite samples (the solutions with high sum-rates). By repeating this procedure for a predefined number of iterations, finally, we can obtain a hybrid precoder close to the optimal solution with an adequately high probability.

We also provide the energy consumption analysis to reveal the fact that the energy consumed by the one-bit PSs based hybrid precoding is much lower than that consumed by the HR-PSs based hybrid precoding. Further, the array gain loss incurred by using one-bits PSs is small and remains constant. Moreover, we also analyze the convergence and complexity of the proposed CEO-based hybrid precoding. Besides, the proposed solution is also extended from the case with one-bit PSs to the case with higher-resolution PSs to show its generality in other scenarios. Finally, the performance evaluation shows that the proposed solution can obtain the near-optimal sum-rate and significantly higher energy efficiency than some existing solutions in frequency-selective wideband mmWave massive MIMO systems.

C. ORGANIZATION AND NOTATION

The rest of the paper is organized as follows. Section II presents the system and channel models. In Section III, we provide the energy consumption analysis of one-bit PSs bases hybrid precoding and then discuss the proposed CEO-based hybrid precoding for one-bit PSs. We illustrate the convergence and complexity analyses of the proposed scheme in Section IV. Section V provides the simulation results in terms of sum-rate and energy efficiency achieved by the proposed CEO-based hybrid precoding. Finally, the paper is concluded in Section VI.

Notations: Upper and lower-case boldface letters represents matrices and vectors, respectively. $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^T$, $\mathrm{tr}(\cdot)$, and $\|\cdot\|_F$ are the Hermitian transpose, inverse, transpose, trace, and Forbenius norm of a matrix, respectively. $\mathbb{E}(\cdot)$ is the expectation. $\otimes$ indicates the Kronecker product. $\mathcal{CN}(0,\sigma^2)$ represents the zero-mean complex Gaussian distribution with zero mean and the variance σ^2. Finally, I_N denotes the $N \times N$ identity matrix.

II. SYSTEM AND CHANNEL MODELS

This section introduces the system and channel models of hybrid precoding for frequency-selective wideband mmWave massive MIMO-OFDM systems.

As depicted in Fig. 1, we consider the hybrid precoding design for frequency-selective wideband mmWave massive MIMO-OFDM system with M subcarriers. The base station (BS) is equipped with N transmit antennas and N_{RF} RF chains ($N_{\mathrm{RF}} \leq N$) to serve K single-antenna users. We assume that each subcarrier carries the equal number of data streams ($N_s = K$) [22].

The transmitted signal vector $\mathbf{s}[m]$ of size $K \times 1$, where $m \in \{1, 2, \cdots, M\}$ at the BS is firstly precoded with a subcarrier-dependent digital precoder $\mathbf{D}_{\mathrm{BB}}[m]$ of size $N_{\mathrm{RF}} \times K$, and then transformed into the time domain by using the M-point inverse fast-Fourier transform (IFFT). We follow the MIMO-OFDM transmission scheme with the common assumption that the maximum delay spread of the channel will fall within the cyclic prefix (CP) interval [7]. After adding the CP of length D to the signal, the whole OFDM signal is fed to the subcarrier-independent analog precoder $\mathbf{A}_{\mathrm{RF}}$ of size $N \times N_{\mathrm{RF}}$.

It should be emphasized that the analog precoder $\mathbf{A}_{\mathrm{RF}}$ is realized in the time domain and it remains common across all M subcarriers. By contrast, the digital precoder $\mathbf{D}_{\mathrm{BB}}[m]$ is subcarrier-dependent and is performed in the frequency domain. This important characteristic differentiates the frequency-selective hybrid precoding from the conventional fully-digital precoding which is subcarrier-dependent.

The transmitted signal vector $\mathbf{s}[m]$ at the mth subcarrier where $m \in \{1, 2, \cdots, M\}$ satisfies the normalized power constraint, i.e., $\mathbb{E}\left[\mathbf{s}[m](\mathbf{s}[m])^H\right] = \mathbf{I}_K$. Further, as the analog precoder $\mathbf{A}_{\mathrm{RF}}$ is realized by one-bit PSs in our considered system so it will bring some additional hardware constraints, which will be explained in Section-III. Furthermore, the digital precoder $\mathbf{D}_{\mathrm{BB}}[m]$ satisfies the total power constraint $\sum_{m=1}^{M} \|\mathbf{A}_{\mathrm{RF}}\mathbf{D}_{\mathrm{BB}}[m]\|_F^2 = \rho$, where ρ is the total transmit power.

At the user side, when perfect synchronization is assumed, the received signal $y_k[m]$ for the kth user at the mth subcarrier

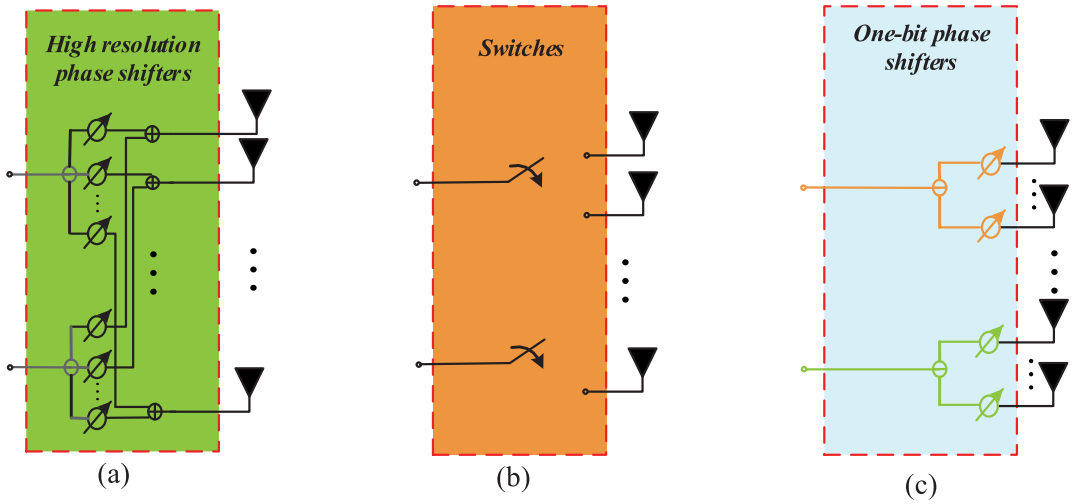

FIGURE 2. Different hybrid precoding architectures: (a) HR-PSs based hybrid precoding; (b) Switch-based (SW) hybrid precoding; (c) One-bit PSs based hybrid precoding.

can be expressed as [7]

$$y_k[m] = \mathbf{h}_k^H[m]\,\mathbf{A}_{\mathrm{RF}}\mathbf{D}_{\mathrm{BB}}[m]\,\mathbf{s}[m] + n_k[m], \tag{1}$$

where $\mathbf{H}[m] = [\mathbf{h}_1[m], \mathbf{h}_2[m], \ldots, \mathbf{h}_K[m]]^H$ is the channel matrix with $\mathbf{h}_k[m]$ $k = 1, 2, \ldots, K$ being the $N \times 1$ channel vector between the kth user and BS at the mth subcarrier, and $n_k[m]$ is the additive Gaussian white noise (AWGN), i.e., $n_k[m] \sim \mathcal{CN}(0, \sigma_k^2)$ with σ_k^2 representing the noise power at the kth user.

A. CHANNEL MODEL

To capture the limited scattering features of frequency-selective wideband mmWave channels, we adopt the widely used geometric delay-d channel model, which can be expressed as [8]

$$\mathbf{h}_k(d) = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \alpha_{k,l}\,\mathbf{a}_{\mathrm{BS}}(\theta_{k,l})\,p_r(dT_s - \tau_{k,l}), \tag{2}$$

where L_{k} is the number of dominant paths, $\alpha_{k,l}$ and $\mathbf{a}_{\mathrm{BS}}(\theta_{k,l})$ are the complex gain of the lth path and the antenna array response vector of the BS with $(\theta_{k,l})$ being the angle of departure (AoD) of the lth path. $p_r(dT_s - \tau_{k,l})$ is the pulse-shaping function for T_s-space signaling at time delay $\tau_{k,l}$. When the widely-used uniform linear array (ULA) is considered, the antenna array response vector $\mathbf{a}_{\mathrm{BS}}(\theta_{k,l})$ in (2) can be expressed as

$$\mathbf{a}_{\mathrm{BS}}(\theta_{k,l}) = \frac{1}{\sqrt{N}}\left[1, e^{\frac{j2\pi}{\lambda} d_s \sin(\theta_{k,l})}, \ldots, e^{j(N-1)\frac{j2\pi}{\lambda} d_s \sin(\theta_{k,l})}\right]^T, \tag{3}$$

where λ is the wavelength of the signal, d_s is the antenna spacing which is usually set as $d_s = \lambda/2$. Given the delay-d channel model in (2), the channel $\mathbf{h}_k[m]$ between the BS and the kth user at the mth subcarrier can be written as [15]

$$\mathbf{h}_k[m] = \sum_{d=1}^{D} \mathbf{h}_k(d)\, e^{-j\frac{2\pi m}{M} d}, \tag{4}$$

where D is the length of the CP.

III. ENERGY EFFICIENT ONE-BIT HYBRID PRECODING

On the basis of Section II, in this section we investigate the sum-rate maximization problem for the energy efficient one-bit PSs based hybrid precoding.[1] At first, we analyze the energy consumption and the array gain loss incurred by using one-bit PSs. Then, we formulate the sum-rate maximization problem for the one-bit PSs based hybrid precoding with some practical constraints, particularly the constraints caused by one-bit PSs. Finally, we present the CEO-based hybrid precoding scheme to solve the sum-rate maximization problem for frequency-selective wideband mmWave massive MIMO systems.

A. ENERGY CONSUMPTION ANALYSIS OF DIFFERENCT HYBRID PRECODING SCHEMES

Fig. 2 shows the difference between different hybrid precoding architectures. Fig. 2 (a) is for the HR-PSs (e.g., 4-bit) based hybrid precoding architecture, where each RF chain is connected with all the antennas through the network of HR-PSs [3]. Note that the power consumed by these HR-PSs is considerably high (e.g., $P_{\text{HR-PS}} = 40$ mW for a 4-bit PS), making this architecture suffer from high hardware cost and power consumption. To overcome this problem, a switch-based hybrid precoding architecture is proposed as shown in Fig. 2 (b) [23], where a network of low-cost switches instead of HR-PSs is used to realize the analog precoder. However, this architecture cannot attain the array gain of mmWave massive MIMO systems since

[1] We will consider Z-bit PSs based hybrid precoding in Section III-D.

only N_{RF} antennas are active simultaneously. To this end, a one-bit PSs based hybrid precoding architecture as shown in Fig. 2 (c) is proposed, which can provide a better trade-off between the performance and hardware cost/power consumption [16]. Specifically, for the one-bit PSs based hybrid precoding architecture, each RF chain is only connected to a sub-antenna array with $T = N/N_{\mathrm{RF}}$ antennas with low-cost one-bit PSs (comparable to switches), which is similar to the sub-connected architecture [5]. This feature enables us to reduce the hardware cost/power consumption, but also obtain the acceptable array gain to achieve a high sum-rate.

The energy consumed by the HR-PSs based hybrid precoding can be expressed as

$$P_{\text{HR-HP}} = \rho + N_{\mathrm{RF}}P_{\mathrm{RF}} + NN_{\mathrm{RF}}P_{\text{HR-PS}} + P_{\mathrm{BB}}, \tag{5}$$

where ρ is the total transmit power, P_{RF}, $P_{\text{HR-PS}}$, and P_{BB} are the energy consumed by an RF chain, a HR-PSs, and a baseband unit, respectively. It is worth emphasizing that HR-PSs-based hybrid precoding can obtain the near-optimal sum-rate performance, but they necessitate a considerable number of PSs to fulfill this purpose, for instance $N \times N_{\mathrm{RF}} = 64 \times 16 = 1024$ [3].

Conversely, the energy consumed by switch based architecture can be expressed as

$$P_{\text{SW-HP}} = \rho + N_{\mathrm{RF}}P_{\mathrm{RF}} + N_{\mathrm{RF}}P_{\mathrm{SW}} + P_{\mathrm{BB}}, \tag{6}$$

where P_{SW} is the energy consumed by a switch, which is lower than HR-PS (e.g., $P_{\mathrm{SW}} = 5$ mW) [23]. However, the SW-based hybrid precoding cannot attain the array gain of mmWave massive MIMO systems, because only N_{RF} antennas are active simultaneously. Hence, an apparent sum-rate performance degradation will be caused [24].

One-bit PSs based hybrid precoding is attractive to provide a favorable trade-off between the two precoding schemes described above. Specifically, for one-bit PSs based hybrid precoding, each RF chain is only connected to a sub-antenna array with $T = N/N_{\mathrm{RF}}$ antennas, which is similar to the sub-connected structure [5]. Thus, the energy consumed by the one-bit PSs based hybrid precoding can be expressed as

$$P_{\text{one-bit-HP}} = \rho + N_{\mathrm{RF}}P_{\mathrm{RF}} + NP_{\text{one-bit-PS}} + P_{\mathrm{BB}}. \tag{7}$$

It is important to mention here that the energy consumed by one-bit PSs ($P_{\text{one-bit-PS}}$) is very low (i.e.,$P_{\text{one-bit-PS}} = 5$mW) [23]. Therefore, by comparing (5) and (7), it can be concluded that the energy consumption of the one-bit PSs based hybrid precoding is considerably lower than that of the HR-PSs based hybrid precoding. Furthermore, compared with the SW-based hybrid precoding, one-bit PSs based hybrid precoding can utilize all antennas to achieve the desired array gain, which is important for mmWave massive MIMO systems.

In **Lemma 1**, we will show that the array gain loss incurred by using one-bit PSs-based hybrid precoding is limited and remains constant as number of antennas goes to infinity.

Lemma 1: Without the loss of generality, the ratio γ between the array gain achieved by one-bit PSs-based hybrid precoding and that achieved by HR-PSs-based hybrid precoding satisfies

$$\lim_{N\to\infty,\,\frac{N}{T}=N_{RF}} |\gamma|^2 = \frac{4}{\pi^2}\sin^2\frac{\pi}{2}. \tag{8}$$

Proof: Refer to Appendix A. ∎

Lemma 1 provides the insight that the performance loss incurred by using one-bit PSs based hybrid precoding is acceptable and remains constant when the number of BS antennas increases. Considering the higher energy efficiency of one-bit PSs, we can easily see that one-bit PSs based hybrid precoding provides a better trade-off between the HR-PSs-based and switch-based hybrid precoding schemes. Next, we formulate the sum-rate maximization problem for one-bit PSs based hybrid precoding under the practical hardware constraints caused by one-bit PSs.

B. PROBLEM FORMULATION

For the existing hybrid precoding schemes as discussed in Section I-A, the non-zero elements of $\mathbf{A}_{\mathrm{RF}}$ are realized by HR-PS as depicted in Fig. 2 (a) [3]. Generally, HR-PSs is employed to ensure a high sum-rate, but the power consumed by these HR-PSs is significantly high. One promising method to solve this problem is to replace the HR-PSs with one-bit PSs. Therefore, in this section, we design an efficient hybrid precoder with one-bit PSs for frequency-selective wideband mmWave massive MIMO systems in this section.

To do this, we first describe the additional hardware constraints incurred by the one-bit PSs, which are different from the conventional constraints. The first constraint is that the analog precoder $\mathbf{A}_{\mathrm{RF}}$ has to be a block diagonal matrix

$$\mathbf{A}_{\mathrm{RF}} = \begin{pmatrix} \bar{\mathbf{a}}_1^{\mathrm{RF}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{\mathbf{a}}_{N_{\mathrm{RF}}}^{\mathrm{RF}} \end{pmatrix}_{N\times N_{\mathrm{RF}}}, \tag{9}$$

where $\bar{\mathbf{a}}_n^{\mathrm{RF}}$ is $T \times 1$ analog precoder on the nth sub-antenna array. Further, all the N nonzero elements of $\mathbf{A}_{\mathrm{RF}}$ should belong to

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

Our goal is to develop an efficient hybrid precoder ($\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}$ and $\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]$) to maximize the achievable sum-rate R. This problem can be formulated as

$$\begin{aligned} &\left(\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}, \left\{\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]\right\}_{m=1}^{M}\right) \\ &\quad = \underset{\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}, \left\{\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]\right\}_{m=1}^{M}}{\arg\max} \mathrm{R}, \\ &\qquad \text{s.t. } \mathbf{A}_{\mathrm{RF}} \in \mathcal{A}, \\ &\qquad \sum_{m=1}^{M} \|\mathbf{A}_{\mathrm{RF}}\mathbf{D}_{\mathrm{BB}}[m]\|_{\mathrm{F}}^2 = \rho, \end{aligned} \tag{11}$$

where $\mathcal{A}$ is the set that contains all the possible analog precoders which satisfies the above two constraints (9) and (10).

Consquently, the sum-rate R can be expressed as

$$\mathrm{R}_k[m] = \frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M}\log_2\left(1+\gamma_k[m]\right), \tag{12}$$

where $\gamma_k[m]$ is the signal-to-interference-plus-noise ratio (SINR) for the kth user at subcarrier m, which is given by

$$\gamma_k[m] = \frac{\left|\mathbf{h}_k^H\mathbf{A}_{\mathrm{RF}}\mathbf{d}_{\mathrm{BB}}^k[m]\right|^2}{\sum\limits_{k'\neq k}^{K}\sum\limits_{m=1}^{M}\left|\mathbf{h}_k^H[m]\mathbf{A}_{\mathrm{RF}}\mathbf{d}_{\mathrm{BB}}^{k'}[m]\right|^2 + K\sigma^2}, \tag{13}$$

where σ^2 is the AWGN power and $\mathbf{d}_{\mathrm{BB}}^k[m]$ is the kth column of $\mathbf{D}_{\mathrm{BB}}[m]$.

The constraints (9) and (10) on the analog precoder $\mathbf{A}_{\mathrm{RF}}$ are non-convex. To solve this challenging problem, we propose the machine learning inspired CEO-based hybrid precoding scheme in the next section.

C. PROPOSED CEO-BASED HYBRID PRECODING

To tackle the above non-convex problem, we first decouple the joint design of the analog and digital precoder. As all the N non-zero elements of the analog precoder belong to (10), hence the number of potential $\mathbf{A}_{\mathrm{RF}}$ is finite. Therefore, we can perceive the problem in (11) as a non-coherent combining problem [25]. We can start by selecting one candidate analog precoder $\mathbf{A}_{\mathrm{RF}}$ and then we can compute the optimal digital precoders $\mathbf{D}_{\mathrm{BB}}[m]_{m=1}^{M}$ by using the effective channel matrix $\mathbf{H}[m]\mathbf{A}_{\mathrm{RF}}$ at each subcarrier m. By searching all possible analog precoders $\mathbf{A}_{\mathrm{RF}}$, finally we can find the optimal analog $\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}$ and digital $\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]_{m=1}^{M}$ precoders at each subcarrier. Unfortunately, such an exhaustive search requires filtering through all 2^N possible combinations, which involves unaffordably high complexity, as N is generally large in mmWave massive MIMO systems (e.g., $N = 64, 2^{64} \approx 1.84\times10^{19}$) [2].

To intelligently search the optimal analog precoder, we utilize the idea of cross-entropy optimization (CEO) which was developed for machine learning applications. In CEO, based on the probability distributions of the elements in the hybrid precoder, several candidate hybrid precoders are generated randomly. Then, based on the achievable sum-rates of these generated candidate hybrid precoders, It improves the probability distributions of the elements in $\mathbf{A}_{\mathrm{RF}}$ by reducing the cross-entropy between the two probability distributions (i.e., the current probability distributions of the hybrid precoders and the experimental probability distributions of elite samples). We have provided the pseudo-code of CEO-based hybrid precoding in **Algorithm 1**. Next, we explain several important steps of CEO-based hybrid precoding in detail.

At first, we collect all of the non-zero elements in $\mathbf{A}_{\mathrm{RF}}$ in an $N\times1$ vector which can be shown as $\mathbf{a} = \left[\left(\bar{\mathbf{a}}_1^{RF}\right)^T, \left(\bar{\mathbf{a}}_2^{RF}\right)^T, \ldots \left(\bar{\mathbf{a}}_{N_{\mathrm{RF}}}^{RF}\right)^T\right]^T$. Then, we define the probability parameter $\mathbf{p} = [p_1, p_2, \ldots, p_N]^T$ as an $N\times1$ vector, where $0 \le p_n \le 1$ represents the probability that $a_n = 1/\sqrt{N}$, where a_n is the nth element of $\mathbf{a}$. Furthermore, $\mathbf{p}^{(0)}$ is initialized as $\mathbf{p}^{(0)} = \frac{1}{2}\times\mathbf{1}_{N\times1}$. After initializing the parameters for CEO-based hybrid precoding with one-bit PSs, we start the search and update process which involves the following key steps.

Algorithm 1 The Proposed CEO-Based Hybrid Precoding With One-Bit PSs

Input: Wideband channel matrix $\mathbf{H}[m]$; Number of subcarriers M; Number of candidates O; Smoothing step size Ω; Number of iterations I; Number of elites $\mathbf{O}_{elite}$;.

1: **Initialize:** $i=0$, $\mathbf{p}^{(0)} = 1/2\times\mathbf{1}_{N\times1}$.
2: **for** $0\le i\le I$ **do**
3: Generate O candidate $\mathbf{A}_{\mathrm{RF}}$ randomly as $\left\{\mathbf{A}_{\mathrm{RF}}^{o}\right\}_{o=1}^{O}$ based on $\Xi\left(\mathcal{A};\mathbf{p}^{(i)}\right)$;
4: **for** $m=1:M$ **do**
5: Calculate O corresponding digital precoders $\mathbf{D}_{\mathrm{BB}}[m]$ based on the effective channel as in (15);
6: **end for**
7: Compute the achievable sum-rate $\left\{\mathrm{R}\left(\mathbf{A}_{\mathrm{RF}}^{o}\right)\right\}_{o=1}^{O}$ using (13);
8: Sort $\left\{\mathrm{R}\left(\mathbf{A}_{\mathrm{RF}}^{o}\right)\right\}_{o=1}^{O}$ in descending as $\mathrm{R}\left(\mathbf{A}_{\mathrm{RF}}^{[1]}\right)\ge\mathrm{R}\left(\mathbf{A}_{\mathrm{RF}}^{[2]}\right)\ge\cdots\ge\mathrm{R}\left(\mathbf{A}_{\mathrm{RF}}^{[O]}\right)$;
9: Select elites as $\mathbf{A}_{\mathrm{RF}}^{[1]}, \mathbf{A}_{\mathrm{RF}}^{[2]}, \ldots, \mathbf{A}_{\mathrm{RF}}^{[O_{elite}]}$;
10: Update the probability $\mathbf{p}^{(i+1)}$ according to $\left\{\mathbf{A}_{\mathrm{RF}}^{[o]}\right\}_{o=1}^{O_{elite}}$ by using (20);
11: $i=i+1$;
12: **end for**

Output: Analog precoder $\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}$, Digital precoder $\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]_{m=1}^{M}$;

In the ith iteration, at **step 3**, we generate O candidate analog precoders $\left\{\mathbf{A}_{\mathrm{RF}}^{o}\right\}_{o=1}^{O}$ depending on the probability distribution $\Xi\left(\mathcal{A};\mathbf{p}^{(i)}\right)$, where we generate $\{\mathbf{a}^o\}_{o=1}^{O}$ according to $\mathbf{p}^{(i)}$, and organize them in matrices form which belongs to $\mathcal{A}$. Next, in **step 5**, by using the effective channel $\mathbf{H}_{\mathrm{eq}}^{o}[m] = \mathbf{H}[m]\mathbf{A}_{\mathrm{RF}}^{o}$ for $1\le o\le O$, we calculate the corresponding digital precoders $\mathbf{D}_{\mathrm{BB}}[m]_{m=1}^{M}$ at each subcarrier m. Here, we employ the zero-forcing digital precoding scheme with low-complexity and near-optimal performance which can be computed as

$$\mathbf{G}^o[m] = \left(\mathbf{H}_{\mathrm{eq}}^{o}[m]\right)^H\left(\mathbf{H}_{\mathrm{eq}}^{o}[m]\left(\mathbf{H}_{\mathrm{eq}}^{o}[m]\right)^H\right)^{-1}, \tag{14}$$

$$\mathbf{D}_{\mathrm{BB}}^{o}[m] = \beta^o\mathbf{G}^o[m], \tag{15}$$

where $\beta^o = \sqrt{\rho}/\left\|\mathbf{A}_{\mathrm{RF}}^{o}\mathbf{G}^o[m]\right\|_F$ denotes the power normalized factor.

After calculating the M digital precoders for each candidate analog precoder, we calculate the sum-rate for O pairs. To calculate the achievable sum-rate $\left\{R\left(\mathbf{A}_{\mathrm{RF}}^{o}\right)\right\}_{o=1}^{O}$ in **step 7**, we substitute $\mathbf{A}_{\mathrm{RF}}^{o}$ and $\mathbf{D}_{\mathrm{BB}}^{o}[m]$ in (13). Subsequently, in **step 8**, the calculated sum-rates are sorted in descending order. In **step 9**, we select the O_{elite} candidate analog precoders with the highest sum-rate as the elite candidates. In **step 10**, these

selected elite candidates are employed to update the probability distribution $\mathbf{p}^{(i+1)}$ to minimize the cross-entropy between the current probability distributions of the hybrid precoder and the experimental distributions of the elite candidates [19] (the solutions with high sum-rates) as

$$\mathbf{p}^{(i+1)} = \arg\max_{p^{(i)}} \frac{1}{O} \sum_{o=1}^{O_{elite}} \ln \Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; \mathbf{p}^{(i)} \right), \tag{16}$$

where $\Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; \mathbf{p}^{(i)} \right)$ is the probability to generate $\mathbf{A}_{\mathrm{RF}}^{[o]}$.

Note that $\Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; \mathbf{p}^{(i)} \right) = \Xi \left(\mathbf{a}^{[o]}; \mathbf{p}^{(i)} \right)$, and the nth element $a_n^{[o]}$ of $\mathbf{a}^{[o]}$ is a Bernoulli random variable, where $a_n^{[o]} = \frac{1}{\sqrt{N}}$ has a probability $p_n^{(i)}$. Similarly, $a_n^{[o]} = \frac{-1}{\sqrt{N}}$ has a probability $1 - p_n^{(i)}$. Therefore, we have

$$\Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; \mathbf{p}^{(i)} \right) = \prod_{n=1}^{N} \left(p_n^{(i)} \right)^{\frac{1}{2}\left(1+\sqrt{N}a_n^{(o)}\right)} \left(1 - p_n^{(i)} \right)^{\frac{1}{2}\left(1+\sqrt{N}a_n^{(o)}\right)}. \tag{17}$$

By substituting (17) into (16), the first order derivative of (17) w.r.t $\mathbf{p}_n^{(i)}$ can be represented as

$$\frac{1}{O} \sum_{o=1}^{O} \left(\frac{1 + \sqrt{N} a_n^{[o]}}{2 p_n^i} - \frac{1 - \sqrt{N} a_n^{[o]}}{2 \left(1 - p_n^i \right)} \right). \tag{18}$$

Setting (18) to zero, ${\mathbf{p}_n}^{(i+1)}$ is updated in **step 10** as

$$\mathbf{p}_n^{(i+1)} = \frac{\sum_{o=1}^{O} \left(\sqrt{N} a_n^{[o]} + 1 \right)}{2}. \tag{19}$$

To avoid the local optimum, we can further modify (19) as

$$\mathbf{p}_n^{(i+1)} = \Omega^{(i+1)} \times \mathbf{p}_n^{(i+1)} + \left(1 - \Omega^{(i+1)} \right) \times \mathbf{p}_n^{(i)}. \tag{20}$$

where $0 < \Omega^{(i+1)} < 1$ is the smoothing step size in current iteration $(i+1)$.

The rationality of (20) will be explained in Section-IV-A. We repeat the **step 3** to **step 10** for the predefined I iterations. Then, the optimal analog precoder $\mathbf{A}_{\mathrm{RF}}^{\mathrm{opt}}$ with best sum-rate performance i.e., $R\left[\mathbf{A}_{\mathrm{RF}}^{[1]}\right]$ is obtained. Finally, the optimal digital precoder $\mathbf{D}_{\mathrm{BB}}^{\mathrm{opt}}[m]_{m=1}^{M}$ at subcarrier m is calculated based on the obtained optimal analog precoder.

D. CEO-BASED HYBRID PRECODING WITH $\mathcal{Z}$-bit PSs

Next, we present the extension of the proposed CEO-based hybrid precoding to the more general case (i.e., Z-bit PSs). To do so, we first generate the parameterized sampling distribution, which generates the candidate solutions for the following iterations. A simple method to generate a random sample $\bar{\mathbf{a}} = \{\bar{a}_n\}_{n=1}^{N}$ is to independently draw from $[\bar{a}_n, \bar{a}_2, \cdots, \bar{a}_N]$, where each $\bar{a}_n$ belongs to a $|\mathcal{Z}|$-point discrete distribution $\left\{ p_{z,n}^{(i)} \right\}_{z=1}^{|\mathcal{Z}|}$. Here, $\left\{ p_{z,n}^{(i)} \right\}$ is the probability of the zth quantized phase in set $\mathcal{Z}$ being selected as $\bar{a}_n$. Then, using the elite candidates to update the probability to minimize the cross-entropy can be written as

$$p_{z,n}^{(i+1)} = \arg\max_{p(i)} \frac{1}{O} \sum_{o=1}^{O_{\mathrm{elite}}} \ln \Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; p_{z,n}^{(i)} \right), \tag{21}$$

where $\Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; p_{z,n}^{(i)} \right)$ is given by

$$\Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; p_{z,n}^{(i)} \right) = \prod_{n=1}^{N} \sum_{z=1}^{|\mathcal{Z}|} p_{z,n}^{(i)} \mathbb{1}_{\{\bar{\mathbf{a}}^{[o]} \in \mathcal{A}_{z,n}\}}, \tag{22}$$

where the indicator function $\mathbb{1}_{\{\cdot\}} = 1$ if the statement $\{\cdot\}$ is true, otherwise $\mathbb{1}_{\{\cdot\}} = 0$. Furthermore, $\mathcal{A}_{z,n} = \left\{ \bar{\mathbf{a}} \in \mathcal{Z}^N : \bar{a}_n = \frac{2\pi z}{|\mathcal{Z}|} \right\}$ where $\mathcal{Z} \triangleq \left\{ \frac{2\pi z}{2^Z} | z = 1, \cdots, 2^Z \right\}$ denotes the set which contains all the possible analog precoders satisfying the given constraint (9).

Here, since only one quantized phase can be assigned to one PS, the probability of each $\left\{ p_{z,n}^{(i)} \right\}$ is constrained to a sum of one, i.e., $\sum_{z=1}^{|\mathcal{Z}|} p_{z,n}^{(i)} = 1$. To meet this constraint, we introduce the Lagrange multiplier $\{\mathcal{L}_n\}_{n=1}^{N}$ into (21) as

$$p_{z,n}^{(i+1)} = \arg\max_{p(i)} \frac{1}{O} \sum_{o=1}^{O_{\mathrm{elite}}} \ln \Xi \left(\mathcal{A}_{\mathrm{RF}}^{[o]}; p_{z,n}^{(i)} \right) + \sum_{n=1}^{N} \mathcal{L}_n \left(\sum_{z=1}^{|\mathcal{Z}|} \left(p_{z,n}^{(i)} - 1 \right) \right). \tag{23}$$

By taking the first derivative of the target in (23) with respect to $\left\{ p_{z,n}^{(i)} \right\}$, and then by setting the result equal to zero, we get

$$\frac{1}{O} \sum_{o=1}^{O_{\mathrm{elite}}} \mathbb{1}_{\{\bar{\mathbf{a}}^{[o]} \in \mathcal{A}_{z,n}\}} + \mathcal{L}_n p_{z,n}^{(i)} = 0. \tag{24}$$

Then, by adding all of (24) for $z = 1, 2, \cdots, |\mathcal{Z}|$, we obtain

$$\mathcal{L}_n = -\frac{1}{O} \sum_{o=1}^{O_{\mathrm{elite}}}. \tag{25}$$

Finally, substituting (25) into (24), we get the probability update as

$$p_{z,n}^{(i+1)} = \frac{\sum_{o=1}^{O_{\mathrm{elite}}} \mathbb{1}_{\{\bar{\mathbf{a}}^{[o]} \in \mathcal{A}_{z,n}\}}}{\sum_{o=1}^{O_{elite}}}. \tag{26}$$

The efficiency of the proposed CEO-based hybrid precoding with Z-bit PSs will be verified in the Section V.

IV. CONVERGANCE AND COMPLEXITY ANALYSIS

This section presents the convergence and computational complexity analyses of the proposed CEO-based hybrid precoding scheme.

A. CONVERGENCE ANALYSIS

Sometimes the local optimum (premature convergence) may be achieved at the beginning of the optimization process by using (19) directly to update the probability distribution [18]. To avoid this premature convergence, updating (16) incrementally for the next iteration is a favorable solution rather than setting up a new probability distribution for the next iteration which can be expressed as [21].

$$\mathbf{p}_n^{(i)} = \Omega^{(i)} \times \mathbf{p}_n^{(i)} + \left(1 - \Omega^{(i)}\right) \times \mathbf{p}_n^{(i-1)}, \tag{27}$$

where $\mathbf{p}_n^{(i-1)}$ is a probability update obtained in the previous iteration and $0 < \Omega^{(i)} < 1$ being the smoothing step size at current iteration i. To avoid the local optimum, choosing the appropriate smoothing step is necessary. To increase the probability to generate the optimal solution smoothing step size selection is a critical parameter, the following theorem can verify this.

Theorem 1: The proposed CEO-based hybrid precoding algorithm can generate the optimal solution A_{RF}^{opt} with the probability approaching to 1*, if the smoothing step size $\Omega^{(i)}$ is set as*

$$\Omega^{(i)} = \frac{\mu}{i \ln (i+1)} \tag{28}$$

where $0 < \mu < 1$ is a smoothing parameter and set as a constant.

Proof: Refer to Appendix B. ■

Remark 1 (Parameter Settings): The parameters (O, μ) must be determined by employing the CEO framework to the considered problem. In this paper, we consider the smoothing parameter ($\mu = 0.8$), as suggested in [25]. Conversely, the value of O is problem dependent. To determine the appropriate number candidate solutions O the simplest way is to evaluate the performance of the algorithm against different values of O and chose the best value for extensive simulations. This fact further demonstrates in the simulation section for clear understanding.

Remark 2 (Stopping Criterion): The most intuitive way to terminate the CEO algorithm is to run the algorithm for a reasonable period until it reaches the predefined maximum number of iterations. Under such a condition, the complexity of the CEO-algorithm can be analyzed by using the number of evaluations of the objective function, which is the product of the candidate solution O and the number of iterations I.

B. COMPUTATIONAL COMPLEXITY ANALYSIS

From **Algorithm 1**, it can be observed that the complexity of the proposed CEO-based hybrid precoding scheme originates from **steps** 5, 7, and 10. As in **step 5**, it is required to calculate the effective channel matrices $\left\{\mathbf{H}_{\text{eq}}^{o}[m]\right\}_{o=1}^{O}$ for each candidate solution O, and the corresponding digital precoder $\left\{\mathbf{D}_{\text{BB}}^{o}[m]\right\}_{o=1}^{O}$ according to (15) and (16). Consequently, this part has the complexity of $\mathcal{O}\left(MONK^2\right)$. Next in **step 7**, the sum-rate achieved by each candidate solution is computed. Here, we apply the classical ZF precoder, the SINR of each user for oth candidate is simplified to $\gamma^o = \left(\beta^o/\sigma\right)^2$. Therefore, it has complexity $\mathcal{O}(MO)$. **step 10** is the updating step of $\mathbf{p}_n^{(i+1)}$ according to (20), which comprises the complexity of $O(MNO_{elite})$.

In conclusion, the total computational complexity of the proposed CEO-based hybrid precoding after running I iterations is $\mathcal{O}\left(MIONK^2\right)$. Since, I and O are not "necessarily very large. Further, M is fixed for mmWave frequency-selective channels. Hence, the complexity of CEO-based hybrid precoding is satisfactory and can be considered comparable to some existing solutions [3], [23].

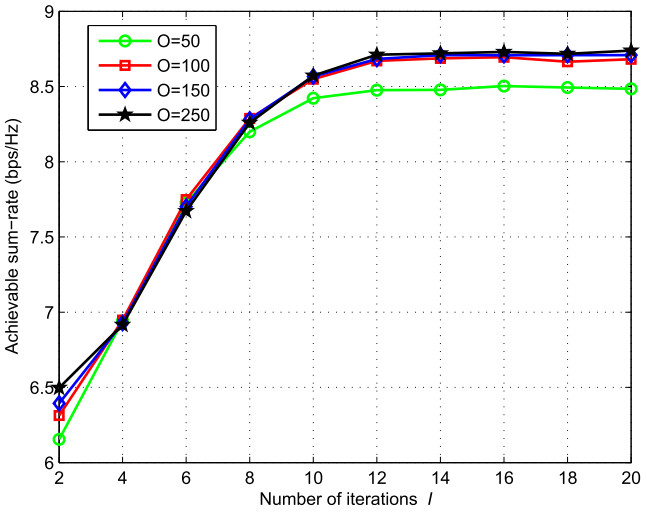

FIGURE 3. Achievable sum-rate of CEO-based hybrid precoding with one-bit PSs against *O* and *I*.

V. SIMULATION RESULTS

This section discusses the performance achieved by the proposed CEO-based hybrid precoding in terms of sum-rate and energy efficiency. We compare our proposed solution with some existing solutions i.e., [3] and [23]. The basic simulation parameters are as follows

We adopt the commonly used geometric channel model as defined in (2) to realize the practical frequency-selective wideband mmWave channel for $K = 4$ users. The bandwidth is set as 4 GHz and the carrier frequency is 60 GHz. The number of paths is set as $L_k = 3$, $\theta_{k,l}$ is assumed to follow the uniform distribution within $[0, 2\pi]$ for $1 \leq l \leq L_k$ and $\alpha_{k,l} \sim \mathcal{CN}(0, 1)$. The number of subcarriers M is 128, and CP length is $D = 64$. The path delay is uniformly distributed within $[0, DT_s]$ with $T_s = 1$. The uniform linear array (ULA) is adopted in simulations [26]. Finally, SNR is defined as ρ/σ^2.

Fig. 3 presents the achievable sum-rate of the proposed CEO-based hybrid precoding with one-bit PSs against the different numbers of iterations I and candidate solutions O, where $N = 64, N_{\text{RF}} = 4$ and SNR $= 10$dB. It can be observed that, as the number of candidate solution O is small (i.e., $O = 50$), the achievable sum-rate is low. Whereas, by increasing O (e.g., $O = 100$ or $O = 150$), there is an

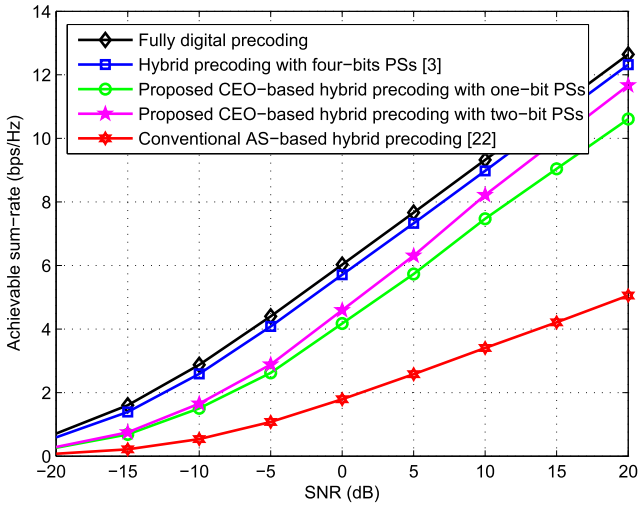

FIGURE 4. Achievable sum-rate of CEO-based hybrid precoding with one-bit PSs, (when $N = 64, N_{RF} = 4, M = 128$) compared to other precoding systems.

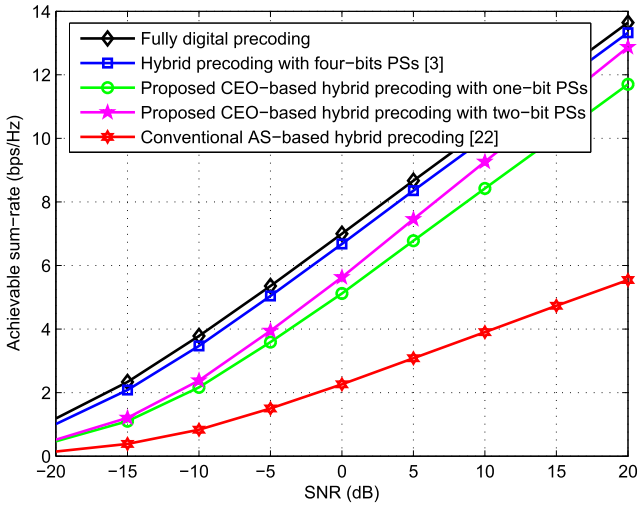

FIGURE 5. Achievable sum-rate of CEO-based hybrid precoding with two-bit PSs, (when $N = 128, N_{RF} = 8, M = 256$) compared to other precoding systems.

obvious enhancement in the sum-rate. Furthermore, if O is adequately large (e.g., $O = 200$), this trend is not apparent, and the performance improvement is negligible. This clarifies that it is not necessary for O to be too large; for instance, $O = 100$ is sufficient to achieve satisfactory results. Additionally, Fig. 3 shows that the slightly small number of iterations (i.e., $I = 20$) is sufficient for the proposed scheme to converge.

Fig. 4 presents the sum-rate achieved by the proposed CEO-based hybrid precoding compared with some existing solutions for the frequency-selective wideband mmWave massive MIMO system. We set $N = 64, N_{\text{RF}} = 4$, $M = 128$, and $K = 4$. Furthermore, the number of candidate solution, the number of elite candidates, and the number of iterations are set as $O = 100$, $O_{\text{elite}} = 40$, and $I = 20$, respectively. We consider the HR-PSs based hybrid precoding using four-bit PSs [3] and the proposed CEO-based hybrid precoding using one-bit PSs and two-bit PSs. Also, the antenna selection (AS)-based hybrid precoding [23] is considered for the SW-based architecture. From Fig. 4 we can observe that the proposed CEO-based hybrid precoding can achieve considerably higher sum-rate than the AS-based hybrid precoding. Furthermore, an obvious performance gap can be observed between the proposed solution using one-bit PSs and HR-based hybrid precoding using four-bit PSs, but this gap can be reduced as the resolution of PSs is increased.

In Fig. 5, we set $N = 128, N_{\text{RF}} = 8, M = 256$, and $K = 4$. Fig 5, clearly shows that the proposed solution with one-bit PSs can achieve a considerably higher sum-rate than the conventional SW-based hybrid precoding and the gap between the HR-based hybrid precoding with four-bit PSs and CEO-based hybrid precoding with two-bit PSs is small as compared to one-bit PSs case.

Fig. 6 compares the achievable sum-rate between the proposed CEO-based hybrid precoding and the exiting hybrid

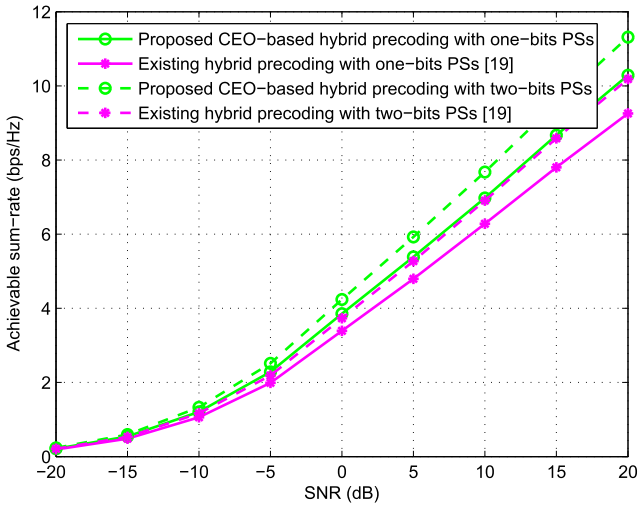

FIGURE 6. Sum-rate comparison of CEO-based hybrid precoding and existing hybrid precoding, when $N = 64, N_{RF} = 4, M = 128$.

precoding scheme proposed for low-resolution PSs [17]. Specifically, in Fig. 6, we consider two cases (i.e., one-bit PSs and two-bit PSs) for both schemes. At first, we set $N = 64, N_{\text{RF}} = 4, M = 128$ for $K = 4$. It can be observed that the performance attained by the proposed solution for both cases as compared to [17] is undeniable. This is due to the reason that the solution in [17] uses the predefined set of phases from codebook which may cause performance loss as compared to our proposed scheme.

Furthermore, in Fig. 7, we employ the same sum-rate comparison with different parameter settings, (i.e., when $N = 128, N_{\text{RF}} = 8, M = 256$ for $K = 4$) for both one-bit and two-bit PSs cases, where we can draw the similar conclusions as in Fig. 6. Therefore, we can conclude that our proposed solution can achieve a better performance in low-resolution PSs cases as compared to the existing solution.

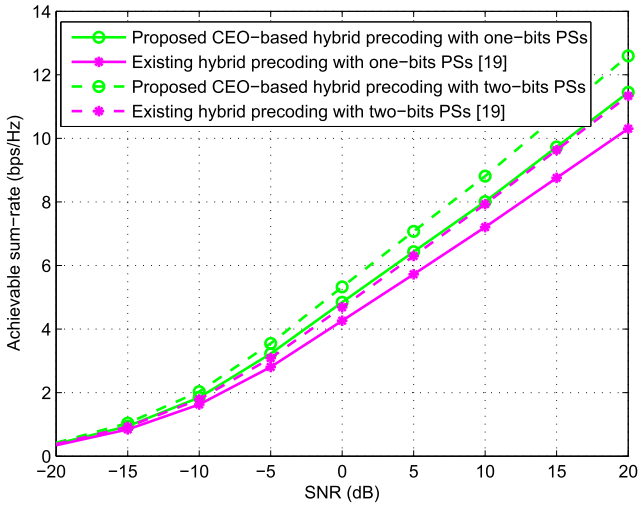

FIGURE 7. Sum-rate comparison of CEO-based hybrid precoding and existing hybrid precoding, when $N = 128$, $N_{RF} = 8$, $M = 256$.

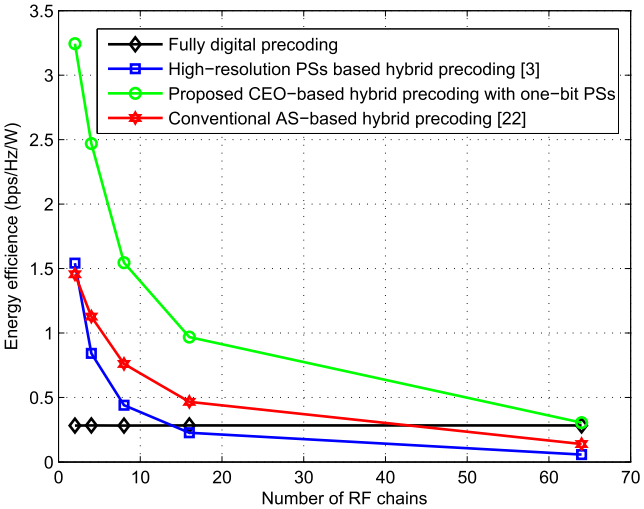

FIGURE 8. Energy efficiency comparison against the number of RF chains N_{RF} when $N = 64$, $M = 128$.

Fig. 8 presents the energy efficiency achieved by the proposed solution in comparison with some existing hybrid precoding schemes when $N = 64$ and the RF chains are varied from 1 to 32. We set the same parameters for **Algorithm 1** as specified in Fig. 4. According to [27], energy efficiency can be well defined by the ratio of the attainable sum-rate and the energy consumption which can be expressed as

$$\eta = \frac{R}{\rho}\,(\text{bps/Hz/W})\,, \tag{29}$$

The energy consumption for HR-PSs based hybrid precoding with 4-bit PSs [3], SW-based hybrid precoding [23], and one-bit PSs based hybrid precodings are given in (6), (7), and (8), respectively. We adopt the practical values in this paper, e.g., $\rho = 30$mW [5], $P_{RF} = 300$mW [23], $P_{BB} = 200$mW [24], $P_{HR\text{-}PS} = 40$mW (4-bit PSs) and $P_{SW} = P_{one\text{-}bit\text{-}PS} = 5$mW [16].

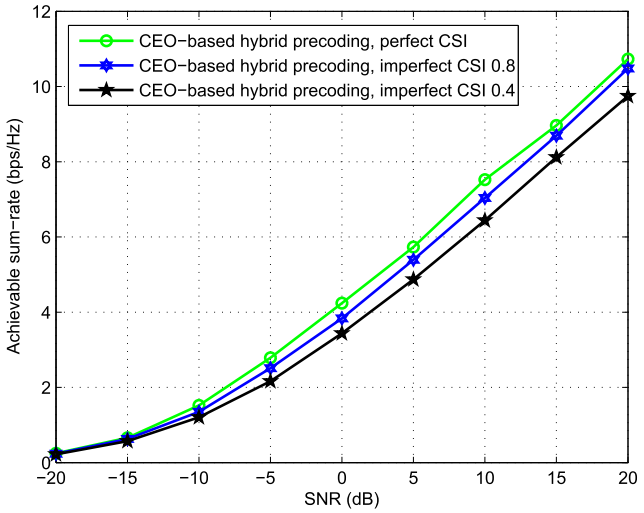

FIGURE 9. CEO-based hybrid precoding with different CSI conditions, (when $N = 64$, $N_{RF} = 4$, $M = 128$).

Fig. 8 illustrates that the proposed CEO-based hybrid precoding scheme with one-bit PSs achieves the highest energy efficiency in comparison to other schemes, especially when the number of RF chains is not extremely large. Furthermore, it can be seen that when $N_{RF} = 8$, the energy efficiency of the HR-PSs based hybrid precoding is much lower than the ZF digital precoding. The reason is that when N_{RF} rises, there is a steep increase in the number of PSs in PS-based hybrid precoding. Consequently, using high-resolution PSs with high power consumption is not suitable in practice.

Since the energy consumption of one-bit PSs compared to high-resolution PSs is much less, we can therefore conclude that CEO-based hybrid precoding scheme with one-bit PSs can accomplish a superior trade-off between the energy efficiency and sum-rate as compared to that achieved by [3], [23].

In the end, the effect of imperfect channel state information (CSI) on the proposed CEO-based hybrid precoding is observed, where $\hat{\mathbf{H}}[m]$ denotes the imperfect CSI as in [28]

$$\hat{\mathbf{H}}[m] = \xi \mathbf{H}[m] + \sqrt{1-\xi^2}\mathbf{E}, \tag{30}$$

where $\xi \in [0, 1]$ is the CSI accuracy, $\mathbf{H}[m]$ is the original channel matrix, and $\mathbf{E}$ denotes the error matrix whose entries followed an i.i.d distribution $\mathcal{CN}(0, 1)$.

Fig. 9 presents the attainable sum-rate of CEO-based hybrid precoding for different CSI conditions when $N = 64$, $N_{RF} = 4$, $M = 128$. We consider a perfect CSI case and several imperfect CSI cases with different values of ξ. The proposed CEO-based hybrid precoding with one-bit PSs is not sensitive to CSI accuracy. The sum-rate achieved by the proposed scheme with the value $\xi = 0.8$ is very near to the perfect CSI scenario. Furthermore, when the CSI accuracy is extremely bad, e.g., $\xi = 0.4$. The CEO-based scheme can still achieve 80% of the sum-rate compared with the perfect CSI scenario. Moreover, in Fig. 10, the same sum-rate comparison with different settings (e.g., $N = 128$, $N_{RF} = 8$, $M = 256$)

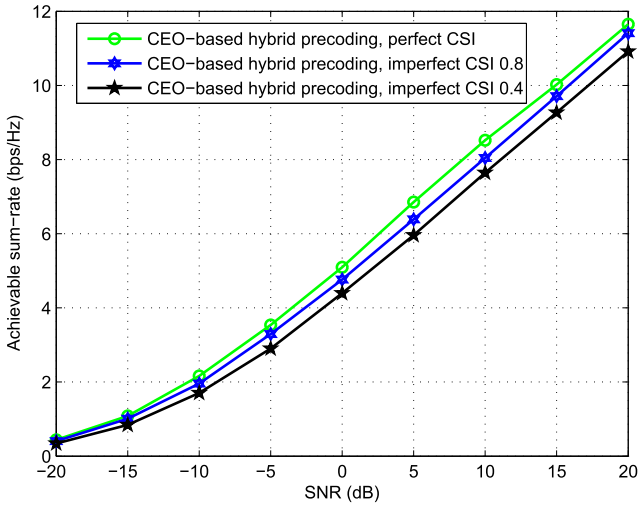

FIGURE 10. CEO-based hybrid precoding with different CSI conditions, (when $N = 128, N_{RF} = 8, M = 256$).

is provided, from which we can draw conclusions similar to Fig. 9.

VI. CONCLUSIONS

This paper proposed the CEO-based hybrid precoding with one-bit PSs for frequency-selective wideband mmWave massive MIMO systems. First, the energy consumption analysis is provided to reveal that one-bit PSs based hybrid precoding has lower power consumption. Further, the array gain loss incurred by using one-bit PSs is limited and remains constant. Then, the sum-rate maximization problem for one-bit PSs based hybrid precoding was formulated under some practical constraints, and a low-complexity algorithm based on CEO was proposed to solve the sum-rate maximization problem. Moreover, we have provided the convergence and complexity analyses of the proposed algorithm. Our results verified that the proposed scheme attain an acceptable sum-rate performance and considerably higher energy efficiency compared to some existing schemes.

APPENDIX A
PROOF OF LEMMA 1

For the HR-PSs, the array gain is

$$\eta_{\infty} = \left|\alpha_{k,l}\right|. \tag{31}$$

On the other hand, the array gain attained by one-bit PSs can be expressed as

$$\eta_{\text{one-bit}} = \frac{1}{N}\left|\alpha_{k,l}\right|\left|\sum_{t=1}^{T} e^{j\bar{\phi}_t}\right|, \tag{32}$$

Then, the ratio γ between the array gain achieved by one-bit PSs and that achieved by HR-PSs is

$$\begin{aligned}|\gamma|^2 &= \left|\frac{\eta_{\text{one-bit}}}{\eta_{\infty}}\right|^2 = \frac{1}{N^2}\left|\sum_{t=1}^{T} e^{j\bar{\phi}_t}\right|^2, \\ &= \frac{1}{N^2}\left(\left|\sum_{t=1}^{T}\cos\left(\bar{\phi}_t\right)\right|^2 + \left|\sum_{t=1}^{T}\sin\left(\bar{\phi}_t\right)\right|^2\right),\end{aligned} \tag{33}$$

Since the non-zero elements in the analog precoder belong to $\frac{1}{N}\{-1,+1\}$, then $\bar{\phi}_t$ denoting the phase quantization error can be assumed to follow the uniform distribution $\mathcal{U}\left(-\pi/2, \pi/2\right)$ for $t = 1, 2, \ldots, T$. Thus, we have

$$\begin{aligned}\lim_{N\to\infty, \frac{N}{T}=N_{RF}} |\gamma|^2 &= \frac{1}{N^2}\left(\left|\sum_{t=1}^{T}\cos\left(\bar{\phi}_t\right)\right|^2 + \left|\sum_{t=1}^{T}\sin\left(\bar{\phi}_t\right)\right|^2\right), \\ &= \left(\mathbb{E}\left[\cos\left(\bar{\phi}_t\right)\right]\right)^2 + \left(\mathbb{E}\left[\sin\left(\bar{\phi}_t\right)\right]\right)^2, \\ &= \frac{4}{\pi^2}\sin^2\frac{\pi}{2},\end{aligned} \tag{34}$$

where (34) is obtained by using the uniform distribution of $\bar{\phi}_t$. Therefore, we reach

$$\lim_{N\to\infty, \frac{N}{T}=N_{RF}} |\gamma|^2 = \frac{4}{\pi^2}\sin^2\frac{\pi}{2}, \tag{35}$$

which completes the proof of **Lemma 1.** ■

APPENDIX B
PROOF OF THEOREM 1

We follow the steps in [21] and [25] to prove the convergence. From [21], we can observe that as $i \to \infty$, we have

$$1 - \frac{\ln i}{\ln(i+1)} = \frac{\ln\left(\frac{1}{i}+1\right)}{\ln(i+1)} \approx \frac{1}{i\ln(i+1)}. \tag{36}$$

Using (36) and by applying the condition $0 < \mu < 1$, it can be observed that

$$\Omega^{(i)} = \mu\frac{1}{i\ln(i+1)} < 1 - \frac{\ln i}{\ln(i+1)}, \tag{37}$$

when i is sufficiently large. According to (27) and using (36), we have

$$\begin{aligned}p_n^{(i)} &\geq \left[\prod_{q=1}^{i}\left(1-\Omega^{(q)}\right)\right].p_n^{(0)} \\ &= \left[\prod_{q=1}^{i-1}\left(1-\Omega^{(q)}\right)\right]\left[\prod_{q=i}^{I}\left(1-\Omega^{(q)}\right)\right].p_n^{(0)} \\ &\geq \left[\prod_{q=1}^{i-1}\left(1-\Omega^{(q)}\right)\right]\left[\prod_{q=i}^{I}\left(\frac{\ln q}{\ln(q+1)}\right)\right].p_n^{(0)} \\ &= \left[\prod_{q=1}^{i-1}\left(1-\Omega^{(q)}\right)\right].\frac{\ln I}{\ln(i+1)}\frac{1}{2} \\ &= \frac{\Gamma}{\ln(i+1)},\end{aligned} \tag{38}$$

where

$$\Gamma \triangleq \left[\prod_{q=1}^{I-1}\left(1-\Omega^{(q)}\right)\right]\frac{\ln I}{2} \tag{39}$$

is a constant, and $p_n^{(0)} = \frac{1}{2}$.

To ensure that CEO will generate the optimal solution, we need to show that the probability of not generating the

optimal solution tends to zero as the number of iteration goes to infinity [21]. To prove that, we define the three events as follows.

First, $\mathcal{X}^{(i)} \triangleq \left\{\mathbf{A}_q^{[o]} \neq \mathbf{A}_{\mathrm{RF}}^{opt}, o=1,2,\cdots,O, q=1,\cdots,i-1\right\}$ is defined as the event in which the optimal solution was never generated before the iteration i. Second, we denote the event in which the optimal solution was not generated in all O candidate solutions at iteration i by $\mathcal{Y}^{(i)} = \left\{\mathbf{A}_q^{[o]} \neq \mathbf{A}_{\mathrm{RF}}^{opt}, o=1,2,\cdots,O\right\}$. Third, $\mathcal{Y}_o^{(i)} = \left\{\mathbf{A}_q^{[o]} \neq \mathbf{A}_{\mathrm{RF}}^{opt}\right\}$ denotes the event in which the optimal solution was not generated in oth sample in iteration i.

Considering the fact that the candidate solutions generated by CEO algorithm at a given iteration are independently and identically distributed. Then, (39) provides the probability to generate the optimal solution for one element. Therefore, the probability to generate the optimal solutions for N_{RF} elements can be shown as $\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{RF}}$.

In contrast, the probability that the optimal solution was not generated in all O candidate solutions at iteration i can be written as

$$\begin{aligned}\Pr\left(\mathcal{Y}^{(i)}|\mathcal{X}^{(i)}\right) &= \prod_{o=1}^{O}\Pr\left(\mathcal{Y}_o^{(i)}|\mathcal{X}^{(i)}\right)\\ &\leq \left[1-\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}1\right]^{O}. \end{aligned} \quad (40)$$

Therefore, the probability that optimal solution was never generated can be bounded as

$$\begin{aligned}\prod_{i=1}^{\infty}\Pr\left(\mathcal{Y}^{(i)}|\mathcal{X}^{(i)}\right) &\leqslant \prod_{i=1}^{I-1}1.\prod_{i=I}^{\infty}\left[1-\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}\right]^{O}\\ &= \prod_{i=I}^{\infty}\left[1-\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}\right]^{O}. \end{aligned} \quad (41)$$

taking the natural log of (41) and considering $\sum_i(\ln(i+1)) \to -\infty$, we have

$$\begin{aligned}&\ln\left\{\prod_{i=I}^{\infty}\left[1-\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}\right]^{O}\right\}\\ &= O\sum_{i=I}^{\infty}\ln\left[1-\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}\right]\\ &\leqslant -O\sum_{i=I}^{\infty}\left(\frac{\Gamma}{\ln(i+1)}\right)^{N_{\mathrm{RF}}}\\ &\to -\infty. \end{aligned} \quad (42)$$

From (42), we know that (41) tends to zero. This means that the probability of generating the optimal solution is equal to 1.

REFERENCES

[1] X. Wang, M. Jia, Q. Guo, I. W.-H. Ho, and J. Wu, "Joint power, original bandwidth, and detected hole bandwidth allocation for multi-homing heterogeneous networks based on cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2777–2790, Mar. 2019.

[2] S. Mumtaz, J. Rodriguez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. New York, NY, USA: Academic, 2016.

[3] 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.

[4] S. Han, C.-I. 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.

[5] 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.

[6] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Select. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.

[7] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[8] O. El 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.

[9] M. Jia, L. Jiang, Q. Guo, X. Jing, X. Gu, and N. Zhang, "A novel hybrid access protocol based on traffic priority in space-based network," *IEEE Access*, vol. 6, pp. 24767–24776, Dec. 2018.

[10] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 340–352, May 2018.

[11] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based CSI feedback approach for time-varying massive MIMO channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 416–419, Apr. 2019.

[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] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[14] C. Kim, J.-S. Son, T. Kim, and J.-Y. Seol, "On the hybrid beamforming with shared array antenna for mmWave MIMO-OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (IEEE WCNC)*, Apr. 2014, pp. 335–340.

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

[16] 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.

[17] S. Gao, Y. Dong, C. Chen, and Y. Jin, "Hierarchical beam selection in mmWave multiuser MIMO systems with one-bit analog phase shifters," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2016, pp. 1–5.

[18] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*. Hoboken, NJ, USA: Wiley, 2016.

[19] Z. Wu, M. Kolonko, and R. H. Möhring, "Stochastic runtime analysis of the cross-entropy algorithm," *IEEE Trans. Evol. Comput.*, vol. 21, no. 4, pp. 616–628, Aug. 2017.

[20] R. Y. Rubinstein, "Optimization of computer simulation models with rare events," *Eur. J. Oper. Res.*, vol. 99, no. 1, pp. 89–112, 1997.

[21] W. J. Gutjahr, "ACO algorithms with guaranteed convergence to the optimal solution," *Inf. Process. Lett.*, vol. 82, no. 3, pp. 145–153, May 2002.

[22] J. Zhang, Y. Huang, J. Wang, and L. Yang, "Hybrid precoding for wideband millimeter-Wave systems with finite resolution phase shifters," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11285–11290, Nov. 2018.

[23] 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, no. 8, pp. 247–267, Dec. 2016.

[24] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. ITA Workshops*, Feb. 2015, pp. 90–97.
[25] A. Costa, O. D. Jones, and D. Kroese, "Convergence properties of the cross-entropy method for discrete optimization," *Oper. Res. Lett.*, vol. 35, no. 5, pp. 573–580, 2007.
[26] 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.
[27] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.
[28] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.

TALHA MIR received the B.S. degree in electronic engineering from the Baluchistan University of IT, Engineering and Management Sciences Pakistan (BUITEMS), in 2007. He is currently pursuing the master's degree from the University of Bradford, U.K., in 2011. He is currently pursuing the Ph.D. degree with Tsinghua University Beijing, China, He is also serving as an Assistant Professor with the Baluchistan University of IT, Engineering and Management Sciences Pakistan (BUITEMS). His current research interests include 5G networks, massive MIMO, mmwaves communications, NOMA, and UDN.

MUHAMMED ZAIN SIDDIQI received the master's degree in electrical engineering from the COMSATS Institute of Information and Technology, Islamabad, Pakistan. He is currently pursuing the Ph.D. degree in electronics engineering from Tsinghua University, Beijing, China. He served as a Lecturer for Lahore Leads University, from 2015 to 2017. His research interests include massive MIMO, mmWave communications, resource, and power allocation in wireless networks, and green communication.

USAMA MIR received the B.S. degree (Hons.) in computer engineering from the Balochistan University of IT, Engineering and Management Sciences (BUITEMS), Pakistan, in 2006, and the master's and Ph.D. degrees in computer science from the Troyes University of Technology (UTT) France, in 2008 and 2011, respectively. He has also served as a Postdoctoral Fellow with Telecom Bretagne France, from 2012 to 2013. He was the Head of the Electronics Engineering Department, Iqra University, Islamabad, Pakistan, from 2013 to 2015. He is currently serving as an Assistant Professor with Saudi Electronic University (SEU), Saudi Arabia. His research interests include resource allocation and handoff management in cognitive radio systems, wireless communications and networking, Markov chains, next-generation networks, multi-agent systems, and big data analysis.

RICHARD MACKENZIE received the M.Eng. degree in electronic engineering from the University of York and the Ph.D. degree in electronic engineering from the University of Leeds.

He joined BT, in 2009. His past work with BT has covered spectrum management, cognitive radio, and next-generation technologies. He has been involved in EU projects, in particular, QoSMOS, where he was a work package leader for the network architecture design. He was a Spectrum Manager for the London 2012 Olympic and Paralympic Games and was part of BT's bidding team for the U.K. 4G spectrum auction, in 2013. He is currently a Principal Researcher with British Telecom Technology. He is also a BT Technical Representative with Next Generation Mobile Networks (NGMN), where he is Chair of the RAN Functional Split and X-Haul Group; Telecom Infra Project (TIP), and is also the Co-Chair of the vRAN fronthaul project; and Small Cell Forum (SCF). His current research interests include 5G, with a particular focus on the air interface, RAN virtualization, network automation, and the use of small cells.

MO HAO received the bachelor's degree from Sichuan University, China, and the master's degree from the University of Southampton, U.K.

He was with Ericsson as a Senior Engineer. He joined the Tsinghua SEM Advanced ICT Lab, in 2015, where he was a Researcher. He is in charge of the 5G related researches and project management. He is participating in the collaboration project on 5G with Tsinghua EE team and also leading the Lab's independent projects on wireless communication technologies. His research interests include the 3GPP LTE technical specification updating, designing and implementing the new features for the Modem product, such as beamforming, carrier aggregation, and 8-antenna MIMO.