IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received July 27, 2019, accepted August 15, 2019, date of publication August 23, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937209

Machine Learning-Based Spectrum Efficiency Hybrid Precoding With Lens Array and Low-Resolution ADCs

LIXIN LI[1], (Member, IEEE), HUAN REN[1], XU LI[1], WEI CHEN[2], (Senior Member, IEEE), AND ZHU HAN[3], (Fellow, IEEE)

[1]School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China
[2]Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
[3]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA

Corresponding author: Lixin Li (lilixin@nwpu.edu.cn)

This work was supported in part by the Aerospace Science and Technology Innovation Fund of the China Aerospace Science and Technology Corporation, in part by the Shanghai Aerospace Science and Technology Innovation Fund under Grant SAST2018045, Grant SAST2016034, and Grant SAST2017049, in part by the National Natural Science Foundation of China under Grant 61671269, in part by Beijing Natural Science Foundation under Grant 4191001, in part by the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University under Grant ZZ2019024, and in part by the U.S. MURI AFOSR under Grant MURI 18RT0073, Grant NSF CNS-1717454, Grant CNS-1731424, Grant CNS-1702850, Grant CNS-1646607, Grant CNS-1350230, Grant CNS-1646607, and Grant CNS-1801925.

ABSTRACT Hybrid precoding is an important issue in millimeter wave (mmWave) massive multi-input and multi-output (MIMO) system. Specially, energy-saving hybrid precoding architectures and efficient hybrid precoding schemes provide ideas for solving this issue. In this paper, we propose a hybrid precoding/combining architecture that is low-cost and easy to implement. Specifically, a hybrid precoding architecture is realized by the lens sub-arrays at the base station (BS). Moreover, a hybrid combining architecture applies the low-resolution analog-to-digital converters (ADCs) at the front end of the radio frequency (RF) chains at the receiving terminal. Based on the hybrid precoding/combining architecture, the hybrid precoder and combiner are jointly optimized to maximize the spectrum efficiency (SE) in the downlink systems, which is a combinatorial optimization problem due to hardware constraints. The cross-entropy (CE) approach in machine learning (ML) is a simple way to solve the combinatorial optimization problem benefiting from its adaptive update procedure. Therefore, we propose an adaptive hybrid precoder/combiner design scheme (AHDS), in which a hybrid precoding algorithm based on the improved CE (ICE) inspired by ML is adopted to design the optimal hybrid precoder, and an approximate optimization method (AOM) is suggested when designing the hybrid combiner. In general, compared with the existing hybrid design schemes, the proposed AHDS is demonstrated to have significant advantage in SE with low computational complexity.

INDEX TERMS Hybrid precoding, machine learning, lens sub-array, low-resolution ADC, multi-input and multi-output.

I. INTRODUCTION

Millimeter (mmWave) frequency between 30 to 300 GHz has been attracting growing attention to enhance the throughput in wireless networks [1], [2]. As for the abundant spectrum resources of the mmWave band, the next generation of wireless communication is developing towards mmWave. In addition, massive multi-input and multi-output (MIMO) is one of the emerging technologies. The combination of mmWave and MIMO can be a great progress in the fifth generation (5G) mobile communication system. Specifically, inspired by the massive MIMO and mmWave technologies, the mmWave massive MIMO is considered as a potential technique to enhance the throughput of wireless communication system [3]. Except the abundant bandwidth resources, the short wavelength of mmWave enables more antennas to

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

be packed in the same physical space, and thus, can better support massive MIMO communication [4].

The mmWave massive MIMO system is capable of enhancing the spectrum efficiency (SE) by providing the large array gain [5]. However, the large-scale antennas, the high-resolution analog-to-digital converters (ADCs)/digital-to-analog converters (DACs), and the fully digital precoding scheme will result in big power consumption and unfordable hardware costs. Therefore, it is a significant challenge to develop the promising hybrid preocoding architecture for overcoming power consumption and hardware costs. At present, the popular hybrid precoding has been proposed for mmWave massive MIMO system [6], which divides spatial processing into radio frequency (RF) and digital baseband (BB) domains [7]. Obviously, the hybrid precoding architecture [8]–[11] can reduce the number of RF chains required for mmWave massive MIMO systems while maintaining efficient performance.

However, most of the existing hybrid precoding architectures [12] have big power consumption and hardware costs. The authors in [13] proposed an effective and energy-saving switches selection network instead of the traditional phase shifters (PSs) network, but the switches selection network required strict channel conditions (independent and identically distributed (i.i.d.) Rayleigh fading channel). The authors in [14] proposed an energy-saving hybrid precoding architecture, in which the analog part was implemented by a small number of switches and inverters. It has proven to have low power consumption and hardware costs with little performance loss. In addition, the authors in [15] proposed new alternatives for traditional large-scale array antennas. Obviously, a wide range of applications of lens-array antennas utilize the sparseness of mmWave channel [16], while the full array multiplexing gain of large-scale antennas can always be realized by low-cost lens.

Although hybrid precoding is beneficial for massive MIMO systems, the number of the RF chains with high-resolution DACs/ADCs is considerable. The utilization of ultra low-resolution DACs/ADCs as alternatives of the high-resolution DACs/ADCs for each RF chain is a potential solution. At present, the application of one-bit ADCs has gained much interest [17]–[19]. However, most relevant performance analysis were discussed in the uplink system with low-resolution ADCs. In [20], the authors considered a massive MIMO relaying system with low-resolution ADCs at both relaying and destination. In conclusion, a small number of contributions were proposed in the downlink system. The authors in [21] proposed an alternate minimization algorithm to solve the hybrid precoding problem. However, there was a bottleneck due to one-bit ADC processing at the receiver. Therefore, the achievable sum rate needs to be improved in massive MIMO systems with low-resolution ADCs.

Deep learning (DL) and machine learning (ML) have attracted much interest in communication networks and system design [22]–[24]. The cross-entropy (CE) [25] approach was firstly introduced in 1997 and developed in ML, which was initially used for estimating probabilities of rare events in complex stochastic network. It is superior to solve combinatorial optimization problems due to its adaptive update procedure. Recently, several applications with CE optimization were discussed, such as the unmanned aerial vehicles (UAVs) task assignment [26], the buffer allocation [27], and ML [28]. All of these works showed the effectiveness and simplicity of CE optimization compared to typical relaxation techniques. CE is a promising method to solve the complicated combinatorial optimization, which is proved to be simple, efficient, and general.

In this paper, we propose an energy-saving hybrid transceiver of the massive MIMO system to realize efficient hybrid precoding/combing with the help of CE optimization. The specific hybrid precoding architecture is as follows: a switch and inverter-based hybrid precoding architecture is proposed, and the transmitter is equipped with the sub-lens antennas. The hybrid combining architecture is realized by the PSs, and low-resolution ADCs are applied at the front end of RF chains. Based on the hybrid precoding/combining architecture proposed in this paper, a hybrid precoder/combiner design problem is formulated to maximize the SE of the downlink communication. To solve the optimization problem, an adaptive hybrid precoder/combiner design scheme (AHDS) is proposed in this paper. Specifically, the first stage is to design the optimal hybrid precoder with a hybrid precoding algorithm based on the improved CE (ICE) in ML, which refines the probability distributions iteratively via CE minimization. The second stage is to design the optimal hybrid combiner with an approximate optimization method (AOM), considering the quantization error caused by the low-resolution ADCs. The simulation results verify that the proposed AHDS in this paper can achieve high SE with low computational complexity.

The major contributions in this paper are summarized as follows:

- A hybrid precoding architecture-based the sub-lens antennas at the base station (BS) and a hybrid combining architecture-based low-resolution ADCs at the user terminal are proposed, which is proved to have low hardware cost and power consumption. In the architecture, the analog precoder is implemented by switches and inverters at the transmitter, and the analog combiner is implemented by PSs. Moreover, a hybrid precoder/combiner optimization problem is formulated aiming to maximize the system SE and energy efficiency (EE).
- We propose an adaptive hybrid precoder/combiner design scheme (AHDS), in which a hybrid precoding algorithm based on the improved CE (ICE) is adopted to design the optimal hybrid precoder, and the approximate optimization method (AOM) is suggested when designing the optimal hybrid combiner. The proposed scheme

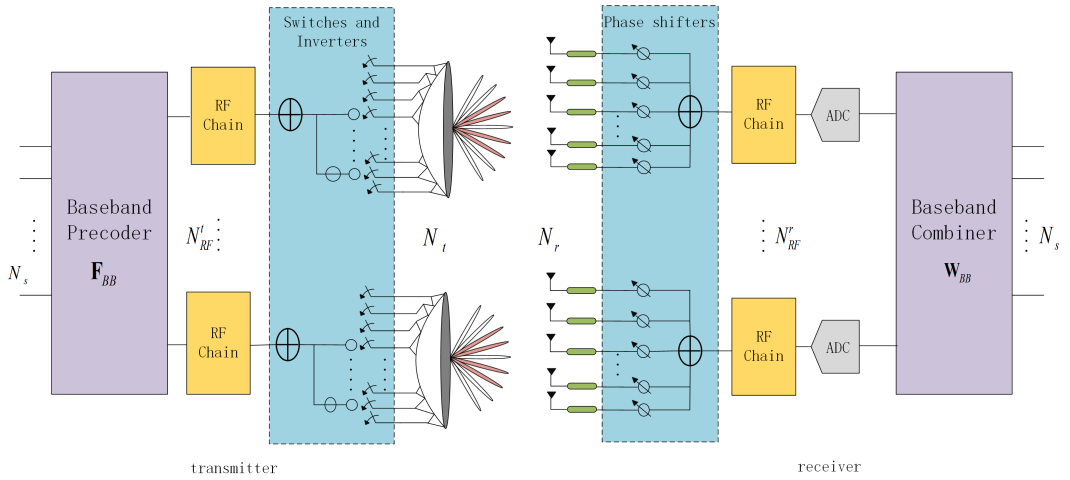


FIGURE 1. mmWave massive MIMO hybrid transmitter/receiver architecture with lens sub-array antennas.

can obtain the optimal precoding matrix and combining matrix to maximize the SE of mmWave massive MIMO system, respectively.

- Simulation results show that the proposed AHDS can achieve high SE and EE with low computational complexity, which can be improved compared with the existing hybrid design schemes significantly. Moreover, simulations are also conducted to verify the feasibility and effectiveness of the CE method in solving the combinational optimization problem.

The remainder of this paper is organized as follows. Section II is a brief overview of the system model and problem formulation. Section III introduces the algorithm ideas and specific steps. In Section IV, we provide simulation results and performance analysis of the proposed algorithm. Finally, the conclusion of this paper is described in Section V.

Notations: In this paper, the boldface uppercase letters denote matrices, and the boldface lowercase letters denote column vectors. $(\cdot)^H$ denotes the conjugate transposition. $(\cdot)^{-1}$ and $(\cdot)^*$ denote the inversion and transposition, respectively. $\|\cdot\|_F$ denotes the Frobenius norm, $|\cdot|$ represents the determinant. $\mathbb{E}(\cdot)$ denotes the expectation operator. $[\mathbf{B}]_{m,n}$ is used to denote the $(m; n)$ component of a matrix $\mathbf{B}$. $\mathbf{I}_{N_r}$ is the identity matrix of the size $N_r \times N_r$. $\boldsymbol{e}_j$ is a zero column vector with a one at the j-th element. $\mathcal{CN}\left(\mu, \delta^2\right)$ denotes a circularly symmetric complex-valued multi-variate Gaussian distribution with a mean of μ and a variance of δ^2. $\mathbf{1}_{M\times N}$ is the $M \times N$ matrix in which all elements are equal to one.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

In this paper, as shown in Fig. 1, a typical mmWave massive MIMO system with multi-antennas users is considered. Specifically, the analog precoder is realized by the switches and inverters at the transmitter. Moreover, the low-cost lens antennas are applied at the transmitter. By employing the lens (an electromagnetic lens with directional energy focusing capability and a matching antenna array with elements located in the focal surface of the lens [15]) sub-array antennas, the signals from different directions can be concentrated on different antennas. The analog combiner is realized by the PSs at the receiver. In addition, the low-resolution ADCs are applied at the front of RF chains at the receiver.

Assuming that there are K users to be served by the BS simultaneously, as shown in Fig. 1. The transmitting terminal is equipped with N_t transmitting antennas (Lens) and N_{RF}^t ($N_{RF}^t \leq N_t$) RF chains, which are used for sending N_s data streams simultaneously. The receiving terminal is also equipped with N_r receiving antennas and N_{RF}^r ($N_{RF}^r \leq N_r$) RF chains. Without loss of generality, there is $N_{RF}^t = N_{RF}^r = K$. However, as shown in Fig. 1, one RF chain connects to a lens sub-array, so we have $N_{RF}^t \leq K$.

The lens antennas are equipped at the BS, so the traditional channel is sparse. According to the sparse beamspace channel matrix, only a small number of beams can be selected to serve K users simultaneously. When N_s data streams are transmitted through the beamspace channel, the receive vector $\mathbf{r} = \left[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_r}\right]^*$ of the user at the receiving terminal can be represented as

$$\mathbf{r} = \sqrt{\rho}\widetilde{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{n}, \tag{1}$$

where ρ is the average transmit power per symbol. $\mathbf{F} \in C^{N_t \times N_s}$ is a linear precoder. $\widetilde{\mathbf{H}} \in C^{N_r \times N_t}$ is the beam spatial channel matrix, which is generated due to the lens antennas. $\mathbf{s} = \left[s_1, s_2, \ldots, s_{N_s}\right]^*$ $(n = 1, 2, \ldots, N_s)$ is the transmission symbol vector with zero mean and normalized power of $\mathbb{E}\left[\mathbf{s}\mathbf{s}^*\right] = \frac{1}{N_s}\mathbf{I}_{N_s}$. $\mathbf{n} \in C^{N_r}$ denotes the zero-mean Gaussian noise vector with $\mathbb{E}\left[\mathbf{n}\mathbf{n}^*\right] = \mathbf{I}_{N_r}$.

We assume that the BS perfectly estimates the beamspace channel matrix $\widetilde{\mathbf{H}}$ by applying an effective channel estimation scheme. Since the lens sub-array antennas and the limited scattering [9], we obtain the beam spatial channel matrix $\widetilde{\mathbf{H}} \in C^{N_r \times N_t}$ from the physical space MIMO channel by discrete

fourier transform (DFT). Specifically, the sparse beamspace channel can be expressed as

$$\widetilde{\mathbf{H}} = \begin{bmatrix} \widetilde{H}_{1,1} & \widetilde{H}_{1,2} & \cdots & \widetilde{H}_{1,N_t} \\ \widetilde{H}_{2,1} & \widetilde{H}_{2,2} & \cdots & \widetilde{H}_{2,N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{H}_{N_r,1} & \widetilde{H}_{N_r,2} & \cdots & \widetilde{H}_{N_r,N_t} \end{bmatrix} = \mathbf{U} \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,N_t} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,N_t} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N_r,1} & g_{N_r,2} & \cdots & g_{N_r,N_t} \end{bmatrix} \tag{2}$$

In (2), $\widetilde{H}_{r,t}$ $(r = 1, 2, \ldots, N_r, t = 1, 2, \ldots, N_t)$ is the channel matrix between the r-th transmit antenna and the t-th receive antenna of a user. $g_{r,t}$ is the spatial domain channel matrix between the BS and the user terminal. $\mathbf{U} \in C^{N_r \times N_r}$ is a DFT matrix corresponding to a carefully designed discrete lens array (DLA) antennas at the BS,

$$\mathbf{U} = \begin{bmatrix} a\left(\varphi_{1,1}\right) & a\left(\varphi_{1,2}\right) & \cdots & a\left(\varphi_{1,N_t}\right) \\ a\left(\varphi_{2,1}\right) & a\left(\varphi_{2,2}\right) & \cdots & a\left(\varphi_{2,N_t}\right) \\ \vdots & & \ddots & \\ a\left(\varphi_{N_r,1}\right) & a\left(\varphi_{N_r,2}\right) & \cdots & a\left(\varphi_{N_r,N_t}\right) \end{bmatrix}, \tag{3}$$

where $a\left(\varphi_{r,t}\right)$ $(r = 1, 2, \ldots, N_r,\ t = 1, 2, \ldots, N_t)$ is the array steering vector. The N_r columns of the matrix in (3) correspond to the orthogonal beamforming vectors of N_r predefined directions covering the entire angular space, respectively, i.e., $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N_r}$. We take the first column in (3) as an example for analysis: $\varphi_{r,1} = \frac{1}{N_r}\left(r - \frac{N_r+1}{2}\right)$ $(r = 1, 2, \ldots, N_r)$ denotes the normalized spatial direction [17]. $a\left(\varphi_{r,1}\right) = \frac{1}{\sqrt{N_r}}\left[e^{-j2\pi\varphi_{r,1}i}\right]_{i\in I}$ $\left(I = \left\{j - \frac{N_r-1}{2} \,|\, j = 0, 1, \ldots, N_r - 1\right\}\right)$ is corresponding $N_r \times 1$ array steering vector. The normalized spatial direction $\varphi_{r,t}$ is related to the physical angle of propagation $\theta_{r,t}$ by $\varphi = \frac{d}{\lambda}\sin\theta$, in which λ is the signal wavelength and d is the antenna spacing. At mmWave frequency, $d = \lambda/2$ is usually to minimize the space of the multi-antennas while achieving the optimal diversity [9].

In this paper, we consider the extended Saleh-Valenzuela channel model [29] for the proposed mmWave massive MIMO system, i.e.,

$$g_{r,t} = \beta_{r,t}^{(0)} a\left(\varphi_{r,t}^{(0)}\right) + \sum_{l=1}^{L} \beta_{r,t}^{(l)} a\left(\varphi_{r,t}^{(l)}\right), \tag{4}$$

where $\beta_{r,t}^{(0)} a\left(\varphi_{r,t}^{(0)}\right)$ and $\beta_{r,t}^{(l)} a\left(\varphi_{r,t}^{(l)}\right)$ represent the channel vector of the line of sight (LoS) and of the non-line of sight (NLoS) link between the BS and the user, respectively. The number of the NLOS links is L. $\beta_{r,t}^{(0)}$ and $\beta_{r,t}^{(l)}$ represent complex gain for LoS and the l-th NLoS channel, respectively. $\varphi_{r,t}^{(0)}$ and $\varphi_{r,t}^{(l)}$ indicate the corresponding spatial direction.

The received signal is processed through a linear combiner $\mathbf{W} \in C^{N_r \times N_s}$. Thus, $\mathbf{y}$ can be expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{W}^*\mathbf{r} \\ &= \sqrt{\rho}\mathbf{W}^*\widetilde{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{W}^*\mathbf{n}, \end{aligned} \tag{5}$$

where the linear precoder $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ is realized by the analog precoder $\mathbf{F}_{RF} \in C^{N_t \times N_{RF}^t}$ and the digital precoder $\mathbf{F}_{BB} \in C^{N_{RF}^t \times N_s}$. The linear combiner $\mathbf{W} = \mathbf{W}_{RF}\mathbf{W}_{BB}$ is implemented by the analog combiner $\mathbf{W}_{RF} \in C^{N_r \times N_{RF}^r}$ and the digital combiner $\mathbf{W}_{BB} \in C^{N_{RF}^r \times N_s}$.

The received signal after the one-bit quantization can be expressed as

$$\begin{aligned} \widetilde{\mathbf{y}} &= Q\left(\mathbf{y}\right) \\ &= Q\left(\sqrt{\rho}\mathbf{W}^*\widetilde{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{W}^*\mathbf{n}\right), \end{aligned} \tag{6}$$

where $Q(\cdot)$ is the one-bit quantization function, which is applied to component-wise and separately to the real and imaginary parts. The digital signal is quantified to form a sequence of specific length.

Although a quantitative process is the quantization operation of ADC, we can convert it to an equivalent linear operation using the Bussgang theorem [30]. The theorem obtains the statistical equivalent linear operator of any nonlinear function of Gaussian signals. Particularly, for the one-bit quantization in (6), the Bussgang theorem can be written as

$$\begin{aligned} \widetilde{\mathbf{y}} &= \mathbf{A}\mathbf{y} + \mathbf{q} \\ &= \sqrt{\rho}\mathbf{A}\mathbf{W}^*\widetilde{\mathbf{H}}\mathbf{F}\mathbf{s} + \mathbf{W}^*\mathbf{n} + \mathbf{q} \\ &= \sqrt{\rho}\mathbf{A}\mathbf{W}^*\widetilde{\mathbf{H}}\mathbf{F}\mathbf{s} + \widetilde{\mathbf{n}}, \end{aligned} \tag{7}$$

where $\mathbf{A}$ is a linear operator, $\mathbf{q}$ is a statistically equivalent quantization noise. $\mathbf{y}$ denotes the received signal vector before quantization. In order to facilitate the calculation, we combine the interference except the transmitted signal, (i.e., $\widetilde{\mathbf{n}} = \mathbf{W}^*\mathbf{n} + \mathbf{q}$). The total noise $\widetilde{\mathbf{n}}$ is composed of Gaussian noise being combinatorial and quantization noise. In addition, $\mathbf{A}$ is be approximated by the existing methods in [31], i.e.,

$$\mathbf{A} = \sqrt{\frac{2}{\pi}}\sqrt{\frac{1}{1+P_t}}\mathbf{I}_{N_r}, \tag{8}$$

where P_t is the total transmission power of all antennas at the transmitter.

B. PROBLEM FORMULATION

Assuming that the transceiver can get perfect channel state information (CSI). The aim of this paper is to maximize the system SE by designing the hybrid precoder and combiner. Firstly, the equivalent channel and noise covariance matrix are defined as follows:

$$\mathbf{H}_u \triangleq \mathbf{A}\mathbf{W}^*\widetilde{\mathbf{H}}\mathbf{F}, \tag{9}$$

and

$$\mathbf{R}_u \triangleq \mathbf{W}^*\mathbf{W}. \tag{10}$$

Thus, the hybrid precoder/combiner design problem can be stated as

$$\begin{aligned}\max_{\mathbf{F},\mathbf{W}} R &= \log_2\left|\mathbf{I}_{N_s}+\frac{\rho}{N_s}\mathbf{H}_u^*\mathbf{R}_u^{-1}\mathbf{H}_u\right|\\ s.t.\ \ &\text{C1:}\boldsymbol{e}_j^*\mathbf{F}_{RF}\mathbf{F}_{BB}(\mathbf{F}_{RF}\mathbf{F}_{BB})^*\boldsymbol{e}_j\leqslant p_j,\\ &\text{C2:}\ \ \left|[\mathbf{F}_{RF}]_{m,n}\right|=1,\\ &\text{C3:}\ \ \left|[\mathbf{W}_{RF}]_{m,n}\right|=1,\\ &\text{C4:}\ \mathbf{F}_{RF}\in C^{N_t\times N_{RF}^t},\\ &\text{C5:}\ \mathbf{W}_{RF}\in C^{N_r\times N_{RF}^r}.\end{aligned}\tag{11}$$

C1 in (11) represents the individual power constraint at each of the N_t transmit antennas, in which p_j is the power of per-antenna at the transmitter. C2 and C3 in (11) denote the elements have unit amplitude in the analog precoding matrix and combining matrix. C4 and C5 in (11) are the hardware-specific constraints of analog precoder and combiner, respectively. Note that $\boldsymbol{e}_j$ is the j-th element of the standard basis.

It is simple to verify that problem (11) is a nondeterministic polynomial-time hard (NP-hard) problem. There are two variables in (11) that need to be optimized: $\mathbf{F}=\mathbf{F}_{RF}\mathbf{F}_{BB}$ and $\mathbf{W}=\mathbf{W}_{RF}\mathbf{W}_{BB}$. Due to the hardware-specific constraints C2-C5, the problem (11) is intractable.

III. TWO-STAGE DESIGN

In this section, we propose AHDS to obtain the optimal solution of problem (11) by solving its two sub-problems, i.e., (14) and (22). Specifically, (14) can be obtained by the hybrid precoding algorithm-based ICE. According to the probability of the elements in the hybrid precoder, the hybrid precoding scheme first generates a hybrid precoder to approximate the optimal one with the optimal probability. Then, AOM is suggested to design the hybrid combiner. Specifically, (22) is solved by determining a reasonable approximation of the analog combiner by the singular value decomposition (SVD) and convert problem (22) to a liner programmer. In the following, we provide the specific algorithm designing process in detail and discuss the computational complexity.

A. HYBRID PRECODER DESIGN

In this subsection, the problem of maximizing the system SE can be formulated as the hybrid precoder optimization problem for designing the optimal digital precoder and the analog precoder. Inspired by the sampling approach developed in ML, the probabilistic model is established and the original problem is reformulated as a CE minimization problem learning the probability distribution of the elements in hybrid precoding matrix.

Theorem 1: Let $\Phi\in C^{N_r\times N_s}$ and $\Gamma\in C^{N_t\times N_s}$ comprise the N_s left and right singular vectors of the channel matrix $\mathbf{H}$ corresponding to the N_s largest singular values, respectively. Then, the solution to the problem (11) with C1 is given by

$$\mathbf{F}=\sqrt{p_0}\Gamma\mathbf{Q},\tag{12}$$

Algorithm 1 Hybrid Precoding Algorithm-Based ICE

Input: Channel matrix $\mathbf{H}$; Number of iterations I; The number of elements in the precoding matrix M.

Output: Analog precoding matrix $\mathbf{F}_{RF}^{[1]}$; Digital precoding matrix $\mathbf{F}_{BB}^{[1]}$.

1: Initialization: $i=0$, $u^{(0)}=\frac{1}{2}\times\mathbf{1}_{N\times 1}$($\mathbf{1}$ is the all-one vector);
2: **for** $i=0$ to I **do**
3: Randomly generate M analog precoders $\{\mathbf{f}^m\}_{m=1}^M$ based on probability model $\Pi\left(F;\mathbf{p}^{(i)}\right)$;
4: Compute corresponding digital precoding matrix by (16) and (17);
5: Calculate the chord distance $\{\Im_m\}_{m=1}^M$ according to an iteratively updated precoding matrix, and sort them in a descend;
6: Select $1\leq m\leq M_{elite}$ as targets;
7: **for** $m=0$ to M_{elite} **do**
8: Update $\mathbf{p}^{(i+1)}$ by (21);
9: $i\leftarrow i+1$.
10: **end for**
11: **end for**

and

$$\mathbf{W}=\Phi\mathbf{T},\tag{13}$$

with $\mathbf{Q}\in C^{N_s\times N_s}$ an arbitrary unitary matrix, and $\mathbf{T}\in C^{N_s\times N_s}$ an arbitrary invertible matrix.

The proof is shown in Appendix A.

According to ***Theorem 1***, any orthogonal basis for the subspace spanned by the columns of Γ is optimal for problem (11) with constraint C1. Inspired by this fact, the hybrid precoder design problem is formulated as

$$\begin{aligned}\max_{\mathbf{F}_{RF},\mathbf{F}_{BB}}\ &\left\|\Gamma^*\mathbf{F}_{RF}\mathbf{F}_{BB}\right\|_F^2\\ s.t.\ \ &\text{C1:}\boldsymbol{e}_j^*\mathbf{F}_{RF}\mathbf{F}_{BB}(\mathbf{F}_{RF}\mathbf{F}_{BB})^*\boldsymbol{e}_j\leqslant p_j,\\ &\text{C2:}[\mathbf{F}_{RF}]_{m,n}\in\frac{1}{\sqrt{N_t}}\{+1,-1\},\\ &\text{C3:}\mathbf{F}_{RF}\in C^{N_t\times N_{RF}^t}.\end{aligned}\tag{14}$$

The problem (14) is non-convex with hardware-specific constraints C2, C3, and the per-antenna constraint C1. Fortunately, since the switches and inverters are used, the elements in the analog precoding matrix are limited, i.e., constraints C2 and C3. Due to the analog precoders are implemented by the inverters and switches, so each element of the analog precoding matrix $\mathbf{F}_{RF}$ is limited to $\{+1,-1\}$. Moreover, $\mathbf{F}_{RF}$ in (14) is a block diagonal matrix because each RF chain is connected to the lens sub-array, as shown in Fig. 1. Therefore, $\mathbf{F}_{RF}$ can be expressed as

$$\mathbf{F}_{RF}=\begin{bmatrix}\mathbf{f}_1^{RF} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{f}_n^{RF} & & \mathbf{0}\\ \vdots & & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{f}_{N_{RF}^t}^{RF}\end{bmatrix}_{N_t\times N_{RF}^t},\tag{15}$$

where $\mathbf{f}_n^{RF}$ represents the $\left(N_t/N_{RF}^t\right) \times 1$ analog precoder on the n-th lens sub-array.

The application of CE in the artificial neural networks (ANNs) [32] provides the solution to (14). The CE can be used to measure the difference between two probability distributions. In this problem, we first take out the non-zero elements in the analog precoding matrix $\mathbf{F}_{RF}$ to form a target vector $\mathbf{f} = \left[\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_{N_{RF}}^*\right]^*$, where the elements in the target vector take values between $+\frac{1}{\sqrt{N_t}}$ and $-\frac{1}{\sqrt{N_t}}$ with equal probability. Next, the probability vector is set as $\mathbf{p} = \left[p_1, p_2, \ldots, p_{N_t}\right]^*$, in which the elements represent the probability that the lens changes the phase of signal. It corresponds to the value of analog precoding matrix elements, i.e., $+\frac{1}{\sqrt{N_t}}$ and $-\frac{1}{\sqrt{N_t}}$. In the following, we initialize $\mathbf{p}^{(0)} = \frac{1}{2} \times \mathbf{1}_{N_t \times 1}$. In step 3, M prediction targets $\{\mathbf{f}^m\}_{m=1}^M$ are randomly generated based on $\Pi\left(F; \mathbf{p}^{(i)}\right)$ (i.e., generate F according to $\mathbf{p}^{(i)}$), and reshape them as matrices belong to). And new vector groups are reshaped into a set of block diagonal matrices belonging to F, where each block is a $\frac{N_t}{N_{RF}^t}$ dimension vector and the off-diagonal blocks are zero vectors. In step 4, the digital precoding matrix can be calculated based on the ZF digital precoding scheme, which corresponds to the analog precoding matrix. In principle, the ZF digital precoding scheme can be obtained by setting the digital precoding matrix, which is equal to the pseudo-inverse of the effective channel matrix. Therefore, the digital precoding matrix can be calculated as

$$\mathbf{G} = \left(\widetilde{\mathbf{H}}\mathbf{F}_{RF}\right)^H \left(\widetilde{\mathbf{H}}\mathbf{F}_{RF}\left(\widetilde{\mathbf{H}}\mathbf{F}_{RF}\right)^H\right)^{-1}, \tag{16}$$

$$\mathbf{F}_{BB} = \frac{\sqrt{\rho}}{\|\mathbf{F}_{RF}\mathbf{G}\|_F}\mathbf{G}, \tag{17}$$

where ρ is the normalized power.

Due to $\Im \triangleq \|\Gamma^*\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2$, we evaluate the objective values by calculating the corresponding $\{\Im_m\}_{m=1}^M$ and sort them in descending order. To select $1 \le m \le M_{elite}$ as the targets, and the weight of the elements in each analog precoding matrix can be obtained by

$$w_m = \frac{\Im_m}{\sum_{m=1}^{Melite} \Im_m}. \tag{18}$$

Based on (18), (19) updates $\mathbf{p}^{(i+1)}$ by minimizing CE, i.e.,

$$\mathbf{p}^{(i+1)} = \arg\min_{\mathbf{p}^{(i)}} \frac{1}{M} \sum_{m=1}^{M_{elite}} w_m \ln \Pi\left(F; \mathbf{p}^{(i)}\right). \tag{19}$$

As for the probability of $f_n = +\frac{1}{\sqrt{N_t}}$ is $p_n^{(i)}$ and the probability of $f_n = -\frac{1}{\sqrt{N_t}}$ is $1 - p_n^{(i)}$. Therefore, we have

$$\Pi\left(F; \mathbf{p}^{(i)}\right) = \prod_{n=1}^{N} \left(p_n^{(i)}\right)^{\frac{1}{2}\left(1+\sqrt{N_t}f_n^{[m]}\right)} \left(1 - p_n^{(i)}\right)^{\frac{1}{2}\left(1-\sqrt{N_t}f_n^{[m]}\right)}. \tag{20}$$

Finally, the following expression by substituting (20) into (19) can be rewritten as

$$\mathbf{p}^{(i+1)} = \underset{\mathbf{p}^{(i)}}{argmin} \sum_{m=1}^{M_{elite}} w_m \left(\left(\frac{1+\sqrt{N_t}f_n^{[m]}}{2\mathbf{p}^{(i)}}\right) - \left(\frac{1-\sqrt{N_t}f_n^{[m]}}{2\left(1-\mathbf{p}^{(i)}\right)}\right)\right) \tag{21}$$

B. HYBRID COMBINER DESIGN

In this subsection, with the aim of maximizing the achievable SE, we formulate the hybrid combiner optimization problem for designing the digital combiner $\mathbf{W}_{BB}$ and the analog combiner $\mathbf{W}_{RF}$. Similar to the design problem of the hybrid precoder, the sub-scheme first designs the optimization problem based on ***Theorem 1***. Therefore, the hybrid combiner design problem can be obtained by solving

Algorithm 2 Approximate Optimization Method (AOM)

1: Set $\widetilde{\phi}_{j,k}$ according to the beam spatial channel matrix $\widetilde{\mathbf{H}}$;
2: Compute $\mathbf{W}_{RF}$ by (23);
3: Set $\mathbf{A} = \Phi^*\mathbf{W}_{RF}$ and SVD $\mathbf{A} = \mathbf{U}_A\Sigma_A\mathbf{V}_A^*$ $\mathbf{W}_{BB} = \mathbf{U}_W\Sigma_W\mathbf{V}_W^*$;
4: Compute Σ_W by solving (30);
5: Compute $\mathbf{W}_{BB}$ by (25).

$$\begin{aligned} \max_{\mathbf{F}_{RF},\mathbf{F}_{BB}} & \left\|\Phi^*\mathbf{W}_{RF}\mathbf{W}_{BB}\right\|_F^2 \\ s.t.\ \ & \text{C1:}\left|[\mathbf{W}_{RF}]_{m,n}\right| = 1, \\ & \text{C2:}\mathbf{W}_{RF} \in C^{N_r \times N_{RF}^r}, \\ & \text{C3:}(\mathbf{W}_{RF}\mathbf{W}_{BB})^*(\mathbf{W}_{RF}\mathbf{W}_{BB}) = \mathbf{I}_{N_s}. \end{aligned} \tag{22}$$

C1, C2, and C3 in (22) are hardware-specific constraints for the hybrid combiner designing problem, which makes problem (22) as an NP-hard problem. Since the analog combiners are realized by the PSs, which only adjust the phases of the signals, each element of the analog combining matrix $\mathbf{W}_{RF}$ is limited to the same norm, i.e., unit modulus constraint C1. As mentioned in [9], maximizing the above objective function is equivalent to maximizing mutual information. This term is related to the chordal distance between Φ and $\mathbf{W}_{RF}\mathbf{W}_{BB}$ in the Grassmann manifold when Φ and $\mathbf{W}_{RF}\mathbf{W}_{BB}$ is made semi-unitary. We use the approximate optimization method to solve problem (22).

Assuming $\widetilde{\Phi} \in C^{N_r \times N_{RF}^r}$ is the N_{RF}^r left singular vectors of $\widetilde{\mathbf{H}}$ corresponding to the N_{RF}^r largest singular values, and $\widetilde{\phi}_{j,k}$ represent the (j,k)-th element of $\widetilde{\Phi}$.

Let

$$(\mathbf{W}_{RF})_{j,k} = \frac{\widetilde{\phi}_{jk}}{\left|\widetilde{\phi}_{jk}\right|}, \tag{23}$$

and we take (23) as the approximate value of $\mathbf{W}_{RF}$. According to the ***Theorem 1***, $\mathbf{W}_{RF}$ have orthonormal columns, there is no loss of the performance in terms of SE. Then, the variable that needs to be optimized in (22) is only $\mathbf{W}_{BB}$ with C3.

Let $\mathbf{A} = \Phi^* \mathbf{W}_{RF}$. By considering the SVD to $\mathbf{A}$ and $\mathbf{W}_{BB}$, the optimization variables in (22) can be expressed as

$$\mathbf{A} = \mathbf{U}_A \Sigma_A \mathbf{V}_A^*, \tag{24}$$

$$\mathbf{W}_{BB} = \mathbf{U}_W \Sigma_W \mathbf{V}_W^*. \tag{25}$$

Then, the optimization problem in (22) is formulated as

$$\left\| \mathbf{A}\mathbf{W}_{BB}^* \right\|_F^2 = tr\left(\mathbf{A}^* \mathbf{A} \mathbf{W}_{BB}^* \mathbf{W}_{BB} \right) \leq tr\left(\Sigma_A^2 \Sigma_W^2 \right), \tag{26}$$

when $\mathbf{V}_A = \mathbf{V}_W$, the "=" is established. We chose $\mathbf{W}_{BB} = \Sigma_W \mathbf{V}_A$ as the digital combiner. Without loss of optimality, we take $\mathbf{U}_W = \mathbf{I}_{N_s}$. Then, (22) is rewritten as

$$\max_{\Sigma_W} tr\left(\Sigma_A^2 \Sigma_W^2 \right)$$
$$s.t.\ \ \mathbf{V}_W \Sigma_W^2 \mathbf{V}_W^* = I_{N_s}. \tag{27}$$

The definitions of Σ_W and $\mathbf{V}_W$ are discretized as follows:

$$\Sigma_W = diag\left\{ \sigma_{W,1}, \sigma_{W,2}, \ldots, \sigma_{W,N_s} \right\}, \tag{28}$$

and

$$\mathbf{V}_W = diag\left\{ \sigma_{V,1}, \sigma_{V,2}, \ldots, \sigma_{V,N_s} \right\}. \tag{29}$$

Then, (27) is rewritten as

$$\max_{\left\{\sigma_{W,k}^2\right\}} \sum_{k=1}^{N_s} \sigma_{W,k}^2 \sigma_{V,k}^2$$
$$s.t.\ \ \sum_{k=1}^{N_s} \left| \sigma_{V,k} \right|^2 \sigma_{W,k}^2 = 1. \tag{30}$$

Foutunately, (30) is a linear program that is easy to solve.

C. COMPLEXITY ANALYSIS

Due to the hardware constraints imposed by the hybrid precoding architecture based-switches and inverters, each antenna is to turn off for transmitting signals. When the switch is "off", the n-th entry of the $\mathbf{f}$ is taken to be 1. And since the inverter ia applied, it can only adjust the phase of the signal, so the value of $\mathbf{f}_n$ is $\{+1, -1\}$. As result, the corresponding search space (complexity) is 2^{N_t} (e.g., $2^{64} \approx 1.84 \times 10^{19}$ when $N_t = 64$). The overall complexity of the algorithm is $O\left(2N_{RF}^t I\right)$.

IV. EXPERIMENT AND DISCUSSION

In this section, we provide simulation results for the proposed AHDS. In the simulations, a typical downlink mmWave massive MIMO system is considered, in which the DLA with transmit antennas configuration of a half wavelength spacing deployed at the BS and the uniform linear arrays (ULA) assumed at the receiver. The simulation parameters are listed in Table 1, in which P_{RF} is the transmitted power of per-antenna, P_{BB} is the power consumption of BB, P_{SW} is the power consumption of the switches, and P_{IN} is the power consumption of the inverters. B is the bandwidth.

TABLE 1. Simulation paramenters.

Parameter	values	Parameter	values
N_s	32	N_t	256
N_r	16	N_{RF}^t	4
K	32	L	2
B	1 MHz	p_j	0.125 mW
P_{RF}	300 mW	P_{BB}	500 mW
P_{SW}	5 mW	P_{IN}	5 mW

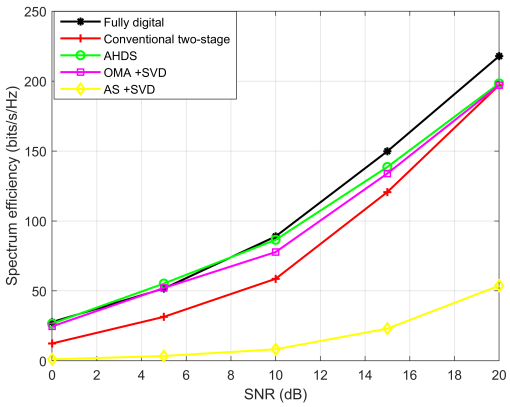


FIGURE 2. Spectrum efficiency against SNR.

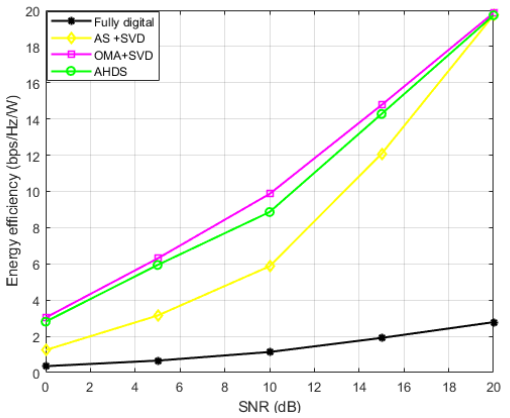


FIGURE 3. Energy efficiency against SNR.

The parameters of the channel model are set as follows [29]: the path gain follows $\beta_{r,t}^{(0)} \sim \mathcal{CN}(0, 1)$ and $\beta_{r,t}^{(l)} \sim \mathcal{CN}\left(0, 10^{-1}\right)$. The φ obeys the uniform distribution within $\left[-\frac{1}{2}, \frac{1}{2}\right]$. The number of LoS link is set to 1 and the number of NLoS links is set to 2.

This paper compares the SE with various schemes: the full-digital solution at the transmitter and receiver, the precoder and combiner are designed by the orthogonal multiplex access (OMA) precoding scheme [33] based on PSs, the conventional antenna selection (AS) precoding scheme and combining scheme-based PSs [34], the conventional two-stage precoding scheme-based PSs [35], and the proposed AHDS including the hybrid precoding

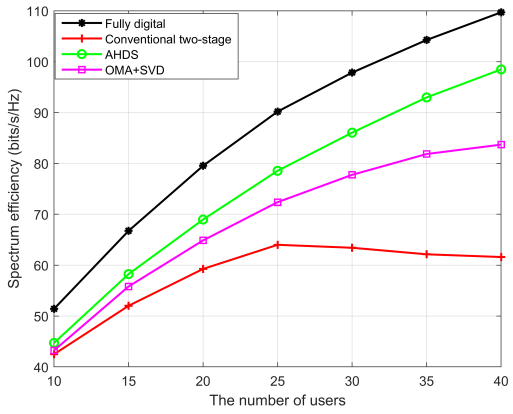


FIGURE 4. Spectrum efficiency against the number of the user, K.

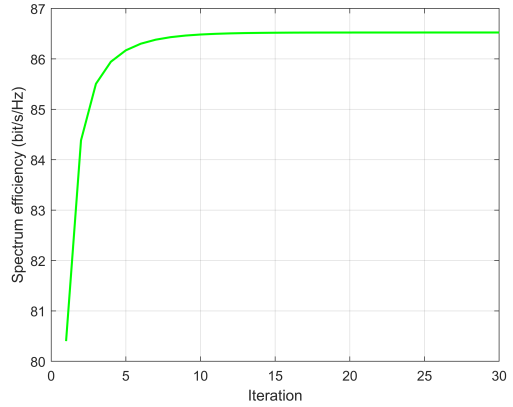


FIGURE 5. Spectrum efficiency against the iteration times.

scheme-based lens array antennas with ICE and the combining scheme-based PSs with AOM.

Fig. 2 shows the SE performance against the signal-to-noise ratio (SNR) of the proposed hybrid precoding/combining architecture. We can observe that the proposed AHDS achieves SE closing to the OMA method with small performance loss. The conventional AS precoding scheme is not ideal because of the uncertainty of antenna selection. Furthermore, when the SNR increases, the performance gap of AHDS and the fully digital scheme is becoming larger. But the gap is less than 20 bits/Hz. This superior performance further demonstrates the effectiveness of the proposed AHDS.

Fig. 3 shows the EE against SNR. The EE is defined as the ratio between the achievable sum-rate and the total power consumption. We can observe that the EE performance of the proposed AHDS-based the proposed hybrid architecture is superior to other schemes, including the AS precoding scheme and the fully digital precoding scheme. Meanwhile, the power consumption of the proposed AHDS is inferior to the OMA scheme. It is intuitive that the fully digital scheme can achieve the worst EE, as shown in Fig. 3, since the fully-connected mapping needs $N_{RF} = N_t = 64$ RF chains to sever all users. The lens antennas take full advantage of the sparseness of the beam-space, so the hardware loss of the proposed AHDS caused by RF chains will reduced.

Fig. 4 shows the performance of the different design schemes in terms of SE against the number of the users for the fixed SNR. In general, the proposed AHDS achieves SE closing to the fully digital design scheme. As the number of the users is increasing, the different users may choose the same beam for transmitting the signal. Therefore, the conventional two-stage hybrid precoding will undergo the performance loss, because some users maybe not be served. Moreover, the OMA scheme preforms for conflicting users, so the SE is not worst.

Fig. 5 shows the iteration times against the SE, where the SNR is set to 10 dB. In this simulation, the iteration times are set to 25. The number of elements in the precoding matrix is set to 250. As shown in Fig. 5, the SE tends to be stable after 10 times of iteration, which verifies the rationality of iteration value.

V. CONCLUSION

In this paper, we propose AHDS with the idea of CE optimization using lens sub-arrays antennas-based hybrid architecture. Moreover, the simulation results demonstrate the proposed AHDS is superior in SE and EE. Furthermore, this paper shows the CE optimization method has an advantage in solving the complicated combinatorial problems, which is proved to be simple, efficient and general. Specifically, the CE-loss function can be used to measure the true distribution and the predicted distribution of the trained model in ML while avoiding the problem of reduced learning rate and sensitivity to outliers. However, the combination of hybrid precoding and CE makes strict hardware on the hybrid precoding architecture, which guarantees the corresponding distribution of specific model. In future work, we plan to develop DL-based precoding algorithms in mmWave massive MIMO systems while extending the applicability of the hybrid precoding architecture.

APPENDIX A

Considering the SVD of precoder and combiner, i.e.,

$$\mathbf{F} = \mathbf{U}_F \Sigma_F \mathbf{V}_F^*, \tag{31}$$

$$\mathbf{W} = \mathbf{U}_W \Sigma_W \mathbf{V}_W^*. \tag{32}$$

We submit (31) and (32) into (11), so the optimization objective function in (10) is rewritten as

$$R = \log_2 \left| I_{N_s} + \frac{\rho}{N_s} \mathbf{U}_W^* \widetilde{\mathbf{H}} \mathbf{U}_F (\mathbf{A}\Sigma_F)^2 \mathbf{U}_F^* \widetilde{\mathbf{H}}^* \mathbf{U}_W \right|. \tag{33}$$

Therefore, (11) is reformulated with constraints (34)-(36)

$$\mathbf{U}_W^* \mathbf{U}_W = \mathbf{I}_{N_s}, \tag{34}$$

$$\mathbf{U}_F^* \mathbf{U}_F = \mathbf{I}_{N_s}, \tag{35}$$

$$p_0 \mathbf{I}_{N_s} \geq \Sigma_F^2, \tag{36}$$

where p_0 is defined as $p_0 \triangleq \min_j \{p_j\}$.

Let

$$\Sigma_F = diag\left\{\rho_1 \rho_2 \cdots \rho_{N_s}\right\}, \tag{37}$$

$$\mathbf{X} = \left(\frac{\rho}{N_s}\mathbf{U}_F^*\widetilde{\mathbf{H}}^*\mathbf{U}_W\mathbf{U}_W^*\widetilde{\mathbf{H}}\mathbf{U}_F\right)^{-1}, \tag{38}$$

$$\begin{aligned} R &= \log_2\left|\mathbf{I}_{N_s} + \mathbf{X}^{-1}(\Sigma_F\mathbf{A})^2\right| \\ &= \log_2\left|\mathbf{X}^{-1}\right| + \log_2\left|\mathbf{X} + (\Sigma_F\mathbf{A})^2\right|. \end{aligned} \tag{39}$$

Denoting $\mathbf{X}_k = \mathbf{X} + \sum_{i\neq k} p_i\mathbf{e}_i\mathbf{e}_i^*$, one has

$$\begin{aligned} \left|\mathbf{X} + (\Sigma_\mathbf{F}\mathbf{A})^2\right| &= \left|\mathbf{X} + \sum_{i=1}^{N_s}\rho_i^2\mathbf{e}_i\mathbf{e}_i^*\right| \\ &= \left(1 + \rho_k^2\mathbf{e}_k^*\mathbf{X}_k^{-1}\mathbf{e}_k\right)\left|\mathbf{X}_k\right|, \end{aligned} \tag{40}$$

where the value of $\rho_k^2 \in [0, p_0]$. When $\rho_k^2 = p_0$ $(k = 1, 2, \ldots, N_s)$ is established, the $\left|\mathbf{X}+(\Sigma_F\mathbf{A})^2\right|$ is maximal. Hence, $(\Sigma_F\mathbf{A})^2 = p_0\mathbf{I}_{N_s}$ is optimal. Thus, we have

$$R = \log_2\left|I_{N_s} + \frac{\rho p_0}{N_s}\mathbf{U}_W^*\widetilde{\mathbf{H}}\mathbf{U}_F\mathbf{U}_F^*\widetilde{\mathbf{H}}^*\mathbf{U}_W\right|. \tag{41}$$

Let $\mathbf{G} = \frac{\rho p_0}{N_s}\widetilde{\mathbf{H}}\mathbf{U}_F\mathbf{U}_F^*\widetilde{\mathbf{H}}^*$, and the optimization function (33) is rewritten as

$$R = \log_2\left|\mathbf{I}_{N_s} + \mathbf{U}_W^*\mathbf{G}\mathbf{U}_W\right|. \tag{42}$$

For $\mathbf{B} \in C^{n\times n}$ Hermite, let $\lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{B}) \geq \cdots \geq \lambda_n(\mathbf{B})$ be its ordered eigenvalues. Using lemma 3 in [36].

$$\begin{aligned} \left|\mathbf{I}_{N_r} + \mathbf{U}_W\mathbf{U}_W^*\mathbf{G}\right| &\leq \prod_{i=1}^{N_r}\left(1 + \lambda_i\left(\mathbf{U}_W\mathbf{U}_W^*\right)\lambda_i(\mathbf{G})\right) \\ &= \prod_{i=1}^{N_s}(1 + \lambda_i(\mathbf{G})), \end{aligned} \tag{43}$$

when the columns of $\mathbf{U}_W$ constitute an orthonormal basis for the subspace spanned by the N_s dominant eigenvectors of $\mathbf{G}$, the upper bounder is achieved.

It remains to maximizing (43) with respect to $\mathbf{U}_F$. Note that

$$\begin{aligned} \prod_{i=1}^{N_s}(1 + \lambda_i(\mathbf{G})) &\leq \prod_{i=1}^{N_r}(1 + \lambda_i(\mathbf{G})) \\ &= \left|\mathbf{I}_{N_r} + \mathbf{G}\right| \\ &= \left|\mathbf{I}_{N_r} + \frac{\rho p_0}{N_s}\widetilde{\mathbf{H}}\mathbf{U}_F\mathbf{U}_F^*\widetilde{\mathbf{H}}^*\right| \\ &= \left|\mathbf{I}_{N_s} + \frac{\rho p_0}{N_s}\mathbf{U}_F\mathbf{U}_F^*\widetilde{\mathbf{H}}^*\widetilde{\mathbf{H}}\right| \\ &\leq \prod_{i=1}^{N_r}\left(1 + \frac{\rho p_0}{N_s}\lambda_i\left(\mathbf{U}_F\mathbf{U}_F^*\right)\lambda_i\left(\widetilde{\mathbf{H}}^*\widetilde{\mathbf{H}}\right)\right) \\ &= \prod_{i=1}^{N_r}\left(1 + \frac{\rho p_0}{N_s}\lambda_i\left(\widetilde{\mathbf{H}}^*\widetilde{\mathbf{H}}\right)\right), \end{aligned} \tag{44}$$

when the columns of $\mathbf{U}_F$ span the subspace of the N_s dominant right singular vectors of $\widetilde{\mathbf{H}}$, the upper bound in (44) is achieved. This yields $\widetilde{\mathbf{H}}\mathbf{U}_F\mathbf{U}_F^*\widetilde{\mathbf{H}} = \Phi\Sigma^2\Phi^*$, where $\Phi \in C^{N_r\times N_s}$ comprises the N_s dominant left singular vectors of $\widetilde{\mathbf{H}}$, and $\Sigma \in C^{N_s\times N_s}$ is diagonal with the corresponding N_s largest singular values.

Hence, the optimum $\mathbf{U}_W$ is of the form $\mathbf{U}_W = \Phi\mathbf{R}$ for any $\mathbf{R} \in C^{N_s\times N_s}$ unitary. The maximize value of R is given as

$$R = \left|\mathbf{I}_{N_s} + \frac{\rho p_0}{N_s}\Sigma^2\right|. \tag{45}$$

REFERENCES

[1] M. R. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[2] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300GHz spectrum," in *Proc. 34th IEEE Sarnoff Symp.*, Princeton, NJ, USA, May 2011, pp. 1–6.

[3] 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.

[4] M. Zeng, W. Hao, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation in uplink mmWave massive MIMO with NOMA," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3000–3004, Mar. 2019.

[5] L. Li, J. He, L.-L. Yang, Z. Han, M. Pan, W. Chen, H. Zhang, and X. Li, "Spectral- and energy-efficiency of multi-pair two-way massive MIMO relay systems experiencing channel aging," *IEEE Access*, vol. 7, pp. 46014–46032, 2019.

[6] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2019.

[7] C. Chen, Y. Dong, X. Cheng, and L. Yang, "Low-resolution PSs based hybrid precoding for multiuser communication systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6037–6047, Jul. 2018.

[8] R. W. Heath, Jr., N. González-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.

[9] 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.

[10] C. Rusu, R. Mèndez-Rial, N. González-Prelcic, and R. W. Heath, Jr., "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.

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

[12] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[13] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

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

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

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

[17] Y. Dong and L. Qiu, "Spectral efficiency of massive MIMO systems with low-resolution ADCs and MMSE receiver," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1771–1774, Aug. 2017.

[18] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, Feb. 2016.

[19] C. Kong, C. Zhong, S. Jin, S. Yang, H. Lin, and Z. Zhang, "Full-duplex massive MIMO relaying systems with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5033–5047, Aug. 2017.

[20] Q. Hou, R. Wang, E. Liu, and D. Yan, "Hybrid precoding design for MIMO system with one-bit ADC receivers," *IEEE Access*, vol. 6, pp. 48478–48488, 2018.

[21] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, Jr., "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[22] L. Li, Y. Xu, Z. Zhang, J. Yin, W. Chen, and Z. Han, "A prediction-based charging policy and interference mitigation approach in the wireless powered Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 2, pp. 439–451, Feb. 2019.

[23] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.

[24] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5g wireless techniques: Opportunities, challenges and solutions," 2019, *arXiv:1904.09673*. [Online]. Available: https://arxiv.org/abs/1904.09673

[25] R. Y. Rubinstein and D. P. Kroese, "The cross-entropy method: A unified approach to combinatorial optimization Monte-Carlo simulation and machine learning," *Technometrics*, vol. 48, no. 1, pp. 147–148, Feb. 2006.

[26] X. Huang, W. Xu, G. Xie, S. Jin, and X. You, "Learning oriented cross-entropy approach to user association in load-balanced HetNet," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 1014–1017, Dec. 2018.

[27] G. Alon, D. P. Kroese, T. Raviv, and R. Y. Rubinstein, "Application of the cross-entropy method to the buffer allocation problem in a simulation-based environment," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 137–151, Feb. 2005.

[28] R. Y. Rubinstein, "Combinatorial optimization, cross-entropy, ants and rare events," in *Stochastic Optimization: Algorithms and Applications*, vol. 54. Boston, MA, USA: Springer, 2001, pp. 303–363.

[29] Z. Lu, Y. Zhang, and J. Zhang, "Quantized hybrid precoding design for millimeter-wave large-scale MIMO systems," *China Commun.*, vol. 16, no. 4, pp. 130–138, Apr. 2019.

[30] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, 1952.

[31] O. Bar-Shalom and A. J. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 868–884, Jul. 2002.

[32] M. Kovaleva, D. Bulger, B. A. Zeb, and K. P. Esselle, "Cross-entropy method for electromagnetic optimization with constraints and mixed variables," *IEEE Trans. Antennas Propag.*, vol. 65, no. 10, pp. 5532–5540, Oct. 2017.

[33] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[34] 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. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2015, pp. 90–97.

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

[36] H. S. Witsenhausen, "A determinant maximization problem occurring in the theory of data communication," *SIAM J. Appl. Math.*, vol. 29, no. 3, pp. 515–522, Nov. 1975.

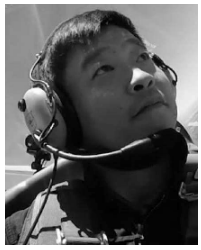
LIXIN LI (M'12) received the B.Sc. and M.Sc. degrees in communication engineering and the Ph.D. degree in control theory and its applications from Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2008, respectively. He was a Postdoctoral Fellow with NPU, from 2008 to 2010. In 2017, he was a Visiting Scholar with the University of Houston, TX, USA. He is currently an Associate Professor with the School of Electronics and Information, NPU. He has authored or coauthored over 100 technical articles in journals and international conferences. He holds ten patents. His current research interests include wireless communications, game theory, and machine learning. He has reviewed articles for many international journals. He received the 2016 NPU Outstanding Young Teacher Award, which is the highest research and education honors for young faculties in NPU.

HUAN REN is currently pursuing the master's degree with the School of Electronics and Information, Northwestern Polytechnical University, China, under the supervision of Prof. L. Li. Her research interests include hybrid precoding in millimeter massive MIMO and machine learning in wireless communication networks.

XU LI received the B.Sc. degree in electronics and information technology, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuits and systems from Northwestern Polytechnical University, Xi'an, China, in 2001, 2004, and 2010, respectively. In 2004, he joined the School of Electronics and Information, Northwestern Polytechnical University, where he is currently an Associate Professor with the School of Electronics and Information. From 2007 to 2008, he was a Visiting Ph.D. Student with Telecom ParisTech (ENST), Paris, France. From 2012 to 2014, he was an Enterprise Postdoctoral Fellow with the Jiangsu RD Center for Internet of Things, Wuxi, China. From 2015 to 2016, he was a Marie Currie Research Fellow with the School of Computer Science, University of Lincoln, Lincoln, U.K. His research interests include image fusion, image enhancement, and image super-resolution.

WEI CHEN (S'05–M'07–SM'13) received the B.S. and Ph.D. degrees (Hons.) from Tsinghua University, in 2002 and 2007, respectively.

From 2005 to 2007, he was also a Visiting Ph.D. Student with The Hong Kong University of Science and Technology. Since 2007, he has been on the faculty of Tsinghua University, where he is currently a tenured Full Professor, the Director of the Degree Office, and a University Council Member. From 2014 to 2016, he served as the Deputy Head of the Department of Electronic Engineering. He visited the University of Southampton, Telecom ParisTech, and Princeton University, in 2010, 2014, and 2016, respectively. His research interests include communication theory, stochastic optimization, and statistical learning. He is a Cheung Kong Young Scholar and a member of the National Program for Special Support for Eminent Professionals, also known as a 10000-Talent Program. He was a recipient of the National May 1st Labor Medal and the China Youth May 4th Medal. He received the IEEE Marconi Prize Paper Award and the IEEE Comsoc Asia–Pacific Board Best Young Researcher Award, in 2009 and 2011, respectively. He has served as a TPC Co-Chair for the IEEE VTC-Spring, in 2011, and the Symposium Co-Chair for the IEEE ICC and GLOBECOM. He has also been supported by the National 973 Youth Project, the NSFC Excellent Young Investigator Project, the New Century Talent Program of Ministry of Education, and the Beijing Nova Program. He serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.



ZHU HAN (S'01–M'04–SM'09–F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an Research and Development Engineer with JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, ID, USA. He is currently a Professor with the Electrical and Computer Engineering Department and the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He received the NSF Career Award, in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society, in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing*, in 2015, the IEEE Leonard G. Abraham Prize in the field of communications systems (Best Paper Award in the IEEE JSAC), in 2016, and several best paper awards in IEEE conferences. He is currently an IEEE Communications Society Distinguished Lecturer.